Proceedings of the 10th International Conference on Wirtschaftsinformatik

WI 2.011

Volume 2

16-18 February 2011
Zurich, Switzerland

Abraham Bernstein and Gerhard Schwabe
(Editors)

Proceedings of the 10th International Conference on Wirtschaftsinformatik

WI 2.011

Volume 2

16-18 February 2011
Zurich, Switzerland

Volume 1 of the conference proceedings contains papers submitted to the following tracks:
-IT in Services
-IS in Industries and Business Applications
-Information Management
-Student track
ISBN: 978-1-4467-9233-9

Volume 2 of the conference proceedings contains papers submitted to the following tracks:
-Development/Architecture of IS
-Adoption, Usage and Impact of IS
-New Technologies and Applications
-Theory and Methodology in Information Systems
ISBN: 978-1-4467-9236-0

Copies of the conference proceedings can be ordered from www.lulu.com

Editorial production by Tom Philip
Cover art production by Natasa Milosevic and Stephanie van Grondel
Printed by Lulu.com

www.wi2011.ch

Table of Contents

Tenth International Conference on Wirtschaftsinformatik

Track 5: Adoption, Usage and Impact of IS

Track 6: New Technologies and Applications

Track 7: Theory and Methodology of Information Systems

Wirtschaftsinformatik 2011: Towards Information Systems 2.011

Preface

Abraham Bernstein
Department of Informatics
University of Zürich
Binzmühlestrasse 14
CH-8050 Zürich, Switzerland
bernstein@ifi.uzh.ch

Gerhard Schwabe
Department of Informatics
University of Zürich
Binzmühlestrasse 14
CH-8050 Zürich, Switzerland
schwabe@ifi.uzh.ch

Welcome to the Wirtschaftinformatik 2011

Information technology is permeating all sectors of both economic and everyday life increasing the importance of the Information Systems field. The goal of the 10^{th} International Conference Wirtschaftinformatik (Information Systems) is to improve our understanding of this development from a scientific perspective as well as to provide practitioners and academics a platform for the discussion of this topic.

In particular, the increasing pace of changes due to the Internet/Business 2.0 phenomenon require reflection on the durability of our findings. Therefore, we chose the conference motto "Information Systems 2.011" to illustrate our needs to think beyond the current phenomena.

Last August 398 authors of papers followed our call for papers and submitted original research papers for the following tracks:

- Track 1: IS in Services
- Track 2: IS in Industries and Business Applications
- Track 3: Information Management
- Track 4: Development / Architecture of IS
- Track 5: Adoption, Usage and Impact of IS
- Track 6: New Technologies and Applications
- Track 7: Theory and Methodology of Information Systems

Each paper was checked by the Track Chairs (TCs) for suitability for the track and then assigned to an Associate Chair (AE) and at least 3 Reviewers. After a first round of reviewing most authors were given a chance to write a rebuttal (or clarification). The AE then lead a discussion with the reviewers resulting—where needed—in additional reviews and ultimately to his or her recommendation to the track meetings. In the week of November 15, each track held a track meeting in Zürich or Berlin, where each paper with a chance to get accepted was discussed throughly and recommended for acceptance, conditional acceptance, or rejection to the final program committee meeting. Authors of conditionally accepted paper got the chance to revise their paper if the track meeting participants felt that such a revision could be successfully achieved in the allotted two weeks. The process resulted in 110 accepted papers and, hence, an overall acceptance rate of a bit below 28% – all of which are included in these proceedings.

In addition to the main paper track, the conference will feature 3 keynotes, 3 panels, a doctoral consortium, and a students track with papers that were submitted by students and reviewed separately.

As always, these proceedings and the associated conference would not have been possible without the diligent work of a large number of individuals. In addition to the 15 Track Chairs, 131 Associate Editors (plus an 21 additional in the Student Track), and 555 Reviewers, each track had a CMT-manager (all of them will be named individually in the next few pages). Furthermore, we were supported by the local organizers—Evelyne Berger and Annouk Bühler—as well as the delegate of the general chairs—Tom Philip. We owe them our deepest gratitude, as the program would not have been possible without them.

Last but not least, we are indebted to our conferences Gold Sponsors Avaloq, IBM, and SAP as well as our other sponsors HILTI and the SI (Swiss Informatics Society). Their generous support allowed us to offer the conference participants the typical conference amenities at a reasonable price.

In closing, we hope that you will enjoy reading the papers in this volume and wish you a thought-provoking conference.

December 2010

Abraham Bernstein & Gerhard Schwabe

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

Conference Organization

General Chairs

Abraham Bernstein, University of Zurich
Gerhard Schwabe, University of Zurich

Track Chairs

Bettina Berendt, Catholic University Leuven
Wynne Chin, University of Houston
Jens Dibbern, University of Berne
Oliver Günther, Humboldt University Berlin
Stefan Klein, University of Munster
Armin Heinzl, University of Mannheim
Gerti Kappel, TU Vienna
Stefan Klein, University of Munster
Helmut Krcmar, TU Munich
Susanne Leist, University of Regensburg
Peter Loos, University of Saarland
Gustaf Neumann, Vienna University of Economics and Business
Markus Nüttgens, University of Hamburg
Detlef Schoder, University of Cologne
Jan vom Brocke, University of Liechtenstein
Robert Winter, University of St. Gallen
Volker Wulf, University of Siegen and Fraunhofer FIT

Delegate of the General Chairs

Tom Philip, University of Zurich

Track Chair Assistance

Silke Balzert, German Research Center for Artificial Intelligence
Thorsten Dollmann, German Research Center for Artificial Intelligence
Kai Fischbach, University of Cologne
Katja Hildebrandt, TU Vienna
Florian Johannsen, University of Regensburg
Paul Meyer, University of Berne
Benjamin Rensmann, University of Munster
Christoph Riedl, TU Munich
Kerstin Schaefer, Humboldt University Berlin
Bernd Schenk, University of Liechtenstein
Roman Tilly, University of Cologne

Local Arrangments

Annouk Bühler, University of Zurich
Evelyne Berger, University of Zurich

Associate Editors

Alessandro Acquisti, Carnegie Mellon University
Heimo Adelsberger, University of Duisburg-Essen
Rainer Alt, University of Leipzig
Michael Amberg, University of Erlangen-Nürnberg
Elske Ammenwerth, UMIT
Hans-Jürgen Appelrath, University of Oldenburg
Hans-Knud Arndt, University of Magdeburg
Andrea Back, University of St. Gallen
Ulrike Baumöl, University of Hagen
Roman Beck, Goethe University Frankfurt
Jörg Becker, University of Munster
Ralph Bergmann, University of Trier
Martin Bichler, TU Munich
Paola Bielli, University of Bocconi
Claudia Müller-Birn, Carnegie Mellon University
Freimut Bodendorf, University of Erlangen-Nürnberg
Tilo Böhmann, ISS Hamburg
Michael Breitner, University of Hannover
Walter Brenner, University of St. Gallen
Ruth Breu, University of Innsbruck
Martin Breunig, University of Osnabrück
Peter Buxmann, University of Darmstadt
Hans Ulrich Buhl, University of Augsburg
Peter Chamoni, University of Duisburg-Essen
Wilhelm Dangelmaier, University of Paderborn
Stefan Eicker, University of Duisburg-Essen
Gregor Engels, University of Paderborn
Werner Esswein, TU Dresden
Torsten Eymann, University of Bayreuth
Carsten Felden, TU Freiberg
Peter Fettke, University of Saarlande
Andreas Fink, Helmut Schmidt University and Bundeswehr University Hamburg
Kerstin Fink, University of Insbruck
Elgar Fleisch, ETH Zurich and University of St. Gallen
Christiane Floyd, University of Hamburg
Ulrich Frank, University of Duisburg-Essen
Mike Gallivan, Georgia State University
Peter Gluchowski, TU Chemnitz
Jorge Marx Gómez, University of Oldenburg
Manfred Grauer, University of Siegen
Norbert Gronau, University of Potsdam
Axel Hahn, University of Oldenburg
Felix Hampe, University of Koblenz
Bernd Heinrich, University of Innsbruck
Georg Herzwurm, University of Stuttgart
Thomas Hess, University of Munich
Oliver Hinz, Goethe University Frankfurt
Roland Holten, Goethe University Frankfurt
Christian Huemer, TU Vienna
Thomas Hummel, Accenture
Hans-Arno Jacobsen, University of Toronto
Norman Johnson, University of Houston
Reinhard Jung, University of St. Gallen
Dimitris Karagiannis, University of Vienna
Hans-Georg Kemper, University of Stuttgart
Stefan Kirn, University of Hohenheim
Gerhard Knolmayer, University of Bern
Michael Koch, Bundeswehr University Munich
Stefan Koch, Bogazici University
Birgitta König-Ries, University of Jena
Lutz Kolbe, Georg-August University Göttingen
Jochen Küster, IBM Zurich Research Laboratory
Dennis Kundisch, University of Paderborn
Karl Kurbel, European University Viadrina Frankfurt (Oder)
Ulrike Lechner, Bundeswehr University Munich
Christine Legner, European Business School
Franz Lehner, University of Passau
Jan Marco Leimeister, University of Kassel
Frank Leymann, University of Stuttgart
Yan Li, ESSEC Business School
Claudia Loebbecke, University of Köln
Wolfgang Maass, Furtwangen University
Alexander Maedche, University of Mannheim
Dirk Mattfeld, TU Braunschweig
Florian Matthes, TU Munich
Marco C. Meier, University of Augsburg
Jan Mendling, Humbold University Berlin
Mirjam Minor, University of Trier
Lars Mönch, University of Hagen
Kathrin Möslein, University of Erlangen-Nürnberg
Günter Müller, University of Freiburg
Ulrich Müller-Funk, University of Munster
Thomas Myrach, University of Bern
Björn Niehaves, University of Münster
Volker Nissen, TU Ilmenau
Jasminko Novak, University of Applied Sciences Stralsund
Markus Nüttgens, University of Hamburg
Stefan Nusser, IBM Research Almaden
Andreas Oberweis, University of Karlsruhe
Erich Ortner, TU Darmstadt
Günther Pernul, University of Regensburg
Niels Pinkwart, TU Clausthal
Volkmar Pipek, University of Siegen
Klaus Pohl, University of Duisburg-Essen
Nancy Pouloudi, Athens University of Economics and Business
Key Pousttchi, University of Augsburg
Kai Rannenberg, Goethe University Frankfurt
René Riedl, University of Linz
Kai Riemer, University of Sydney
Susanne Robra-Bissatz, TU Braunschweig
Markus Rohde, University of Siegen

Franz Rohloff, University of Mainz
Michael Rosemann, Queensland University of Technology
Kjeld Schmidt, Copenhagen Business School
Mareike Schoop, University of Hohenheim
Michael Schrefl, JKU Linz
Guido Schryen, RWTH Aachen
Petra Schubert, University of Koblenz-Landau
Matthias Schumann, Georg-August University Göttingen
Christa Schwanninger, Siemens
Andrew Schwarz, Louisiana State University
Stefan Smolnik, European Business School
Gunnar Stevens, University of Siegen
Susanne Strahringer, TU Dresden
Stefan Strohmeier, University of Saarland
Leena Suhl, University of Paderborn
Martin Spann, Ludwig Maximilians University München
Myra Spiliopoulou, Otto von Guericke University Magdeburg
Steffen Staab, University of Koblenz-Landau
Dirk Stelzer, TU Ilmenau
Eberhard Stickel, University of Applied Sciences Bonn
Jens Strüker, University of Freiburg
York Sure, Leibniz Institute for the Social Sciences
Alfred Taudes, Vienna University of Economics and Business
Alexander Teubner, University of Munster
Stephanie Teufel, University of Freiburg
Frank Teuteberg, University of Osnabrück
Bernhard Thalheim, Christian Albrechts University Kiel
Jason Thatcher, Clemson University
Frederic Thiesse, University of St. Gallen
Oliver Thomas, University of Osnabrück
A Min Tjoa, TU Vienna
Klaus Tochtermann, TU Graz
Klaus Turowski, University of Augsburg
Axel Uhl, SAP
Axel Uhl, University of Applied Sciences Northwestern Switzerland
Stefan Voss, University of Hamburg
Ina Wagner, TU Vienna
Christof Weinhardt, TU Karlsruhe
Tim Weitzel, University of Bamberg
Isabell Welpe, TU Munich
Oliver Wendt, TU Kaiserslautern
Hannes Werthner, TU Vienna
Rüdiger Zarnekow, TU Berlin
Uwe Zdun, TU Vienna
Peter Zencke, SAP

Reviewers

Rainer Alt, University of Leipzig
Joerg Ackermann, SAP
Christoph Adolphs, University of Koblenz
Andreas Albers, Goethe University Frankfurt
Michael Altenhofen, SAP
Jörn Altmann, Seoul National University
Nadine Amende, University of Passau
Chad Anderson, Georgia State University
Henning Baars, University of Stuttgart
Hillol Bala, Indiana University
Bastian Bansemir, Friedrich Alexander University Erlangen-Nürnberg
Thomas Barth, University of Siegen
Stefano Basaglia, Bocconi University
Christian Bauer, nVista technologies
Oliver Baumann, LMU Munich
Ulrike Baumöl, University of Hagen
Roman Beck, Goethe University Frankfurt
Jörg Becker, University of Munster
Alexander Benlian, LMU Munich
Michael Berger, DocuWare
Marina Berkovich, TU Munich
Martin Bertramj, Commerzbank
Markus Böhm, TU Munich
Martin Böhringer, TU Chemnitz
Markus Bick, ESCP Europe
Christian Bischof, RWTH Aachen
Philipp Bitzer, University of Kassel
Stephan Bloehdorn, Karlsruhe Institute of Technology
Alexander Boden, University of Siegen
Tilo Boehmann, ISS Hamburg
Franczyk Bogdan, University of Leipzig
Bernhard Böhm, TU Vienna
Susanne Boll, University of Oldenburg
Jens Borchers, Schufa Holding
Michael Breitner, Leibniz University Hannover
Carsten Brockmann, University of Potsdam
Steffen Budweg, Fraunhofer FIT
Hans Ulrich Buhl, University of Augsburg
Angelika Bullinger, Friedrich Alexander University Erlangen-Nürnberg
Lars Burmester, Cundus
Cinzia Cappiello, Politecnico di Milano
Michelle Carter, Clemson University
Frank Chan, Hong Kong University of Science and Technology Hong Kong
Yi-da Chen, University of Arizona
Jyoti Choudrie, University of Hertfordshire
Kevin Craig, Clemson University
Mandy Dang, University of Arizona
Bruce Dehning, Chapman University
Gernot Dern, SEB Bank
Barbara Dinter, University of St. Gallen
Daniel Dünnebacke, RWTH Aachen
Sebastian Dünnebeil, TU Munich
Andreas Drechsler, University of Duisburg-Essen
Thomas Dreier, Karlsruhe Institute of Technology
Alexander Dreiling, SAP
Roland Dösing, Ruhr University Bochum
Michael Durst, itonics
Rajeev Dwivedi, Howe School of Technology Management
Stefan Eckstein, Cologne University of Applied Sciences
Philip Effinger, University of Tuebingen
Ulrich Egle, Lucerne University of Applied Sciences and Arts
Roman Englert, Deutsche Telekom
Martin Engstler, HDM Stuttgart
Waltraud Ernst, University of Linz
Werner Esswein, TU Dresden
Adir Even, Ben-Gurion University
Torsten Eymann, University of Bayreuth
Jens Fähling, TU Munich
Thomas Fehlmann, Euro Project Office
Carsten Felden, TU Freiberg
Stefan Figge, Deutsche Telekom
Kathrin Figl, Vienna University of Economics and Business
Agata Filipowska, Poznan University of Economics
Hans-Georg Fill, University of Vienna
Ralf Finger, Information Works
Kai Fischbach, University of Cologne
Daniel Fischer, TU llmenau
Sabine Fliess, University of Hagen
Ulrich Frank, University of Duisburg-Essen
Bernd Freisleben, Philipps University Marburg
Michael Freundlieb, University of Osnabrueck
Karsten Frey, University of Bern
Norbert Frick, University of Koblenz
Gilbert Fridgen, University of Augsburg
Burkhardt Funk, University of Lueneburg
Andreas Gadatsch, Bonn-Rhine-Sieg University of Applied Sciences
Martin Gaedke, TU Chemnitz
Uri Gal, University of Sydney
Sandro Georgi, University of St. Gallen
Gerhard Satzger, Karlsruhe Institute of Technology
Daniel Gille, University of Freiburg
Peter Gluchowski, TU Chemnitz
Lakshmi Goel, University of North Florida
Thomas Goldschmidt, ABB
Peter Gomber, Goethe University Frankfurt
Suparna Goswami, TU Munich
Joern Grahl, University of Mainz

Henner Graubitz, Otto von Guericke University Magdeburg
Robert Gregory, Goethe University Frankfurt
Bernhard Groene, SAP
Norbert Gronau, University of Potsdam
Tom Gross, Bauhaus University Weimar
Jens Grossklags, Princeton University
Volker Gruhn, University of Duisburg-Essen
Marcel Grüter, University of Berne
Maik Guenther, Stadtwerke Munich
Daniel Gull, University of Augsburg
Dorina Gumm, Effective Webwork
Kai Gutenschwager, University of Applied Sciences Ulm
Joerg Haake, University of Hagen
Hagen Habicht, Leipzig Graduate School of Management
Werner Hackl, UMIT
Bjoern Haeckel, University of Augsburg
Axel Hahn, University of Oldenburg
Michael Hahsler, Southern Methodist University
Jörg Haller, Friedrich Alexander University Erlangen-Nürnberg
Ulrich Hasenkamp, Philipps University of Marburg
Wilhelm Hasselbring, University of Kiel
Helmut Hauptmeier, University of Siegen
Iris Hausladen, Leipzig Graduate School of Management
Markus Heckner, Accenture
Julia Heidemann, University of Augsburg
Bernd Heinrich, University of Innsbruck
Lutz Heinrich, Johannes Kepler University Linz
Priscilla Heinze, University of Potsdam
Markus Heller, SAP Research
Erik Hemmer, University of Mannheim
Matthias Henneberger, Siemens
Alexander Herzfeldt, TU Munich
Otthein Herzog, Jacobs University Bremen and University of Bremen
Jan Hess, University of Siegen
Thomas Hess, LMU Munich
Lorenz Hilty, University of Zurich
Knut Hinkelmann, University of Applied Sciences Northwestern Switzerland
Oliver Hinz, Goethe University Frankfurt
Eike Hirdes, University of Kassel
Martin Hochmeister, TU Vienna
Axel Hoffmann, University of Kassel
Holger Hoffmann, University of Kassel
Birgit Hofreiter, University of Liechtenstein
Jens Hogreve, University of Paderborn
Roland Holten, Goethe University Frankfurt
Alexander Hörbst, UMIT
Bettina Hoser, Karlsruhe Institute of Technology
Andreas Hotho, University of Wuerzburg
Volker Hoyer, University of St. Gallen
Hagen Höpfner, University of Weimar
Michael Huber, TU Munich
Thomas Huber, University of Berne
Christian Huemer, TU Vienna
Ulrike Hugl, University of Innsbruck
Thomas Hummel, Accenture
Dean Jacobs, SAP
Christian Janiesch, SAP Research
Marc Jansen, IT.NRW
Nils Joachim, University of Bamberg
Florian Johannsen, University of Regensburg
Isabel John, Aristotle University Thessaloniki
Reinhard Jung, University of St. Gallen
Fischbach Kai, University of Cologne
Marcus Kaiser, Senacor Technologies
Axel Kalenborn, University of Trier
Dimitris Karagiannis, University of Vienna
Dimitris Karaiskos, Athens University of Economics and Business
Alexandra Kees, Bonn-Rhine-Sieg University of Applied Sciences
Matthias Kehlenbeck, Leibniz University Hannover
Thorben Keller, IBM
Gudrun Kellner, TU Vienna
Hans Kemper, University of Stuttgart
Dogan Kesdogan, University of Siegen and Norwegian University of Science and Technology
Frank Kühne, Viadee
Marc Klages, Leibniz University Hannover
Ralf Klamma, RWTH Aachen
Christian Klein, University of Munster
Mathias Klier, University of Innsbruck
Ralf Klischewski, German University of Cairo
Birgitta König-Ries, Friedrich Schiller University Jena
Gerhard Knolmayer, University of Berne
Achim Koberstein, University of Paderborn
Maximilian Kobler, University of Applied Sciences Burgenland
Wolfgang Koenig, University of Frankfurt
Sabine Koeszegi, TU Vienna
Lutz Kolbe, Georg-August University Goettingen
Oliver Kopp, University of Stuttgart
Emanuel Kopp, Austrian National Bank
Roland Koppe, OFFIS Oldenburg
Veit Köppen, Otto von Guericke University Magdeburg
Axel Korthaus, Queensland University of Technology
Thomas Koslowski, University of Freiburg
Panagiotis Kourouthanassis, Athens University of Economics and Business
Tim Kraemer, Goethe University Frankfurt
Nane Kratzke, University of Applied Sciences Lübeck
Zuzana Kristekova, TU Munich

Ioannis Krontiris, Goethe University Frankfurt
Lars Krüger, Otto von Guericke University Magdeburg
Uwe Kubach, SAP Research
Herbert Kuchen, University of Munster
Maurice Kuegler, EBS Business School
Josef Küng, University of Linz
Dennis Kundisch, University of Paderborn
Renata Kunstov, University of Economics Prague
Fabian Lang, Helmut Schmidt University Hamburg
Matthias Lange, McKinsey
Tobias Langkau, Georg-August University Goettingen
Sven Laumer, University of Bamberg
Nguyen-Thinh Le, TU Clausthal
Ulrike Lechner, Bundeswehr University Munich
Jong Seok Lee, Georgia State University
Christine Legner, EBS Business School
Franz Lehner, University of Passau
Jan Leimeister, University of Kassel
Uwe Leimstoll, University of Applied Sciences Northwestern Switzerland
Stefan Lessmann, University of Hamburg
Joerg Leukel, University of Hohenheim
Katrina Leyking, IDS Scheer
Philipp Liegl, TU Vienna
Markus Lilienthal, Goethe University Frankfurt
Aleck Lin, Australian National University
Claudia Loebbecke, University of Cologne
Jan Löhe, EBS Business School
Frank Loll, TU Clausthal
André Ludwig, University of Leipzig
Stephan Lukosch, Delft University of Technology
Wolfgang Maass, Furtwangen University
Ronald Maier, University of Innsbruck
Bernd Markscheffel, TU Ilmenau
Oliver Marschollek, Goethe University Frankfurt
Dirk Mattfeld, TU Braunschweig
Florian Matthes, TU Munich
Mauricio Matthesius, Sones
Andreas Mattig, University of St. Gallen
Athanasios Mazarakis, FZI Karlsruhe
Jochen Meis, Fraunhofer ISST
Martin Meissner, University of Bielefeld
Werner Mellis, University of Cologne
Jan Mendling, Humbold University Berlin
Severino Meregalli, SDA Bocconi
Dieter Merkl, TU Vienna
Christian Messerschmidt, Goethe University Frankfurt
Manfred Meyer, Universiy of Applied Sciences Gelsenkirchen
Florian Michahelles, ETH Zurich
Anton Michlmayr, Mercatis
André Miede, TU Darmstadt
Martin Mikusz, University of Stuttgart
Claudia Müller-Birn, Carnegie Mellon University
Günther Müller-Luschnat, Pharmatechnik
Martin Mocker, Reutlingen University
Felix Moedritscher, Vienna University of Economics and Business
Lars Moench, University of Hagen
Dominik Molitor, LMU Munich
Daniel Möller, EBS Business School
Jörgen Moormann, Frankfurt School of Finance and Management
Kathrin Möslein, Friedrich Alexander University Erlangen-Nürnberg
Juergen Mueller, Hasso Plattner Institute at the University of Potsdam
Benjamin Müller, EBS Business School
Jörg Müller, TU Clausthal
Jan Muntermann, Goethe University Frankfurt
Bela Mutschler, University of Applied Sciences Ravensburg-Weingarten
Bernhard Nett, University of Siegen
Dirk Neuhaus, HSF
Markus Neumann, Leibniz University Hannover
Daniela Nicklas, University of Oldenburg
Björn Niehaves, University of Munster
Markus Nüttgens, University of Hamburg
Andreas Oberweis, Karlsruhe Institute of Technology
Sebastian Olbrich, Cundus
Hannes Olivier, TU Clausthal
Jan Ondrus, ESSEC Business School
Frank Ortmeier, University of Magdeburg
Erich Ortner, TU Darmstadt
Philipp Osl, University of St. Gallen
Boris Otto, University of St. Gallen
Sven Overhage, University of Augsburg
Barbara Paech, University of Heidelberg
Immanuel Pahlke, Goethe University Frankfurt
Wolfgang Palka, TU Munich
Dominik Papies, University of Hamburg
Jan Paralic, Technical University of Kosice
Susanne Patig, University of Berne
Jan Pawlowski, University of Jyväskylä
Gertraud Peinel, Frauenhofer FIT
Helmut Petritsch, SAP
Ilia Petrov, TU Darmstadt
Mathias Petsch, TU Ilmenau
Jella Pfeiffer, University of Mainz
Reinhard Pfliegl, GF Austria Tech
Maciej Piechocki, TU Freiberg
Volkmar Pipek, University of Siegen
Joachim Plumbaum, SAP
Margit Pohl, TU Vienna
Jens Pöppelbuß, University of Munster
Annika Pötter, University of Munster
Sören Preibusch, University of Cambridge
Torsten Priebe, Teradata
Andreas Prinz, University of Kassel

Wiedemann Proventa, Proventa
Thomas Puschmann, Direct Management Institute St.Gallen
Corina Raduescu, University of Sydney
Philip Raeth, EBS Business School
Leonardo Ramirez, Fraunhofer FIT
Kai Rannenberg, Goethe University Frankfurt
Andreas Rathgeber, University of Augsburg
Karl-Heinz Rau, Pforzheim University
Jan Recker, Queensland University Brisbane
Tim Reichling, University of Siegen
Ulrich Reimer, University of Applied Sciences St. Gallen
Andreas Reiser, University of Hohenheim
Christopher Rentrop, HTWG Konstanz
Jonas Repschlaeger, TU Berlin
Werner Retschitzegger, Johannes Kepler University Linz
Elmar Reucher, University of Hagen
Sebastian Richter, Bundeswehr University Munich
Alexander Richter, Bundeswehr University Munich
Matthias Riebisch, TU Ilmenau
Christoph Riedl, TU Munich
Rene Riedl, University of Linz
Kai Riemer, University of Sydney
Ryan Riordan, Karlsruhe Institute of Technology
Thomas Ritz, University of Applied Sciences Aachen
Susanne Robra-Bissantz, TU Braunschweig
Maximilian Röglinger, University of Augsburg
Friedrich Roithmayr, Johannes Kepler University Linz
Florian Rosenberg, CSIRO ICT Centre
Christoph Rosenkranz, Goethe University Frankfurt
Heiko Rossnagel, Fraunhofer IAO
Michael Röthlin, University of Applied Sciences Berne
David Rückel, Johannes Kepler University Linz
Roland Ruhwandl, Accenture
Maurice Ruiu, Daimler
Stefan Sackmann, University of Halle-Wittenberg
Thorben Sandner, Leibniz University Hannover
Gerhard Satzger, Karlsruhe Institute of Technology
Juergen Sauer, University of Oldenburg
Jörg Schäfer, University of Applied Sciences Frankfurt
André Schäfferling, German Graduate School of Management and Law
Matthieu Schapranow, Hasso Plattner Institute at the University of Potsdam
Roland Schegg, University of Applied Science Western Switzerland
Sven Scheibmayr, University of Mannheim
Stefan Schellhammer, University of Munster
Joachim Schelp, RTC
Michael Schermann, TU Munich
Christian Schieder, TU Chemnitz
Christian Schlaeger, Ernst&Young
Thomas Schlegel, University of Stuttgart
Johann Schlichter, TU Munich
Christian Schloegel, Wincor Nixdorf
Frank Schlosser, University of Bamberg
Daniel Schmalen, University of Trier
Klaus Schmid, University of Hildesheim
Alexander Schmidt, Universtity of St. Gallen
Roman Schmidt, Swiss Post Solutions
Florian Schnabel, University of St. Gallen
Michael Schneider, TU Kaiserlautern
Carsten Schoepp, University of Koblenz
Michael Scholz, University of Passau
Eric Schoop, TU Dresden
Hagen Schorcht, TU Ilmenau
Michael Schrefl, Johannes Kepler University Linz
Petra Schubert, Copenhagen Business School
Sarah Schuelke, University of Hagen
Christian Schulze, Goethe University Frankfurt
Matthias Schumann, Georg-August University Goettingen
Wieland Schwinger, Johannes Kepler University Linz
Martin Sedlmayr, Friedrich Alexander University Erlangen-Nürnberg
Stefan Seidel, University of Liechtenstein
Stefan Seifert, University of Karlsruhe
Juergen Seitz, DHBW Heidenheim
Bernd Simon, University of Economics and Business Vienna
Carlo Simon, Provaids School of International Management and Technology
Bernd Simon, Vienna University of Economics and Business
Alexander Simons, University of Liechtenstein
Matthias Söllner, University of Kassel
Stefan Smolnik, EBS Business School
Harry Sneed, AneCon
Stefan Sobernig, Vienna University of Economics and Business
Thorsten Spies, University of Duisburg-Essen
Thorsten Spitta, University of Bielefeld
Timm Sprenger, TU Munich
Shirish Srivastava, HEC Paris
Michael Stal, Siemens
Ioannis Stamelos, Aristotle University of Thessaloniki
Dirk Staskiewicz, University of Hohenheim
Douglas Steel, University of Houston
Stefan Stein, University of Koblenz
Hans-Peter Steinbacher, University of Applied Science Kustein
Gunnar Stevens, University of Siegen
Stefan Stieglitz, University of Munster
Steffen Stock, OPITZ CONSULTING Gummersbach
Stefan Stoeckl, University of Augsburg

Margareth Stoll, University of Innsbruck
Susanne Strahringer, TU Dresden
Steffen Strassburger, TU Ilmenau
Christine Strauss, University of Vienna
Stefan Strecker, University of Duisburg-Essen
Mark Strembeck, Vienna University of Economics and Business
Jens Strüker, University of Freiburg
Christian Suchan, University of Bamberg
Eldar Sultanow, Producto
Alfred Taudes, Vienna University of Economics and Business
Gerrit Tamm, SRH University of Applied Sciences
Stefan Tams , Clemson University
Barney Tan, University of Sydney
Wolfgang Theilmann, SAP
Frederic Thiesse, University of St. Gallen
Christof Thim, University of Potsdam
Oliver Thomas, University of Osnabrueck
A Min Tjoa, TU Vienna
Kowatsch Tobias, University of St. Gallen
Huy Tran, TU Vienna
Horst Treiblmaier, Vienna University of Economics and Business
Matthias Trier, University of Amsterdam
Sascha Uelpenich, Evosoft
Axel Uhl, University of Applied Sciences Northwestern Switzerland
Rainer Unland , University of Duisburg-Essen
Nils Urbach, EBS Business School
Frank van der Linden, Philips
Uwe van Heesch, University of Groningen
Daniel Veit, University of Mannheim
Rudolf Vetschera, University of Vienna
Johannes Vetter, LMU Munich
Bernhard Volz, University of Bayreuth
Jan vom Brocke, University of Liechtenstein
Hans-Jörg von Mettenheim, Leibniz University Hannover
Alexander von Stetten, University of Bamberg
Stefan Voss, University of Hamburg
Jens Vykoukal, Goethe University Frankfurt
Johannes Watzl, LMU Munich
Ingo Weber, University of New South Wales
Barbara Weber, University of Innsbruck
Edzard Weber, Universität Potsdam
Christian Weber, Goethe University Frankfurt
Rene Wegener, University of Kassel
Alexander Wehrmann, Senacor Technologies
Matthias Weidlich, HPI Potsdam
Christof Weinhardt, Karlsruhe Institute of Technology
Edgar Weippl, TU Vienna
Christof Weinhardt, TU Karlsruhe
Anette Weisbecker, Fraunhofer IAO
Michael Weiss, Carleton University
Ruediger Weissbach, Hamburg University of Applied Sciences
Oliver Wendt, TU Kaiserslautern
Birgit Wenke, Bundeswehr University Munich
Stefan Wesner, University of Stuttgart
Felix Wex, University of Freiburg
Thomas Wieland, HS Coburg
Martin Wiener, Friedrich Alexander University Erlangen-Nürnberg
Stephan Wildner, University of Passau
Andreas Will TU Ilmenau
Manuel, Wimmer TU Vienna
Axel Winkelmann, University of Munster
Sven Wohlgemuth, National Institute of Informatics
Robert Woitsch, BOC
Ryan Wright, University of San Francisco
Jochen Wulf, TU Berlin
Michael Wurst, IBM
Xue Yang, Nanjing Unviersity
Fahri Yetim, University of Siegen
Panagiotis Zaharias, Athens University of Economics and Business
Marco Zapletan, TU Vienna
Novica Zarvic, University of Osnabrück
Uwe Zdun, TU Vienna
Alexander Zeier, University of Potsdam
Jan Zibuschka, Fraunhofer IAO
Frank Zickert, Goethe University Frankfurt
Jürgen Ziegler, University of Duisburg-Essen
Holger Ziekow, Humbold University Berlin
Hans-Dieter Zimmermann, University of Applied Sciences St. Gallen
Steffen Zimmermann, University of Innsbruck
Andreas Zolnowski, ISS Hamburg

Acknowledgments

The 10th International Conference on Wirtschaftsinformatik would have been impossible without the generous support of the following sponsors:

Gold sponsors:

avaloq
essential for banking

IBM

SAP

Bronze sponsors:

HILTI

Schweizer Informatik Gesellschaft
Société Suisse d'Informatique
Società Svizzera per l'Informatica
Swiss Informatics Society

Track 4: Development/Architecture of IS

Gerti Kappel
Business Informatics Groups
TU Vienna
gerti@big.tuwien.ac.at

Gustaf Neumann
Institute for Information Systems
and New Media, WU Vienna
gustaf.neumann@wu.ac.at

Editorial

In most of today's business processes, information systems (IS) play a crucial role. Thereby, information systems have emerged from isolated, company-internal solutions to distributed and open systems. At the same time the business significance of information systems has constantly increased. The responsibility in regard to mission critical decisions such as the availability or performance of information systems has moved from the level of IT departments to the executive level. Furthermore, information systems are no longer regarded as support technologies, but as mission critical assets. This requires a higher flexibility of the different IT systems in order to be able to realize business process related decisions faster and demands as well for higher transparency and traceability, required for IT Governance and Controlling. In particular the combination of methods and systems is crucial in order to deliver high-quality, reliable, and effective information systems.

This track focuses on five topics, namely Software Development Processes (4 papers), IT Support for Business Services (3 papers), Business Process Management (4 papers), IS Architectures (7 papers), and Reusability and Integration (2 paper).

The overall quality of the track's submissions (67 papers) has been very high. Based on at least three reviews per paper, one meta-review and a detailed discussion among the conference chairs, the track chairs and the associate editors, the program committee selected finally the 20 best papers, giving an acceptance rate of less than 30%.

Many thanks to the 19 associate editors and the 88 reviewers for the excellent work.

Software Development Processes

Understanding Acceptance of Information System Development and Management Methodologies by Actual Users: A Review and Assessment of Existing Literature (Kunal Mohan, Frederik Ahlemann)

How Agile Are You Thinking? – An Exploratory Case Study (Roy Wendler, Andre Gräning)

Analysis of Two Theoretical Perspectives on Information Systems Development: Towards an Integrated Perspective (Frank Zickert)

A Longitudinal, Multi-Project Study of Bug Tracking Productivity and Learning in Open Source Software Development (Janina Matz, Arne Beckhaus, Dierk Erdmann, Peter Buxmann)

IT Support for Business Services

A Cost-Benefit-Based Analytical Model for Finding the Optimal Offering of Software Services (Jörn Altmann, Khin Swe Latt)

Towards an Ontological Foundation of Services Science: The General Service Model (Roberta Ferrario, Nicola Guarino, Christian Janiesch, Tom Kiemes, Daniel Oberle, Florian Probst)

The BabelNEG System - A prototype Infrastructure for protocol-generic SLA Negotiations (Sebastian Hudert, Torsten Eymann)

Business Process Management

Requirements of Process Modeling Languages – Results from an Empirical Investigation (Susanne Patig, Vanessa Casanova-Brito)

Usability of Modelling Languages for Model Interpretation: An Empirical Research Report (Christian Schalles, Michael Rebstock, John Creagh)

On Measures of Behavioral Distance between Business Processes (Jörg Becker, Philipp Bergener, Dominic Breuker, Michael Räckers)

Sind Ereignisgesteuerte Prozessketten besser für Fachanwender geeignet als UML Aktivitätsdiagramme? Eine empirische Untersuchung (Sven Overhage, Sebastian Schlauderer, Dominik Birkmeier)

IS Architectures

Integrating Innovation into Enterprise Architecture Management (Michael Rohloff)

Goal-oriented requirements modeling as a means to address stakeholder-related issues in EA (Carsten Lucke, Ulrike Lechner)

Towards the Analysis of Flexibility of Information Systems: a Proposition of a Method (Daniel Wagner, Christian Suchan, Benjamin Leunig, Jochen Frank)

Construction and Evaluation of a Meta-Model for Enterprise Architecture Design Principles (Stephan Aier, Christian Fischer, Robert Winter)

Understanding Enterprise Architecture Management Design – An Empirical Analysis (Bettina Gleichauf, Stephan Aier, Robert Winter)

Prioritizing Strategic IT Projects with Tropos (Andreas Gehlert)

Reinventing the Wheel?! Why Harmonization and Reuse Fail in Complex Data Warehouse Environments and a Proposed Solution to the Problem (Torsten Priebe, Andreas Reisser, Duong Hoang)

Reusability and Integration

Similarity Determination in Activity Sequences - A Supportive Framework (Jörg Schmidl, Holger Wittges, Helmut Krcmar)

A portal-based approach for user-centric legacy application integration in collaborative environments (Oliver Gmelch, Günther Pernul)

Construction and Evaluation of a Meta-Model for Enterprise Architecture Design Principles

Stephan Aier
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 3360
stephan.aier@unisg.ch

Christian Fischer
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 3877
christian.fischer@unisg.ch

Robert Winter
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 2190
robert.winter@unisg.ch

ABSTRACT

Based on an IEEE definition and its adaptation by The Open Group enterprise architecture (EA) is often defined as the fundamental organization of a company or government agency and the principles governing its design and evolution [16, 25]. While design representation issues like meta-modeling or notations have been discussed in EA literature [3], design activity issues and design principles in particular are often neglected. This is surprising because EA principles play an important role in practice. As a contribution towards a clear definition of EA principles, we analyze state-of-the-art EA principle definitions first. Based on this state-of-the-art analysis already discussed in [11] the original contribution of this paper is the construction and evaluation of a meta-model defining EA design principle as a second step. Our proposal differentiates a core definition of EA principle dealing with a principle itself (statement, rationale, implications, key actions, and measures) and an extended definition taking the use and impact of an EA principle in its environment into account. Important elements of an EA principle's environment are corporate strategy, the constructional view on EA transformation projects, and EA itself including its layers and its dimensions in time like as-is and to-be EA. We evaluate our meta-model in two case studies. Our consolidated meta-model provides the basis to analyze phenomena of actual EA principles in practice and thus to uncover the latent structures of EA principle taxonomies.

Keywords

enterprise architecture, principle, meta-model.

1. INTRODUCTION

One of the most often cited publications for defining Enterprise Architecture (EA) is that of the IEEE standard 1471-2000 [16] and its adaptation to EA by The Open Group [25]. Architecture is defined there as (1) "[t]he fundamental organization of a system embodied in its components, their relationships to each other, and to the environment", and as (2) "the principles guiding its design and evolution" [16]. In the field of EA, 'system' is then substantiated by The Open Group as an enterprise meaning "any collection of organizations that has a common set of goals" e.g. a company or government agency [25]. The (1) *fundamental organization* of a system is often represented by models of the as-is state or the to-be state of a system. For these purposes, meta-models, methods, and frameworks have been developed and extensively discussed in literature [22, 23]. However, (2) *activities*, *rules*, and particularly *principles* guiding an architecture's design and evolution from an as-is state into a to-be state are often neglected and thus are hardly covered in literature.[1] Stelzer's [24] review of EA literature identifies only six publications that specifically address EA design principles.

In practice, many companies' EA departments formulate principles and some review projects based on these principles.[2] For this purpose, documentation and communication of EA principles is essential. The fundament for such documentation is a clear definition of the principle's structure and of its relations to its environment. However, our practitioner interviews in a non-representative sample, aiming at a deeper understanding of EA principles' use in practice shows that except for a few cases EA principles are unequally and only selectively defined (regarding the scope of EA) and that their impact varies significantly.

Our in-depth analysis of different notions of EA principles, from scientific as well as from practitioner's literature [11], reveals that there is no consensus on a definition of the term *EA principle*. The aim of this research is therefore to analyze these different notions of EA principle in order to derive a consolidated understanding. Thus, this paper aims at defining and evaluating a construct which forms the vocabulary of a domain [19].

The paper at hand is structured as follows: In section two, different notions of EA principle are analyzed. In section three, these notions are discussed and consolidated into a meta-model for EA principle. In section four, we use two case studies to evaluate the

[1] Exceptions to this generalized observation are [11, 24] as far as EA *rules* and *principles* are concerned as well as [1, 2, 7].as far as *activities* guiding an architecture's design and evolution from an as-is state into a to-be state are concerned.

[2] Cf. for instance the Open Group's architecture compliance review method proposed in TOGAF 9 [25].

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

proposed meta-model. In section five, we discuss our contribution and give an outlook on further research.

2. RELATED WORK

In this section, we analyze related work dealing with EA principles in order to develop a consolidated understanding of what an *EA principle* is.[3] Therefore we analyze six approaches with respect to their understanding of EA principle. The selection of the papers analyzed is based on Stelzer's [24] literature review. He selects relevant literature by applying Weber and Watson's [26] guidelines: (1) IS journals and conference proceedings are analyzed using the search term: *"enterprise architecture" AND ("principle" OR "design" OR "rule" OR "guideline")*. (2) Stelzer extends his research to further sources and ensures that all top 20 IS journals and the top IS conferences (e.g. ICIS, AMCIS, ECIS, HICCS, and Wirtschaftsinformatik) are included. In total, 42 relevant articles are identified. (3) Each of these articles is analyzed in detail. Based upon this analysis, 27 articles are excluded. (4) The citations of the remaining 15 articles are analyzed; this way, four further articles are added. (5) These 19 articles are analyzed in detail. Articles from related research areas such as software engineering, organizational design, and engineering are excluded. Principles for designing or evaluating architecture frameworks and principles for service oriented architectures are excluded, too.

As a result of his analysis Stelzer identifies eleven articles on EA principles. His analysis differentiates EA *design* principles from EA *representation* principles. EA design principles refer to the design of EA while EA representation principles refer to its representation. Lindström [18] makes a similar distinction by differentiating syntactic (i.e. representation) and semantic (i.e. design) principles. Examples for representation (or syntactic) principles are understandability, consistency, and unambiguousity [18, 24]. As EA representation principles are out of scope of this publication, we exclude all papers that solely refer to EA representation principles.

Table 1: EA design principles according to Stelzer [24]

Reference	Method	Principle definition
Richardson, 1990 [21]	case study	"Principles are an organization's basic philosophies that guide the development of the architecture. … Principles provide guidelines and rationales for the constant examination and re-evaluation of technology plans." (p. 389)
Armour, 1999 [5]	conceptual	"… simple, direct statements of how an enterprise wants to use IT. These statements establish a context for architecture design decisions by translating business criteria into language and specifications that technology managers can understand and use. Architecture principles put boundaries around decisions about system architecture." (p. 38)
Hoogervorst, 2004 [14]	conceptual	*no explicit definition,* "collectively the design principles are identified as enterprise architecture" (p. 217)
Chen, 2004 [8]	conceptual	"Architecting principles are rules to use when elaborating enterprise architectures." (p. 1214)
Wilkonson, 2006 [27]	case study	*no explicit definition*
Lindström, 2006 [18]	case study	"Architectural principles define the underlying general rules and guidelines for the use and deployment of all IT resources and assets across the enterprise …" (p. 2)

The characteristics of the six remaining articles are summarized in Table 1 and analyzed in the following subsections. The reconstructed meta-models of the analyzed articles have been adopted from [11].

We have verified Stelzer's literature review and found it to hold very well. However, in some cases we have added further literature by the same author or school in order to further clarify the respective position.

2.1 Richardson et al.,1990 [21]

Richardson et al. [21] document EA principles which they have extracted from a case study of Star Enterprise. The principles are attributed to different layers: organization, applications, data, and infrastructure.

For each principle, Star Enterprise documents (1) the principle itself, (2) a rationale explaining how the principal is assumed to work, and (3) concrete implications (Figure 1).

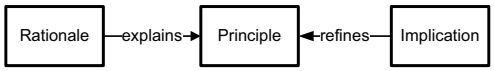


Figure 1: Meta-model of *EA principles* according to Richardson et al. [21]

2.2 Armour et al., 1999 [5]

Armour et al. [5] take a "big picture look at enterprise architectures" [5] from a practitioner's perspective and mainly develop an EA framework.

For this framework, they propose five views: (1) business view, (2) work view, (3) function view, (4) information view, and (5) infrastructure view. The framework "begins with a business vision—including the IT vision—which determines IT goals and objectives. Together, the business and IT visions drive the business view and architecture principles. […] To provide the structure and guidelines for EITA [enterprise information technology architecture] development, most frameworks will include a set of architectural principles, architectural views, a technical reference model, and a standards profile" [5]. Standards and technical reference model are meant to "make sure everyone has a common understanding of function and term" [5]. The meta-model of the principle definition by Armour et al. is illustrated in Figure 2.

[3] The analysis itself has already been published in [11]. The analysis in [11] provides the foundation for the artifact design which is an original contribution of the paper at hand.

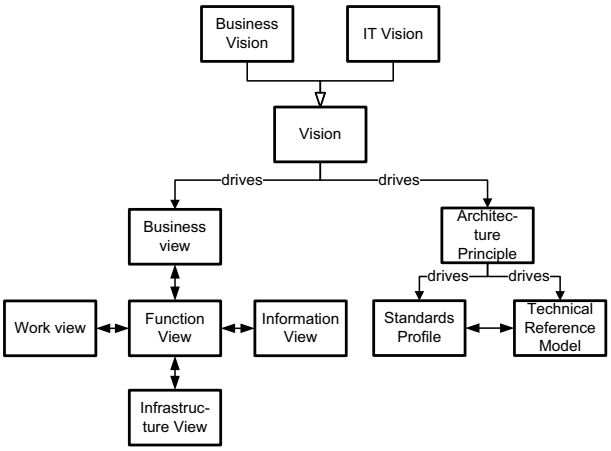


Figure 2: Meta-model of *EA principle* according to Armour et al. (1999) [5][4]

2.3 Hoogervorst, 2004 [14]

Hoogervorst [14, 15] understands architecture solely as a prescriptive concept comprising "a set of design principles and standards that guide design" [14]. In accordance with Dietz [10], Hoogervorst argues that architecture normatively restricts design freedom. For Hoogervorst [14, 15] and Dietz [10], EA is hence limited to the second part of the architecture definition by IEEE Std. 1471-2000 [16], i.e. principles governing the architecture's design and evolution; they explicitly exclude its first part, i.e. representations of "the fundamental organization of a system". Hoogervorst's understanding of EA principles is illustrated in Figure 3.

Also in accordance with Dietz [9, 10], Hoogervorst [14, 15] differentiates a functional view and a constructional view on an enterprise. Whilst the functional view (teleological view, black box view) deals with the purpose or goal of a system, the constructional view (ontological view, white box view) is about how the system's functions are brought to life [9]. For Hoogervorst [14], design principles refer to the constructional view. In contrast, requirements refer to the functional view on a system [15].

Hoogervorst [14] differentiates between four types of architecture: (1) business architecture, (2) organizational architecture, (3) information architecture, and (4) technology architecture. For each type, he proposes an architecture framework highlighting the main areas of the respective architecture type.[5] Each of these architecture types contains "a logically consistent and coherent set of principles and standards that guide" [14]

- "how a particular field of (commercial) endeavor will be exploited and explored" [14] (business architecture),
- "how the purposeful activities are to be organized" [14] (organization architecture), and
- "how information is to be managed" [14] (information architecture).

Besides the principle statement, Hoogervorst [15] claims for documenting its rationale(s), its implication(s) and its key action(s). The *rationale* says why the principle is defined. The *implication* states how relevant system stakeholders are affected by the principle. The definition of *key actions for effectuating the architecture* is implied by the fact that not all architecture principles can be applied immediately, but can only be used under certain conditions. The key actions ensure these conditions, such that the architecture principles can be followed" [15].

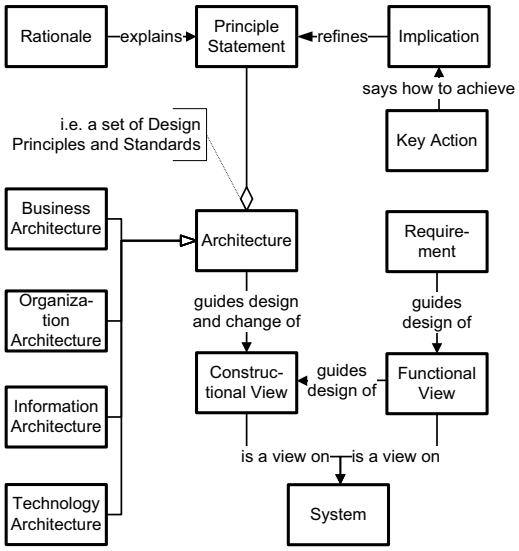


Figure 3: Meta-model of *EA principle* according to Hoogervorst (2004, 2009) [14, 15]

2.4 Chen & Lillehagen, 2004 [8]

Chen & Lillehagen [8] review literature and reveal the different authors' understanding of *architecture* and *architecture principles* in particular. Their literature review is mostly based upon practitioner sources like homepages of consultancy companies. Chen & Lillehagen differentiate between *generic* EA principles, i.e. principles that "apply to all enterprises" [8], and *specific* principles "reflecting a level of consensus among the various elements of a particular enterprise, and form[ing] the basis for making future decisions" [8]. They point out that EA principles are meant to facilitate architecture decisions.

Chen & Lillehagen [8] do not explicate a clear definition of components of EA principle. We therefore cannot derive a meta-model from this particular source.

2.5 Wilkinson (2006) [27]

Wilkinson [27] has been Chief Technology Officer at Hewlett Packard (HP) and reports on his experiences at HP.

For him, it is important for enterprises (1) to understand what and how IT is being used and to get control of existing IT assets (stability), (2) to leverage best practice and automation of aspects of IT processes (efficiency), and (3) to align IT governance and business strategy such that IT can rapidly react on business changes (agility). According to Wilkinson, architecture principles and IT governance are a means for realizing an adaptive enter-

[4] The double-arrows "←→" indicate an interdependency between the two entities concerned.

[5] For instance, the business architecture comprises principles concerning the enterprise's mission, its strategy, its market, its competitors, its product services, its key resources, its operating method(s), its economic and revenue model, its customers, its stakeholders, and its environment [14].

prise. In an ideal world, IT governance and IT strategy are connected to corporate strategy. Different frameworks such as ITIL, ITSM, or COBIT help implementing IT governance.

Wilkinson names two main areas for implementing an adaptive enterprise: IT organization and technology. IT organization (1) should focus on innovation in order to support business and (2) should be optimized in order to save costs. A project management office can help realizing these goals by assuring the conformity of projects to corporate strategy. On the technology layer, an adaptive infrastructure should be aimed at.

Wilkinson describes some EA principles at HP although he does not explicitly call them "principle": *modularity*, *simplification*, *integration*, and *standardization*. He does not explicate a definition of what a principle is and what it is composed of. Nevertheless, we tried to reconstruct Wilkinson's notion of EA principle in the meta-model illustrated in Figure 4.

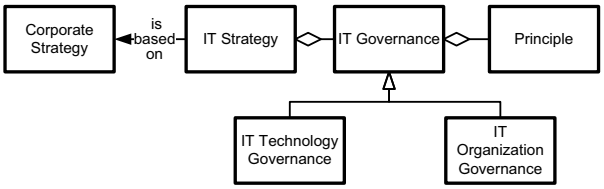


Figure 4: Meta-model of EA principle according to Wilkinson [27]

2.6 Lindström, 2006 [18]

Lindström [18] reviews literature on EA principles. "Principles respresents [sic!] a shared understanding on what needs to happen if the organization is to successfully execute the strategies" [18]. For Lindström, architectural principles are important for the transition of today's architecture to the desired target architecture. This transition is driven by business strategy and business principles. Architectural principles are a tool for supporting this transition process. Therein, "architectural principles can justify architecture activities by showing the rationale for the investment" [18].

Referring to Broadbent et al. [6], Lindström [18] states that IT strategy is based on IT governance, that IT governance is based on architectural principles, that architectural principles are based on business principles, and that business principles are based on business strategy. Business strategy "tells us how an organization is going to compete in a chosen market" [18].

She mainly describes an architectural review of EA principles at Vattenfall in a case study. For this purpose, she defines syntactic and semantic characteristics of good principles. As syntactic quality criteria, she names *consistency*, *verifiability*, *unambiguousity*, and *modifiability*; as semantic quality criteria she names *stability*, *verifiability*, *modifiability*, *correctness*, and *completeness*. Moreover, she recommends a syntax for architectural principles which is summarized in Table 2.

Table 2: Components of EA principle according to Lindström [18]

Name	Definition	Example
Statement	What to improve	IT system's fit to business
Motivation	Why this is important	Increase the effectiveness in the business organization
Implication	What must be done and when, and who is responsible	Investigate the influence on the business processes when a new system is acquired. The project manager is responsible.
Measure	How the fulfillment of the principles is measured. Both for long-term and short-term, e.g. after an investment.	Time to perform a business process

In Figure 5, her notion of EA principle is illustrated as a meta-model.

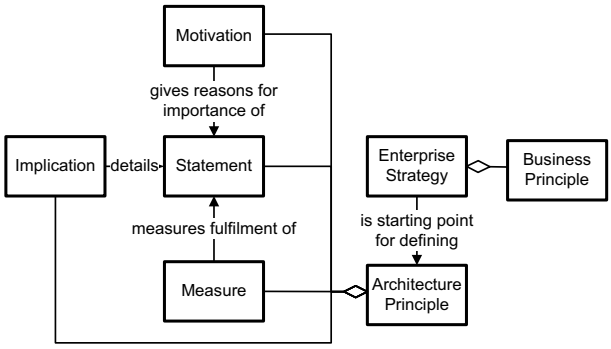


Figure 5: Meta-model of EA principle according to Lindström [18]

3. A CONSOLIDATED EA PRINCIPLE DEFINITION

The discussion in section 2 shows that authors hold different views on what an EA principle is. Nevertheless, their definitions have several aspects in common. In this section, we aim at uncovering these communalities and constructing a consolidated meta-model.

Our analysis also shows that except Richardson et al. [21], authors mix definitions of an EA principle itself with definitions of an EA principle in its environment. For reasons of transparency we will first consolidate definitions of an EA principle itself (*core definition*, section 3.1) and will then extend the core definition by covering the impact that it has on its environment (*extended definition*, section 3.2).

3.1 Core Definition

Richardson et al. [21] mention (1) a rationale explaining how the principle is meant to work and (2) implications that it has to the enterprise. Hoogervorst [14, 15] re-uses the components defined by Richardson et al [21] and adds key actions, i.e. concrete guidelines for implementing the principle. Armour et al. [5] proposes another way to refine architectural principles. As far as the standards profile is concerned, we prefer the more extensive proposition by Richardson et al. [21] and Hoogervorst [14, 15]. Furthermore Hoogervorst [14, 15] as well as Lindström [18] introduce the principle statement. Their description implies that the statement is part of the principle. Lindström [18] also mentions measures as an important part of an EA principle in order to be able to evaluate a principle's efficacy, thus the fulfillment of the statement, and finally to support the process of managing (introducing,

evaluating, changing, and revoking) EA principles. Figure 6 illustrates the core definition of EA principle in a meta-model.

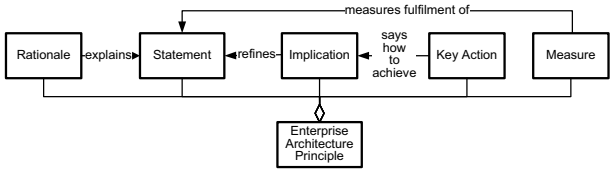


Figure 6: Core meta-model of EA principle

3.2 Extended Definition

To further specify the nature of an EA principle, it is helpful to understand an EA principle's impact on its environment. We will therefore extend our core definition based on our literature analysis as well as on results from focus groups. Our meta-model construction for the extended definition of EA principle is based on a two-step process. In step one we build the basic structure of the extended meta-model. In step two this basic structure is refined based on the different foci taken by the authors of the analyzed related work.

3.2.1 Step 1: Basic Extensions

As with every design the question of where to begin is a difficult one. While we have analyzed various related work none of these contributions provided a substantial starting point since the respective contributions either follow a general but incomplete or a very specific but unbalanced approach. Therefore we have chosen to start with an exploratory focus group [13, 20]. The aim of the focus group has been to identify the main products of EA as well as their relationships. The rationale behind this approach is that EA principles should have a prominent role in the development and advancement of EA. Thus understanding the products of EA as well as its relations should provide a valuable understanding of EA principles' environment from an EA perspective.

The participants of the focus group are practitioners that are experts in the field of EA and have experience managing and using EA principles. The focus group has nine participants from seven different companies from Germany and Switzerland plus the focus group's moderator.

As a result the focus group identified five main products of EA (*EA transparency*, *EA guidelines and principles*, *EA plans*, *EA implementation*, and *EA education*) as well as their respective components. The focus group has also been asked to discuss the relations among the five main products. The analysis of these results focusing EA principles leads to the basic extended meta-model illustrated in Figure 7.

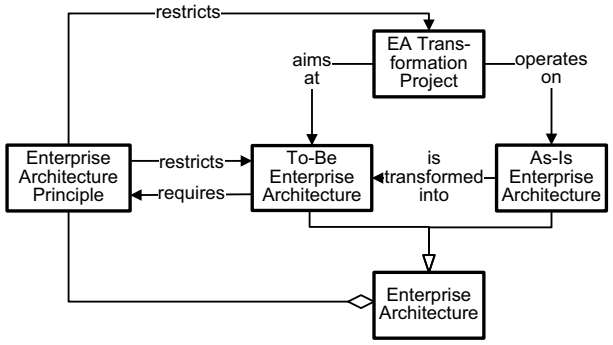


Figure 7: Basic EA principle meta-model extensions

In this basic extended meta-model, we have omitted EA education. While EA education is an important product of EA, it does not belong to the same level of abstraction as the other EA products. Furthermore we have substituted EA transparency by *as-is enterprise architecture* and EA plans by *to-be enterprise architecture* for internal consistency of the model.[6] The substitution of EA transparency by as-is EA is valid since the focus group members discussed transparency as the transparency of the current EA. The substitution of EA plans by to-be EA is valid since the focus group members discussed EA plans as the future state of EA resulting from EA implementation transforming the current as-is state. This argumentation also accounts for the relations between *EA transformation project*, *to-be EA* and *as-is EA*. A defined to-be EA requires for certain *EA principles* in order to be achieved by EA transformation projects. Therefore EA principles also have to restrict an EA transformation project's freedom of choosing a design in order to ensure a development towards the to-be EA. Finally there also may be EA principles that restrict the possible to-be EAs.

3.2.2 Step 2: Refinement of Basic Extensions

With the background of the basic extended meta-model resulting from the focus group, we aim at consolidating the different definitions of EA principle in the analyzed publications.

Armour et al. [5] introduce the notion of a model of a system. In order to cover the differences between an enterprise and its model, we have fundamentally extended the basic meta-model and put the enterprise into its centre. We differentiate between an existing enterprise and a desirable enterprise. Both, existing and desirable enterprise, are possible enterprises, being part of possible future worlds.[7] A special desirable enterprise is the enterprise that has been defined as a target enterprise for an EA transformation project. An EA transformation project begins to operate on the existing enterprise while aiming at changing it such that the target enterprise is realized.

EA principles give advice how to design target architecture by restricting the design freedom of EA transformation projects [9, 10, 14, 15]. In contrast to business requirements referring to the functional view of projects[8], architecture principles refer to the constructional view of the project[9].

---

[6] A reviewer of an earlier version of this paper has remarked that the common differentiation in *as-is, intermediate*, and *target architecture (vision)* [1, 7, 17] might be appropriate here. Following this terminology our to-be EA is equivalent to *intermediate architecture* because these architectures are actually realized and restricted by EA principles. The *target architecture* may influence EA principle definition but it will most probably never be achieved. Instead it will constantly be reformulated and/or become the next *intermediate architecture*.

[7] Cf. Frank [12] for further information on possible worlds in IS; for the term *world* cf. Wittgenstein [29].

[8] Functional view on projects: "What functionality of the enterprise does the project change?"

[9] Constructional view on projects: "How must the elements be changed that provide the enterprise's functionality?"

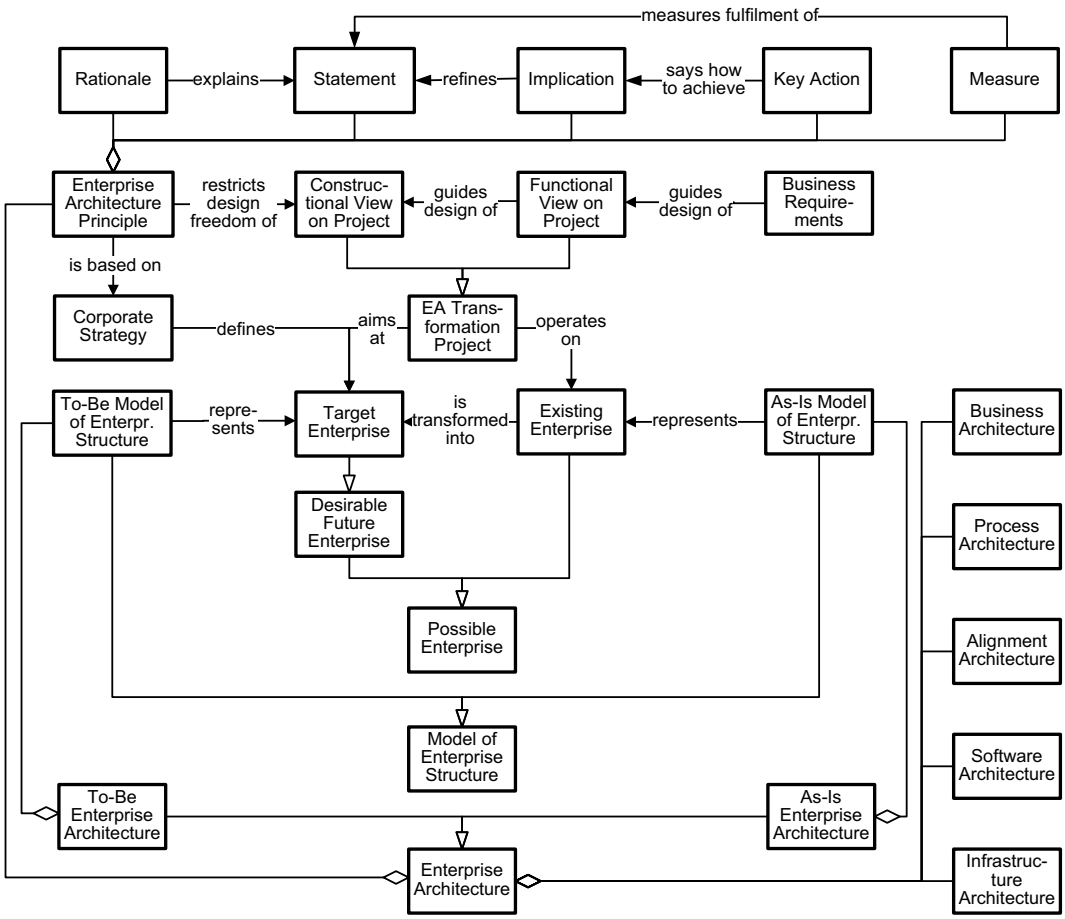


Figure 8. Consolidated meta-model of EA principle

Different authors propose different architecture layers. Whilst Hoogervorst [14, 15] proposes a business architecture, an organization architecture, an information architecture, and a technology architecture, Armour et al. [5] propose a business view, a function view, a work view, an information view, and an infrastructure view. Winter and Fischer [28] consolidated a variety of EA frameworks and identified four essential architectures: a business architecture, a process architecture, an alignment architecture,[10] a software architecture, and an infrastructure architecture. As Fischer and Winter [28] already consolidated the understanding of different layers, we adopt their proposition in our meta-model.

The structure of an enterprise can be represented in a model: the existing enterprise in an as-is EA model, the target enterprise in a to-be EA model. In accordance with the architecture definition in the IEEE Std. 1471-2000 [16], both the models of the enterprise structure and the architecture principles form the architecture.

As the main input for an EA principle Armour as well as Wilkinson [27] and Lindström [18] highlight the influence of corporate strategy. Figure 8 illustrates our consolidated meta-model of EA principles.

4. META-MODEL EVALUATION

In the following subsections, the case studies of two companies are presented in order to evaluate the applicability of our proposed meta-model in practice. These companies have introduced EA principles for guiding evolution of EA several years ago. Data for the case studies have been collected in interviews with representatives. In accordance with Yin [30], our case study is based upon different data sources. We have not only analyzed internal documents, but also, as described by Hevner and Chatterjee [13], have performed a focus group workshop with representatives from IT management and EA management in order to gather additional information and to ensure the elimination of misunderstandings.

4.1 Selection of Evaluation Cases

Company A is a major transportation and logistics service provider. It offers both cargo and passenger transportation and provides rail infrastructure. A couple of years ago, the inauguration of a new CIO resulted in renewed architecture efforts including the creation of a corporate EA team. The EA team is complemented by domain architecture teams, which are changing their focus from a domain and software centered perspective to an EA pers-

[10] While Winter and Fischer [28] initially named the central layer *integration architecture* Aier and Winter [4] renamed it to *alignment architecture*.

pective. EA processes have been set up altering existing development processes to reflect architectural issues, e.g. by defining quality gates, which projects cannot surpass without fulfilling EA principles. This change in processes is fostered by a broad range of efforts to enhance EA attention, knowledge, and skills throughout the company. Therefore a broad training program, addressing architects as well as non-architects, was set up. In addition to that, further initiatives were set up. For example (1) EA communication has been advanced by an EA tool providing a broad set of EA artifacts in an easy-to-use web interface, (2) all information required to meet EA principles in the quality gates is available through a well-organized intranet web application.

Company B is an IT service provider for a large banking network. In its current form, the network is the result of several mergers of formerly independent, regional IT service providers. Every formerly independent company had its own, evolutionary grown banking solution. However, none of these solutions had a predominant position within the network. Therefore the network decided to implement a new and common system as their core banking solution. The development started in 2002 and was finished in 2005 for the time being. The new system design follows a service oriented paradigm in its alignment and software architecture in order to adapt and to consistently provide the implemented functionality to every partner. The business architecture design of company B follows the process reference model which has been defined for the banks belonging to the network. For alignment, software, and infrastructure architecture; strict principles are defined. These EA principles are enforced through tools, repositories, and processes (e.g. for release management) which are the basis of company B's development. Because of this highly structured and tool supported processes, any development outside this environment is almost impossible and thus non-existent. At the same time company B has no explicit EA roles except for an EA board. Instead EA principles and their enforcement are built into the highly standardized change and production processes.

We have chosen these two particular companies because their culture of defining and enforcing EA principles is very different. While company A follows a very participative, grass roots democracy like approach, company B follows a very strict top-down driven approach – resulting from the necessity to efficiently and effectively manage several post-merger integration scenarios.

None of the two companies have participated in the focus groups used for constructing the extended EA principle definition.

4.2 Company A

The EA division of company A's information technology (IT) department defines several architecture design principles. Before such principles have been defined, most architectural decisions were taken ad-hoc. As a consequence, (1) architecture decisions of different projects were inconsistent and (2) architecture decisions were often intensively discussed, took a long time, and bound many resources.

In order to overcome these shortcomings, company A has defined a set of EA design principles. These principles are formulated [statement][11] such that they correspond to corporate strategy [is based on]. By means of concrete guidelines [implications/key actions], the principles are refined. Both principles and guidelines are intended to guide architectural decisions in projects [restricts design freedom of].

Every employee of company A is allowed to propose an architectural principle or a guideline. Therefore the principle or guideline needs to be well founded [rationale]. An architectural board elaborates theses proposals, declares proposals to be valid principles, and revises them, based upon the experience and feedback in projects. If a principle does not lead to the desired effects, it is revoked by the architectural board. Therefore the effectiveness of a principle is measured [measure].

All principles are available in the company's intranet and all projects are obliged to respect them when taking an architecture decision. Projects at company A are based upon the waterfall model and are structured in six phases. After each phase, projects must pass a quality gate. In each of these quality gates, the quality gate committee evaluates whether the principles and guidelines are respected. Thus company A successfully ensures that projects having impact on EA transform EA towards a defined to-be state [target enterprise].

4.3 Company B

In company B, EA principles focus on alignment, software, and infrastructure architecture only. Business and process architecture are based on a reference model which is defined outside of company B [EA layers]. EA principles are documented [statement] in the tools, repositories, and workflows which implement the standardized project procedures. These workflows clearly advise how to perform certain development tasks and thus how to observe EA principles [implications/key actions]. In order to foster these's workflows' acceptance, it is also explained why the respective steps are necessary [rationale]. As the entire company is driven by performance figures, the principles implemented in workflows are evaluated on a regular basis [measure].

In this case, the individual banks are the owners and customers of company B that need their business requirements to be implemented by the common banking solution. The main strategic proposition of company B, however, is that company B can serve each and every of the network's banks (currently more than 400 banks with more than 10,000 branch offices) using the very same (maybe differently configured) banking solution. Therefore the banking solution needs to be highly standardized, but configurable. EA principles guiding the evolution of the banking solution must reflect this strategic proposition [is based on corporate strategy]. If company B fails in enforcing these principles, it loses its right to exist.

5. Discussion and Outlook

Apart from Chen & Lillehagen [8], all existing publications on EA principles allow for reconstructing their EA definition in a meta-model. The analysis shows that while authors focus different aspects of an EA principle definition, they do not contradict each other's definitions.

For the purpose of constructing a consolidated meta-model of EA principle, we have differentiated a core definition (dealing with the EA principle itself) and an extended definition (dealing with the impact of an EA principle on its environment).

The evaluation of our consolidated EA meta-model shows that despite the cultural differences of the cases illustrated above, our meta-model for EA principle serves both cases well. The terminology used in each case is specific to the respective company and therefore is not identical with the terminology used in our meta-

[11] The terms in square brackets refer to the respective meta-model element illustrated in Figure 8.

model. The structure of meta-model elements, however, fits the situation of each case very well. Our case studies have also shown the importance of EA principles in practice. EA principles serve to constantly guide the evolution of an EA definition to a to-be EA.

Our proposed meta-model of EA principle can serve as a basis to systematically analyze existing EA principles in practice. The aim of such analyses is to understand the latent structure of EA principles and to derive a corresponding taxonomy. Possible (hierarchies of) dimensions in such taxonomy could be degree of generality, architectural layers concerned, life cycle dependencies, stakeholders etc. Once this underlying structure of EA principle in practice is understood, it will be possible to more systematically construct EA principle instantiations for specific situations (i.e. comparable to reference models) or even for particular application cases.

6. REFERENCES

[1] Aier, S., Gleichauf, B. Application of Enterprise Models for Engineering Enterprise Transformation. *Enterprise Modelling And Information Systems Architectures,* 5, 1 (2010), 58–75.

[2] Aier, S., Gleichauf, B.: Applying Design Research Artifacts for Building Design Research Artifacts: A Process Model for Enterprise Architecture Planning. In: *Proceedings of the Global Perspectives on Design Science Research, 5th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2010)* (St. Gallen, 04.06.2010). Springer, 2010, 333–348.

[3] Aier, S., Riege, C., Winter, R. Unternehmensarchitektur – Literaturüberblick und Stand der Praxis. *Wirtschaftsinformatik,* 50, 4 (2008), 292–304.

[4] Aier, S., Winter, R. Virtual Decoupling for IT/Business Alignment – Conceptual Foundations, Architecture Design and Implementation Example. *Business & Information Systems Engineering,* 51, 2 (2009), 150–163.

[5] Armour, F.J., Kaisler, S.H., Liu, S.Y. A Big-Picture Look at Enterprise Architectures. *IEEE IT Professional,* 1, 1/2 (1999), 35–42.

[6] Broadbent, M., Kitzis, E.S. Interweaving business-driven IT strategy and execution: Four foundation factors. *Ivey Business Journal,* 69, 3 (2005), 1–6.

[7] Buckl, S., Ernst, A.M., Matthes, F., Schweda, C.M.: An Information Model for Landscape Management – Discussing Temporality Aspects. In: *Proceedings of the Service-Oriented Computing – ICSOC 2008 Workshops* (Berlin, 01.12.2008). Springer, 2009, 363–374.

[8] Chen, D., Lillehagen, F.: Enterprise Architectures – Review on Concepts, Principles and Approaches. In: *Proceedings of the 10th International Conference on Concurrent Engineering (ISPE CE 2004)* (Beijing, Tsinghua University Press, 2004, 1211–1216.

[9] Dietz, J.L.G. *Enterprise Ontology – Theory and Methodology.* Springer, Berlin, Heidelberg, 2006.

[10] Dietz, J.L.G. *Architecture. Building strategy into design.* Academic Service, The Hague, 2007.

[11] Fischer, C., Winter, R., Aier, S. What is an Enterprise Architecture Design Principle? Towards a Consolidated Definition. In: Lee, R. (ed.) *Computer and Information Science 2010*. 193–205. Springer, Berlin, Heidelberg, 2010.

[12] Frank, U. *Towards a Pluralistic Conception of Research Methods in Information Systems Research*. ICB Research Reports 7, Institut für Informatik und Wirtschaftsinformatik (ICB), Universität Duisburg-Essen, Essen, 2006.

[13] Hevner, A.R., Chatterjee, S. *Design Research in Information Systems: Theory and Practice.* Springer US, Dordrecht, Heidelberg, London, New York, 2010.

[14] Hoogervorst, J.A.P. Enterprise Architecture: Enabling Integration, Agility and Change. *International Journal of Cooperative Information Systems,* 13, 3 (2004), 213–233.

[15] Hoogervorst, J.A.P. *Enterprise Governance and Enterprise Engineering.* Springer, Berlin, 2009.

[16] IEEE IEEE Recommended Practice for Architectural Description of Software Intensive Systems (IEEE Std 1471-2000). (2000)

[17] Lankhorst, M. *Enterprise Architecture at Work: Modelling, Communication and Analysis.* Springer, Berlin et al., 2005.

[18] Lindström, Å.: On the Syntax and Semantics of Architectural Principles. In: *Proceedings of the 39th Annual Hawaii International Conference on Systems Sciences* (Los Alamitos, CA, USA, 04.01.2006). IEEE Computer Society, 2006.

[19] March, S.T., Smith, G.F. Design and Natural Science Research on Information Technology. *Decision Support Systems,* 15, 4 (1995), 251–266.

[20] Morgan, D.L. *Focus Groups as Qualitative Research.* Sage Publications, Thousand Oaks, 1997.

[21] Richardson, G.L., Jackson, B.M., Dickson, G.W. A Principle-Based Enterprise Architecture: Lessons From Texaco and Star Enterprise. *MIS Quarterly: Management Information Systems,* 14, 4 (1990), 285–403.

[22] Schelp, J., Winter, R.: Language Communities in Enterprise Architecture Research. In: *Proceedings of the Diversity in Design Science – 4th Conference on Design Science Research in Information Systems and Technologies (DESRIST2009)* (Philadelphia, PA, USA, 07.05.2009). ACM, 2009, 1–10.

[23] Schönherr, M.: Towards a Common Terminology in the Discipline of Enterprise Architecture. In: *Proceedings of the Service-Oriented Computing – ICSOC 2008 Workshops* (Sydney, 01.12.2008). Springer, 2009, 400–413.

[24] Stelzer, D.: Enterprise Architecture Principles: Literature Review and Research Directions. In: *Proceedings of the 4th Workshop on Trends in Enterprise Architecture Research (Pre-Proceedings)* (Stockholm, 23.11.2009). 2009, 21–35.

[25] The Open Group *TOGAF Version 9 – The Open Group Architecture Framework (TOGAF).* The Open Group, 2009.

[26] Webster, J., Watson, R.T. Analyzing the Past to prepare for the Future: Writing a Literature Review. *MIS Quarterly,* 26, 2 (2002), 13–23.

[27] Wilkinson, M. Designing an "Adaptive" Enterprise Architecture. *BT Technology Journal,* 24, 4 (2006), 81–92.

[28] Winter, R., Fischer, R. Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. *Journal of Enterprise Architecture,* 3, 2 (2007), 7–18.

[29] Wittgenstein, L. *Tractatus logico-philosophicus.* Routledge & Kegan Paul, London, 1981.

[30] Yin, R.K. *Case Study Research. Design and Methods.* Sage Publications, Thousand Oaks, London, New Delhi, 2003.

Understanding Enterprise Architecture Management Design – An Empirical Analysis

Stephan Aier
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 3360
stephan.aier@unisg.ch

Bettina Gleichauf
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 3812
bettina.gleichauf@unisg.ch

Robert Winter
Institute of Information Management
University of St.Gallen
Müller-Friedberg-Strasse 8
9000 St. Gallen, Switzerland
+41 (0) 71 224 2190
robert.winter@unisg.ch

ABSTRACT
Enterprise architecture management (EAM) is expected to provide business value by guiding the continuous development and transformation of an enterprise. Based on the approach we strive for constructing useful artifacts that guide the successful and situational design of EAM. In order to do so we argue for a thorough analysis of the design problem in advance. This is realized by a two-step survey conducted on EAM practices. The empirical analysis reveals eight determining design factors of EAM, a delineation of three different types of EAM design in the form of clusters as well as insight about the successfulness of the different types.

Keywords
Enterprise architecture management, design science research, problem analysis, empirical analysis

1. INTRODUCTION
Enterprise architecture (EA) describes the fundamental structures of a company (or government agency) and enables its transformation by bridging the gap between business and information technology (IT). Enterprise architecture management (EAM) is concerned with the establishment and continuous development of EA. As such, the notion of EAM goes beyond EA modeling and includes the management tasks of planning and controlling business change from an architectural perspective.

EAM is established in a growing number of companies and government agencies but there is only limited transparency about which conditions have a positive effect on the successful adoption of EAM as a holistic approach. EAM is always comprised of a multitude of complementary approaches such as architecture design, modeling, graphical representation, change management or stakeholder management, to name just a few. In scientific literature as well as in EAM practice, such solutions have been developed individually. A combination of useful solutions for partial problems might not necessarily constitute a suitable overall EAM approach, though. Our aim is to look at EAM as one complex approach, i.e. we do not want to focus on specific aspects or views but on the entirety of EAM.

Methodically, we refer to the design science research (DSR) approach as described by Hevner et al. [12]. DSR is a vividly discussed and applied research approach and is concerned with problem solving and developing useful artifacts. DSR aims at designing general design solutions that address a class of problems instead of a specific problem. Literature reflecting DSR deals with research processes [24], the role of theory [33] and with the evaluation of artifacts [5].

However, not many publications within the DSR community put emphasis on problem analysis. This is surprising taking the wickedness of problems that are subject of DSR research into account: While there may be rather narrowly defined problems where artifact building is either trivial or can employ design methods, e.g. from computer science, a large number of problems in DSR, including EAM, involve some kind of management activities and therefore tend to be more complex. As Pries-Heje and Baskerville argue, complex or *wicked* problems need to be treated differently that simple problems because they lead to asymmetric criteria decision situations [25]. Typical topics that have been addressed in the conference series on Design Science Research in Information Systems and Technology (DESRIST) involve organizational change, workflow systems, knowledge management, innovation management, business process design, and requirements engineering. All these topics include management aspects and fulfill the criteria of wicked problems: “poorly formulated, confusing, and permeated with conflicting values of many decision makers or other stakeholders” [25].

We argue that for wicked problems the step of understanding and maybe theorizing the problem is vital before starting to actually build the artifact. Therefore in this paper we focus on the step of problem analysis in a DSR process: In order to effectively understand the problems in EAM it is necessary to gain insight into the fundamental structure and characteristics of EAM design that can

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

be distinguished. Therefore we aim at answering the following research questions:

- *What factors describe the design of EAM?*
- *What types of EAM design can be distinguished?*

We will ground our problem analysis on empirical data about existing, implemented EAM approaches in companies. Thereby, we aim at a thorough understanding of their inherent design and usefulness in terms of realization of their potential value. Our study is based on a questionnaire survey conducted with EAM practitioners (see section 3.1). Based on this sample we are able to explore inherent structures of EAM design. Factors influencing the design can then be used to develop useful artifacts that give guidance on the situational design of EAM.

The paper is structured as follows. In chapter 2 we outline conceptual foundations of the task of problem analysis within DSR and of EAM. Our empirical analysis is then presented in chapter 3. Based on a questionnaire (3.1) we have conducted a factor analysis (3.2) and a cluster analysis (3.3). In order to gain insight into the utility of the different EAM approaches found in the first analysis, we have conducted a second survey on the realization of EAM use potentials, which is described in chapter 4. Finally, chapter 5 integrates and discusses the findings of both surveys and gives an outlook on further research activities that may tie in with our results.

2. CONCEPTUAL FOUNDATIONS

2.1 Problem Analysis in DSR

DSR is engaged with the rigorous construction of useful IS artifacts, i.e. constructs, models, methods, or instantiations [20]. Artifact development in DSR is a problem solving process [12] and starts off with the identification of a problem that is going to be addressed [e.g. 4, 24, 29, 35]. A useful artifact is expected to solve a relevant business problem and to provide utility to the organization applying it. At the same time the artifact should be sufficiently general and address a class of design problems. As utility and generality can be conflicting goals, the concept of situational artifacts has been introduced [36]. In order to develop artifacts that are adaptable to different design problems within a problem class, the delineation of the design problem becomes a crucial task.

Considering the complexity of the underlying design problem and with the goal of situational artifact construction in mind, we argue that understanding and analyzing the problem is an essential part of building a useful artifact as a design solution. In doing so, it is important to understand the dimensions/design factors, parameters, generality and granularity of problems and possible solutions [36]. Winter proposes a procedure for developing situational artifacts based on a thorough problem analysis [36]: (1) Initial delineation of the design problem class, (2) identification of potential contingency factors based on literature analysis, (3) field study based analysis of design problems in practice in order to derive design factors, (4) refined specification of the design problem class, (5) calculation of the similarity of different design solutions, (6) determination of a useful level of generality, (7) specification of design situations. We will apply this procedure in this article by conducting an empirical analysis in order to identify design factors of EAM design and to achieve a more detailed specification of the design problem class EAM.

2.2 Enterprise Architecture Management

Most authors agree that EA targets a *holistic scope* and therefore provides a broad and aggregate view of an entire company or government agency [26, 31]. The ANSI/IEEE Standard 1471-2000 defines architecture as ”the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution” [13]. Relevant architectural views are strategic positioning, organizational structure, process organization, information flows, and implementation by means of software systems and data structures [15, 17, 38]. EA can provide systematic support to organizational change that affects business structures as well as IT structures by providing constructional principles for designing the enterprise [7]. In order to provide support for transformation in an efficient way, EA has to be driven by business and/or IT oriented application scenarios [37] based on stakeholders concerns [22, 23, 39] (*goal orientation*) [17, 26]. Since the involvement of heterogeneous stakeholder groups may create conflicting requirements in a complex environment, an appropriate documentation and communication of EA is vital. A suitable degree of *formalization* is needed in order to ensure traceable and repeatable results. Furthermore (semi) formalized models and well structured methods are needed to enable division of labor among the stakeholder groups [10, 14].

In the field of EAM a lot of methods have been developed [2, 3, 6, 30, 34]. These methods typically comprise the following EAM processes: (1) strategic design of an architectural vision, (2) development and maintenance of as-is architecture models, (3) development and maintenance of to-be architecture models, (4) migration planning, (5) implementation of EA, and (6) analysis of EA on the basis of architecture models. Furthermore, communication and lobbying of architectural guidelines and principles are part of EAM processes [28].

Aiming at a deeper understanding of the constituent factors that influence EAM, there has been some scientific effort to analyze contingency factors of EAM. Aier et al. [1] have identified three factors that describe three aspects of EA focusing on models, data, and organizational penetration. However, they did not explicitly consider management aspects of EAM. Leppänen et al. [18] took a first step towards a complex contingency framework for an engineering method for EA. Ylimäki [39] conducted several studies in order to identify potential critical success factors for EA. Ylimäki found the following factors: commitment, governance, methodology, EA models, project management, training and education, organizational culture, IT investment strategy, assessment and evaluation, business-driven approach, communication, and scope. These success factors give a first insight into possible design factors of EAM. Therefore, we have used this set as a starting point for our study design.

3. EXPLORING EAM DESIGN

Following the problem analysis approach described by Winter [36], we have conducted an empirical analysis that examines EAM design approaches (our design problem) currently applied in practice. This analysis allows for a refined specification of different EAM designs based on descriptive factors (see section 3.2) and a specification of different types of EAM designs (see section 3.3). The analysis has been carried out via a questionnaire-based survey to address a large number of different organizations.

3.1 Questionnaire and Data Set

The questionnaire has been designed to describe EAM approaches by means of constituting aspects which have been identified in advance based on literature analysis. In order to distinguish different EAM approaches, the first part of the questionnaire has asked for the organization's understanding of the notion "EAM". The understanding in terms of the tasks and results of EAM can be manifold and is crucial to EAM design within the respective organization [32]. Second, the positioning of EAM within an organization is expressed by its integration into the organizational structure. The way organizational units, teams and roles are involved in the EAM processes is an expression of this aspect [21, 32]. Other important criteria in this context are the scope of EAM processes, the penetration of EAM processes and EAM results throughout the organization as well as the level of continuity and controlling of EAM processes. Finally, the types of EAM results that are used by different organizational units play an important role in EAM design.

Table 1. Absolute Number of Returned Questionnaires Grouped by Industry and Size of Organization

	Number of Employees						
Industry	**20-49**	**50-99**	**100-249**	**250-499**	**500-1000**	**>1000**	**Σ**
Manufacturing	0	0	0	0	0	6	6
Retail	0	0	0	0	1	4	5
Telecom.	0	0	0	0	1	8	9
Financial	0	1	1	4	3	21	30
Insurance	1	0	1	0	0	6	8
Public Admin.	1	0	1	3	1	8	14
Software/IT	4	0	3	2	3	3	15
Other	1	2	2	1	4	22	32
Total	**7**	**3**	**8**	**10**	**13**	**78**	**119**

Table 2. Factor Analysis Results

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Item 1.1	0.8005	0.1414	0.0584	0.1482	0.1741	0.0005	0.1250	0.1625
Item 1.2	0.7344	0.1076	0.0963	0.3178	0.2019	0.0961	0.1093	0.1141
Item 1.3	0.6653	0.0131	0.2617	0.3951	0.1362	0.1635	0.0107	0.1141
Item 1.4	0.5700	0.3222	0.0473	-0.1345	0.0492	0.0976	0.4473	-0.1242
Item 1.5	0.5103	0.2223	0.3147	0.0360	0.0224	0.2129	0.1296	0.3467
Item 1.6	0.5041	0.4041	0.1099	-0.0587	0.3947	-0.0328	0.2592	-0.0511
Item 2.1	0.1543	0.6984	0.0718	0.1489	0.1571	-0.0091	0.1917	-0.1022
Item 2.2	0.1662	0.6790	0.2116	0.1783	0.1376	0.1001	-0.0594	0.0959
Item 2.3	0.2192	0.6727	0.2571	0.1340	-0.0934	0.1851	0.2520	0.1019
Item 2.4	0.1005	0.6171	0.1932	0.0900	0.2565	0.0309	0.1163	0.4120
Item 2.5	0.0504	0.5903	0.0612	-0.0424	0.0523	0.2221	0.1369	0.4311
Item 2.6	0.1472	0.5445	0.0860	0.4527	-0.0052	0.0370	0.0584	0.2522
Item 2.7	-0.0470	0.5422	0.0985	0.2524	0.3363	0.2609	-0.1153	-0.0254
Item 3.1	0.0408	0.0405	0.7838	0.0045	0.0500	0.0420	0.2519	0.2246
Item 3.2	-0.0369	0.0399	0.7828	0.0237	0.1358	0.0413	0.1264	0.3787
Item 3.3	0.2156	0.1512	0.7503	0.0456	0.2513	-0.0068	-0.0626	0.1290
Item 3.4	0.1423	0.1925	0.6954	0.1726	0.1079	0.1310	0.2415	-0.2124
Item 3.5	0.2834	0.4356	0.5383	0.0916	0.0872	0.1084	-0.0446	0.0741
Item 3.6	0.3595	0.2582	0.5229	0.0661	-0.1861	0.2791	0.0940	0.2506
Item 3.7	0.1328	0.2919	0.5215	0.1928	0.1498	0.2199	-0.1639	-0.1297
Item 4.1	0.0425	0.1193	-0.0869	0.7153	-0.0010	-0.0447	0.3171	0.1112
Item 4.2	0.1381	0.1993	0.2465	0.6490	0.1004	0.1694	0.2107	0.0299
Item 4.3	0.3159	0.2069	0.0679	0.6006	0.0122	0.0122	-0.1052	0.3857
Item 4.4	0.4137	0.2716	0.2632	0.4316	0.2543	0.0597	0.0045	-0.1618
Item 5.1	0.3226	0.3338	0.2408	0.1384	0.6098	0.2141	0.0378	0.1362
Item 5.2	0.5143	0.2658	0.3642	0.0594	0.5146	0.0871	0.0524	0.1610
Item 5.3	0.1810	0.4383	0.2223	-0.0427	0.5080	0.1378	0.3287	0.1531
Item 5.4	0.4091	0.0184	0.2955	0.0409	0.4839	0.0180	0.1672	0.2445
Item 5.5	0.2382	0.0966	0.1071	0.4501	0.4635	0.1426	0.2576	0.2482
Item 5.6	0.4081	0.4095	0.0890	0.3137	0.4424	0.2017	0.0116	0.1022
Item 6.1	-0.0504	-0.0335	0.1975	0.0048	0.1466	0.8264	0.1755	0.1309
Item 6.2	0.1754	0.3235	0.0705	0.0613	0.1660	0.7595	0.0698	0.0980
Item 6.3	0.4109	0.3142	0.0491	0.2182	-0.1043	0.6578	0.2125	0.0363
Item 7.1	0.0474	0.1081	0.0865	0.1935	0.0899	0.2076	0.7206	0.2501
Item 7.2	0.2778	0.1232	0.2307	0.3425	0.1531	0.1181	0.6506	0.0201
Item 7.3	0.4072	0.0864	0.1453	0.3974	0.0762	0.1575	0.5480	0.0709
Item 8.1	0.1240	0.1582	0.2604	0.1934	0.1663	0.0956	0.1205	0.7207
Item 8.2	0.2578	0.1152	0.1583	0.3391	0.1612	0.1705	0.1204	0.5719
Cronbach's Alpha	**0.858**	**0.852**	**0.852**	**0.744**	**0.875**	**0.794**	**0.820**	**0.706**

The questionnaire reflects these considerations by listing the mentioned characteristics of an EAM approach as questions. The resulting questionnaire contains 54 items asking for the as-is state of EAM in the companies. The respondents have been asked to assess the current degree of realization on a 5-point Likert scale [19]. The minimum value (1) represents "not realized", whereas the maximum value (5) represents "fully realized".

Empirical data has been collected at four events on the topic of EAM. The events focused on EAM only and were attended by EAM experts from both IT and business departments, as well as IT management executives, IT service providers, and other IT users concerned with the task of EAM. The events took place in Germany and Switzerland between June and September 2009. A total of 119 data sets were collected that did not reveal substantial extent of missing data (10% at maximum).

More than half of the respondents are corporate users (53.8%), besides consulting firms or vendors (36.2%) and others. The companies that participated in the survey are mainly mid-size and large, most of them being active in the financial industry. Table 1 shows the distribution of industry and size of organization the respondents stem from.

3.2 Factor Analysis: Descriptive Factors of EAM Design

In order to identify common underlying dimensions characterizing EAM, we have applied an exploratory factor analysis using the principal component analysis. A factor analysis involves extracting a small number of latent factors among the variables in the data set. It is necessary to test the adequacy of the data set prior to applying a factor analysis. To form an adequate foundation, the data set has to meet two criteria. The first criterion is derived from the variables' anti image covariance. The anti image covers the part of the variance that cannot be explained by the remaining variables in the data set. As factor analysis aims at finding latent factors based on the data set, a data set is suitable for factor analysis if the anti image is rather low: According to Dziuban and Shirkey [8], the percentage of none diagonal elements of the anti image covariance matrix, which are non-zero (>0.09), should not exceed 25%. This holds true for the data set at hand. The second criterion involves the computation of the Kaiser-Meyer-Olkin measure of sampling adequacy. In the data set at hand, the measure is 0.87. According to Kaiser and Rice [16], this characterizes the intercorrelation among the variables within the factors found as "meritorious". In this case, the results prove that the data set is generally appropriate for factor analysis.

The factor analysis, using Varimax rotation with Kaiser normalization, have led to eight factors, including 38 items of the questionnaire (cf. Table 2). 16 items have been deleted because they were intentionally designed as control items or did not seem to contribute to the factor identification [11]. Due to some incomplete questionnaires, missing values have been excluded pair wise during factor analysis. This resulted in a total number of 109 cases contributing to the factor analysis. The items selected for the factor analysis explain 67.63% of variance in total.

In order to test the reliability of the factor scale, Cronbach's Alpha has been calculated for each factor (Table 2). For Cronbach's Alpha a value above 0.7 indicates an adequate reliability, which holds true for our data set. It must be noted that a large number of items may increase the value artificially [9]. However, in our case, the reliability analysis showes values mostly between 0.4 and 0.5 for the inter-item correlation within the factors, which indicates that the high value of Cronbach's Alpha can be regarded as valid for our purposes.

With regards to the interpretation of factors factor loadings from 0.3 to 0.4 are considered a minimal level [11]. Generally, factor loadings from at least 0.5 are considered sufficient for an unambiguous assignment to one factor. Some items in our data set show identically high factor loadings for more than one factor (Table 2). In these cases, factor assignment has been based on practical considerations. These considerations are explained in detail in conjunction with the factor descriptions below.

Table 3. Factor 1: IT Operations Support

Item No.	Item Description
Item 1.1	Results of EAM are used for IT development
Item 1.2	Results of EAM are used for coordination of IT development products
Item 1.3	Results of EAM are used for IT planning and infrastructure design
Item 1.4	IT departments perceive EAM as a useful device
Item 1.5	Results of EAM are used for analyses on architecture models (e.g. dependence analyses)
Item 1.6	IT departments use EAM results for their daily job

Factor 1 comprises items describing the concern *IT operations support* within the EAM approach. The use of results for IT operation tasks and by IT departments for their daily job characterizes this factor. Considering the items' loadings on this factor it becomes obvious that usage of EAM results as well as the perception of EAM within the organizational units concerned with IT operations exert a conjoint effect on overall EAM assessment.

Table 4. Factor 2: Enterprise Focus and Management Support

Item No.	Item Description
Item 2.1	Business and IT departments actively seek advice from architects
Item 2.2	Results of EAM are used for communications with management functions
Item 2.3	EA stakeholder are involved in EAM
Item 2.4	Management board uses EAM results for management tasks
Item 2.5	Management board perceives EAM as a useful device
Item 2.6	EAM is aligned with business objectives
Item 2.7	Architects have an extensive network within the company

Factor 2 summarizes items related to the *support of management tasks by EAM*. This is again expressed by the usage of EAM results by management tasks as well as by the perception of EAM in the management board. This factor constitutes the *antipole* to factor 1 and reveals that EAM can serve both IT and management purposes, but that these purposes are most probably not highly interrelated.

Table 5. Factor 3: EAM Governance

Item No.	Item Description
Item 3.1	EA models are assessed and evaluated regularly
Item 3.2	EAM processes are assessed and evaluated regularly
Item 3.3	There are defined maintenance processes for EA models and EA data
Item 3.4	There are defined EAM processes
Item 3.5	Results of EAM are used for documentation and tracking of EA models
Item 3.6	There is one consistent, enterprise-wide effective architecture model
Item 3.7	Architecture data is centralized with the EAM department

Factor 3 denotes the *governance of EAM* itself. Thereby, it is a characteristic not describing the execution of EAM but the maintenance of EAM. EAM governance consists of model and process assessment and maintenance and a central supervision of EA models and data.

Table 6. Factor 4: IT Strategy and IT Governance Support

Item No.	Item Description
Item 4.1	EAM is essential part of IT strategy development
Item 4.2	EAM is essential part of IT governance
Item 4.3	Results of EAM are used for IT strategy development
Item 4.4	Results of EAM are used for IT governance

Supplementing factor 1, factor 4 characterizes the *support of IT strategy and governance tasks* by EAM. Item 4.4 shows almost equal values for factors 1 and 4 (0.41 and 0.43), which may lead to the conclusion that the use of EAM for IT governance purposes may be highly correlated with operational IT tasks. Due to the contents of items 4.1 through 4.3 we decided to assign item 4.4 to factor 4.

Table 7. Factor 5: Information Supply

Item No.	Item Description
Item 5.1	Results of EAM are used for supply of information for business departments (service function)
Item 5.2	Results of EAM are used for supply of information for IT departments (service function)
Item 5.3	Business departments use EAM results for their daily job
Item 5.4	Results of EAM are used for operations and maintenance
Item 5.5	Results of EAM are used for Business/IT Alignment
Item 5.6	Results of EAM are used for moderation between business and IT departments (and among them)

Factor 5 again characterizes a support aspect of EAM: the *information supply* with EAM results, independent of business or IT purposes. The closeness to IT operations support (factor 1) is also backed by the fact that item 5.2 shows equal values for factors 1 and 5 (0.51). Factor 5 reflects the service function EAM can fulfill both for business and IT departments. Moreover the support of business/IT alignment is an essential part of this factor.

Table 8. Factor 6: Integrative Role

Item No.	Item Description
Item 6.1	EAM takes place in an interdisciplinary team
Item 6.2	EAM team and business departments continuously exchange information (e.g. in architecture boards)
Item 6.3	EAM team and IT departments continuously exchange information (e.g. in architecture boards)

Table 9. Factor 7: Design Impact

Item No.	Item Description
Item 7.1	EAM has an impact on IT (infrastructure) architecture design
Item 7.2	EAM has an impact on application architecture design
Item 7.3	EAM has an impact on business architecture design

Factors 6 and 7 comprise items describing the role of EAM within the organization. While factor 6 summarizes aspects expressing the *integrative role*, factor 7 focuses on the *design impact*. The integrative role of EAM can be realized by interdisciplinary teams and a continuous exchange between EAM roles. It can be assumed that the existence of an architecture board is part of such an organizational structure for EAM. The design impact can describe EAM's impact on IT or infrastructure, application or business architecture. The degree of design impact most probably reflects the penetration of the EAM approach throughout the organization as well as its active role.

Table 10. Factor 8: Business Strategy Support

Item No.	Item Description
Item 8.1	Results of EAM are used for enterprise development
Item 8.2	Results of EAM are used for strategic planning (e.g. product planning)

Finally, factor 8 again describes a support concern of EAM: *business strategy support*. In contrast to factor 2, items in factor 8 describe the support of explicit strategic tasks like enterprise development and product planning. Most probably, high degrees of realization of this factor correspond to a high realization of factor 2.

To summarize the results of the factor analysis three different groups of characteristics of EAM were found: Factors 1, 2, 4, 5 and 8 characterize the concern of EAM, i.e. whether EAM supports IT operations, management tasks, IT strategy, Business/IT alignment or business strategy. Factors 6 and 7 describe the role of EAM within the company (as moderator or designer). Finally, factor 3 describes the governance of EAM itself.

3.3 Cluster Analysis: Specific Types of EAM Design

In order to further specify the design problem class EAM and to assess the similarity of design problems within the class a cluster analysis has been performed upon the eight design factors found by our exploratory factor analysis. Cluster analysis aims at finding groups of respondents that apply similar EAM approaches. As they are more common and do not bear the risk of constraining the possible clusters by ex-ante presumptions [11], hierarchical

clustering algorithms have been used. Those algorithms produce all possible clustering results, so the final solution must be identified based on the agglomeration schedule and the dendrogram.

Considering results from preliminary cluster analyses on the data, one case has been eliminated as it showed heavy outlier behavior [11]. Excluding cases with missing factor loadings, 94 cases could be used for the cluster analysis. For the final cluster analysis, which is presented below, we have used the *Average Within-Group Linkage* cluster algorithm provided by SPSS and *Squared Euclidean Distance* as the distance measure. The results of the agglomeration schedule and the dendrogram have led to the identification of three clusters. Based on this empirical evidence the level of optimal granularity of the design problem class EAM can be determined as three.

In order to interpret the clusters with respect to the underlying characteristics, the cluster centroids have been analyzed. These can be identified by the mean factor values within each cluster, enabling a characterization of the clusters. The mean factor values for each cluster are depicted by the net diagram in Figure 1. The clusters can be described as follows.

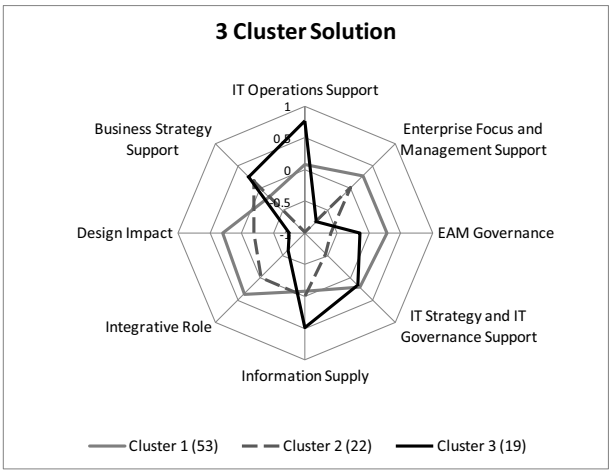


Figure 1. Net Diagram for the 3 cluster solution

Cluster 1: Balanced, active approach

The first cluster (solid line) presents a rather balanced approach to EAM. For most factors this cluster shows the highest or at least average values. Especially the similar values for the factors IT Operations Support and Enterprise Focus and Management Support lead to the conclusion that organizations within this cluster do not focus neither on IT support nor on management support with their EAM approach.

In contrast to other clusters, the high support of IT operations, management, IT strategy as well as the focus on design impact, the integrative role and EAM Governance point to a high degree of integration within the organization. In particular values for Design Impact, Integrative Role and EAM Governance are by far the highest between all three clusters. It can therefore be presumed that these organizations have a rather high level of maturity in their EAM approach.

It should be noted that this cluster includes 53 out of 94 organizations, which lead to the supposition that this cluster represents a "mainstream" approach. This can also be ascribed to the fact that the majority of participants in the survey are from large organizations (>1000 employees, cf. Table 1).

Cluster 2: Business-oriented approach

The second cluster (dashed line) groups 22 organizations that have an apparent focus on business support in their EAM approach. The factors IT Operations Support as well as IT Strategy and IT Governance Support are clearly assigned with comparatively low values. Comparing mean factor values to those of cluster 1, the overall low values imply that the organizations in this cluster do not show a high degree of EAM implementation in any dimension. Two conclusions can be derived from this fact: First, the organizations could have decided to apply a minimalist EAM approach, focusing on management support without putting resources in EAM governance or an active role of EAM. Second, the introduction of EAM could only recently be initiated by management and is not very mature yet. For both cases, literature suggests that a sustainable EAM approach can only be established by realizing an effective EAM governance [2, 3].

Cluster 3: IT-oriented, passive approach

Organizations assigned to this cluster (dotted line) clearly emphasize the use of EAM for IT operations as well as the information supply by EAM. In contrast, values for management support are by far the lowest compared to the other clusters. As the factors Design Impact as well as Integrative Role are not focused in this approach, it can be described as a passive approach that is most probably realized very locally within the organization.

Obviously, this small cluster, which includes only 19 of 94 organizations, represents a specialized IT-centered EAM approach that primarily takes a documentation role. It can be presumed that the EAM approach was initiated by IT departments and has not been disseminated throughout the organization yet.

4. UTILITY OF EAM APPROACHES

4.1 Questionnaire and Data Set

The factor and cluster analysis provide a basis for understanding different types EAM design and their constituting design factors that should be respected by a situational artifact. However, it can only be concluded that the described types of EAM design can be distinguished. The results of the analyses do not allow assessing whether a certain EAM design is "good" or "appropriate" or even "successful". Therefore, we conducted a subsequent analysis that asked for the realization of use potentials that are typically assigned to the adoption of EAM. Thereby, we aimed at assessing if a certain EAM approach is more successful than another.

A second questionnaire has been distributed at a subsequent EA expert event within the same series of events where the first survey had been conducted. The event took place in Switzerland in February 2010. 98 questionnaires were returned, with 94 questionnaires holding less than 10% missing data. Hence, 94 data records could be included in the analysis. The respondents of the second questionnaire were asked to assess the current realization as well as their satisfaction with the current degree of realization of 14 use potentials [21, 27, 32, 34]:

- Business/IT alignment
- Consistent implementation of business strategy
- Improved complexity management
- Higher flexibility in reaction to external changes

- Higher efficiency in reaction to customer and market needs, and pressure to be innovative by proactive action
- Lower risk by preparing for unplanned changes
- Fewer inconsistencies and redundancies through transparent IT functionalities
- Adoption of modern technologies
- Integration of business activities across business units
- Dissolution of information silos (e.g. CRM information)
- Lower heterogeneity of technologies in use
- Lower support costs
- Improved reusability of technologies, information, and functions
- Lower development costs

4.2 Profiling Respondents

In a first step, the respondents of the second questionnaire were assigned to EAM clusters found in the explorative analysis of the first questionnaire. In order to do so, the respondents were asked about their current realization of the eight factors upon which the clusters are based. In a first step, standard scores for these values were calculated, so that they are comparable with the factors values, i.e. the cluster profile line values from the first analysis. Standard scores have a mean μ of zero and a standard deviation σ of 1. The standard score Z was calculated using the following formula:

(1) $Z = (x-\mu)/\sigma$, while x is original value of variable

In a second step the "distance" between each respondent's profile and the three clusters' profiles were calculated. The distance was measured by the method of least squares, which is also used in regression analysis [11]:

(2) $\sum (z - c)^2 \rightarrow \min$, while z is standard score of the variable and c is the cluster profile point

This distance calculation method was performed for each respondent and each cluster. The least sum of squares designates the cluster profile the respondent is most similar to. Hence, each respondent could be assigned to cluster 1, 2, or 3. Finally, 50 cases could be assigned to cluster 1, 38 cases could be assigned to cluster 2, and 6 cases could be assigned to cluster 3.

4.3 Results: Realization of Use Potentials

By comparing the standard mean values of the respondents that are assigned to the same cluster, the realization of the use potentials can be analyzed dependent on the type of EAM design applied. In Figure 2 realization and the satisfaction of the use potentials are depicted in one diagram, divided by the three clusters the respondents were assigned to. While satisfaction with the realization is given on the vertical axis, the degree of realization itself is given on the horizontal axis. It shows that especially in cluster 1, i.e. the balanced, active approach the organizations are comparably highly satisfied with their high realization of use potentials (upper right area). In contrast to this, organizations in cluster 2, i.e. business-oriented approach show low realization and low satisfaction with it (lower left area). In addition Table 11 shows the standard means of the use potential realization variables across the groups. It becomes obvious that cluster 1 has the highest realization for each use potential. From there it can be concluded that the EAM approach represented by cluster 1 is more successful than the other.

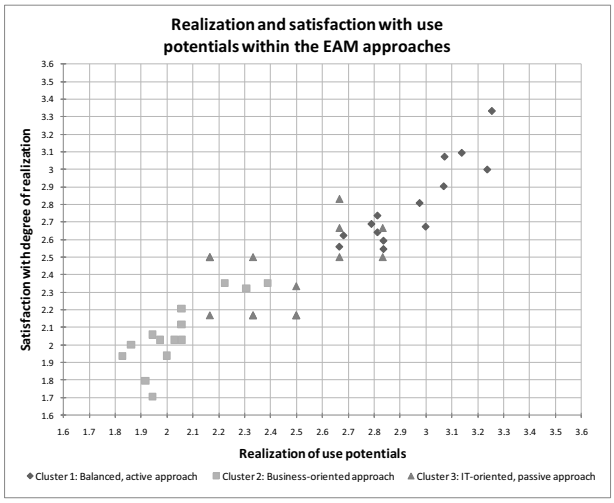


Figure 2. Realization and satisfaction with use potentials across all three clusters

Table 11. Standard Means of Use Potential Variables across all Three Clusters

	Cl. 1	Cl. 2	Cl. 3	F	Sig.
Business/IT alignment	3.07	1.92	2.50	26.039	.000
Consistent implementation of business strategy	2.84	2.03	2.33	9.446	.000
Improved complexity management	3.00	1.94	2.33	13.286	.000
Higher flexibility in reaction to external changes	2.68	1.83	2.17	8.468	.000
Higher efficiency in reaction to customer and market needs, and pressure to be innovative by proactive action	2.67	1.86	2.67	9.835	.000
Lower risk by preparing for unplanned changes	2.81	1.94	2.33	10.196	.000
Fewer inconsistencies and redundancies through transparent IT functionalities	3.07	2.00	2.67	14.723	.000
Adoption of modern technologies	3.26	2.39	2.83	6.897	.002
Integration of business activities across business units	2.98	2.06	2.83	11.344	.000
Dissolution of information silos (e.g. CRM information)	2.81	2.31	2.67	2.848	.064
Lower heterogeneity of technologies in use	3.14	2.22	2.50	10.592	.000
Lower support costs	2.79	2.06	2.17	9.889	.000
Improved reusability of technologies, information, and functions	3.24	2.06	2.50	20.869	.000
Lower development costs	2.84	1.97	2.17	12.454	.000
Legend: Cl. = Cluster, Sig. = Significance					

In order to further examine the differences between the three clusters, ANOVA was performed to analyze if there are significant differences in the realization of use potentials between the groups (clusters) [9]. ANOVA tests whether group means are

equal by comparing variations within and between the groups. The comparison is expressed by the F ratio, which is the higher the more likely the difference of the values between groups is not a chance result. In our analysis, F values and significance values of ANOVA show that the null hypothesis that the standard means are equal across all groups must be rejected for all use potentials (cf. Table 11). On exception is the use potential "Dissolution of information silos" with a significance value of 0.064.

Consequently, it can be stated that the realization of use potentials is significantly different between the clusters. This means that the employment of a certain EAM design approach has an effect of the realization of the use potentials, i.e. that the three different EAM approaches are unequally successful. In the following sections we describe the three different EAM approaches regarding the degree of realization of use potentials in detail.

4.3.1 Business-Oriented Approach

Overall the business-oriented approach shows the lowest values of realization and satisfaction with use potentials and therefore seems to be the least successful approach. Among the use potentials, the variables showing the highest realization values are:

- Adoption of modern technologies
- Dissolution of information silos
- Lower heterogeneity of technologies in use

The variables showing high realization but still low satisfaction values are:

- Higher flexibility in reaction to external changes
- Higher efficiency in reaction to customer and market needs, and pressure to innovate
- Improved complexity management
- Business/IT alignment
- Fewer inconsistencies and redundancies

It is remarkable that technology-related use potentials seem to be realized more successfully although the approach is characterized by business orientation. In comparison to the other approaches that also show high values in these use potentials, it can be assumed that these use potentials can be realized by different approaches.

The list of use potentials with high satisfaction values shows that the approach lacks the realization of business-related goals. Especially an increased flexibility and efficiency in reaction to external changes and customer needs are use potentials that are often expected from business-oriented EAM. However, in this cluster the approach does not seem to be successful in this matter. Hence, this is the crucial aspect that needs to be improved and respected by an appropriate method how to design EAM.

4.3.2 IT-Oriented, Passive Approach

The IT-oriented, passive approach shows intermediate values for both realization and satisfaction with use potentials of EAM. By further analyzing this cluster it has to be considered that this can only be based on 6 cases found in the second survey. Within this cluster, the variables showing the highest realization values are:

- Adoption of modern technologies
- Dissolution of information silos
- Fewer inconsistencies and redundancies through transparent IT functionalities

The variables showing high realization but still low satisfaction values are:

- Higher flexibility in reaction to external changes
- Improved complexity management
- Lower risk by preparing for unplanned changes
- Lower support costs

The use potentials with high realization values seem to be compatible with the IT-oriented approach in this cluster. On the other side, there is a lot of improvement potential regarding reduced risks, costs and improved complexity management and flexibility. The pursuit for reaching these goals reflects the IT-orientation of this approach. At the same time, the gap between current and aspired realization leads to the assumption that the IT-oriented, passive EAM approach is not very mature yet.

4.3.3 Balanced, Active Approach

The majority of the organizations in both our surveys can be assigned to the balanced, active approach cluster. In addition, this cluster shows the highest values for realization and satisfaction with use potentials by far. The variables showing the highest realization as well as satisfaction values in this cluster are:

- Adoption of modern technologies
- Lower heterogeneity of technologies in use
- Improved reusability of technologies, information, and functions
- Fewer inconsistencies and redundancies through transparent IT functionalities

The variables showing high realization but still low satisfaction values are:

- Consistent implementation of business strategy
- Lower development costs
- Higher flexibility in reaction to external changes
- Higher efficiency in reaction to customer and market needs, and pressure to innovate

The high realization values show that although the approach is characterized as balanced between IT and business goals, IT-related use potentials are realized better than business-related ones. This can be ascribed to the high fraction of IT experts among the respondents in the survey or to the fact that many EAM initiatives in organizations are driven by IT departments. Therefore, there is still improvement potential regarding business-related goals like flexibility and efficiency.

5. DISCUSSION AND OUTLOOK

Our analysis provides the means for specifying the design problem EAM: It shows that three different types of EAM design can be distinguished. The approaches differ regarding the focus of EAM (IT or business) as well as regarding their emphasis on an active design role of EAM. The separation of one *mainstream* and two *specialized* approaches that can reliably be distinguished leads to the conclusion that there are significant differences in realizing EAM in practice. In conjunction with the analysis of the realization of use potentials, i.e. the achievement of typical EAM goals, the empirical analysis of EAM practices furthermore reveals some important details about how EAM is done and what improvement potentials still exist. These potentials should be target of future research activities in order to develop useful artifacts for the design and adoption of EAM in practice.

However, there are still some limitations in our work. First, the design factors are most probably not stable over time but will change, either over the time of application within one organization or in dependency of some other aspect. In order to ensure the reliability of the eight design factors it would also be helpful to gain information in what way they depend on other typical contingency factors that we did not examine in our research, e.g. culture, organizational structure etc. Further research towards a specification of EAM design situations should thus aim at a better understanding of the relevant contingency factors and their combination.

Moreover, it needs to be noted that the probably complex set of design factors of EAM cannot be explained fully by our research results. Although we found eight constituting factors that determine the design of EAM as it is done in practice, there is no information about how these factors interact with each other. The emphasis on one factor, e.g. regarding design impact, might limit the possible design scope of another factor, e.g. EAM governance. From our results it cannot be inferred that high values of all factors correspond to the "best" EAM design.

Our analysis clearly structured the problem domain of EAM. This is a valuable basis for a situational artifact construction in DSR. The approach of understanding a certain problem domain – like EAM in this analysis – might also provide value for other wicked problems in IS. Therefore, a more systematic in-depth analysis of the actual problems might lead to a more structured build process in DSR. As a long term goal the DSR process models should reflect such a systematic in-depth problem analysis as we have performed here.

Instead of striving for a complete understanding of the interplay of all possible contingency factors of EAM we put our focus on exploring the EAM approaches with regard to their success. From our analysis it can be deduced that the balanced, active approach showed the highest satisfaction and realization values for possible use potentials. Hence, a general conclusion is that even EAM initiatives that are at the very beginning or are intended to start within a small scope should aim at a vision that is in line with this approach. Furthermore, this result encourages the opinion that EAM should not be considered as an IT or business approach only. The comparison between the three clusters showed that a high emphasis on one of these extremes is not assessed as equally successful as the balanced approach.

Regarding the achievement of goals that are expected from EAM, the adoption of modern technologies is the highest realized use potential. By contrast the use potentials that are still not satisfied are: flexibility to external changes and reaction to customer and market needs. This result is remarkable as these are goals usually assigned to EAM and often serve as a main selling point for adopting EAM in an organization. This mismatch indicates once more that EAM cannot be realized by applying a standard or "one-size-fits-all" approach but that situational design guidance is necessary. It also shows a common dissatisfaction with the way EAM is currently done in practice. Maybe these goals can only be achieved in the long run when an EAM initiative "grows in an organization". Such a dependency on other influencing factors should be subject of further research activities.

6. REFERENCES

[1] Aier, S., Riege, C., Winter, R. 2008. Classification of Enterprise Architecture Scenarios – An Exploratory Analysis. *Enterprise Modelling and Information Systems Architectures,* 3, 1 (2008), 14–23.

[2] Aziz, S., Obitz, T., Modi, R., Sarkar, S. 2005. *Enterprise Architecture: A Governance Framework - Part I: Embedding Architecture into the Organization*. Technical Report Infosys.

[3] Aziz, S., Obitz, T., Modi, R., Sarkar, S. 2006. *Enterprise Architecture: A Governance Framework - Part II: Making Enterprise Architecture Work within the Organization*. Technical Report Infosys Technologies Ltd.

[4] Baskerville, R.L., Pries-Heje, J., Venable, J. 2007. Soft Design Science Research: Extending the Boundaries of Evaluation in Design Science Research. In: *Proceedings of DESRIST 2007,*

[5] Baskerville, R.L., Pries-Heje, J., Venable, J. 2008. Evaluation Risks in Design Science Research: A Framework. In: *Proceedings of the 3rd International Conference on Design Science in Information Systems and Technology*, 329-334.

[6] Department of Veterans, A. 2001. *Enterprise Architecture: Strategy, Governance, & Implementation*. Technical Report.

[7] Dietz, J.L.G. 2006. *Enterprise Ontology – Theory and Methodology.* Springer, Berlin, Heidelberg.

[8] Dziuban, C.D., Shirkey, E.C. 1974. When is a Correlation Matrix Appropriate for Factor Analysis? *Psychological Bulletin,* 81, 6 (1974), 358-361.

[9] Field, A. 2009. *Discovering Statistics Using SPSS.* SAGE, London.

[10] Frank, U. 2002. Perspective Enterprise Modeling (MEMO) – Conceptual Framework and Modeling Languages. In: *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35),* Los Alamitos, CA, IEEE Computer Society, 1258–1267.

[11] Hair Jr, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L. 2006. *Multivariate Data Analysis.* Pearson Prentice Hall, Upper Saddle River, New Jersey.

[12] Hevner, A.R., March, S.T., Park, J., Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly,* 28, 1 (2004), 75-105.

[13] IEEE 2000. *IEEE Recommended Practice for Architectural Description of Software Intensive Systems (IEEE Std 1471-2000)*. Technical Report IEEE Computer Society, New York, NY.

[14] Jonkers, H., Lankhorst, M., van Buuren, R., Hoppenbrouwers, S., Bonsangue, M., van der Torre, L. 2004. Concepts for Modelling Enterprise Architectures. *International Journal of Cooperative Information Systems,* 13, 3 (2004), 257–287.

[15] Jonkers, H., Lankhorst, M.M., ter Doest, H.W.L., Arbab, F., Bosma, H., Wieringa, R.J. 2006. Enterprise architecture: Management tool and blueprint for the organisation. *Information Systems Frontiers,* 8, 2 (2006), 63–66.

[16] Kaiser, H.F., Rice, J. 1974. Little Jiffy, Mark Iv. *Educational and Psychological Measurement,* 34, 1 (1974), 111–117.

[17] Lankhorst, M. 2005. *Enterprise Architecture at Work: Modelling, Communication and Analysis.* Springer, Berlin et al.

[18] Leppänen, M., Valtonen, K., Pulkkinen, M. 2007. Towards a Contingency Framework for Engineering an Enterprise Architecture Planning Method. In: *Proceedings of the 30th Information Systems Research Seminar in Scandinavia (IRIS 2007),* Tampere, Finland, 1–20.

[19] Likert, R. 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology,* 22, 140 (1932), 1–55.

[20] March, S.T., Smith, G.F. 1995. Design and Natural Science Research on Information Technology. *Decision Support Systems,* 15, 4 (1995), 251-266.

[21] Niemann, K.D. 2006. *From Enterprise Architecture to IT Governance. Elements of Effective IT Management.* Vieweg, Wiesbaden.

[22] Niemi, E. 2007. Enterprise Architecture Stakeholders – A holistic view. In: *13th Americas Conference on Information Systems (AMCIS 2007),* Keystone, CO

[23] Op 't Land, M., Proper, E., Waage, M., Cloo, J., Steghuis, C. 2009. *Enterprise Architecture – Creating Value by Informed Governance.* Springer, Berlin.

[24] Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems,* 24, 3 (2007), 45–77.

[25] Pries-Heje, J., Baskerville, R.L. 2008. The Design Theory Nexus. *MIS Quarterly,* 32, 4 (2008), 731-755.

[26] Rohloff, M. 2008. Framework and Reference for Architecture Design. In: *Proceedings of the 14th Americas Conference on Information Systems (AMCIS 2008),* Toronto, 1–14.

[27] Ross, J.W. 2006. Enterprise Architecture: Driving Business Benefits from IT.

[28] Schelp, J., Winter, R. 2008. Management der Unternehmensarchitektur. In: Reussner, R., Hasselbring, W. (Eds.) *Handbuch der Software-Architektur*. 249–266. dpunkt, Heidelberg.

[29] Takeda, H., Veerkamp, P., Tomiyama, T., Yoshikawa, H. 1990. Modeling design processes. *AI Magazine,* 11, 4 (1990), 37-48.

[30] The Open Group 2009. *TOGAF (The Open Group Architecture Framework) Version 9*. Technical Report.

[31] Tyler, D.F., Cathcart, T.P. 2006. A Structured Method for Developing Agile Enterprise Architectures. In: *Proceedings of the International Conference on Agile Manufacturing (ICAM 2006),* Norfolk, Virginia, 1–8.

[32] van den Berg, M., van Steenbergen, M. 2006. *Building an Enterprise Architecture Practice: Tools, Tips, Best Practices, Ready-to-Use Insights.* Springer, Berlin.

[33] Venable, J. 2006. The Role of Theory and Theorising in Design Science Research. In: *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006),* Claremont, CA, Claremont Graduate University, 1–18.

[34] Wagter, R., van den Berg, M., Luijpers, J., van Steenbergen, M. 2005. *Dynamic Enterprise Architecture: How to Make It Work.* John Wiley & Sons, Hoboken, New Jersey.

[35] Wieringa, R.J. 2009. Design Science as Nested Problem Solving. In: *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST 2009),* New York, NY, ACM, 1–12.

[36] Winter, R. 2010. Problem Analysis for Situational Artefact Construction in Information Systems. In: tbd (Ed.) *A Tribute to Marco de Marco (forthcoming)* Springer

[37] Winter, R., Bucher, T., Fischer, R., Kurpjuweit, S. 2007. Analysis and Application Scenarios of Enterprise Architecture – An Exploratory Study. *Journal of Enterprise Architecture,* 3, 3 (2007), 33–43.

[38] Winter, R., Fischer, R. 2007. Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. *Journal of Enterprise Architecture,* 3, 2 (2007), 7-18.

[39] Ylimäki, T. 2006. Potential Critical Success Factors for Enterprise Architecture. *Journal of Enterprise Architecture,* 2, 4 (2006), 29–40.

A Cost-Benefit-Based Analytical Model for Finding the Optimal Offering of Software Services

Khin Swe Latt
Technology Management, Economics, and Policy Program (TEMEP)
College of Engineering, Seoul National University
599 Gwanak-Ro, Gwanak-Gu
Seoul 151-744, South-Korea
khinswelatt@gmail.com

Jörn Altmann
Technology Management, Economics, and Policy Program (TEMEP) & Department of Industrial Engineering
College of Engineering, Seoul National University
599 Gwanak-Ro, Gwanak-Gu
Seoul 151-744, South-Korea
jorn.altmann@acm.org

ABSTRACT
In this paper, we introduce an analytical model for maximizing social welfare, which can be used for finding the optimal offering of a set of software services. The analytical model also explains the impact of service flexibility on customer's selection of business services and on the revenue of service providers. The analytical model is based on a utility model and a cost model. The cost model uses the number of lines of code as the basic measure for cost and applies linear and polynomial cost functions. The utility model is derived from a customer-provider relationship model, which relates the user's utility to the functionality of business services. The result of the analytical model shows that the distribution of functions of an existing business service to a large number of new business services does not generate any additional revenues for the service provider from existing customers. Instead, additional revenue is generated through the offering of business services with fewer functions at lower price. This business services attract customers, which could not afford the original software service of the provider. The result of the analytical model also shows that there is an optimal number of business services that maximizes the net utility of customers.

Keywords
Service-oriented architectures, economics of digital products, business process analysis, business service, customer satisfaction model, software services, service science, customer preferences, social welfare maximization and profit maximization of software service development, cost modeling, economics of service decomposition and service composition.

1. INTRODUCTION
The development of new software services is challenging in three ways: First, software service development is usually costly to the software vendor, who also may face the risk of increasing the complexity for the user and the risk of not achieving service flexibility at all [8]. Service flexibility is defined here as the possibility of users to adapt their business process according to their needs. Second, the existing service systems in enterprises are composed of complicated processes, which are interdependent to each other, making it difficult to separate these service systems into software services [11]. Third, the individually provided services are strongly tied to a specific service provider, which limit service composition (i.e., reduce service flexibility) for customers. This situation is known as the business process silo problem [6].

To address these challenges as much as possible, enterprises seek to apply the concept of software reuse. This concept reduces cost, the time to market, and the response time to changes in customer demand. However, it requires the decomposition of existing software into modular software components. To support this, Bennett et al. propose a dynamic service composition architecture [4]. It supports the development of software that is capable to meet changing business needs. An analysis of further software component concepts has been conducted by Kraemer [14].

Service decomposition is useful if a customer requests a new service, which requires a fraction of what the original software service can deliver. It offers an option for substituting a complex service with a set of simple services, potentially reducing costs and improving service flexibility. Decomposition of services into basic services can also benefit service providers. Any change in customer demand does not increase the business risk for the service provider and, therefore, has a low impact on the service provider's business. At the same time, by using these modular software components as basic building blocks, several new and innovative services can be composed [8].

As a preferred technology that supports these concepts, Web services have been chosen. Web services allow dynamic service compositions [15][18]. In a wider context, Web services belong to an emerging technology concept, which is called service-oriented computing (SOC) [14][20].

In order to enable customers of software services to align their business processes with these set of IT services, a new discipline called service science emerged. Besides the technology suites (e.g., Web services), it comprises business process management and performance assessment. Service science also deals with the formalization of interactions between services, allowing enterprises to evaluate the impact of business services on their business processes [23].

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

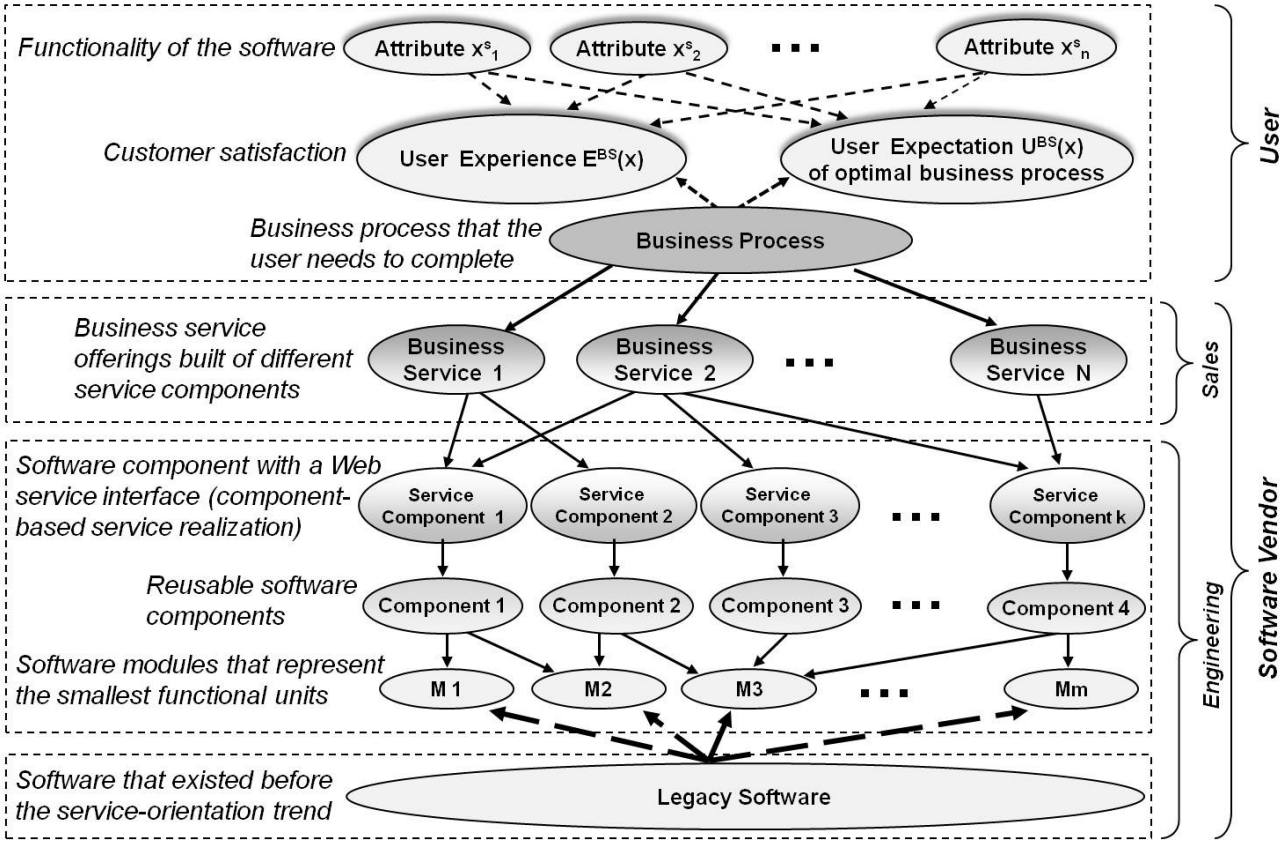


Figure 1. Service composition model.

Within this service science framework, we analyze the costs and the value creation of software services. In particular, we propose a conceptual model for software composition and software decomposition (called service composition model), which allows analyzing the value creation (i.e., user utility) in terms of flexibility and cost (Figure 1). In particular, the model conceptualizes the value creation of business processes. The model also exhibits the dependencies of business services on the composition of service components. In particular, the business service functionality depends on the service components composed and their modularized software components. The software modules are the result of the decomposition of legacy software.

As Figure 1 shows, this conceptual model can be divided into a user side and a software vendor side. From the user point of view, the user selects the business services, which generate the highest utility, and integrates (composes) these business services into his desired business process. The selection of business services is based on the user's requirements, which could be a set of functions (attributes). The user's satisfaction is determined by subtracting the user's experience of the delivered business services from his expectation of the business services offered.

From the software vendor (provider) perspective, the provider tries to fulfill the user requirements by designing business services accordingly. This task is part of the sales/marketing department of a software company. These business services are created from service components (Figure 1). A business service is a workflow of service components. The pricing of the business services is up to the sales department [21]. In order to maximize profit, the provider has to set these prices carefully and has to lower the cost of the service components, which are created by the engineering department.

This separation between business service creation and service component creation provides two advantages. First, it enables the provider to respond quickly to changes in user requirements. Second, the decomposition of existing software into services allows the provider to lower the cost of service creation. By decomposing software functionality of existing software into small, atomic service components, these service components can work as basic building blocks for new business services and, therefore, reduce development costs.

In this paper, the service composition model is used to describe the cost and benefits of service composition. We assume that the software vendor decomposes existing complex software into small, independent units (modules), representing a unique and single function. Then, the provider combines these small units into more complex, composite software components in accordance with software development requirements (e.g.,

software reusability, cost). Consequently, this process incurs costs for service decomposition, costs for combining modules into software components, and cost for wrapping these software components into service components (e.g., Web services), and the costs of building service component workflows. This paper explains these costs in detail and describes how the cost for services, which are composed of small and independent modules of legacy software, can be calculated. Furthermore, this paper explains the costs for composing business services (i.e., the cost for integrating business services into business processes).

The objective is to find a balance between the costs of offering a large number of business services and meeting the requirements of customers, which is a high flexibility in adapting their business processes (i.e., in adapting the workflow of their business services). Within this paper, we provide a solution to this problem by introducing an analytical model for optimizing service offerings.

The remainder of the paper is organized as follows: Section 2 describes the relationship between the customer and software vendor (provider) and explains how customer satisfaction influences the revenues of providers within the software services area. Based on this, utility functions for business services and business processes are introduced. In section 3, we introduce the cost estimation model for software services. It includes the definition of cost functions for business service components, business services, and business processes. The model can deal with workflows. Using this model, section 4 describes the net utility maximization problem, the provider profit maximization problem as well as the social welfare maximization problem for software services. The final section briefly discusses the results and concludes this paper.

2. UTILITY MODEL

2.1 Customer-Provider Relationship

In order to understand the value chain of the service composition model, we develop and analyze a customer-provider relationship model that is based on works of [2][10][13][25]. In particular, it helps defining the framework for the utility model and the cost model.

Vargo & Lusch (2004) proposed the service-dominant logic (S-D Logic) [25], which defines services as the application of specialized competences (i.e., knowledge and skills) through deeds, processes, and performances for the benefit of another entity or the entity itself. S-D logic further assumes that all economies are service economies, all businesses are service businesses, and customer and provider always co-create value (i.e., customers participate in the service creation process). Since the customer-provider relationship can be described as a long-term and dynamic process, the interactions (e.g., pre-sale and post-sale) between customer and provider is very important [13]. Alter (2008) stated that customer satisfaction is affected by the complete set of activities, responsibilities, and experiences that typical customers associate with acquiring, receiving, and benefiting from a particular service [2]. Heskett et al. (1994) discussed about the service-profit chain model, which exhibits the relationships between profitability, customer loyalty, employee satisfaction, loyalty, and productivity [10]. By using these concepts, we develop the customer-provider relationship model for the software services area as shown in Figure 2. It is a more detailed view of the relationship between the user and the software vendor (i.e., sales department), which is shown in Figure 1.

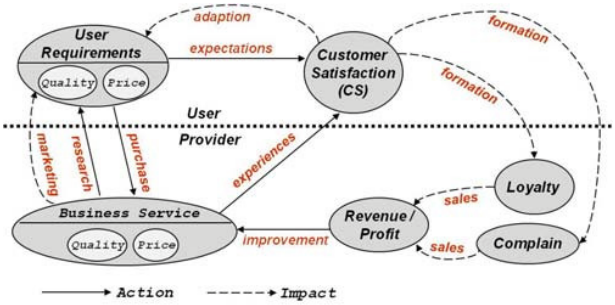


Figure 2. Customer-provider relationship model.

Within this model, customer satisfaction is defined as the satisfaction towards the business services the customer has consumed. The satisfaction is impacted by the user's experience of the consumed business service and the user's requirements for the service. The different components of the general customer-provider relationship model, as shown in Figure 2, are defined as follows:

User requirements are the needs of the user expected to be fulfilled by the provider (El-Kiki & Lawrence [9], and Lee & Ahn [16]). In our model, the user requirement is the degree of flexibility at a certain price. Price is the price of the business service, which is the cost of the service to the customer. Flexibility is defined as the ability to adapt the business process to changes in the business.

Customer satisfaction describes how the customer is satisfied with a business service provided. Customer satisfaction is defined as the difference between the utility from consuming a business service with a certain set of attributes and the expected utility from this business service.

Customer loyalty refers to a consumer's commitment to repurchase a preferred service of the same provider in the future again.

Customer complaint refers to the user's dissatisfaction with a service provided (i.e., the service did not meet the user's requirements).

Business service is the service (product) created by the provider to fulfill user requirements. The user experiences the quality of the service.

Revenue is the income from the business service sold. **Profit** is the gain from selling the services after deducting all expenses incurred through the creation of the service.

Figure 2 shows that providers identify user requirements for business services through market research. Based on this information, providers try to address users' requirements for business services by producing and marketing services in accordance with the user requirements. Created services satisfy the user, if services meet the customer's expectations about quality of service and price.

Customer satisfaction is largely influenced by the net value provided to the consumer. A high customer satisfaction leads to user loyalty to services and providers. Customer loyalty makes the customer decide to repurchase services of the same provider in the future. Consequently, it increases the revenues of providers and retains customers. Customer satisfaction also impacts the user requirements, since satisfaction of a customer determines its expectations of the quality of future services. Customer expectation increases, if user requirements have been met in terms of price and flexibility. Customer dissatisfaction, however, can also lead to complaints of customers. This will happen, if services do not fulfill customer expectations. As a consequence, customers might not continue to purchase the services, reducing the revenue of the provider. However, if customers explicitly express their dissatisfaction, it is an opportunity for providers to improve their service offerings. Concluding, customer satisfaction concerns the improvements of business services and maintenance of customer loyalty in competitive markets.

Based on this model, our customer satisfaction model for software services is defined as the difference between the utility E^{BS}, which the customer gains from a service, and the utility U^{BS}, which the customer expected to get from that service:

$$CS = E^{BS}(X) - U^{BS}(X). \qquad (1)$$

The utility U^{BS} is assumed to be 1. The variable X refers to the vector of all possible attributes x^s_j, where j presents a specific attribute of the business service s. These attributes are assumed to be functional attributes (e.g., business functions) or non-functional attributes (e.g., security, quality of service) and independent of each other.

If the customer satisfaction CS = 0, then the customer is fully satisfied. The provider has exceeded customer expectation, if CS is greater than 0. This is possible, if a provider delivers better quality on the non-functional attributes. For functional attributes, the maximum value is CS = 0, showing that the functionality has been delivered. If CS lies between -1 and 0, the customer expectation has not been fulfilled at all. That means, certain functional attributes have not been delivered or some non-functional attributes have a lower quality than expected by the customer.

2.2 Customer Utility Obtained from Business Services

Using utility functions is most appropriate in this context, since it helps identifying the value proposition of software services. In the past, researchers used utility functions for evaluating resource management approaches [22]. In particular, they used utility functions to measure the performance of management systems. Utility functions also have been applied for achieving QoS-aware service composition. The utility functions were used to select the most appropriate services (Alrifai & Risse [1]). Besides, the utility concept has also been used in decision support systems for scheduling tasks (Yang [26], Jimenez A. et al. [12]). In this paper, we use utility functions to describe the functionality of business processes and business services. The business service attributes experienced by the user determine the overall utility obtained from the business service.

To estimate the overall utility obtained from a business service, we need to define the shape of the utility functions, the range of possible service outcomes, and the weighting of attributes, which expresses the relative importance of an attribute of the business service to a customer. The weights for expressing the relative importance are normalized and the sum of all weights is equal to 1. The relative weights could be determined by the user, using, for instance, AHP or SAW [1][27]. The input parameters (i.e., attributes) are assumed to be independent to each other. Then, after having obtained the weights, the overall value of the utility function for a business service can be estimated, using an additive function.

Customer utility $E^{BS}(X)$ for a business service s, which consists of an vector X of n attributes, can be calculated by multiplying the preference weights h^s_j with the utility $v_j(x^s_j)$ of the service attribute x^s_j. A functional attribute x^s_j is 1, if the business service includes this functionality. Otherwise, it is 0. For functional attributes, the utility $v_j(x^s_j)$ is 1, if $x^s_j = 1$. Otherwise, it is 0. A non-functional attribute x^s_j, which may represent response time or throughput (Menasce & Dubey [19]), is expressed as a real number. The utility function $v_j(x^s_j)$ maps the attribute value onto a scale between 0 and 1. Based on these definitions, the customer utility E^{BS}, can be expressed as:

$$E^{BS}_s(X) = \sum_{j=1}^{n} h^S_j v_j(x^S_j), \quad \text{where} \sum_{j=1}^{n} h^S_j = 1. \qquad (2)$$

Looking at the current situation in the software industry, it becomes clear that customers demand more flexibility in the way how they can use their software purchased from a software vendor. To address this need, software vendors create services that have a reduced number of functions. The functions of those business services can easily be combined. Therefore, assuming the total number of functional attributes to be constant, it can be stated that the higher the number of business services is, the higher the flexibility for the customer is. Equivalently, it can be stated that the utility obtained from the flexibility of a set of business services with the same functions as one single business service is higher than the utility obtained from the flexibility of the single business service. The following equation gives an example, in which a single business service s is split into two business services s1 and s2:

$$E^{BS}_S(X_S) + flx(s) < E^{BS}_{S1}(X_{S1}) + E^{BS}_{S2}(X_{S2}) + flx(s1, s2), \quad \text{where } E^{BS}_S(X_S) = E^{BS}_{S1}(X_{S1}) + E^{BS}_{S2}(X_{S2}). \qquad (3)$$

In equation 3, the functions (i.e., functional attributes) offered by the business service s is equal to the functions of s1 and s2. The only difference is in the utility flx(s) and flx(s1,s2) that are obtained from flexibility. The utility of flexibility is larger for s1 and s2 than the utility of flexibility for s.

For calculating the customer utility $E^{BP}(X)$ that can be obtained from a set of business services s involved in the business process BP, we add the utility of all business services, which are purchased by the customer, and the value R that is obtained from executing the business process. $E^{BP}(X)$ also considers the flexibility flx(), which is expressed as the ratio of the number of

functions n and the average number of functions per business process. Consequently, $E^{BP}(X)$ can be written as shown in equation 4:

$$E^{BP}(X) = \sum_{s=1}^{m} E_s^{BS}(X) + [\frac{n}{\frac{1}{m}\sum_{s=1}^{m}\sum_{j=1}^{n} x_j^S} - 1] + R\,, \qquad (4)$$

where m represents the number of business services that are involved in the business process BP. The total number of functions (functional attributes) offered by the service provider is denoted as n. It is assumed to be larger than 0 ($n > 0$) and larger or equal to m ($n \geq m$). The sum of x_j^s calculates the total number of functions of a business service s. If all functions are included in one business service then the service flexibility is 0. If one function is included, then the flexibility equals the maximum, which is n - 1.

3. COST ESTIMATION MODEL FOR SOFTWARE SERVICES

The costs of business services and business processes depend on software components (i.e., service components, or business services) and the composition of those components. Therefore, based on Figure 1, a more detailed illustration of the service composition model, indicating that business processes BP are workflows of business services and that business services BS are workflows of service components, has been developed (Figure 3).

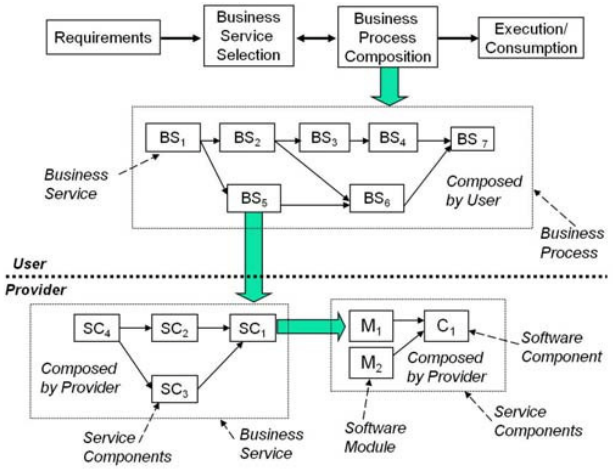


Figure 3. Cost estimation model for a business process.

A service component SC can be obtained by creating a service from a software components C. A software component C is the result of a composition of software modules, which have been obtained from a decomposition of legacy software or have been programmed from scratch. The cost calculation is explained in detail in the following sub-sections.

3.1 Background on Cost Functions

The objective of this cost model is to base the cost estimation on the programmers' effort for developing software services (i.e., for developing modules, software components, service components, business services, and business processes).

Software development cost estimation comprises the entire process of predicting the effort required to develop a software system (Leung & Fan [17]). A precise estimation of cost of a software project can help managers to manage projects adequately.

There are different metrics to size software. Each of those can be used as input to a cost model for software services. The line of code (LOC) is the most popular software sizing metric. For example, Uysal (2008) proposed a COCOMO-based equation that uses the number of lines of code as a parameter [24]. Line of code is also used in a Fuzzy logic model for measuring the software development effort (Attarzadeh & Ow [3]). In this paper, we also use the number of lines of code as the software sizing metric. By counting the number of lines of code, we estimate the provider's effort in providing software services.

A linear cost function expresses cost as a linear function of the number of lines of code. The proposed cost function C for our model is as follows:

$$C(LOC) = A + B * LOC\,, \qquad (5)$$

where A is the fixed cost of producing software (i.e., provisioning a software service), B is the marginal cost of an additional line of code, and LOC refers to the lines of code in programs [17].

In general, the cost function can also be a polynomial function of the number of lines of code. In this case, the cost function would be defined as:

$$C(LOC) = A + B * LOC^{K}\,, \qquad (6)$$

where A is the fixed cost of producing software, B is the marginal cost of an additional line of code, and LOC refers to the number of lines of code in the software. K is an empirically derived constant, which has been identified to be in the range between 1.05 to 1.2 [17]. Within this paper, we set the constant to 1.15.

The difference between a linear cost function and a polynomial cost function is the weighting of the number of lines of code. Because of this fact, there is no difference in choosing one over the other. The only implication is its impact on the selection of the appropriate mathematical method for solving the optimization problems that will be introduced in section 4.

Therefore, these cost functions can be applied to calculate the cost for programming modules, software components, service components, business services, and business processes. These cost functions are the basis for providers to decide on how to structure their software. For example, it can be used to decide on how many service components should be created in order to be able to offer many different business services to consumers. Ultimately, these cost functions are the basis for finding a balance between the costs of offering a large number of business services and meeting the requirements of customers, which is a high flexibility in combining services (i.e., the possibility to re-arrange business processes). In the following subsections, the cost functions for service components, business services, and business processes are described in detail.

3.2 The Cost Function of Service Components

Service components are the basic units for composing a business service. They are created by wrapping software components with service interfaces (e.g., Web services). Software components are, as mentioned, a set of software modules that work together (Figure 1 and Figure 3). These different units of software code are considered for defining the cost function of service components.

The cost of a service component is the sum of the cost for wrapping a software component with a service interface (e.g., Web service), the cost for programming the component based on modules, and the cost of the modules themselves. Therefore, the cost of the service component C^{SC} can be expressed as shown in equation 7:

$$C^{SC} = x^{M} * \sum_{i=1}^{k} LOC_i^{M} + x^{C} * \sum_{j=1}^{l} LOC_j^{C} + x^{WC} * LOC^{WC}, \quad (7)$$

where LOC_i^M is defined as the total number of lines of code of module i, LOC_j^C as the total number of lines of code for component j, and LOC^{WC} as the total number of lines of code needed to wrap a software component with a service interface. x^M represents the cost per module line of code (i.e., the unit is [$ / LOC^M]), x^C represents the cost per component line of code (i.e., the unit is [$ / LOC^C]), and x^{WC} denotes the cost per wrapping line of code (i.e., the unit is [$ / LOC^{WC}]).

3.3 The Cost Function of Business Services

The cost function of business services is defined as the sum of the service components costs and the cost for programming the service components workflow. The cost of constructing a business service C^{BS} can be described with the following equation:

$$C^{BS} = C^{SCWF} = \sum_{l=1}^{d} C_l^{SC} + x^{SCWF} * LOC^{SCWF}, \quad (8)$$

where C^{SCWF} denotes the cost of programming a workflow of service components, representing the cost of a business service. LOC^{SCWF} represents the total number of lines of code used in programming the workflow of service components. x^{SCWF} is defined as the cost per service component workflow line of code. The unit is [$ / LOC^{SCWF}]. C^{SC} represents the total cost of a service component that is used within the business service (equation 7).

3.4 The Cost Function of Business Processes

The cost function of a business process is defined as the sum of the costs of all business services involved and the cost for programming the workflow with business services. This cost for constructing a business process C^{BP} can be expressed with the following equation:

$$C^{BP} = C^{BSWF} = \sum_{s=1}^{m} C_s^{BS} + x^{BSWF} * [LOC^{BSWF}]^{1.15}, \quad (9)$$

where C^{BP} equals C^{BSWF}, representing the cost of programming the workflow of business services. LOC^{BSWF} denotes the total number of lines of code used in programming a workflow of business services. x^{BSWF} represents the cost per business service workflow line of code. The unit is [$ / LOC^{BSWF}]. C^{BS} is the cost of providing a business service, as defined in equation 8. In this case, it is assumed that all business services are build from a disjoint set of service components. In case that business services are build on some identical service components, equation 9 has to be modified such that the C^{SC} of those service components is not counted twice.

Besides, the cost function of business processes is a polynomial function, whereas the two cost functions for programming service components (C^{SC}) and for programming business services (C^{BS}) used linear functions (equation 7 and equation 8). The reason is that the number of service components and software modules is assumed to be fixed. The number of business services used for programming the business process, however, is assumed to be flexible. This is necessary as this paper investigates the effect of the number of business services on the utility of customers and the revenue of providers. A higher number of business services increases the complexity of programming a business process.

4. ANALYTICAL MODEL FOR OPTIMIZING SERVICE OFFERINGS

After having introduced the different cost functions in the previous section, we introduce three optimization problems for software services. The solutions to these optimization problems help answering questions like how many business services should be offered by a software vendor in order to maximize the profit of the vendor.

Economics-based optimization has been proposed by many researchers. Derbel et al. proposed an optimization approach considering user preferences in multi-services IP networks [7]. Yang (2008) proposed a utility-based decision support system, using separate utility functions for time and cost [26]. The objective of optimization in our research is similar. The objective is to maximize the net utility of the parties involved by minimizing the cost and maximizing the flexibility in creating new business services and business processes. In particular, we formulate the following three optimization problems: customer net utility maximization, provider profit maximization, and social welfare maximization.

Within the following optimizations, we only calculate the additional costs that incur if additional business services are offered in addition to an existing software solution. The basis for calculating the additional costs is the cost for providing the software (e.g., a monolithic software solution) that solves the entire business process need of a user.

4.1 Customer Net Utility Maximization

The calculation of the customer net utility that is gained from buying software services can help customers (i.e., businesses) to understand how many business services they should buy. Since a business process generates utility to a customer, we calculate the net utility at the business process level. Therefore, the customer net utility U^{net} is calculated as the utility E^{BP} gained from the business process minus the cost of all business services being

used within the business process and the cost for programming the workflow of business processes, as shown in the following equation:

$$U^{net} = E^{BP}(X) - C^{BP}. \quad (10)$$

Consequently, the net utility maximization problem can be expressed as:

$$\max[E^{BP}(X) - C^{BP}], \quad (11)$$

where $E^{BP}(X)$ denotes the customer utility from consuming the business process BP. C^{BP} represents the cost for provisioning the business process. X is the vector of all service attributes. Using equations 4 and equation 9, equation 11 can be expressed as:

$$\max_m [\sum_{s=1}^{m} E_s^{BS}(X) + [\frac{n}{\frac{1}{m}\sum_{s=1}^{m}\sum_{j=1}^{n} x_j^s} - 1] + R - [\sum_{s=1}^{m} C_s^{BS} + p^{BSWF} * [LOC^{BSWF}]^{1.15}]]. \quad (12)$$

In our model, the cost of programming the business process workflow is only incurred for the customer. Besides, it is assumed that the number of attributes n is fixed. For simplification, we assume that n attributes represent n functional attributes (e.g., software functions). That means that the n attributes are distributed across m business services. Therefore, the more business services exist, the less attributes (i.e., functions) the business service contains on average.

Figure 4 illustrates the different cost factors, the utility, and the customer net utility U^{net}, depending on the number of business services m. It shows that the cost of programming shifts from the software vendor to the user with increasing flexibility.

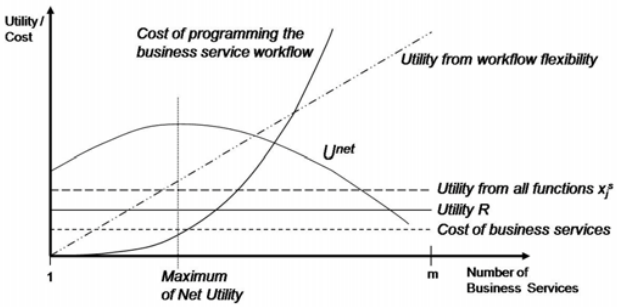


Figure 4. Customer net utility.

In particular, Figure 4 illustrates that there is a net utility maximum with respect to the number of business services. This point is indicated with a vertical line crossing the U^{net} curve. This means that the customer should try to reach this number of business services. Only in this case, the customer can obtain the maximum benefit from service flexibility. The reason is the cost of programming of the business process workflow. It is a significant cost factor that reduces the benefit obtained from service flexibility.

4.2 Provider Profit Maximization

The calculation of the provider profit helps providers to understand the impact of adapting their existing software to a service environment on their revenues. In general, the profit F' of a service provider is defined as the revenue H_S from selling one business service minus the cost C^{BS} of provisioning the business service:

$$F' = H_S - C^{BS}. \quad (13)$$

By applying equation 13 to all business services, the profit F can be calculated by adding the revenue H_s from each business services sold minus the costs for provisioning the business services s:

$$F = \sum_{s=1}^{m} [H_s - C_s^{BS}]. \quad (14)$$

A business service comprises different functionality with different costs. In order to increase customer satisfaction (i.e., improve the flexibility for the customer in creating his business processes), the service provider should produce as many business services as possible. However, since the cost for creating business services also increases, an optimum has to be found. The question is how many business services should be produced in order to maximize the profit. Thus, the provider profit maximization problem is:

$$\max_m \sum_{s=1}^{m} [H_s - C_s^{BS}]. \quad (15)$$

Applying equation 8, equation 15 can be re-written as:

$$\max_m [\sum_{s=1}^{m} H_s - \sum_{s=1}^{m} [p^{SCWF} * LOC_s^{SCWF}] - \sum_{l=1}^{d} C_l^{SC}], \quad (16)$$

where m represents the number of business services and d the total number of service components that have been created. H_s is the revenue from business service s. p^{SCWF} denotes the cost for one line of code and LOC^{SCWF} the cost for programming a service component workflow. C_l^{SC} is the cost of service component l.

For our analysis, we assume that the cost of all service components is equal. This is justified if we assume that the entire software is split into code segments of equal size and that the existing software had been structured into components following software management principles from the very beginning. Furthermore, we assume that the total number of service components is fixed. Any business service is created based on these fixed number of service components.

The costs for programming a business service (i.e., a workflow of service components) is assumed as high as the cost for providing the monolithic solution. Because of that, the cost of business service programming increases linearly. (Note: The cost would

actually be decreasing with increasing number of business services, since programming effort is shifted to the user. Therefore, this programming effort of the user is depicted as a polynomial increasing curve in Figure 4.) We consider this an upper bound.

The cost is incurred once for each business service, independent of the number of sales of the business service. Therefore, assuming that there is only one customer, who would have belonged to the customer base of the monolithic software solution, the profit from all business services is the same, Y. That means that Y is independent of the number of business services offered by the provider:

$$\sum_{s=1}^{m} H_s = Y \quad \forall m . \qquad (17)$$

These discussion results are illustrated in Figure 5.

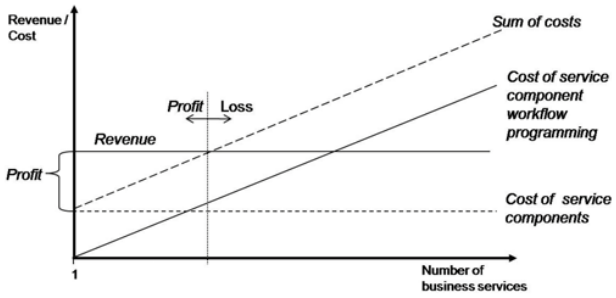


Figure 5. Provider profit in case of a single user.

In particular, Figure 5 depicts that the provider does not get any additional revenue from offering more business services (i.e., from a set of business services with the same functionality as the monolithic software) to its existing customer base. After creating a certain number of business services, the provider would even incur some losses. Even if we would assume that the provider sets the price such that it catches the net utility U^{net} that the customer gets from service flexibility (Figure 4), business service creation were finally to result in losses. Therefore, we can state that adapting existing software to a service environment does not allow extracting any additional surplus from existing customers. It simply helps to increase customer satisfaction and stay competitive within the software market.

However, the revenue of providers H_s will change, if we assume that the software vendor could attract new customers. Those new customers are businesses, who could not afford to buy business services that provide the full functionality but have sufficient funds available to buy business services with less functionality. Figure 6 illustrates the customer distribution with respect to the budget and the number of business services consumed.

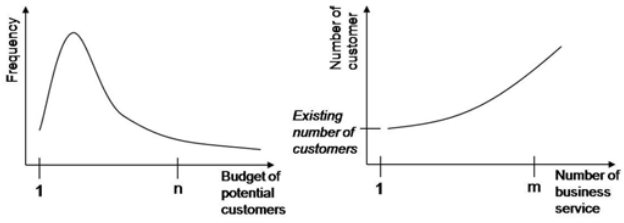


Figure 6. Budget distribution and customer distribution

Although we can only assume the amount of increase in customers through offering new business services (Figure 6), i.e., business services with less functionality than the monolithic software, any increase will have an impact on the revenue of the provider. Figure 7 illustrates this impact. It shows the revenue of the provider, the profit of the provider, and the same cost as the cost shown in Figure 5.

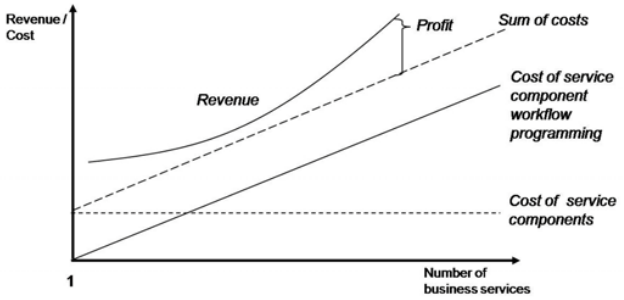


Figure 7. Provider profit in case of many customers.

Figure 7 shows that additional revenue is generated from offering different business services, which differ in their combination of functions. The revenue curve, which is based on the one in Figure 6, is justified, since a large variety of business services can address a larger variety of needs of customers.

4.3 Social Welfare Maximization

Social welfare SW is the sum of all user's benefits (Courcoubetis & Weber [5]), i.e., the sum of all customer surplus and producer surplus. It is the sum of the net utility of all customers and the provider profit PP:

$$SW = \sum_{i=1}^{u} U^{net} + PP . \qquad (18)$$

The equation 18 is equivalent to the utility of all customers minus the costs of programming a business service workflow for each customer, the cost of programming all service component workflows, and the cost of producing all service components. Since the revenue of the provider is equal to the costs of business services for all customers, they are not shown in the social welfare equation:

$$SW = \sum_{i=1}^{u}[\sum_{s=1}^{m} E_s^{BS}(X) + [\frac{n}{\frac{1}{m}\sum_{s=1}^{m}\sum_{j=1}^{n} x_j^s} - 1] + R_u] - \sum_{i=1}^{u}[p^{BSWF} * [LOC^{BSWF}]^{1.15}] - \sum_{s=1}^{m}[p^{SCWF} * LOC_s^{SCWF}] - \sum_{l=1}^{d} C_l^{SC}, \quad (19)$$

where the first term of the equation represents the utility of the customers, while the second, third, and fourth term shows the three cost factors. Note, the utility used is the sum of the utility of all u customers that are served by the service provider. R is assumed to be an average utility across all customers. Then, the social welfare maximization can easily be formulated as:

$$\max_m SW. \quad (20)$$

Since the utility function is concave and the cost function is convex, the social welfare maximization problem can be solved by applying the Lagrange approach.

Having the same assumptions as in the previous two sections, Figure 8 illustrates the social welfare from offering business services within this service system.

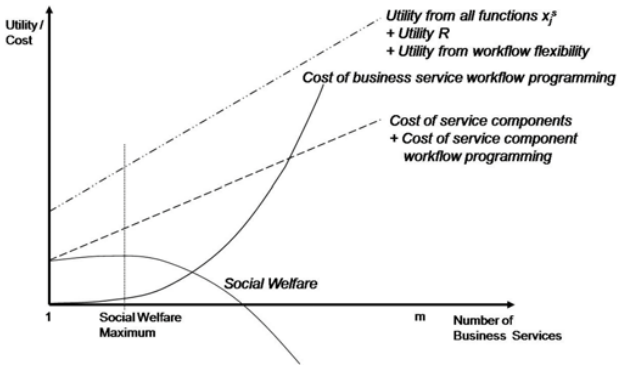


Figure 8. Social welfare.

In particular, it illustrates that in a service-based industry, the flexibility of services and costs of delivering the flexibility impact the social welfare of the system of customers and provider. In order to improve the social welfare of customers and provider, it is necessary to balance the amount of functionality offered with business services. That means, on the one hand, customer satisfaction and profit increase can be achieved by splitting functionality into many different business services. On the other hand, the cost for splitting the existing software into many business services increases the cost for consumers. Therefore, the provider has to find a set of business services that satisfies the customers' demand for flexibility at a reasonable cost. Only then, a pre-requisite for a successful software service industry has been fulfilled.

5. DISCUSSION AND CONCLUSION

In this paper, we introduced a cost-based analytical model for analyzing software service provisioning in a service system. The benefit of this study is the gain in understanding of how customers and service provider interact in an on-demand service-oriented business environment, how costs vary with respect to the customer requirement in flexibility, and how to optimize the consumption and provisioning of business services.

The analytical model is based on a customer-provider relationship model, a utility model, and a cost model for service provisioning. The customer-provider relationship model describes the interdependencies between customer satisfaction, user requirements, business services, and service quality. The utility model defines the customer utility functions for business services and business processes. The utility function allows describing service composition with respect to customer satisfaction for functional and none-functional attributes. The cost model describes in detail the cost of business processes, business services, and business service components. It explains how the costs of services are incurred, especially focusing on costs of service workflow creation. The service provisioning cost is assumed to depend on the programming effort, which is measured in units of line of code.

In particular, the analytical model describes the interdependency between revenue and service provisioning. It explains the conditions under which an increase in the number of service offerings would be profitable and under which an increase would incur losses. The model also explains how to determine the maximum of the customer net utility.

Finally, the analysis of the social welfare of the service system shows that it is necessary to balance the amount of functionality offered per business service. On the one hand, customer satisfaction and profit increase can be achieved by splitting functionality into many different business services, meeting customers' demand for flexibility. On the other hand, for consumers, the cost of integrating business services into business process has to be limited by offering consumers a minimum set of business services. Only if both is achieved, social welfare is maximized.

6. ACKNOWLEDGMENTS

This research was supported by the International Research & Development Program of the National Research Foundation of Korea of the Ministry of Education, Science and Technology of Korea (Grant number: K21001001625-10B1300-03310).

7. REFERENCES

[1] Alrifai, M. and Risse, T., 2009. Combining global optimization with local selection for efficient QoS-aware service composition. *18th International Conference on World Wide Web, WWW2009*.

[2] Alter, S., 2008. Service system fundamentals: work system, value chain, and life cycle, *IBM Systems Journal*, 47, 1.

[3] Attarzadeh, I. and Ow, S.H., 2009. Software development effort estimation based on a new fuzzy logic model. *International Journal of Computer Theory and Engineering*, 1, 4.

[4] Bennett, K., Layzell, P., Budgen, D., Brereton, P., Macaulay L., and Munro, M., 2000. Service-based software: the future for flexible software. *IEEE APSEC 2000, Asia-Pacific Software Engineering Conference*, IEEE Computer Society, 214-221, Singapore.

[5] Courcoubetis, C., Weber, R., Coe, M., 2003. Pricing communication networks: economics, technology, and modeling, Wiley-Interscience Series in Systems and Optimization, Wiley, ISBN 978-0470851302.

[6] Demirkan, H., Kauffman, R. J., Vayghan, J. A., Fill, H.-G., Karagiannis, D., and Maglio, P. P., 2008. Service-oriented technology and management: perspectives on research and practice for the coming decade. *Electronic Commerce Research and Applications*, 7.

[7] Derbel, H., Agoulmine, N., and Salaun, M., 2007. Service utility optimization model based on user preferences in multiservice IP networks. *IEEE DANMS*, Washington, D.C., USA.

[8] Dörr, E and Winiwarter, W., 2004. Decomposition and reuse of mobile services. *6th International Conference on Information Integration and Web-based Applications and Services*, OCG Press, Vienna, Austria.

[9] El-Kiki, T. and Lawrence, E., 2007. Mobile user satisfaction and usage analysis model of mGovernment services. *Second European Conference On Mobile Government*, Brighton, UK.

[10] Heskett, J. L., Jones T. O., Loveman G. W., Sasser, W. E., and Schlesinger, L. A., 1994. Putting the service-profit chain to work. *Harvard Business Review*. 164–174.

[11] IBM SSME Symposium, 2007. Succeeding through service innovation: a service perspective for education, research, business and government. *Cambridge Service Science, Management and Engineering Symposium*, Churchill College, Cambridge, Uk.

[12] Jimenez, A., Rıos-Insua, S., and Mateos A., (2002). A decision support system for multiattribute utility evaluation based on imprecise assignments. *Decision Support Systems* 36, 65–79.

[13] Kowalkowski, C. and Malmgren, M., 2008. Dynamics of value co-creation in buyer-supplier relationships. *Australian & New Zealand Marketing Academy Conference, ANZMAC*, Sydney, Australia.

[14] Krämer, B. J., 2008. Component meets service: what does the mongrel look like? *Innovations Syst Softw Eng*, 4, 4, 385-394.

[15] Krämer, B. J., Papazoglou, M. P., Schmidt, H. W., 1998. Information systems interoperability. *Advanced software development series (6)*. Research Studies Press, Taunton, Somerset, England. ISBN 0-86380-228-1.

[16] Lee, D.-J. and Ahn, J.-K. ,2007. Factors affecting companies' telecommunication service selection strategy. *Omega*, 35, 5, 486-493. doi:10.1016/j.omega.2005.09.004.

[17] Leung, H., and Fan, Z., 2002. Software cost estimation. *Handbook of Software Engineering*, Hong Kong Polytechnic University.

[18] Liu, D. and Deters, R., 2008. Management of service-oriented systems. *SOCA*, 2, 51–64. DOI 10.1007/s11761-008-0028-1.

[19] Menasce, D. A and Dubey, V., 2007. A heuristic approach to optimal service selection in service oriented architectures. *IEEE International Conference on Web Services (ICWS)*.

[20] Papazoglou, M. P., Traverso, P., Dustdar, S., Leymann, F. Krämer, B.J., 2006. Service-oriented computing: a research roadmap. *Dagstuhl Seminar Proceedings*, 05462, Schloss Dagstuhl, Germany.

[21] Rohitratana, J. and Altmann, J., 2010. Agent-based simulations of the software market under different pricing schemes for software-as-a-service and perpetual software. *Workshop on the Economics of Grids, Clouds, Systems, and Services, GECON 2010*, Springer LNCS, Ischia, Italy.

[22] Rohloff, K., Ye, J., Loyall, J., and Schantz, R., 2006. A hierarchical control system for dynamic resource management. In: *2006 IEEE Real-Time and Embedded Technology and Applications Symposium*, San Jose, CA, USA.

[23] Spohrer, J., Anderson, L. C., Pass, N. J., Ager, T., and Gruhl, D., 2008. Service science. *Journal of Grid Computing*, 6, 3, 313–324 . DOI 10.1007/s10723-007-9096-2.

[24] Uysal, M., 2008. Estimation of the effort component of the software projects using simulated annealing algorithm. *World Academy of Science, Engineering and Technology*, 41.

[25] Vargo, S. L. & Lusch R. F. 2004. Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68, 1–17.

[26] Yang, I.-T., (2008). Utility-based decision support system for schedule optimization. *Decision Support Systems*, 44, 595–605.

[27] Zahedi F., and Ashrafi, N., 1991. Software reliability allocation based on structure, utility, price, and cost. *IEEE Transactions on Software Engineering*, 17, 4.

On Measures of Behavioral Distance between Business Processes

Jörg Becker
University of Münster
Leonardo-Campus 3
D-48149 Münster
+4902518338100
becker@ercis.de

Philipp Bergener
University of Münster
Leonardo-Campus 3
D-48149 Münster
+4902518338067
bergener@ercis.de

Dominic Breuker
University of Münster
Leonardo-Campus 3
D-48149 Münster
+4902518338062
breuker@ercis.de

Michael Räckers
University of Münster
Leonardo-Campus 3
D-48149 Münster
+4902518338075
raeckers@ercis.de

ABSTRACT
The desire to compute similarities or distances between business processes arises in numerous situations such as when comparing business processes with reference models or when integrating business processes. The objective of this paper is to develop an approach for measuring the distance between Business Processes Models (BPM) based on the behavior of the business process only while abstracting from any structural aspects of the actual model. Furthermore, the measure allows for assigning more weight to parts of a process which are executed more frequently and can thus be considered as more important. This is achieved by defining a probability distribution on the behavior allowing the computation of distance metrics from the field of statistics.

Keywords
Business Process Modeling, Process Similarity Measurement, Business Process Distance Measurement, Behavioral Process Similarity

1. INTRODUCTION
Business Process Models (BPMs) enable organizations in Public and Private Sector to get a transparent overview over the relevant extracts of their organization. BPMs are used to gain clarity about the logical sequence of activities in an organization. They are also applied to describe the resulting products and services, the required resources and data, as well as the involved organizational units. They are discussed in Information Systems (IS) literature as a tool to evaluate the overall performance of an organization [1] and to support business process reorganization and optimization by both capturing the as-is situation and designing the to-be process.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

The comparison of business processes with the help of a similarity measure is often an important component in approaches supporting business process management. Examples are the integration of business processes in scenarios of distributed modeling, the identification of similar processes in a huge set of company process models, e.g. to leverage synergy effects, or the benchmarking of processes between organizations [2]. Furthermore, it can be applied to control reorganization projects by comparing to-be and implemented as-is processes or in the context of process mining [3], where the actual behavior of a business process is compared with process models or certain business rules to check for process compliance [4].

The contribution of this paper is to apply distance measure approaches from statistics on the field of business process management. Though a distance measure can be of use in many different contexts, an application scenario in which our approach is of particular interest is the comparison of designed to-be process models with their actual implementations to check for conformance.

In our approach, we introduce a set of related distance measures for business processes. There are two basic characteristics underlying these distance measures. First, it aims at measuring the distance with respect to the behavioral aspects of the business processes. Aspects regarding the modeling language employed to represent the process and the constructs defining this behavior shall be excluded from consideration. This is especially applicable in situations where the information about processes is taken from log files, e.g. to check for conformance or compliance. Second, it takes into account the frequency of the observed behavior, i.e. the number of executions for activities. This allows weighing the important (more frequent) parts of the process stronger than the unimportant ones.

These goals are achieved by taking a probabilistic perspective on the behavior of the process. All the different sequences of activities that may be observed are extracted from the process together with the corresponding probabilities. This delivers a probability distribution on these sequences. Then, using a well-known distance measure from the field of statistics that is based on the so called Bhattacharyya coefficient, our notions of distance between business processes are defined and illustrated by examples.

Often, measures of distance and similarity can be used interchangeably as high distance means low similarity. In our case, the distance will lie between zero and one, making the

question of whether we propose distance or similarity measures a matter of definition. Since the measure we use is often defined as a distance we stay with this convention.

The remainder of this work proceeds as follows. Section two discusses different notions of similarity between business processes by providing an overview of the related work. Section three then introduces our behavioral representation of business processes and derives a probability distribution over its behavior. Following that, section four defines our distance measures and provides examples on how to compute them. Section five then illustrates the defined measures by applying them to an example. Finally, section six concludes and gives an outlook in future research.

2. NOTIONS OF SIMILARITY - RELATED WORK

The first problem arising in business process similarity calculation is matching of elements in different models. Usually, this is based on the labels assigned to the elements, which is why this problem is often referred to as label matching [5]. Due to the use of natural language in these labels, matching them is by no means a trivial task. Simple methods like computing the distances between strings can be employed here. A good overview on such methods is given in [6]. More advanced techniques employ for instance machine learning algorithms or lexical systems such as WordNet [7] to identify higher level relation between words used in the labels. A good survey on approaches to this problem can be found in [8]. In this work, we take the label matching as given. Any of the above mentioned methods could be combined with our approach.

Once a matching of elements is achieved, the computation of business process similarity can be done with respect to two different aspects. On the one hand, the focus can be laid on the actual graphs by which the business processes are represented. This is called the structural aspect of similarity. On the other hand, one can abstract from the particularities of the graphs and restrict the comparison to the interplay of the activities performed in the processes. This is called the behavioral aspect of similarity.

Approaches on structural similarity naturally lead to the well-known field of graph matching, which has a longstanding tradition in computer science [9]. An important concept in that research area is the edit distance of two graphs. It is defined as the cost of transforming one graph into the other by means of elementary change operations like inserting, substituting or deleting nodes. Applications of this concept to the area of business processes can be found for instance in [10, 11].

However, not all approaches focusing on structural aspects use graph matching techniques. In [12], so called features are extracted from the business processes under consideration. Based on the arising feature space, a similarity metric is being derived. In [13] a similarity flooding algorithm is applied to match one process graph onto the other.

One characteristic of these purely structural measures of similarity is that they can identify differences between models even if they describe exactly the same behavior, which may be wanted or unwanted depending on the context of use. Nevertheless, approaches using structurally inspired techniques can also address behavioral aspects. This can, for example, be achieved by building a graph of the process behavior in such a way that ideas from traditional graph matching like edit distances can then be applied to these representations [14, 15].

When viewing business processes – in contrast to the approaches described above – from an entirely behavioral point of view, one is interested in the sequences of activities that a particular business process allows. A widely known approach addressing this aspect is based on a causal footprint representation of the process behavior [16, 17]. This is a graph capturing the possible ordering relations, which means that it specifies which activities can follow on each other and which cannot. The similarity of two processes is then calculated by embedding the causal footprint into a vector space and computing the cosine of the angle between these vectors. A comparable representation of the process behavior is used in [18], where a matrix of so called transition adjacency relations is build. It contains one if a transition can be observed directly after the other and zero if not. The similarity of the behavior is then measured by the similarity of these matrices.

Other approaches taking a behavioral view utilize another traditional field of computer science, namely automata theory [9]. In this area, automata are used to describe languages consisting of words over an alphabet of symbols. Applied to BPM, the behavior of a business process can be understood as a set of activity sequences.

A fundamental concept to compare automata is that of bisimulation [19], which effectively means that, when two automata are bisimilar, their behavior cannot be distinguished by an external observer. Many different notions of bisimulation have been developed over time, but most of them deliver binary yes/no answers only. However, methods for computing the similarity of automata have also been developed that can be interpreted as fuzzy versions of bisimulation, measuring the degree to which the relation holds. See [20, 21] for examples as well as [22] for an application to workflow modeling.

Instead of comparing the automata of languages, one can directly compare the languages themselves, i.e. the sets of possible words [23, 24]. This again involves a notion of edit distance, but this time between languages. In a very rigid case, the distance between two languages is defined as the lowest distance between any of their words. In the context of business processes, this would already result in 100% similarity if there is a single activity sequence shared by the models. To relax this, probabilities can be assigned to each of the words of a language, in which case the comparison can be based on all words of a language, weighted by their probabilities.

The introduction of probabilities assigned to words is, in a sense, closely related to an approach of business process similarity calculation that is, in contrast to any other approach discussed so far, based on observed instances of business processes [25]. The aim of this method is to explicitly address frequent aspects of a business process stronger than infrequent ones. In contrast to our work, it computes two one-sided measures of similarity, called behavioral precision and recall. They measure how well the behavior of one process fits to the other and vice versa.

3. PROBABILITY-WEIGHTED LABELED TRANSITION SYSTEMS

As a model of the business process behavior, we will use a labeled transition system [26] equipped with probabilities on the transitions. We will call the model a probability-weighted labeled transition system (PLTS). Any business process that is supplemented with probabilities on the paths between activities could be transformed into such a representation. The advantage of the PLTS is that the possible paths through the process and the probability of taking it can easily be seen.

A probability-weighted labeled transition system shall be defined as the 6-tuple $PLTS = (S, T, \Sigma, s_o, S_F, p)$ with

- S being a finite set of states
- $T \in (S \setminus S_F) \times \Sigma \times S$ being a finite set of transitions between states emitting an activity label $x \in \Sigma$
- Σ being a finite set of activity labels
- $s_o$being the unique initial state
- S_F being the set of final states which cannot be left by a transition
- $p: T \rightarrow]0,1]$ being a function assigning a positive probability to each of the transitions.

The probability weighting function p is defined such that, for each state s, the sum of the probabilities of transitions leaving this state sums to one:

$$\forall s \in S: \sum_{t=s \xrightarrow{x} \acute{s} \in T} p(t) = 1 \quad .$$

Here, $s \xrightarrow{x} \acute{s}$ denotes the transition $t = (s, x, \acute{s})$. Given this definition of a PLTS, we can define a path of length $n \in \mathbb{N}$ through it, which shall be an n-tuple of transitions $t^{(n)} = \left(s_0 \xrightarrow{x_1} s_1, s_1 \xrightarrow{x_2} s_2, \ldots, s_{n-1} \xrightarrow{x_n} s_n\right)$ with $s_n \in S_F$. The path starts at the initial state, wanders through the PLTS and ends in one of the final states.

A path of length n gives rise to a sequence of activity labels $X = (x_1, \ldots, x_n) \in \Sigma^n$ having length n by reducing the path to the activity labels the transitions emit. The enumerated behavior can then be defined as the set $EB(PLTS)$ of all possible activity sequences:

$$X = (x_1, \ldots, x_n) \in EB(PLTS) \Leftrightarrow$$
$$\exists t^{(n)} = \left(s_0 \xrightarrow{x_1} s_1, s_1 \xrightarrow{x_2} s_2, \ldots, s_{n-1} \xrightarrow{x_n} s_n\right) \in T^n : s_n \in S_F$$

This means that the enumerated behavior of a PLTS consists of all the activity sequences that can arise from taking any path through the PLTS. Note that the set can be infinite in the case that the PLTS has loops.

The probabilities $p(t)$ assigned to the transitions t induce an assignment of probabilities $\hat{p}(t^n)$ to paths. It is defined as the product of all the probabilities of the transitions that belong to this path.

$$\hat{p}\left(t^{(n)}\right) = \prod_{i=1}^{n} p\left(s_{i-1} \xrightarrow{x_i} s_i\right)$$

This assignment of probabilities actually induces a probability distribution over paths, as one can see from the following inductive argument. Assume the simplest PLTS is given, consisting of only the initial state $s_o$and a set of final states S_1. As the transitions are arbitrary, this PLTS can have any number of paths with length one. All these paths will leave $s_o$and enter one of the states S_1. Since the probabilities for all transitions leaving s_o must sum to one, they define a distribution over all possible paths. Now assume that, for any PLTS having paths of at most length n, a probability distribution over the paths is induced. Then, by adding transitions from the final states of this model to new states, any PLTS can be created that has a path length of $n + 1$. All paths previously having length n and now having length $n + 1$ will be multiplied by the respective probability of an

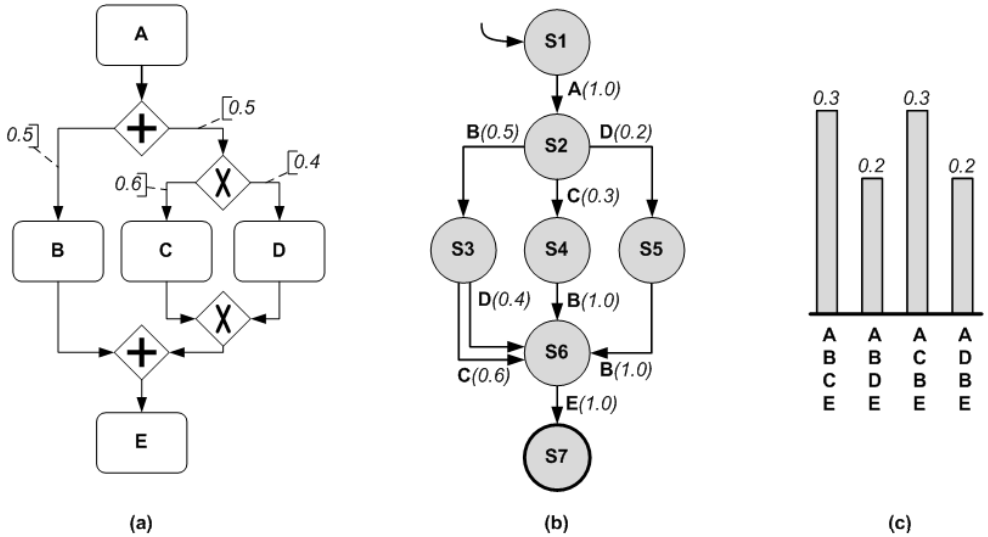


Figure 1: (a) defines a business process in BPMN notation, (b) represents the same process as a PLTS, (c) illustrates the corresponding distribution over activity sequences.

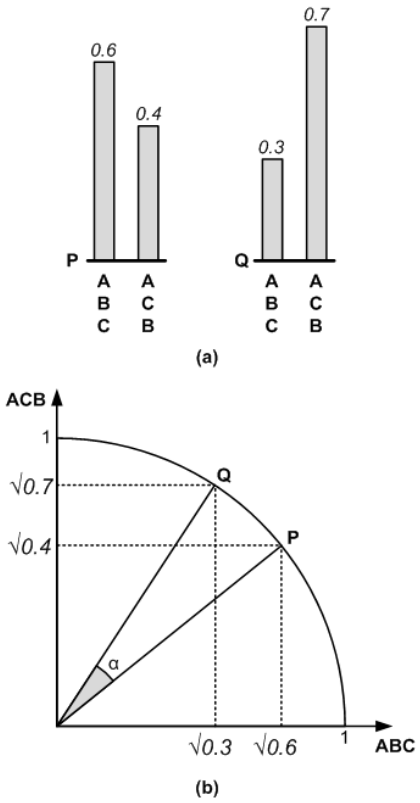


Figure 2: (a) defines two distributions over activity sequences, (b) represents the Bhattacharyya coefficient as the angle between the distributions projected onto a circle.

additional transition to a new state, which sum, for each of the final states of the model with maximum length *n*, to one. Thus, when summing over all the paths, the sum will again turn out to be one.

Finally, given the probability distribution $\hat{p}\ \left(t^{(n)}\right)$ over paths, we can define a probability distribution $P(X)$ over the set Σ^* of all possible activity sequences of any length. This is easily accomplished by assigning to each sequence X the sum of all probabilities of paths that emit exactly the activity symbols of this sequence:

$$P: \Sigma^* \to [0,1], with\ P(X) = P(x_1, \dots, x_n) = \sum_{t^* emitting\ X} \hat{p}(t^*)$$

Any other of the infinitely many paths in Σ^* is assigned zero probability. This distribution shall be the probabilistic behavior of a business process.

As an example, consider the business process in figure 1 (a) and (b), giving rise to the distribution shown in figure 1 (c).

4. MEASURES OF BEHAVIORAL SIMILARITY

4.1 The Bhattacharyya Coefficient

The main idea of our behavioral similarity measures will be to measure the difference of distributions over activity sequences as defined in the previous chapter. In statistics, there are numerous different notions of distance between distributions that may be used for this purpose [27]. In our case, two requirements need to be fulfilled. First, we do not want the distance to depend only on extreme values of the distributions such as the maximum distance between the probabilities of two corresponding activity sequences. Rather, all of the sequences shall be taken into account. Second, it must be possible to compare zero probability activity sequences since there will most certainly be sequences in one business process that are completely impossible in others. However, many popular distance measures on distributions, like the Kullback-Leibler divergence or the χ^2 distance, do not handle such singularities.

Having in mind these two requirements, the Bhattacharyya coefficient seems to be a reasonable choice [28]. It is a quantity that measures the similarity of two distributions, i.e. is assumes values between one and zero with one if the distributions are equal. The definition of the Bhattacharyya coefficient is as follows

$$\rho = \sum_{X \in \Sigma^*} \sqrt{P(X) \cdot Q(X)}$$

with P and Q being the distributions to compare. While the summation here is over all of the infinitely many activity sequences of any length, the actual computation has to be done only for sequences to which a positive probability is assigned by both of the distributions, which is unproblematic if this set is finite. The case of infinite sets due to loops will be dealt with later.

The Bhattacharyya coefficient has a straight forward geometric interpretation [29]. Consider a space over $\mathbb{R}$ having a dimension for each of the possibly observable activity sequences $X \in \Sigma^*$. Also assume that there are two distributions P and Q over these activity sequences, assigning probabilities $P(X_1), P(X_2), \dots, P(X_n)$ and $Q(X_1), Q(X_2), \dots, Q(X_n)$ to all the activity sequences. Note that the probabilities are allowed to be zero. Here, $\{X_1, X_2, \dots, X_n\}$ shall be the set of sequences to which at least one distribution assigns a positive probability. Within the space, one can interpret the vectors $\left(\sqrt{P(X_1)}, \sqrt{P(X_2)}, \dots, \sqrt{P(X_n)}\right)$ and $\left(\sqrt{Q(X_1)}, \sqrt{Q(X_2)}, \dots, \sqrt{Q(X_n)}\right)$ as representations of the distributions. Since the vectors contain the square roots of the probabilities and the distributions sum up to one, the vectors will always lie on the unit hypersphere. The Bhattacharyya coefficient can now be interpreted as the cosine of the angle between the two vectors corresponding to the distributions.

For cases in which only two different possible activity sequences are observable, a graphical representation like the one in figure 2 can be given. Here, two distributions P and Q over two activity sequences ABC and ACB respectively are given, as illustrated figure 2 (a). Then, a space having one dimension for each of the activity sequences is given in figure 2 (b) and the relevant part of the unit circle is drawn. As it can be seen, the vectors $\left(\sqrt{0.6}, \sqrt{0.4}\right)$ and $\left(\sqrt{0.3}, \sqrt{0.7}\right)$ lie on that circle. The Bhattacharyya coefficient then calculates to be roughly 0.95, which is the cosine of the angle $\alpha \approx 17{,}56°$ between the vectors.

Bearing in mind this geometrical interpretation, one can reason easily about extreme cases. If two distributions are identical, they will be assigned to exactly the same point, making the angle between them be equal to zero. Thus, the Bhattacharyya

coefficient will be equal to $\cos(0) = 1$, expressing the intuition that the distributions are 100% similar. Contrary, when the one distribution distributes its probability mass on only those activity sequences the other distributions assigns zero probability to, the two vectors will be perpendicular to each other and the coefficient calculates to $\cos(0.5 \cdot \pi) = 0$, expressing that the distributions are entirely different.

Based on the Bhattacharyya coefficient, a distance measure on distributions could be defined as $d(P,Q) = -\ln\big(\rho(P,Q)\big)$ [30]. It follows from the properties of the Bhattacharyya coefficient that this distance satisfies several desirable properties. Those are:

- $d(P,Q) \geq 0$ *non-negativity*
- $d(P,Q) = d(Q,P)$ *symmetry*
- $d(P,Q) = 0 \Leftrightarrow P = Q$ *identity*

for any choice of distributions *P* and *Q*. However, there is a fourth property this quantity does not satisfy which can be of advantage in various applications. This property is:

- $d(P,R) + d(R,Q) \geq d(P,Q)$ *triangle inequality*

for any choice of distributions *P*,*Q* and *R*. A distance measure fulfilling the triangle inequality is called a distance metric [31]. The important difference of such a metric as compared to non-metric quantities is that it allows sorting the entities being compared by it in a consistent way. When entities are compared by a non-metric distance measure, one could pick an arbitrary reference entity, compare it to all other entities and then sort the entities for instance with increasing distance to the reference entity. The problem is that, for a different reference entity, a new sorting has to be computed separately, whereas a distance metric allows embedding the entities in a metric space in such a way that the distance of entities in that space is consistent for any arbitrary reference point. This allows, for example, the use of powerful search algorithms [32].

Consider for instance the artificial example given in figure 3. For an arbitrary non-metric measure used to sort the distributions in figure 3 (a), a different sorting has to be created for each of the two reference distributions *P* and *Q*. In particular, one can only reason about relations of distributions to the reference distribution. Knowing the distance of *P* to *R* and *P* to *T* implies nothing about the distance of *R* to *T*. It could happen that the total distance of *P* to *T* is actually higher than the sum of the distances of *P* to *R* and *R* to *T*, which is counterintuitive. In figure 3 (b), a distance metric was used such that the distributions can be embedded into a two-dimensional space. In that case it is easy to see that the distance of *P* to *T* cannot be bigger than the sum of the distances *P* to *R* and *R* to *T*, which is due to the triangle inequation.

Luckily, a small modification to the Bhattacharyya coefficient gives rise to a quantity satisfying the triangle inequality [33]. We define

$$\hat{d}(P,Q) = \sqrt{1 - \rho(P,Q)}$$

to be the Bhattacharyya distance metric on distributions.

4.2 Strict Match Distance Measures

In this section, we will define our first two distance measures on business process. We name them the strict match measures since they treat any activity sequences arising from the business processes as being completely different when only a single discrepancy is found. For instance, the sequences ABC and ACB

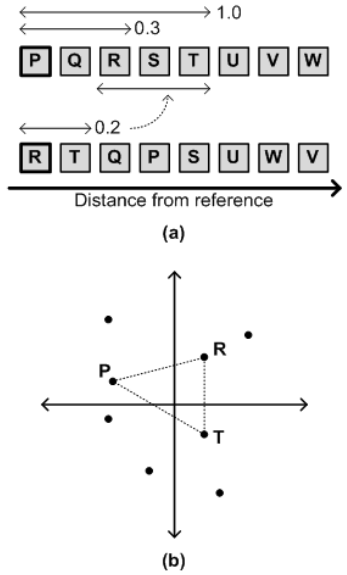


Figure 3: (a) represents two orderings of distributions with respect to a non-metric distance measure (b) represents an embedding of distributions into a space with respect to a metric distance measure.

are treated as being different and the sequences ABC and CBA are treated as being equally different. No distinction based on the similarity of the sequences is made.

We can now define the strict match distance of two business processes BP_1 and BP_2 by calculating the Bhattacharyya distance metric between the distributions on activity sequences $P(X_1)$ and $P(X_2)$ of those two processes to be

$$dist^{strict}_{complete}(BP_1, BP_2) = \sqrt{1 - \rho\big(P(X_1), P(X_2)\big)} \ .$$

This distance metric should be used whenever small discrepancies between rather similar processes shall be measured. The order in which the activities are performed should be critical for the processes as differences in this order are strongly penalized by this distance metric.

As an example, consider the two business processes in figure 4 given as PLTSs. They only have two activity sequences, namely ABCE and ACBE in common. All other activity sequences have probability zero in one of the processes. Furthermore, the probability of the sequence ACBE being observed in the first process differs from that of the second. The Bhattacharyya coefficient computes to $\sqrt{0.3 \cdot 0.3} + \sqrt{0.3 \cdot 0.1} = 0.47$. Thus, the above defined distance metric in this example is equal to $\sqrt{1 - 0.47} = 0.73$.

For some applications, one might not be interested into the behavioral distance with regard to the entire behavior of two processes but rather with regard to the overlaps that exist between the two. In the example of figure 4, the first process contains activity D, while the second does not, and the second contains activity F which is not found in the first process. In such cases, the distance measure can be computed in a slightly different way. Any transition emitting a symbol that is specific to only one the processes is then switched to a “silent mode” which means that it still belongs to the path but its symbol does not appear in the

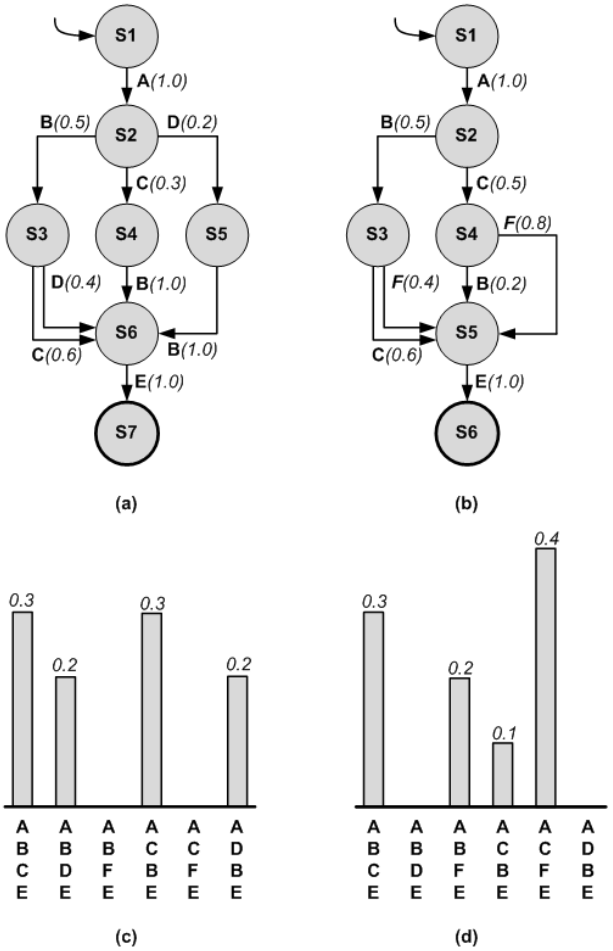


Figure 4: (a) and (b) represent two business processes as PLTSs, (c) and (d) represent the corresponding distributions over activity sequences.

activity sequences anymore. This ensures that all activity sequences only contain activities common to both processes.

We define the rigid distance of the overlap of two business processes BP_1 and BP_2 as

$$dist_{overlap}^{strict}(BP_1, BP_2) = \sqrt{1 - \rho\left(P(\hat{X}_1), P(\hat{X}_2)\right)}$$

where $\hat{X}_1$ and $\hat{X}_2$ are the activity sequences observed when the activities unique to either of the processes are silent.

Figure 5 illustrates this concept using the previous example. The grey shaded transition symbols D and F denote that these transitions are currently silent. Thus, the activity sequences ABDE and ADBE both merge to the single sequence ABE and the sequence ACFE reduces to ACE. The Bhattacharyya coefficient in this example is equal to $\sqrt{0.3 \cdot 0.3} + \sqrt{0.4 \cdot 0.2} + \sqrt{0.3 \cdot 0.1} = 0.76$ giving a distance of $\sqrt{1 - 0.76} = 0.49$.

4.3 Fuzzy Match Distance Measures

The following section is devoted to more relaxed measures of distance. The main difference from the strict match ones is that we will abstain from the assumption that all activity sequences having small differences are already treated as completely different. Rather, we will use the similarity of these sequences to identify them with each other, thereby introducing new notions of distance being more appropriate for application scenarios in which small differences in the order of the activities or the exclusion of some of the activities of a sequence should result in small distances of the behavior.

To derive these measures, consider again the example of figure 4. Both processes have two activity sequences ABCE and ACBE in common. Since we can directly identify them with each other, not special treatment is necessary. The activity sequences ABDE and ADBE however are unique to the first process, but we are now interested in whether we can associate them with similar sequences of the other process.

In general, we can quantify our belief that a particular activity sequence being unique to one business process belongs to any sequence of the other process by calculating any kind of string similarity between the sequences. Usually, these string similarities are based on calculating the minimum number of elementary operations required for transforming one string into the other and summing up the costs of all the operations [6]. The basic operations vary among the algorithms, but possible operations are

- Insertions: Insert one symbol into the string
- Deletions: Remove one symbol from the string
- Substitutions: Replace one symbol with another
- Transpositions: Swap two symbols with each other

In the remainder of this work, we will use the popular Levenshtein distance to compute a similarity of two activity sequences, which uses insertions, deletions and substitutions. However, other choices might be suitable as well. For instance, the Damerau-Levenshtein distance [34], adding transpositions to the set of operations, could be used with the effect that variations in the order of activities would result in less cost. The activity sequence ABC requires two substitutions to be transformed into the sequence ACB, but only one transposition.

To transform the Levenshtein distance $l(X_1, X_2)$ of two activity sequences X_1 and X_2 into a similarity measure, we simply normalize the distance by the maximum distance that could be observed between the sequences, which is $\max(|X_1|, |X_2|)$ with $|X_1|$ and $|X_2|$ being the lengths of the sequences. Thus, we define the sequence similarity to be

$$sim(X_1, X_2) = 1 - \frac{l(X_1, X_2)}{\max(|X_1|, |X_2|)}$$

This definition results, for the example of the sequence ABDE, in a similarity of 0.75 compared to the sequence ABFE and 0.5 compared to ACFE.

The idea of creating the associations among the unique sequences is very simple. First, the similarity of each pair of unique activity sequences is computed where the first sequences stems from one process and the second sequence from the other. Second, associations *R* between activity sequences are created in a greedy way, creating an association between those sequences having the highest similarity. In the example, the first association to be created is the one between sequences ABDE of the first process and ABFE of the second process, leaving the association between ADBE and ACFE as the only possibility for the second association. This procedure is illustrated in figure 6.

In the general case, we will end up with the following relation:

$$R \subset \Sigma^*_{u1} \times \Sigma^*_{u2}$$

$$with\ (X_1, X_2) \in R \wedge (X'_1, X'_2) \in R \Rightarrow X_1 \neq X'_1 \wedge X_2 \neq X'_2$$

where Σ^*_{u1} and Σu^*_2 denote the sets of unique activity sequences in the first and second process respectively. The condition ensures that no activity sequence is mapped to more than one other sequence.

The relation enables us to define a modified version of the Bhattacharyya coefficient which computes the similarity of the distributions according to the created associations but correcting for the dissimilarities of the associated sequences. It shall be

$$\dot{\rho} = \sum_{X \in X_c} \sqrt{P(X) \cdot Q(X)} + \sum_{(X_1, X_2) \in R} sim(X_1, X_2) \cdot \sqrt{P(X_1) \cdot Q(X_2)}$$

where X_c denotes the set of activity sequences common to both processes.

It is then easy to define the fuzzy match distance metrics based on this modified Bhattacharyya coefficient, the first of which is:

$$dist^{fuzzy}_{complete}(BP_1, BP_2) = \sqrt{1 - \dot{\rho}\ \left(P(X_1), P(X_2)\right)}\ .$$

For the example, the Bhattacharyya coefficient computes to $0.5 \cdot \sqrt{0.2 \cdot 0.4} + \sqrt{0.3 \cdot 0.1} + 0.75 \cdot \sqrt{0.2 \cdot 0.2} + \sqrt{0.3 \cdot 0.3} = 0.76$, which gives a fuzzy distance of $\sqrt{1 - 0.76} = 0.49$.

In a fashion similar to the previous section, we can also define a fuzzy match distance metric that removes all activities unique to one of the processes from the activity sequences. The associations are then created based on this already reduced distribution over sequences. For the sake of completeness, we define it to be

$$dist^{fuzzy}_{overlap}(BP_1, BP_2) = \sqrt{1 - \dot{\rho}\ \left(P(\hat{X}_1), P(\hat{X}_2)\right)}\ .$$

In our example, no difference can be observed compared to the rigid case. This is due to the fact that there is only one unique sequence, namely ACE in the first process. As there is no other sequence to assign it to, the relation remains empty and $\dot{\rho}$ assumes the same value as ρ, resulting in equal distances.

4.4 Distances in Presence of Loops

Although the distance measures presented in this paper are defined for the case that there are infinitely many activity sequences that can possibly be observed, the actual computation will be infeasible in such cases. This problem arises in all business processes having loops, as for example in the process represented in figure 7 in which the sequence BD may be executed arbitrarily often.

While this is a well-known problem of all approaches to similarity measurement that rest upon activity sequences, our current setting allows circumventing this problem. Since any decision on entering or not entering a loop is weighted by a certain probability, sequences with more loop iterations tend to become more and more unlikely. In the example of figure 7, the probability of the sequence having one iteration of the loop still

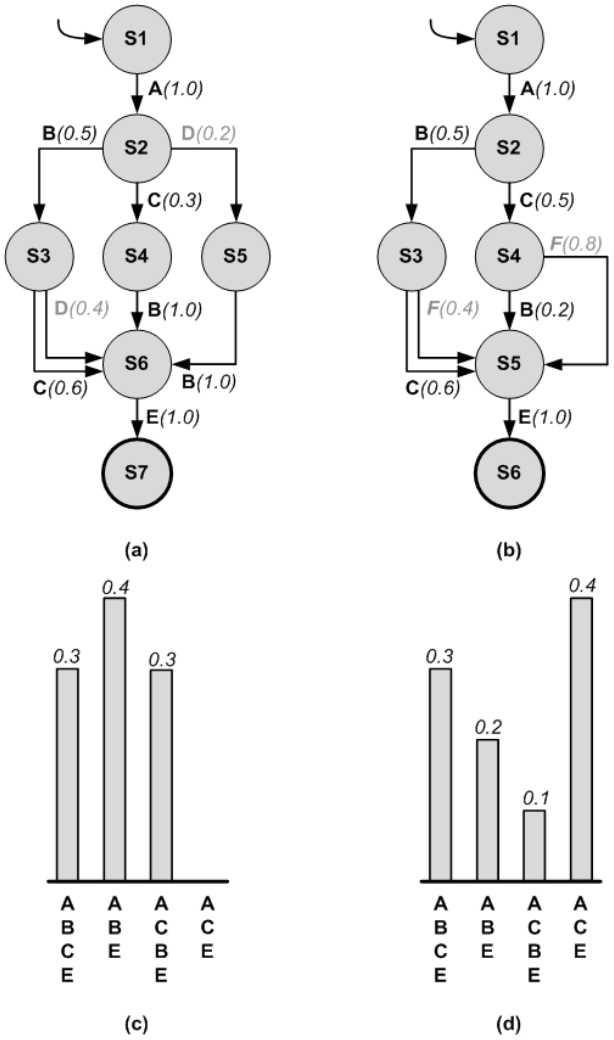


Figure 5: (a) and (b) represent two business processes as PLTSs, (c) and (d) represent the corresponding distributions over activity sequences where activities unique to one of the processes are silent.

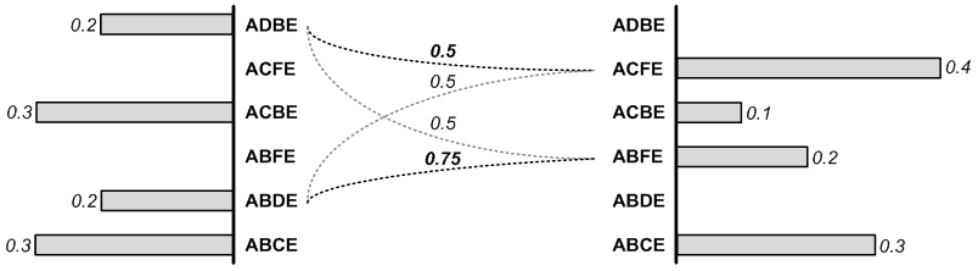


Figure 6: similarities of unique activity sequences, with dominant relations being highlighted.

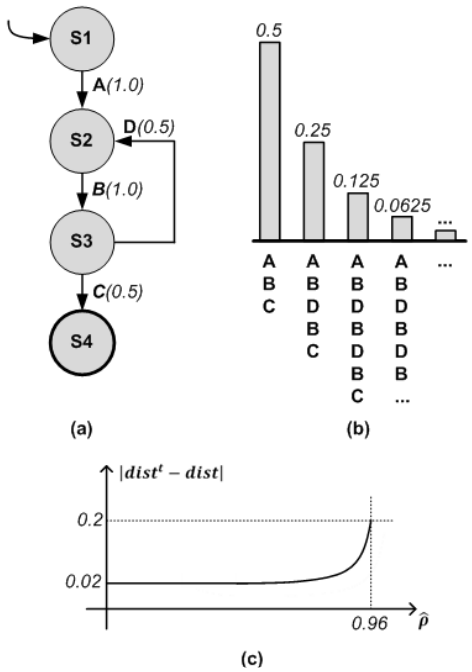


Figure 7: (a) represents business process as PLTS having a loop, (b) represents the corresponding infinite distribution, (c) is a graph of the maximal error of a distance at truncation level 0.96.

amounts for 0.25, while for three iterations it is already down to 0.0625. Since unlikely sequences are of low relevance for the value of the Bhattacharyya coefficient, the remedy to the problem is to just truncate the sum when a certain amount of probability mass is covered.

To formalize this, we first need to define the truncation level *t*. It shall be the amount of probability mass we require to use in the computation of the Bhattacharyya coefficient. Thus, it should assume a value close to one. The set of activity sequences required to fulfill this level shall be named X_t. Thus, the approximated Bhattacharyya coefficient is

$$\hat{\rho} = \sum_{X \in X_t} \sqrt{P(X) \cdot Q(X)} + error \ .$$

In the worst case, the error term comprises only one missing activity sequence X_m to which both processes assign probability $1 - t$. Hence, the error is bounded by *t*:

$$0 \leq error \leq \sqrt{P(X_m) \cdot Q(X_m)} \leq \sqrt{(1-t) \cdot (1-t)} = 1 - t \ .$$

Now let $dist$ be the true distance to compute and $dist^t$ be the truncated approximation. Then we get

$$|dist^t - dist| = \left|\sqrt{1-\hat{\rho}} - \sqrt{1-\hat{\rho}-error}\right|$$

$$= \sqrt{1-\hat{\rho}} - \sqrt{1-\hat{\rho}-error} \leq \sqrt{1-\hat{\rho}} - \sqrt{1-\hat{\rho}-(1-t)} \ .$$

This upper bound on the deviation of true and truncated distance is maximal for cases in which $\hat{\rho}$ is maximal. As $\hat{\rho}$ is bounded above by t, the upper bound on the deviation is:

$$|dist^t - dist| \leq \sqrt{1-\hat{\rho}} - \sqrt{1-\hat{\rho}-(1-t)} \leq \sqrt{1-t} \ .$$

How the upper bound on the deviation develops with respect to the approximated Bhattacharyya coefficient can be seen in figure 7 (c), where the example of a truncation level $t = 0.04$ is illustrated. As one can see, for the maximum Bhattacharyya coefficient $\hat{\rho} = 0.96$, the deviation is maximal. After that, it rapidly decays due to the square root used in computing the distance.

5. APPLICATION EXAMPLE

To further illustrate the definitions we have made in the previous chapter, consider the rather simple set of business processes shown in figure 8 (a). It contains the processes P3 to P8 in form of PLTSs. Processes P1 and P2 shall be the ones already known from the last chapter. They can be seen in figure 4 (a) and (b) respectively. We now want to analyze this set of processes to investigate how the similarity measures perform on this example.

First, we take a look at the processes themselves to get a broad overview of their characteristics. All of them define activity sequences over the activities A,B,C,D,E,F,G,H, and I, but not all of them include the entire set of activities. Two of the eight processes, namely P3 and P6, define infinite sequences since they contain loops. Therefore, an approximate similarity calculation will be necessary. We used a truncation level of $t = 0.99999$ for all our calculations in this section, giving a maximum deviation from the true distance equal to 0.01.

One directly sees from figure 8 (a) that some processes seem to be very similar. For instance, processes P7 and P8 look identical on the first sight. A closer look, however, reveals that the left branch is much more likely to be taken in P7 than in P8 and vice versa. Also the models P5 and P6 seem to be quite similar, as the only difference between P6 as compared to P5 are the two additional loops. Also model P4 defines behavior that is very similar to P5 and even P6. Process P3 on the other hand does not have much in common with the other processes in the set.

We now want to use the distance metrics to derive and visualize the thoughts we have just made. For the first analysis we choose the metric $dist^{fuzzy}_{complete}$ for comparison and compute a complete distance matrix for the set of business processes. This information on how close the processes are can then be used to represent this distance graphically. The result of this computation is shown in figure 8 (b). It represents the business processes as points in a two-dimensional space. All the points are fitted into this space such that distances between them reflect the distances of the processes they represent. The technique we have used to create this picture is called Multidimensional Scaling (MDS). From this picture, one can easily see groups of similar processes being close to each other. The groups we would suggest based on the picture are indicated by dashed circles around the points.

Another analysis that can be performed on the models is to cluster them with respect to their distance. For this experiment, we have chosen the metric $dist^{strict}_{overlap}$. We then applied an agglomerative clustering algorithm to the distance matrix computed from that metric to find possible clusters of processes. The dendrogram visualizing the results can be seen in figure 8 (c). When compared to the results of the MDS analysis, one can see that, although a simpler metric was used that does not identify similar activity sequences with each other, the result is rather similar. The same groups are suggested to the analyst.

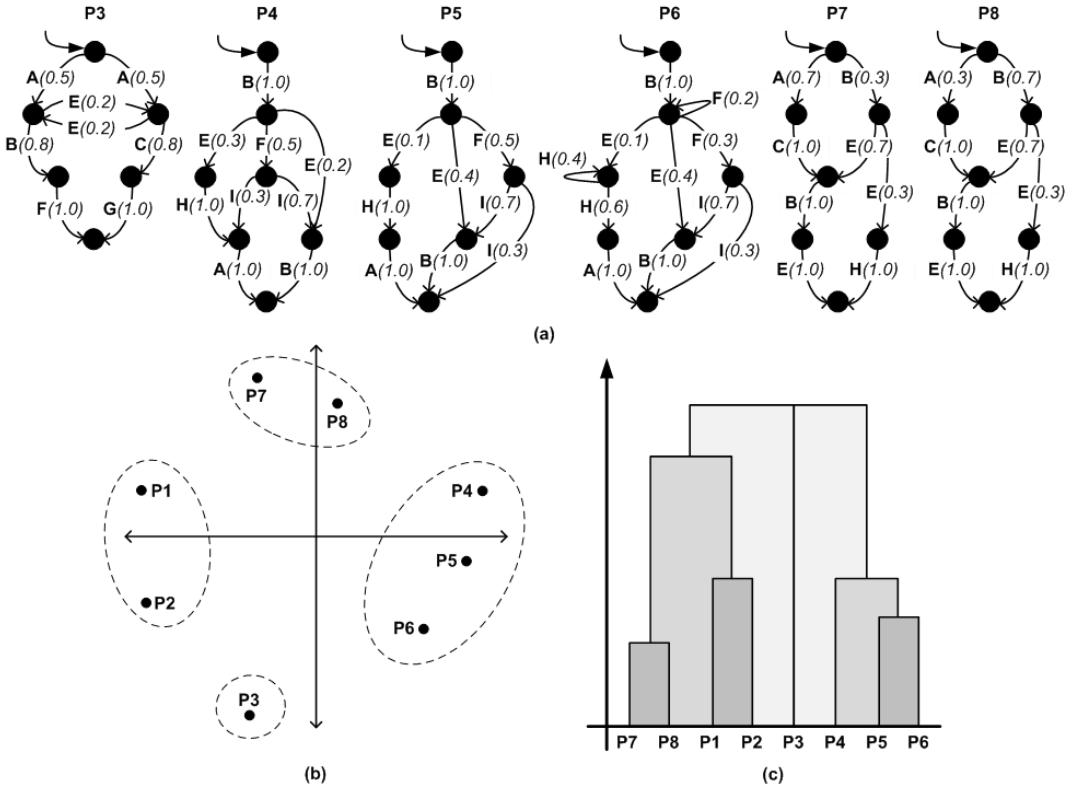


Figure 8: (a) defines some business processes as PLTSs, (b) represents the result of running MDS on these processes using fuzzy complete distance, (c) represents the dendrogram after running agglomerative clustering using strict overlapping distance. Processes P1 and P2 are shown in figure 4 (a) and (b).

6. CONCLUSION AND OUTLOOK

In this paper we presented a set of distance measures for business processes having the following distinct properties:

- *Behavioral*: The emphasis lies upon the behavior of the business process in form of the possible activity sequences the processes allow. No structural aspects are considered. The approach abstracts from the modeling language used to represent to process.
- *Probabilistic*: The measures explicitly incorporate the probabilities observing certain behavior and weights more probable behavior stronger.
- *Approximate*: The incorporation of probabilities allows approximating the distance of processes having infinitely many possible activity sequences.
- *Customizable*: Various different notions of distance measures are proposed ranging from very strong to more relaxed versions.
- *Metric*: The distance measures can be interpreted as metric distances, making them suitable for algorithms exploiting such structures.

The approach builds on the idea that business process behavior is not only defined by the possible activity sequences being compliant with a process model but also by the frequencies with which these activity sequences occur in the real world. It is, however, not restricted to cases in which explicit annotations on decision probabilities are given since it may be valid to make certain assumptions. For instance, one could assume that any branching of the control flow is equally likely, which would reduce our metrics to quantities very similar to other approaches in the literature.

Furthermore, one does not even have to assume the existence of an explicit process model to calculate distances. In many cases, process-aware information systems like ERP or CRM systems provide event logs documenting the past behavior of a possibly unknown process [35]. Clearly, such event logs can be used to approximate a distribution over activity sequences that can be used in a distance calculation.

In the future, we plan to intensively evaluate the metrics we have proposed, especially with respect to the conformance with human judgment. Several studies in literature evaluated similarity measures by experimentally counterchecking them with human opinions on similarity [16, 17, 35].

7. REFERENCES

[1] Kueng, P. 2000. Process performance measurement system: a tool to support process-based organizations. *Total Quality Management & Business Excellence* 11, 1, 67-85.

[2] van Dongen, B., Dijkman, R., Mendling, J. Year. Measuring Similarity between Business Process Models. In *Proceedings of the 20th International Conference on Advanced Information Systems Engineering, 2008*, 450-464.

[3] van der Aalst, W. M. P., Reijers, H. A., Weijters, A., van Dongen, B., de Medeiros, A. K., Song, M., Verbeek, H. M.

W. 2007. Business process mining: An industrial application. *Information Systems* 32, 5, 713-732.

[4] Namiri, K., Stojanovic, N. Year. Pattern-based design and validation of business process compliance. In *Proceedings of the 2007 OTM Confederated International Conference on the Move to Meaningful Internet Systems: CoopIS, DOA, ODBASE, GADA, and IS, 2007*, 59-76.

[5] Dumas, M., Garcıa-Banuelos, L., Dijkman, R. 2009. Similarity search of business process models. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 32, 3, 23-28.

[6] Navarro, G. 2001. A Guided Tour to Approximate String Matching. *ACM Computing Surveys* 33, 1, 31-88.

[7] Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Massachusetts.

[8] Navigli, R. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)* 41, 2, 10.

[9] Hopcroft, J. E., Motwani, R., Ullman, J. D. 2001. *Introduction to automata theory, languages, and computation*. 2 ed., Addison-Wesley Longman, Amsterdam.

[10] Dijkman, R., Dumas, M., García-Bañuelos, L. Year. Graph matching algorithms for business process model similarity search. In *Proceedings of the 7th International Conference on Business Process Management (BPM), 2009*, 48-63.

[11] Minor, M., Tartakovski, A., Bergmann, R. Year. Representation and structure-based similarity assessment for agile workflows. In *Proceedings of the 7th International Conference on Case-Based Reasoning, ICCBR, 2007*, 224-238.

[12] Lu, R., Sadiq, S. Year. On the Discovery of Preferred Work Practice Through Business Process Variants. In *Proceedings of the 26th international conference on Conceptual modeling (ER'07), 2007*, 165-180.

[13] Madhusudan, T., Zhao, J. L., Marshall, B. 2004. A case-based reasoning framework for workflow model management. *Data & Knowledge Engineering* 50, 1, 87-115.

[14] Grigori, D., Corrales, J., Bouzeghoub, M. 2008. Behavioral matchmaking for service retrieval: Application to conversation protocols. *Information Systems* 33, 7-8, 681-698.

[15] Li, C., Reichert, M., Wombacher, A. Year. On measuring process model similarity based on high-level change operations. In *Proceedings of the 27th International Conference on Conceptual Modeling (ER'08), 2008*, 248-264.

[16] Dijkman, R., Dumas, M., Dongen, B. V. 2009. Similarity of business process models: Metrics and evaluation. *Information Systems* 36, 2, 498-516.

[17] van Dongen, B., Dijkman, R., Mendling, J. Year. Measuring Similarity between Business Process Models. In *Proceedings of the 20th International Conference on Advanced Information Systems Engineering, 2008*, 450-464.

[18] Zha, H., Wang, J., Wen, L., Wang, C., Sun, J. 2010. A workflow net similarity measure based on transition adjacency relations. *Computers in Industry* 61, 5, 463-471.

[19] Park, D. Year. Concurrency and automata on infinite sequences. In *Proceedings of the 5th GI-Conference on Theoretical Computer Science, 1981*, 167-183.

[20] Nejati, S., Sabetzadeh, M., Chechik, M. Year. Matching and merging of statecharts specifications. In *Proceedings of the 29th International Conference on Software Engineering, 2007*, 54-64.

[21] Sokolsky, O., Kannan, S., Lee, I. Year. Simulation-based graph similarity. In *Proceedings of the 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS), 2006*, 426-440.

[22] Wombacher, A., Rozie, M. 2006. Evaluation of workflow similarity measures in service discovery. *Service Oriented Electronic Commerce*, 51-71.

[23] Mohri, M. 2003. Edit-distance of weighted automata: General definitions and algorithms. *International Journal of Foundations of Computer Science* 14, 6, 957-982.

[24] Mohri, M. Year. Edit-distance of weighted automata. In *Proceedings of the 7th International Conference on Implementation and Application of Automata (CIAA), 2003*, 1-23.

[25] de Medeiros, A. A., Aalst, W. V. D. 2008. Quantifying process equivalence based on observed behavior. *Data & Knowledge Engineering* 64, 1, 55-74.

[26] Winskel, G., Nielsen, M. 1995. *Models for concurrency*. Oxford University Press, Oxford, UK.

[27] Gibbs, A., Su, F. 2002. On choosing and bounding probability metrics. *International Statistical Review* 70, 3, 419-435.

[28] Bhattacharyya, A. 1943. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Sociecty* 35, 4.

[29] Thacker, N., Aherne, F., Rockett, P. 1997. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika* 34, 4, 363-368.

[30] Rauber, T., Conci, A., Braun, T., Berns, K. Year. Bhattacharyya Probabilistic Distance of the Dirichlet Density and its Application to Split-And-Merge Image Segmentation. In *Proceedings of the 15th International Conference on Systems, Signals and Image Processing (IWSSIP), 2008*, 145-148.

[31] Zezula, P., Amato, G., Dohnal, V., Batko, M. 2006. *Similarity search: the metric space approach*. 1st edn ed., Springer, New York.

[32] Hjaltason, G. R., Samet, H. 2003. Index-Driven Similarity Search in Metric Spaces. *ACM Transactions on Database Systems (TODS)* 28, 4, 517-580.

[33] Comaniciu, D., Ramesh, V., Meer, P. 2003. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 5, 564-575.

[34] Damerau, F. J. 1964. A technique for computer detection and correction of spelling errors. *Communications of the ACM* 7, 3, 171-176.

[35] van der Aalst, W. M. P., Weijters, A. 2004. Process Mining: A Research Agenda. *Computers in Industry* 53, 3, 231-244.

Towards an Ontological Foundation of Services Science: The General Service Model

Roberta Ferrario
Laboratory for Applied Ontology
ISTC-CNR
Trento, Italy
+390461314841
ferrario@loa-cnr.it

Nicola Guarino
Laboratory for Applied Ontology
ISTC-CNR
Trento, Italy
+390461314871
guarino@loa-cnr.it

Christian Janiesch
SAP Research
SAP Australia Pty Ltd
South Brisbane, Australia
+61732599617
c.janiesch@sap.com

Tom Kiemes
SAP Research
SAP AG
Karlsruhe, Germany
+496227752518
tom.kiemes@sap.com

Daniel Oberle
SAP Research
SAP AG
Karlsruhe, Germany
+496227752585
d.oberle@sap.com

Florian Probst
SAP Research
SAP AG
Darmstadt, Germany
+496227768850
f.probst@sap.com

ABSTRACT
Despite the ubiquity of services, there is still no consensus on their exact nature and structure. This consensus, however, is necessary to unambiguously describe and trade services physically as well as in an Internet of Service. To make some progress towards a shared conceptualization, we present in this paper a general service model based on the DOLCE foundational ontology. In our understanding, a service is essentially composed of a service process whose core actions (delivered by a service producer to a service consumer) comply with the exposed description of a service provider's commitment. Each service belongs to a larger service system process which obeys legal and pricing constraints. We illustrate the model's usefulness and relevance by the means of a continuous example.

Keywords
Service ontology, service model, foundational ontology, conceptual model, service science

1. INTRODUCTION
Nowadays services are everywhere. There are public services conceived to make life easier to the members of a certain community, like public residences for elderly. There are private services providing actions or processes we want to be performed, but we are unable or not willing to perform ourselves, so we pay someone else to do them, like car repairs. There are information services, which people pay for in order to be informed about something that they would not come to know otherwise, like online license registries. Further there are services which are sold in combination with products to augment the value of such products (e.g. a car combined with free check-up services) or even enable the use of these products in the first place (e.g. a mobile phone).

However, despite the pervasiveness of the term *service* in the ordinary discourse, there is no wide consensus on the meaning of such term; not only it is used in different ways across disciplines (e.g. economics vs. computer science), but even within the same discipline confusions and inconsistencies predominate.

Given such a situation, interoperability across services becomes a myth, since as service designers do not share a common semantic background, they may use the same terms to express different concepts or different terms to refer to the same concept [39; 50].

Our claim is that, in order to overcome the problem of service interoperability, we need a unified, rigorous, and principled reference ontology of services, able to clarify the intended meaning of the terms used and to make explicit how the domain of services can be structured.

Since many different perspectives on the services domain may be adopted, our choice is to build a reference ontology based on a rigorous ontological analysis, anchoring the primitives of the service domain to more fundamental primitives taken from a top level ontology, which is in our case the foundational ontology DOLCE [28]. We deem such a foundational perspective is necessary to substantiate a *services science* [7].

In the following section we review related work on service description efforts and how they can be categorized. In doing so, we highlight the shortcomings which motivated our research. In Section 3 we introduce the theory and approach we use throughout the remainder of the paper. We then elaborate on our general service model and substantiate it with the foundational primitives of DOLCE to demonstrate the validity of the constructs. We summarize the paper's contribution and put the model in relation to the emerging standard of the Unified Service Description Language (USDL) [42] to further underpin the necessity of the research.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

2. RELATED WORK

There exists a plethora of service description efforts that can be grouped into different strands. Each of these strands has its own motivation and representation needs for capturing service information. Each strand consists of standardized, academic, or proprietary efforts. The efforts can be roughly compared according to *scope*, e.g., whether the effort captures IT or business aspects of services or the whole service system. Another relevant criterion is the effort's *purpose*: is the effort geared towards normative data exchange, is it there to facilitate software engineering, is it there to automate a specific task, or is it there to act as reference model?

The first strand of service description efforts is the field of *Service-oriented Architectures (SOA).* Typically their scope concerns the IT aspects of services only, e.g., the interface description. Different standards bodies specified several dozens of different aspects which are collectively known as WS-* (incl. WSDL, WS-Policy, WS-Security, etc.) mainly for the purpose of exchanging such information over the Web. Another effort in this strand is the Service-oriented architecture Modeling Language (SoaML) by (OMG) [36]. Its purpose is to support model-driven software engineering for services. Finally, there emerged the need to establish a Reference Model for Service Oriented Architecture (SOA-RM) which was published by OASIS [37]. An alternative reference model in the form of an ontology for SOA (SOA Ontology) is available by The OpenGroup [51]. Current research in the SOA strand mainly concerns RESTful services and their description (cf. WADL [19]). Oberle et al. [35] provide an ontological account of Web services according to the principles of ontological analysis on top of the DOLCE foundational ontology. The ontology can be regarded as a reference model with the scope limited to IT aspects.

The second strand consists mainly of ontologies in the field of *Semantic Web Services*. The main goal of Semantic Web Services approaches is automation of discovery, composition, and invocation of services in an SOA by ontology reasoners and planning algorithms. The most prominent efforts within this strand are OWL-S [27] and WSMO [40]. Many surrounding and similar efforts have surfaced in academia and most of them are geared at automation and limit their scope to IT and non-functional properties. Along the lines of SOA-RM, the community is working on a Reference Ontology for Semantic Service Oriented Architectures (RO-SOA) which is available as a draft by OASIS.

The third strand is rooted in the rise of on-demand applications that led to the notion of *Software-as-a-Service (SaaS).* Here, the emphasis of service implies that the consumer gets the designated functionality. Thus, SaaS is not synonymous with SOA. The strand of SaaS contains a standard, namely, the W3C Service Modeling Language (SML) [56]. The anticipated purpose of SML is to define a consistent way for exchanging information about computer networks, applications, or servers so businesses can more easily manage the services that are built on these resources. Current research is represented by the Software-as-a-Service Description Language (SaaS-DL). SaaS-DL builds on WS-* to capture SaaS specificities in order to support model-driven engineering [49].

The fourth strand focuses on capturing the purely *economic* aspects of services regardless of their nature. The DIN PAS 1018 standard essentially prescribes a paper form for the description of services for tendering [10]. The structure is specified in a non-machine-readable way by introducing mandatory and optional non-functional attributes, such as, classification, resources, location, etc. O'Sullivan [33] adopts a wider scope and contributes a domain independent taxonomy which is capable of representing the non-functional properties of conventional, electronic, and web services. Toma [53] presents a syntactic translation of O'Sullivan's work in the proprietary WSML language. Emmich [13] focuses on product-related services, such as maintenance, and is specified in UML. He basically merges existing standards and models for products, companies, organization, and resources. Finally, the Unified Service Description Language (USDL) is a proposal to unify the business and technical scope of services [5; 6].

The fifth strand is also focused on economic aspects but draws attention mainly to describing *service networks*, i.e., the ecosystem and value chain relationships between services of economic value. So far, this strand is represented by academic approaches mainly by Akkermans research group. The latter brought forth several ontologies, among them the Obelix and Serviguration ontologies which exhibit a clear business scope. The latest effort is the e^{3}Service ontology which models services from the perspective of the user's needs [9]. The main purpose is to generate service bundles under the consideration of customer needs. The Service Network Notation (SNN) captures similar aspects to the e³Service ontology [4]. However, SNN is a UML model for the purpose of analyzing measurements of added value for each single participant as well as for the whole network optimization of value flows.

Finally, there are overarching efforts that concentrate on the bigger picture of *service systems* or service science also taking into account value co-creation, i.e., the sharing and distribution of labor, investments, expertise, risk, and – most of all – knowledge. In the last few years the studies dedicated to this new field have multiplied [25; 26; 47]. One example in this strand is Alter [1] who contributes three informal frameworks as a first attempt to define the fundamentals of service systems. Another effort is the *OASIS Reference Architecture Foundation for SOAs* [38]. Although the background is SOA, the specification argues that SOA-based systems are better thought of as ecosystems rather than stand-alone software products. Therefore, we classify this effort into the service system strand. It is directly related to our effort. However, the reference architecture foundation is not based on ontological analysis but takes the OASIS SOA-RM as its starting point by building on its vocabulary of important terms and concepts. Another effort considering the wider scope of the service system is the *Service Design Model* of Dhanesha et al. [11]. It is geared at a software engineering purpose and essentially comes in the form of UML. The model's scope takes into account the business organization, the customer, and the delivery organization during service design.

Our ontological foundations of service science represent a reference model with the scope being the whole service system which is common to the various strands described so far. It is based on and formalizes earlier ideas of Guarino and Ferrario [14]. Our approach is therefore mainly related to the efforts in the service system strand. Our approach differs in that it is explicitly built using the DOLCE foundational ontology. This means relating core classes and relations to proposed invariant categories of human cognition (which are reflected in the foundational ontology itself). This prompts the modeler to sharpen his notions

with respect to the distinctions made in the foundational ontology. What is typically gained is an increased understanding of the modeled domain as well as a cleaner design.

3. BACKGROUND

3.1 Theory

Conceptual models typically condense multiple people's perceptions of a matter into a shared representation. Thereby, the models drawn upon always result in an abstract account of reality.

Conceptual models are usually graphical, i.e. semi-formal, representations [cf. also 24] and can be applied to static (e.g., data models) and dynamic (e.g., process models) states of affairs in some domain [55]. Generally they are used to structure and systematize problems and thereby used to omit irrelevant aspects of the surrounding scenario and help focus on the key problem at hand. Thus, a conceptual model is the representation of an application domain for the ends of a subject which is commonly based on a semi-formal language with a graphical representation [57].

According to Wand and Weber [55], conceptual modeling serves in particular to support communication between developers and users, to help analysts understand a domain, to provide input for the design process, and to document the original requirements for future reference. Usage in the early stages of information systems development is considered to be particularly beneficial, since the efforts for resolving mistakes made in this stage increase exponentially as time passes and subsequent project stages commence [cf. 30].

Evidently these different purposes require different modeling processes. If, for instance, conceptual models are intended to serve as an input for the design process, a formal and unambiguous grammar has to be used in order to map concepts precisely to implementation artifacts. Support for the communication process, however, can be achieved by less formal means, e.g. in tabular form [for other representation forms cf. e.g. 52].

We use UML class diagrams of the Technical Architecture Modeling (TAM) Standard [41] to visualize the model of the ontology. TAM represents a pragmatic combination of conceptual and formal modeling methods.

The starting point of the construction of a conceptual model is a result of perception and preexisting knowledge of an individual about phenomena in the application domain. This result of perception and cognition is represented as a mental model (or conceptualization) by the subject. Based on preexisting knowledge the mental model organizes perception into a coherent structure and establishes internal connections among them [12; 29; 31; 32]. The mental model is the basis for comprehension of the real world as well as its elements [48]. Its content is influenced by the intentions of the subject and the objectives of the conceptual modeling project. It reflects the pragmatic aspects of modeling and reduces as well as combines knowledge and perception accordingly.

These insights about a mental model lead to the following assumptions: A subject represents the results of perception and cognition as mental models. A mental model has a specific structure consisting of elements and relations. Also, the structure of a mental model is influenced by preexisting knowledge.

In a conceptual modeling process at least two different types of actors are involved. First, the model creators are the subjects who observe a material domain and explicate the conceptual model based on their insights. Second, the model users are the individuals who try to comprehend the conceptual model in order to learn about the material domain.

Based on this knowledge about the mental model and the corresponding roles, the conceptual modeling process [23; 32; 44] can be derived (cf. Figure 1):

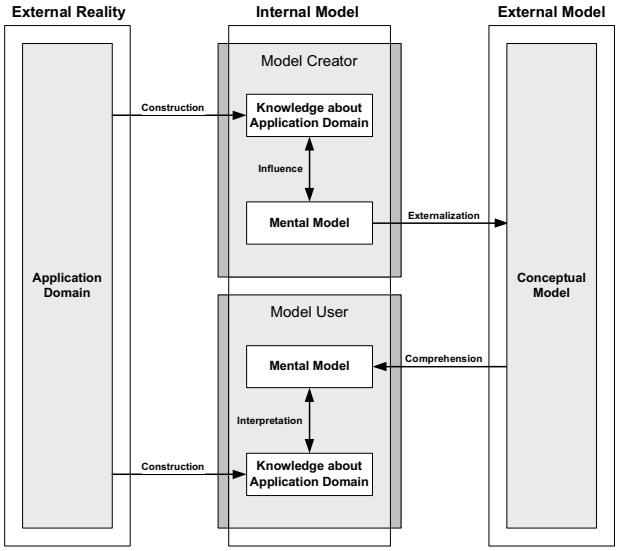


Figure 1. General Conceptual Modeling Process [cf. 22].

The external reality (the original) is perceived by the model creator and (re-)constructed in the form of a mental model [31]. This internal structure represents the intended comprehension of the conceptual model from its creator's perspective. The (re-)construction process is influenced by preexisting knowledge. Subsequently, the mental model is explicated as a conceptual model by the model creator [43]. Afterwards, the model user tries to comprehend the conceptual model in the form of an own mental model [17]. To accomplish that, knowledge about the application domain and the modeling method are required. The new mental model embodies the subjective comprehension of the conceptual model from the model user's perspective.

Thus, a conceptual model can only emerge and be interpreted successfully when the model creator and the model user share common knowledge. Only then the conceptual model can be properly encoded and decoded. If such a conceptual model is supported by an ontology (intended as a partial account of a conceptualization, i.e. a mental model), the alignment of the mental models of model creator and model user is greatly simplified. Aligning such domain-dependent mental models to general notions such as those specified in an upper-level ontology like DOLCE further contributes to make the semantics and the intended meaning of the terms used in the model more explicit, therefore reducing ambiguities and misunderstandings.

3.2 Research Approach

The origin of this research can be traced back to collaboration between the ISTC-CNR Laboratory of Applied Ontology and the office of social and housing policies of the Autonomous Province

of Trento. The latter was seeking help from the former to conceptually "clean up" their catalogue of services.

During a series of interviews, Ferrario and Guarino found that people who have created or were using the catalogue had issues with different understandings of the same fundamental terms which co-existed in the catalog. Often people in the same business context were wrongly assuming to share the same conceptualization of what they were talking about. Thee most striking issue was the use of the term service itself: Some of them used the term *service* to refer to a series of actions, others to some kinds of actions, others to some capability to execute an action, while still others were called an office in a public administration a service, or the people working in it.

It was concluded that a foundational analysis was needed. First insights were presented in [14; 15]. The main feature of such an approach is that it adopts the *glass box view* [34] instead of the more traditional *black box view*. According to the latter, services are described as transfer functions from an input to an output state, with a strong focus on the external interface, as opposed to the internal view, which is kept separated. Hence the metaphor of the black box which does not allow others to understand how the service internally works.

If, on the one hand, this approach seems to work well from a technological perspective, on the other hand there is a well known gap between the business perspective on services and the IT perspective, which determines a difficulty of usage on the business stakeholder's side. Business applications need not only specify what the service does, but also how the service is performed and when the various processes involved in a service occur (and this means reference to internal details). Still from another point of view, also contracts and service level agreements need to refer to internal and contextual details (thus how the service interacts with its outer environment). In other terms, one needs to be able to look inside the box and out of the box, i.e. one needs to have a glass box.

We chose a rather high level of abstraction as the main purpose of the model is to facilitate the understanding of (concrete) services and their facets and not their automated invocation and adaptation. When discussing examples, we found that the majority of differences in services can already be distinguished at this level of detail and drilling further into the model does not add sufficient surplus to justify the effort. It rather makes the discussion too complex for people not familiar with the model. Consequently, the general service model is supposed to be a baseline on which more concrete efforts such as USDL can build.

The general service model presented in this paper is a *design artifact* in the sense of the design science-based approach to IS research as described in Hevner et al. [20]. IS research accordingly is concerned with two design processes, i.e. to *build* purposeful artifacts to address heretofore unsolved problems, and to *evaluate* these artifacts with respect to the utility provided in solving those problems. Based on a thorough review of related work, we build a service meta model and validate its constructs through DOLCE and illustrate its usefulness and relevance by the means of examples.

In the following, we further embrace this view, with a special emphasis on the environment which the service belongs to, or, in other words, we take the socio-technical system into account which the service is a part of [2; 3].

4. GENERAL SERVICE MODEL

4.1 Overview of Service Activities

In this section we outline the central service activities and introduce the notions of service commitment, service process, and the service value exchange as central concepts of a general conceptual model of a service. We introduce in detail the core concepts of the service model and provide an alignment to the DOLCE foundational ontology.

Keeping in mind the wider perspective of socio-technical systems, we start by analyzing the internal structure of a *service system process*, consisting of different interconnected processes and events, resulting from complex interactions involving intentional agents and technological artifacts. Cf. Figure 2 for an overview. A *service system process* as such is composed by three main elements which are always present: the *service commitment*, the *service process*, and the *service value exchange*. Between the service commitment and both the service process and the service value exchange there is a relation of ontological dependence. The first dependence can be deduced by the informal definition given in [14]:

"A service commitment is an agent's explicit commitment to guarantee the execution of some type of actions, on the occurrence of a certain triggering event, in the interest of another agent and upon prior agreement, according to a certain specification (service description) which constraints the way service actions will be performed (service process)."

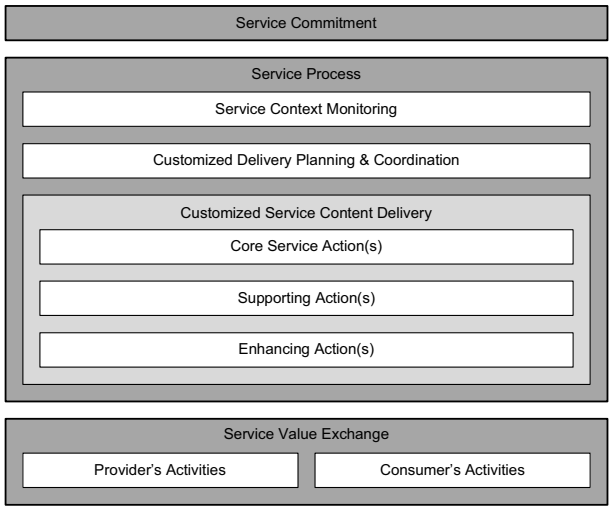


Figure 2. Activities in the Service System Process [cf. 14].

To better illustrate the concepts of the model, we introduce an example that we will use continuously throughout the paper. In order to reduce complexity, we use a service most people should be reasonably familiar with: a car wash. As with most service examples, there will be some cases where such an example is not the most appropriate. This is either because too simple or because it is out of focus. In those cases, we reinforce the explanation by using further alternative examples.

In the car wash example, we start with the event of the service commitment, when the owner of the car wash goes to the chamber of commerce to attend to all bureaucratic practices which are necessary to start the commercial activity. Among these practices,

there will be some signed official declaration in which the main features of the service are described. It is such description the car wash owner commits to.

What actually happens in the service process is constrained by what is written in the service description which defines the actions that must and/ or can be executed in the service process and the range within which a certain parameter specifying the individual actions of the service process can vary.

The service value exchange is also ontologically dependent on the commitment, as the co-creation of value can in a sense be seen as the result of a more or less specific compliance of the service actions being performed during the service process and those defined in the service description that the provider has committed to.

The central part of the service process is given by the *customized service content delivery*, which is the actual event in which one executes what has been promised in the service commitment, it is composed by core service actions, that are those actions that, in a sense, characterize a service for what it is and must necessarily be exposed to the customer, and supporting and enhancing actions, that may or may not be visible.

The *service process* presupposes two other events, namely *context monitoring* and *customized delivery planning and coordination.* The former is necessary in order to detect whether the events triggering the execution of the service are occurring, the latter comprises all the organizational activities aimed at translating into practice for a specific customer the offer contained in the service description.

As mentioned above, the service process is composed of various sub-processes. Service context monitoring is hardly explainable through the example, as the event which usually triggers the car wash service is a request by the customer who shows up with his or her car at the car wash. But there are other cases in which this activity is much more important. Take for example a firefighting service; here the triggering event is the detection of a fire in the area of responsibility of that specific department. Such detection is the result of a monitoring activity of the area.

In customized delivery planning and coordination, a car wash offers a range of different possible implementations of the service, like washing only the outside of the car, or cleaning also the inside, using particular products, like specific shampoos or waxes, etc. In the customized delivery planning phase the customer and the provider at the car wash negotiate all these details. In more complex cases customer and provider may also negotiate other parameters such as the duration of the service (e.g. for a car repair) or what kind of resources necessary for the service will be paid for (e.g. new tires).

With respect to the service delivery as such, the core action here is washing the car; singling out supporting actions is more difficult in this example, as there are not many actions that are necessary to the service but are not explicitly mentioned as constituting the service. The procedure of taking out all the contents of the car in order to be able to clean the inside could be considered a supporting action. Similarly, a complementary coffee for the waiting consumer could be a supporting action. In other examples this becomes clearer. For a firefighting service, the action of driving to the place, where the fire is, is necessary to be able to extinguish the fire. But it does not fight the fire itself as a core action does. Enhancing actions, instead, are actions which are meant to augment the value of the service. Here we could think about an additional service that is connected but not strictly included in the service, like hand polishing the car.

The *service value exchange* is a complex process involving two agents, the service provider and the service customer who, through complementary activities, contribute to developing the value chain.

Note that the *service value exchange* is not a proper part of the service process, as the latter presupposes a commitment on the side of the service provider, while the *service value exchange* may actually start before the service provider has committed to have the service executed. The first phase of the *service value exchange* is the service awareness/ need awareness phase, which might be followed by the decision – on the provider's side – to begin building the service. In the next phase offer and demand meet, as the service provider advertises her service and the service customer searches for a suitable service; these are a bundling, presentation and pricing phase on the side of the service provider and a discovery and readiness to pay phase on the side of the customer. Subsequently, the service provider and service customer negotiate according to the respective expected benefits and bearable sacrifices. Then, sometimes before the service has been delivered, at other times right after the delivery, there is the payment phase. Finally, there is the follow up phase, in which positive or negative feedback should ideally guide the provider in ameliorating the service for the future. We leave a more detailed analysis of the service value exchange to the next Section.

4.2 Service Model

We now drill a level deeper into the nature of a service and its components. While Figure 2 gave a semi-formal overview of the activities that constitute a service, Figure 3 aims at characterizing in more detail the main concepts of our general service model.

Starting from the top, we see that the main element is the *service system process*, which can contain as part one or many *services* (and each service may be part of one or more service system processes) and complies with a *service system description*, namely an abstract representation of how the whole system should behave and how the service should interact with other elements of the environment. The description of this interaction can be given (possibly among other things) by a *price plan* (which is the value that can be ascribed to that service in the market or economic system in which the service occurs) and *legal constraints*, that are the consequence of the obvious fact that a service always operates in a legal system which can limit or regulate its range of applicability.

Participants to the service system process are the *service system context* (for instance the surrounding economic, legal, and social systems) and the actors, such as the service provider, service customer, service producer, and service consumer.

With respect to the car wash example, we already introduced the service commitment and the service process. Again, we will deal with the service value exchange later. The service system description can include details on maximum liabilities during the car wash and price plans for one or multiple visits or corporate plans. The service system context description includes all contextual information which is a given and not explicitly covered by the description. For a car wash this may include that you need

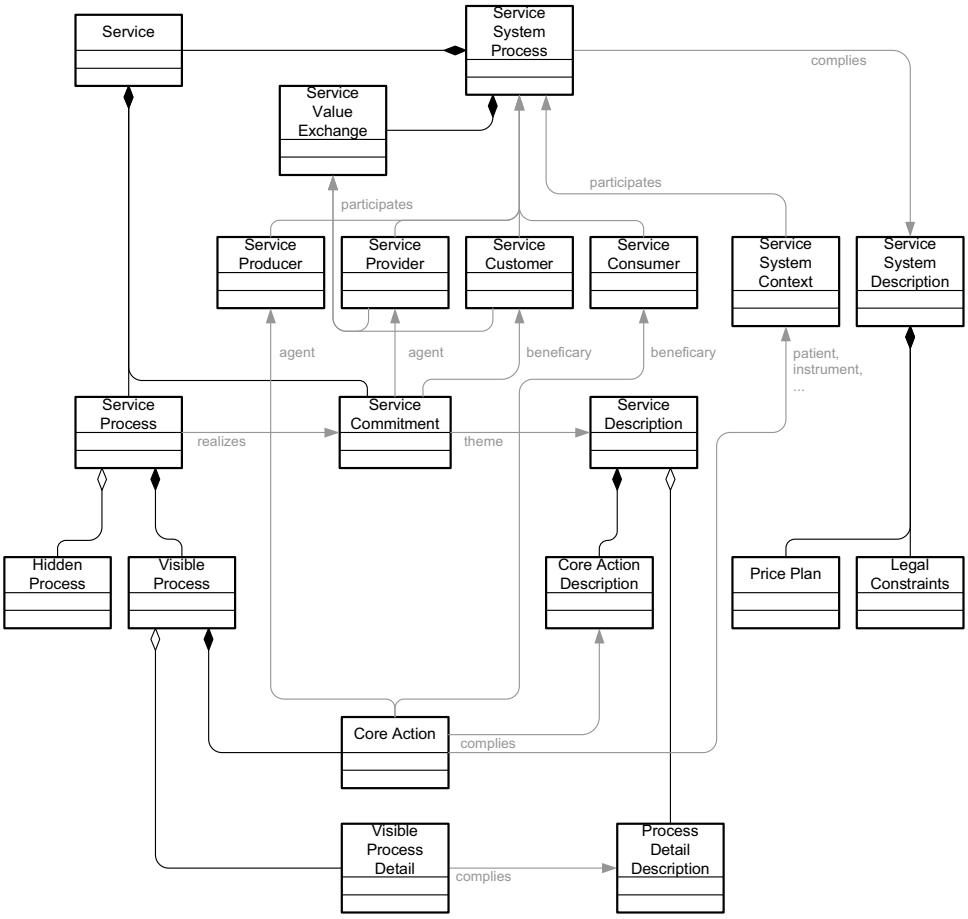


Figure 3. General Service Model.

to obtain a trade certificate before providing the service in a specific country or that you only take cleaning requests in the native language of the country you operate in.

Another participant to the service system process is, obviously, the service system itself. It has been left implicit and which can be defined as the mereological sum of all entities that participate in a service process (i.e. actors, but also resources and artifacts). There are more elaborate actor frameworks available but for sake of simplicity we refrain from expanding these entities.

The *service* has two essential parts, the service commitment and the service process. The latter should not be confused with the service system process (for instance, the way in which the price of a service changes belongs to the service system process and not to the service process).

The *service commitment* is connected through thematic relations (cf. also Section 4.3) to its components: service provider and customer, who participate in the commitment event as agent and beneficiary respectively, while the service description is the commitment's theme, in the sense that it is what the commitment is about, i.e. the provider commits to respect what is written in the description.

The *service process* realizes the commitment, i.e. it is the execution of the actions described in the service description, according to the constraints there stated and is composed by two parts: the *visible process* (mandatory) and the *hidden process* (optional); these two can be roughly identified with the front-end and the back-end processes. The visible process has some mandatory *core action* (those that in a sense define the service for what it is, i.e. the core action is what the service fundamentally does) and some optional *visible process detail*. These are usually enhancing or supporting actions which are performed in the back-end. The core action has to comply with the *core action description*, while the visible process detail has to comply with the *process detail description*. The core action description and process detail description are both part of the service description (though only the former is necessary). The hidden process does not have a correspondent in the description because it contains all those actions that are performed but not constrained by the description, i.e. the provider is free to perform such actions as he or she wishes since they are not ruled by the commitment.

Most of the above has already been exemplified in Section 4.1. As mentioned above, hidden process details are usually related to back-end activities. For example, the cleaning of car cleaning equipment after each fifth wash could be a hidden service. Hidden services are more common and also more notorious with smartphones due to their behavior to send data without the explicit consumers consent.

The agent who commits to the execution of a service process is called *service provider*, while the agent who actually executes the service is called *service producer*. These two may incidentally coincide, but this is not always the case. The *service customer* is the one who requires the service and hence also negotiates it and pays for it. Conversely, the agent who (actively or passively) participates to the service as the one whom the service is directed to is called *service consumer*. He may or may not coincide with the service customer. Service producer and service consumer both participate in the core action, the former as agent and the latter as beneficiary.

In the car wash example, the provider is the owner of the car wash, the producer is the person who washes the car and, if the driver of the car has borrowed it from someone else, the former is the customer, the one who pays for the service, while the latter, as ultimate beneficiary, is the consumer.

It is very important to conceptually distinguish the four roles, even though it can happen in practice that two or more of them coincide. In the car wash example, we can have cases in which the person who actually works in the car wash is also the owner (in this case provider and producer coincide) and, the most common case, when the driver of the car is also its owner. In this case, customer and consumer coincide. In other, more particular cases, we can have consumer and/ or customer coinciding with the producer, for instance when the car wash is an automatic one and the driver is the one who actually washes the car by driving through it.

Figure 4 shows the composition of the *service value exchange process* and how this interacts with other parts of the service system process.

Service value exchange is part of the service system process, not of the service itself. This choice is motivated by the fact that the components of the service value exchange, e.g. pricing, depend not only on elements which are intrinsic to the service, but also on things belonging to the service system context, such as laws that regulate the service or particular cultural and social traits that can make the result of a service more or less desirable.

The service value exchange is composed of five phases: *awareness*, *initiation*, *negotiation/ agreement*, *settlement*, and *after sales*. The only phase which has to be present in the service value exchange is negotiation/ agreement. Implicitly, the service provider and service customer are participants of all five phases.

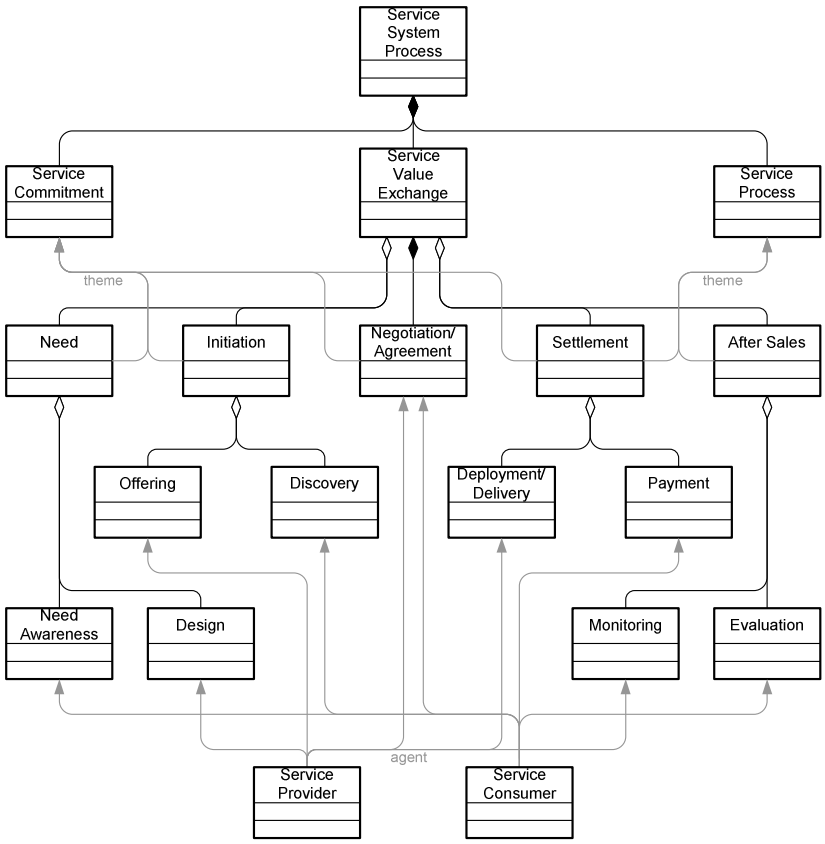


Figure 4. Service Value Exchange.

More specifically, four of these phases are composed by two complementary events, one in which the provider is the *agent* and the other in which the customer is the *agent*. During negotiation/ agreement, provider and customer, both act as agents. Furthermore, the figure details that awareness, initiation, and negotiation /agreement are all about the service description (which is what is exposed and negotiated between the parties). Settlement is relative to the visible process and the service result (which is what the customer ultimately pays for). The after sales process (both monitoring and evaluation) is about the compliance between service description and the actual service result and visible process. In order to render all these connections, we used the *theme* thematic relation.

The exchange usually starts with awareness, either of the service customers for a need they want to fulfill or the service providers in terms of an innovation they conceptualize and design. In the subsequent initiation, service providers make an offering of a service which can be discovered by service consumers. The exchange between the two parties is negotiated until an agreement is reached. During settlement the service provider invoices the service constomer who pays for the service. At this stage we exclude service delivery/ deployment as it may be considered as a service of its own to provide the service (similar to a customer buying a car wash ticket online and then buying a valet service to deliver his car to the car wash facility and back). The exchange continues after the settlement in an after sales process which allows the service provider to monitor the service use and the service consumer to evaluate the service.

In the car wash example, we can think of the owner paying for some market study, in order to understand what the needs of the customers are that should drive the innovation that may be introduced in the service. This would be the awareness phase. For what concerns the initiation phase, we may think about the many ways in which the car wash can be advertised. In the negotiation/ agreement phase, the car driver and the owner discuss the price the former will pay for the service with some customized features which are also discussed and agreed. The driver will then pay, after the car has been washed and receives an invoice. The last phase is not really typical of this kind of service, but we can suppose that every customer/ consumer can be allowed to answer to a customer satisfaction questionnaire or post a feedback on a review website.

4.3 Ontological Foundation of the General Service Model

In contrast to philosophical ontology, Information system (IS) research has inherited and altered the idea of ontology. One can speak of informational ontologies, which are partial, domain specific, and committed to an epistemological constructivism [18; 45]. This plurality substantiates the introduction of different levels to structure different ontologies according to their specificity. Most classifications distinguish top-level or foundational, domain or task, and application

ontologies [18]. Foundational ontologies are intimately related to the philosophical notion of ontology and are based on generic categories [8; 46; 54].

One of such foundational ontologies is DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [28]. Its main characteristic – which may be inferred from its very name – is to be descriptive, rather than prescriptive. It tries to describe the surface structure of language and cognition: The focus is on making explicit already existing conceptualizations, rather than prescribing how a correct representation of a conceptualization should look like. It is an ontology of particulars, whose categories are taken from the mesoscopic level.

The choice of adopting DOLCE has been determined by the features of DOLCE mentioned above that make it especially appropriate for representing services according to a commonsensical perspective that could be congenial to all different stakeholders involved.

In DOLCE there is a primary distinction between endurants (roughly speaking objects that endure in time) and perdurants (things that occur in time, like events). For the sake of this paper, we just distinguish, among endurants, agentive physical objects (APO) and non agentive physical objects (NAPO), based on the fact that these objects display intentionality or not, and among perdurants, we distinguish states (stative perdurants), events and processes (both dynamic, but the latter having a behavior characterized by repetition).

In order to anchor the concepts just introduced, we refer to this foundational ontology. Figure 5 relates the uppermost elements of the general service model to the top categories taken from DOLCE.

Thus, service provider, service producer, service customer, and service consumer are all *agentive physical objects* (APO), while the service system context is a *non agentive physical object* (NAPO). The service commitment is a *state* (the state of being committed, to be kept distinct from the commitment act, which is an instantaneous event). The core action is an *event*, while the service system process, the service value exchange, and the service in its entirety including the service process are *processes*.

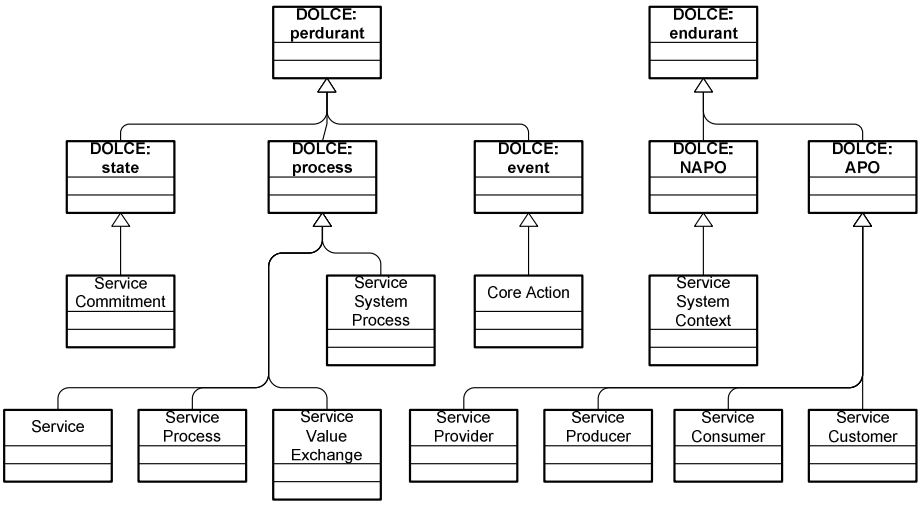


Figure 5. Relation of General Service Model and DOLCE.

The model currently lacks an *is-a relation* for entities such as the service description, all its parts, the service system description and its parts. Intuitively, we could say that these are all descriptions and, thus, social entities. For the moment, we assume that these are *non agentive non physical objects*. Descriptions, though being in DOLCE, do not appear in its stable version. Thus, we have not included them in the Figure.

We used relations like *agent*, *beneficiary*, *theme* in the Figures above. In linguistics, these are called *thematic roles* or *thematic relations* [16; 21] and they are usually meant to express the relation between a certain element of a sentence and the action expressed by the main verb of the sentence.

Our choice is to employ thematic relations to describe the relations between the core action of a service and the other elements involved.

5. CONCLUSION

By elaborating on the current state-of-the-art and research issues in the area of service science and service description, we highlighted the necessity of a shared understanding, a shared conceptualiziation, of what a service is. We also provided evidence that there is preliminary research on this topic which needs to be extended to provide a sound basis for the engineering and brokering of services in particular and in order to serve as a foundation of the service science discipline in general.

We introduced a general service model based on the foundational ontology DOLCE, characterizing services in terms of endurants, perdurants, and their relationships. Using DOLCE on the one hand ensures that all entities used in the model correspond to a well-founded primitive and, thus, are meaningful. On the other hand it guarantees that all relevant ontological primitives have been taken into consideration and an ontological completeness of the model can be assumed.

The content of the model has been derived by considering related work and through research in large-scale research projects on service such as Theseus/ TEXO which involved face-to-face methodology workshops, conference call workshops, prototyping, and validation through small and medium enterprises as use case partners.

In our understanding, a service is essentially composed of service processes whose core actions delivered by a service producer to a service consumer provide the capabilities that fulfill a service provider's commitment to a service customer. The service's description explicates these capabilities through the visible process details of the service process. We acknowledge that this part of it can be hidden in parts. Services compose a service system process which complies with legal and pricing constraints. It has a context which provides a frame for all performed actions.

The model is not yet stable. Next steps include dedicated work on special aspects such as legal aspects,

service pricing or the description of hybrid services (i.e. product accompanying/ enabling services), the complete axiomatization of the model in the Web Ontology Language (OWL), and the closer alignment with a service description language such as USDL. If the model was serializable in an interchange format, the model can provide an actionable frame for communication purposes between service provider and service customer to describe business aspects as well as capabilities of services. The service description could be used in data exchange on services for discovery or service bundling.

6. ACKNOWLEDGMENTS

The project was funded by means of the German Federal Ministry of Economy and Technology under the promotional reference "01MQ07012" and the BMBF funded project InfoStrom (13N10713). The authors take the responsibility for the contents.

REFERENCES

[1] Alter, S. 2008. Service System Fundamentals: Work System, Value Chain, and Life Cycle. *IBM Systems Journal* 47, 1, 71-85.

[2] Alter, S. 2010. Viewing Systems as Services: A Fresh Approach in the IS Field. *Communications of the Association for Information Systems* 26, 11, 195-224.

[3] Becker, J., Niehaves, B., and Janiesch, C. 2007. Socio-Technical Perspectives on Design Science in IS Research. In *Advances in Information System Development: New Methods and Practice for the Networked Society, Vol. 2*, G. Magyar, G. Knapp, W. Wojtkowski, W. G. Wojtkowski, and J. Zupančič (eds.). Springer, New York, NY, 127-138.

[4] Bitsaki, M., Danylevych, O., van den Heuvel, W., Koutras, G., Leymann, F., Mancioppi, M., Nikolaou, C., and Papazoglou, M. 2008. An Architecture for Managing the Lifecycle of Business Goals for Partners in a Service Network. In *Proceedings of the 1st European Conference ServiceWave. Lecture Notes in Computer Science Vol. 5377* (Madrid), P. Mähönen, K. Pohl, and T. Priol (eds.). 196-207.

[5] Cardoso, J., Barros, A., May, N., and Kylau, U. 2010. Towards a Unified Service Description Language for the Internet of Services: Requirements and First Developments. In *Proceedings of the 2010 IEEE International Conference on Services Computing (SCC)* (Miami, FL). 602-609.

[6] Cardoso, J., Winkler, M., and Voigt, K. 2009. A Service Description Language for the Internet of Services. In *Proceedings of the 1st International Symposium on Services Science (ISSS)* (Leipzig), R. Alt, K.-P. Fahnrich, and B. Franczyk (eds.). 229-240.

[7] Chesbrough, H. and Spohrer, J. 2006. A Research Manifesto for Services Science. *Communications of the ACM* 49, 7, 35-40.

[8] Chisholm, R. M. 1996. *A Realistic Theory of Categories*. Cambridge University Press, Cambridge.

[9] de Kinderen, S. and Gordijn, J. 2008. e^{3}Service: A Model-based Approach for Generating Needs-driven E-service Bundles in a Networked Enterprise. In *Proceedings of the 16th European Conference on Information Systems (ECIS)* (Galway). 1-12.

[10] Deutsches Institut für Normung e. V. (DIN). 2002. Grundstruktur für die Beschreibung von Dienstleistungen in der Ausschreibungsphase. PAS 1018. Beuth, Berlin.

[11] Dhanesha, K. A., Hartman, A., and Jain, A. N. 2009. A Model for Designing Generic Services. In *Proceedings of the 7th IEEE International Conference on Services Computing (SCC)* (Bangalore). 435-442.

[12] Eberts, R. E. and Bittianda, K. P. 1993. Preferred Mental Models for Direct Manipulation and Command-Based Interfaces. *International Journal of Man-Machine Studies* 35, 5, 769-785.

[13] Emmrich, A. 2005. *Ein Beitrag zur systematischen Entwicklung produktorientierter Dienstleistungen*. Dissertation, University of Paderborn.

[14] Ferrario, R. and Guarino, N. 2008. Towards an Ontological Foundation for Services Science. In *Proceedings of the 1st Future Internet Symposium (FIS). Lecture Notes in Computer Science Vol. 5468* (Wien). 152-169.

[15] Ferrario, R., Guarino, N., and Fernández Barrera, M. 2010. Towards an Ontological Foundation for Services Science: The Legal Perspective. In *Approaches to Legal Ontologies*, G. Sartor, P. Casanovas, M. Biasiotti, and M. F. Barrera (eds.). Springer, Berlin.

[16] Fillmore, C. 1971. Types of Lexical Information. In *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, D. Steinberg and L. A. Jacobovitz (eds.). Cambridge University Press, London.

[17] Gemino, A. and Wand, Y. 2003. Evaluating Modeling Techniques Based on Models of Learning. *Communications of the ACM* 46, 10, 79-84.

[18] Guarino, N. 1998. Formal Ontology and Information Systems. In *Proceedings of the 1st International Conference on Formal Ontology in Information Systems (FOIS)* (Trento), N. Guarino (ed.). 3-15.

[19] Hadley, M. 2009. Web Application Description Language, W3C Member Submission 31 August 2009. Retrieved 2010-12-06 from http://www.w3.org/Submission/wadl/.

[20] Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28, 1, 75-105.

[21] Jackendoff, R. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

[22] Janiesch, C. 2007. *Contextual Method Design: Constructing Adaptable Modeling Methods for Information Systems Development*. Dissertation, University of Münster.

[23] Krogstie, J., Sindre, G., and Jørgensen, H. 2006. Process Models Representing Knowledge for Action: A Revised Quality Framework. *European Journal of Information Systems* 15, 1, 91-102.

[24] Larkin, J. H. and Simon, H. A. 1987. Why a Diagram Is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11, 1, 65-99.

[25] Maglio, P. and Spohrer, J. 2008. Fundamentals of Service Science. *Journal of the Academy of Marketing Science* 36, 1, 18-20.

[26] Maglio, P., Srinivasan, S., Kreulen, J., and Spohrer, J. 2006. Service Systems, Service Scientists, SSME, and Innovation. *Communications of the ACM* 49, 7, 81-85.

[27] Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B.,

Payne, T., Sirin, E., Srinivasan, N., and Sycara, K. 2004. OWL-S: Semantic Markup for Web Services. W3C Member Submission. Retrieved 2010-08-23 from http://www.w3.org/Submission/OWL-S/.

[28] Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. 2003. WonderWeb Deliverable D18. Ontology Library (final). Retrieved 2010-08-10 from http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf.

[29] Mayer, R. J. 1989. Models for Understanding. *Review of Educational Research* 59, 1, 43-64.

[30] Moody, D. L. 1998. Metrics for Evaluating the Quality of Entity-Relationship Models. In *Proceedings of the 17th International Conference on Conceptual Modeling (ER). Lecture Notes in Computer Science Vol. 1507* (Singapore), T. W. Ling, S. Ram, and M.-L. Lee (eds.). 211-225.

[31] Neisser, U. 1987. From Direct Perception to Conceptual Structure. In *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, U. Neisser (ed.). Cambridge University Press, Cambridge, 11-24.

[32] Norman, D. A. 1983. Some Observations on Mental Models. In *Mental Models*, D. Gentner and A. L. Stevens (eds.). Erlbaum, Hillsdale, NJ, 7-14.

[33] O'Sullivan, J. 2006. *Towards a Precise Understanding of Service Properties*. Dissertation, Queensland University of Technology, Brisbane.

[34] Oberle, D., Bhatti, N., Brockmans, S., Niemann, M., and Janiesch, C. 2009. Countering Service Information Challenges in the Internet of Services. *Business & Information Systems Engineering* 1, 5, 370-390.

[35] Oberle, D., Lamparter, S., Grimm, S., Vrandečić, D., Staab, S., and Gangemi, A. 2006. Towards Ontologies for Formalizing Modularization and Communication in Large Software Systems. *Applied Ontology* 1, 2, 163-202.

[36] Object Management Group (OMG). 2009. Service oriented architecture Modeling Language (SoaML). 1.0 - Beta 2. Retrieved 2010-08-23 from http://www.omg.org/spec/SoaML/1.0/Beta2/PDF.

[37] Organization for the Advancement of Structured Information Standards (OASIS). 2006. Reference Model for Service Oriented Architecture 1.0. OASIS Standard. Retrieved 2010-08-23 from http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf.

[38] Organization for the Advancement of Structured Information Standards (OASIS). 2009. Reference Architecture Foundation for Service Oriented Architecture. Version 1.0. Retrieved 2010-07-17 from http://docs.oasis-open.org/soa-rm/soa-ra/v1.0/soa-ra.html.

[39] Petrie, C. and Bussler, C. 2008. The Myth of Open Web Services: The Rise of the Service Parks. *IEEE Internet Computing* 12, 3, 94-96.

[40] Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., and Fensel, D. 2005. Web Service Modeling Ontology. *Applied Ontology* 1, 1, 77-106.

[41] SAP AG. 2007. Standardized Technical Architecture Modeling: Conceptual and Design Level. Version 1.0. Retrieved 2010-08-10 from http://www.fmc-modeling.org/download/fmc-and-tam/SAP-TAM_Standard.pdf.

[42] SAP Research. 2009. USDL Specifications. Retrieved 2010-08-23 from http://www.internet-of-services.com/index.php?id=54.

[43] Schütte, R. and Zelewski, S. 2002. Epistemological Problems in Working with Ontologies. In *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI)*, *Vol. VII* (Orlando, FL). 161-167.

[44] Shaft, T. M. and Vessey, I. 1998. The Relevance of Application Domain Knowledge: Characterizing the Computer Program Comprehension Process. *Journal of Management Information Systems* 15, 1, 51-78.

[45] Smith, B. 1998. Basic Concepts of Formal Ontology. In *Formal Ontology in Information Systems*, N. Guarino (ed.). IOS Press, Amsterdam, 19-28.

[46] Sowa, J. F. 1999. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing, Pacific Grove, CA.

[47] Spohrer, J., Maglio, P., Bailey, J., and Gruhl, D. 2007. Steps Toward a Science of Service Systems. *Computer* 40, 1, 71-77.

[48] Stachowiak, H. 1973. *Allgemeine Modelltheorie*. Springer, Wien.

[49] Sun, W., Zhang, K., Chen, S.-K., Zhang, X., and Liang, H. 2007. Software as a Service: An Integration Perspective Service-Oriented Computing. In *Proceedings of the 5th International Conference on Service-oriented Computing (ICSOC). Lecture Notes in Computer Science*, *Vol. 4749* (Wien). 558-569.

[50] Sycara, K. 2007. Unthethering Semantic Web Services. *IEEE Intelligent Systems* 22, 6, 11-13.

[51] The OpenGroup. 2010. SOA Ontology. Public Draft 3.1. Retrieved 2010-08-23 from https://www.opengroup.org/projects/soa-ontology/ uploads/40/22324/SOA_ontology_public_draft_3_1.pdf.

[52] Tolvanen, J.-P. 1998. *Incremental Method Engineering with Modeling Tools: Theoretical Principles and Empirical Evidence*. Dissertation, University of Jyväskylä.

[53] Toma, I. 2010. *Modeling and Ranking Semantic Web Services Based on Non-Functional Properties*. Dissertation, University of Innsbruck.

[54] Wand, Y. and Weber, R. 1993. On the Deep Structure of Information Systems. *Information Systems Journal* 5, 3, 203-223.

[55] Wand, Y. and Weber, R. 2002. Research Commentary: Information Systems and Conceptual Modeling: A Research Agenda. *Information Systems Research* 13, 4, 363-376.

[56] World Wide Web Consortium (W3C). 2009. Service Modeling Language, Version 1.1. W3C Recommendation. Retrieved 2010-08-23 from http://www.w3.org/TR/sml/.

[57] Wyssusek, B. and Klaus, H. 2005. On the Foundation of the Ontological Foundation of Conceptual Modeling Grammars: The Construction of the Bunge-Wand-Weber-Ontology. In *Proceedings of the 1st International Workshop on Philosophical Foundations of Information Systems Engineering (PHISE)* (Porto). 583-593.

Prioritizing Strategic IT Projects with Tropos

Andreas Gehlert
Federal Ministry of the Interior
Alt-Moabit 101D
10559 Berlin
++49 (0) 30 18681 1743
andreas.gehlert@bmi.bund.de

ABSTRACT

One of the daily tasks of an enterprise architect is to prioritize strategic IT projects. To achieve a business-IT alignment, this prioritization needs to be based on business strategies and goals. Therefore, business goals and their traceability to strategic IT projects are relevant for the enterprise architect. However, surprisingly little formalisations and reasoning techniques have been developed in the enterprise architecture domain. In this paper we show that the popular goal modelling technique Tropos together with its formal reasoning techniques can support the enterprise architect when prioritizing strategic IT projects. We prove the feasibility of our work with a tool implementation of the proposed modelling language and its corresponding algorithms; and demonstrate their usefulness with the help of an example taken from the enterprise architecture literature.

Keywords

Enterprise Architecture, Prioritization, Goal Modelling, Tropos, Formal Methods

1. Introduction

Enterprise architecture (EA) is a young discipline and has been recently recognised not only by researchers but also by practitioners. In this paper we are interested in the link between strategic IT projects, initiated by the enterprise architect to strengthen the company's architecture, and the company's business goals (business-IT-alignment [7, p. 77, 12-13, 21, 26]). Since the enterprise architecture should always serve the business [7, pp. 72], the enterprise architect should be able to demonstrate the value of each strategic IT project to the business. This demonstration can be achieved by showing how a strategic IT project contributes to business goals of the company. Linking each strategic IT project with the business goals enables the enterprise architect, additionally, to prioritize strategic IT projects according to the importance of the business goals.

This paper concentrates on the latter aspect and shows how enterprise architects can prioritize their strategic IT projects according to the business goals of their company. We show that current goal modelling techniques used in the EA field are not formal enough to allow for such a prioritization. The research question of this paper is, therefore, to provide a goal modelling technique suitable for the EA domain, which allows prioritizing strategic IT projects efficiently.

The construction of such a technique is a design science activity and we, therefore, use a design science research method [9, 19]. Such a method starts with requirements for the envisioned artefact. These requirements do not only drive the development of the artefact but are also used to evaluate existing artefacts (cf. Section 2). We use the following requirements for our paper:

- *Req-1* *The goal model must provide concepts to represent goals, relations between goals and relation between goals and strategic IT projects.*
- *Req-2* *The goal modelling technique is formal enough to apply formal reasoning techniques.*
- *Req-3* *A formal reasoning technique is desirable which calculates the impact of a strategic IT project on all goals of the goal model.*
- *Req-4* *The additional information needed to use the formal reasoning technique should be minimal.*

Req-1 is derived from the fact that the enterprise architect may need to explain the value of strategic IT projects to the business. Therefore, the goal modelling technique needs to allow modelling goals, strategic IT projects and relations between these elements (see motivation above). Req-2 and Req-3 enables to use the goal modelling technique efficiently by applying formal reasoning techniques. These formal reasoning techniques are a prerequisite to execute parts of the reasoning process on a machine and, therefore, release the enterprise architect from manual work. This is especially interesting for large goal models. Req-4 states that the usage of formal reasoning techniques should not come at the cost of higher modelling efforts for the enterprise architect.

In the sequel we demonstrate how existing goal modelling techniques from the requirements engineering (RE) domain and its reasoning techniques can be applied to the problem of prioritizing strategic IT projects. The paper is structured as follows: Section 2 describes the related work in EA and RE on goal modelling. We show that Tropos – a goal modelling language from the RE domain – is a good candidate, which could be extended to suit our requirements. In section 3 Tropos is briefly described, applied to an example from the EA domain and extended by an algorithm. This algorithm generates a prioritisation of strategic IT projects based on a given goal model. The section also contains a lightweight evaluation of the proposed technique and the algorithm using a tool implementation. We discuss important design decisions and assumptions in Section 4 and summarise our findings in Section 5.

2. Related Work

In this section we review goal modelling techniques found in the EA and RE disciplines. The aim of this section is to identify interesting approaches, whose fragments can be used to construct a goal modelling technique, which fulfils our requirements.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

2.1 Goal Modelling in EA

The popular Zachman framework did not include goals from the very beginning [26]. Sowa and Zachman introduced them as the "why" perspective on enterprise architectures five years later [21]. Goals are part of business strategies and are modelled as trees using goal-sub-goal relations. Formal reasoning techniques for analysing the goal tree are not included in the Zachman framework (violations of Req-1 and Req-3).

The same can be said for ARIS [17-18]. In ARIS, goals and relations between goals can be modelled. ARIS' goal trees share the same shortcomings with the goal trees of the Zachman framework (violations of Req-1 and Req-3).

Goals are also mentioned in TOGAF's Architecture Development Method (ADM), Phase A [12]. The so called architecture vision is a set of goals for that architecture. It is used to communicate and to agree upon the future enterprise architecture. This vision should be based on an existing business strategy. Despite the importance of the architecture vision, no methodological support can be found in TOGAF (violations of Req-1, Req-2 and Req-3).

QUASAR Enterprise prescribes the step-wise refinement of goals from business goals to IT-related goals [7, pp. 72]. The relations between goals should serve as traceability links from business to IT. These traceability links can be exploited to explain the value the IT provides to the business. Although the modelling technique is not described in detail, it seems that a tree structure is assumed between the goals. This tree structure shares the same properties with the goal trees of the Zachman Framework and ARIS (violations of Req-1 and Req-3).

In Archimate goals are not seen as core construct of the framework. The authors argue that this concept and its relations can be added to the framework as needed ([3, p. 7]: violations of Req-1, Req-2 and Req-3).

2.2 Goal Modelling in RE

Goal modelling in RE can be divided into two main research lines: research on i* introduced by Yu [25] and research on KAOS introduced by Dardenne [6]. The i* approach was later extended by Tropos [4-5]. Based on these goal modelling approaches, formal reasoning techniques were developed to allow choosing between different designs of a software system.

Mylopoulos introduced was the first who introduced a formal reasoning technique in the goal modelling domain [15]. The author describes qualitative propagation rules, which explain how an evidence for a satisfied (denied) goal can be propagated in a goal graph. The problem with such qualitative reasoning techniques is that they become quickly inconclusive [23, p. 389]. Therefore, the use of qualitative techniques for EA is limited since the reasoning algorithm may not produce valid results (violation of Req-3).

Mylopoulos' algorithm was later extended by Letier and van Lamsweerde [14], van Lamsweerde [23] and Sebastiani et al. [20] to support quantitative reasoning in KAOS and Tropos: However, the approaches rest on the introduction of additional variables for each goal (quality variables in [14]; gauge variables in [23] and costs in [20]). In addition, rules must be assigned to each relation between goals, to propagate these variables along the goal graph structure. These approaches clearly provide higher accuracy and interpretability of the results but at the cost of higher efforts for eliciting the required information. This property hinders the applicability of these approaches in the EA domain (violation of Req-4).

The most promising algorithm was introduced in Tropos by Giorgini et al. [10-11]. The authors describe a formal extension to Tropos, which can propagate qualitative and quantitative satisfaction rates of goals in a goal model. The algorithm computes satisfiability and deniability values of all goals in the goal model by analysing the relationships between these goals. Since the approach supports quantitative reasoning, the inconclusiveness problem does not occur. In addition, the algorithm uses only information, which is already encoded in goal models and, therefore, does not require additional information to use the formal reasoning techniques (fulfilment of Req-2 and Req-4). However, Tropos's algorithm is not capable of computing a prioritization of strategic IT projects (modelled as plans in Tropos) as the algorithm does not compute a quantitative number expressing how one goal influences all other goals in the goal model. The Tropos' algorithm needs to be extended for this purpose (partial fulfilment of Req-3).

2.3 Conclusion for this Paper

As we have seen, goal modelling is an important topic in the EA literature. However, surprisingly little formalisations have been developed in this area. The focus is more on simple and informal goal trees, which hinder the usage of formal reasoning techniques.

In line with previous research by Bleistein et al., who applied RE goal modelling techniques to modelling business strategies, we conclude that using RE goal models and their formal reasoning techniques will be beneficial for the enterprise architect [1-2]. The reason for this conclusion is threefold:

1. Goal models in RE are used to prioritize requirements and therewith to choose between different system designs [24]. The task of making an informed choice between different system designs is very similar to the task of the enterprise architect to choose between different strategic IT projects.
2. Goal modelling in requirements engineering is used to align requirements to initial stakeholder goals and, therefore, allows tracing requirements to business goals [22]. This traceability is very similar to tracing strategic IT projects to business goals.
3. Goal modelling techniques from the RE field are well developed, e. g. they support different types of relations between goals and are not limited to goal trees. In addition, these techniques are formal enough to apply formal reasoning techniques to them (cf. Section 2.2). Since goal modelling techniques in the EA domain are not yet very well developed, it is reasonable to transfer existing knowledge from the RE domain to the EA domain.

Since RE and goal models are used for similar purposes (prioritisation of requirements vs. prioritisation of strategic IT projects; traceability of requirements to business goals vs. traceability of EA goals to business goals) we conclude that RE goal modelling techniques are *applicable* to EA problems as well. Together with their formal focus, which enables formal reasoning and the empirical findings we propose to use these techniques in the EA domain (fulfilment of Req-1).

We chose Tropos from the list of goal modelling techniques discussed in Section 2.2 because it is equipped with formal reasoning techniques, which can be used to construct an algorithm for prioritizing strategic IT projects (Req-2; partial fulfilment of Req-3). Finally, the formal reasoning technique is based on the information in the goal model only; no further elicitation activities are needed (fulfilment of Req-4).

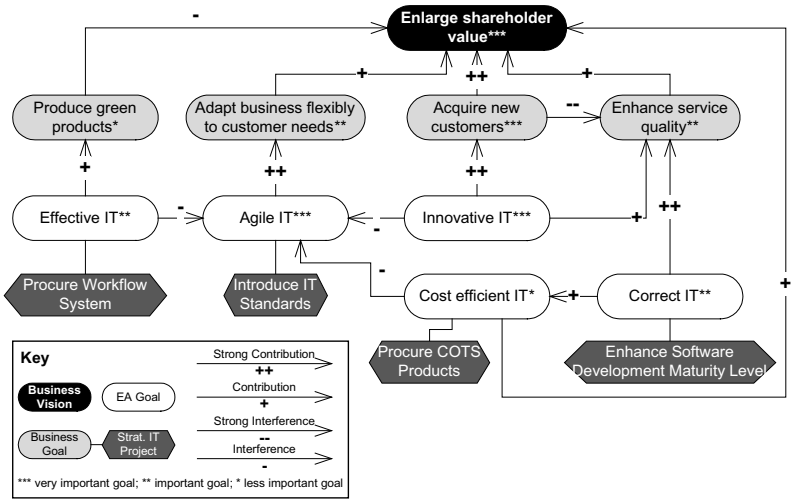


Figure 1: Tropos Example (adapted from [7])

3. Goal Modelling with Tropos

In this section we develop an algorithm, which satisfies requirement Req-3 and provides decision support for prioritizing strategic IT projects. We ground this algorithm in Tropos – an established goal modelling technique. We start our investigation by introducing Tropos (Section 3.1). Then we explain how Tropos can be used to describe and analyse EA problems (Section 3.2). We develop the algorithm in Section 3.3. Finally in Section 3.4, we provide a lightweight evaluation of our technique by applying it to an example taken from the literature. This evaluation also demonstrates how the approach can be used to prioritize strategic IT projects.

3.1 Introduction to Tropos

Tropos is a RE technique, which rests on the agent oriented paradigm and uses goal modelling techniques known from i* for analysing early and late requirements. These early requirements are documented as actor models and goal models [for the following explanations see 4, pp. 206]. *Actor models* include stakeholders of the later system modelled as actors and describe the actors' goals and dependencies. The actor model is complemented by a goal model for each actor. This *goal model* shows the decomposition of the actor's goals into sub-goals.

Sub-goals in goal models can be derived by decomposing super-goals with And/Or decomposition links or by using contribution links. Decomposition links are used to hierarchically decompose a goal into sub-goals. In case of an *Or decomposition*, the super-goal is satisfied if at least one sub-goal is satisfied (modelling alternatives); in case of an *And decomposition* all sub-goals must be satisfied to satisfy the super-goal. *Contribution links* are further described with a strength, which specifies how much a sub-goal contributes to a super-goal. These strengths could also be negative to describe an interference of goals.

Goals are further distinguished in soft-goals and hard-goals. *Hard-goals* have clear cut criteria to decide whether the goal is satisfied. *Soft-goals* do not have such clear cut criteria.

Each goal-model can be complemented with plans. *Plans* describe tasks or activities to be carried out to achieve a certain goal. Plans are connected to goals with *means-end links*, where the plan represents the "end" and the goal represents the "mean".

Figure 1 demonstrates the elements of Tropos. This example used throughout this paper. The model contains a part of a business strategy of a company and its related EA goals. The EA goals are taken from [7, p. 77]. The top goal of this model is to "Enlarge the shareholder value". It is supported by the goals "Adapt business flexibly to customer needs" and "Enhance service quality"; it is strongly supported by the goal "Acquire new customers" and interfered by the goal "Produce green goods". The EA goals are interpreted analogously. For instance, the goal of an "Agile IT" supports the goal "Adapt business flexibly to customer needs" strongly. Strategic IT projects are modelled in Tropos as plans (hexagon). Linking these strategic IT projects to goals means that the goal is satisfied once the project is realised. For instance, implementing the strategic IT project "Procure Workflow System" satisfies the goal "Efficient IT".

3.2 Applying Tropos to EA Problems

In this subsection we describe important restrictions on using Tropos. These restrictions are expressed as design decisions and assumptions. These design decisions and assumptions allow a) tailoring Tropos to EA problems by choosing a relevant part of the Tropos language; and b) focussing the paper. The consequences of all design decisions and assumptions are discussed in Section 4.

We concentrate on goal models here because we are interested in breaking down business goals to strategic IT projects and not in analysing the stakeholders involved in this activity. Therefore, we deal with goal models only.

DD-1 Actor models are disregarded.

In addition, we restrict ourselves to soft-goals. In RE, hard goals describe (mainly) functional requirements. In this sense it can be decided using clear-cut criteria whether a software system has a certain functionality or not. Soft-goals do not have such clear cut criteria. They are used to describe quality requirements, e. g. usability. Using soft-goals as concept for describing high-level business and EA goals seems reasonable since there are no clear-cut criteria for goals such as agility, efficiency, time-to-market, etc.

DD-2 Hard-goals are disregarded.

It is unlikely that the remaining soft-goals are equally important [cf. 23, p. 390]. Therefore, we introduce the new meta-property importance and assign it to the soft-goal concept. We distinguish between very important, important and less important goals and visually mark them as "***", "**" and "*" respectively.

DD-3 The goal concept is extended by the property importance. The property values are in the range {very important, important, less important}.

Furthermore, we do not use decomposition links here. The reason for this design decision is to reduce the complexity in the later model. For the remaining contribution links we have to decide whether we use a quantitative or qualitative notation for the strengths of the links [10]. Using quantitative reasoning real numbers are assigned to the strength of a contribution link whereas using qualitative reasoning the contribution link is usually annotated as $\{++, +, --, -\}$ meaning strong contribution, contribution, strong interference and interference respectively. We use the qualitative model here to reduce the effort for specifying the contribution link strength.

DD-4 Decomposition links are disregarded.

DD-5 A qualitative notation for contribution link strengths is used.

We assume that strategic IT projects are modelled as plans and are assigned to goals using means-end links. This model fragment means that an EA activity contributes to the satisfaction of the assigned goal. We assume that each plan is assigned to exactly one goal and that the implementation of this plan satisfies this goal completely.

A-1) Plans represent strategic IT projects and are assigned to exactly one goal. Realising this plan means fully satisfying the assigned goal.

3.3 Impact Analysis in Tropos

Once the goal model is developed and the relevant strategic IT projects are assigned to these goals, the question about the priorities of strategic IT projects arises. When selecting strategic IT projects, the enterprise architect wants to achieve a positive impact on many goals while avoiding a negative impact on goals at the same time [16, p. 95]. Therefore, an impact analysis based on the dependencies between the goals and their importance is a good starting point for this analysis.

Since each plan is assigned to exactly one goal, we can disregard the plans and can concentrate on the goals only (assumption A-1). A goal model is then a directed, weighted graph $DG = (G, C)$ where G represents a set of goals and C represents a set of contribution links with strength ω. The graph is connected (e. g. there are no lose goals), incomplete and may contain cycles as well as multiple edges.

In the following we distinguish between a positive and a negative impact of a goal g ($I^+(g)$ and $I^-(g)$ respectively). Informally the positive impact $I^+(g)$ describes the contribution of g to all connected goals. The negative impact $I^-(g)$ describes the interference of g with all connected goals. In addition, an overall impact $I(g) = I^+(g) - I^-(g)$ is used for an initial prioritization of goals. By connected goals we mean all goals to (with) which g contributes (interferes) including all transitively connected goals. The contribution (interference) to (with) transitively connected goals should degrade the longer the path to the connected goal is.

Let P(g) describe the relative importance of goal g in comparison with all other goals in the goal graph. It describes the importance of g in isolation, e. g. without considering its relations with other goals. This importance value needs to be considered when calculating the impact values and deriving the prioritization of goals.

For the following formalisations of the term "impact" we extend an algorithm proposed by Giorgini et al. [10-11]. The authors introduce two variables $Sat(g)$ and $Den(g)$ for each goal, which describe the evidence that goal g is satisfied or denied. The authors also define rules, which propagate these evidence values along the contribution links in the goal graph. Here, we use the probabilistic model described by the authors and define: Let a contribution link $c \in C = (g_1, g_2)$ with strength ω be represented as $g_1 \xrightarrow{\omega} g_2$. The propagation rules can then be described as:

$g_1 \xrightarrow{\omega} g_2$:

$$\begin{cases} Sat(g_2) \geq Sat(g_1) * \omega;\ Den(g_2) = Den(g_1) * \omega & \text{if } \omega > 0 \\ Sat(g_2) \geq Den(g_1) * |\omega|;\ Den(g_2) = Sat(g_1) * |\omega| & \text{if } \omega < 0 \end{cases}$$

The first line of the formula states that the evidence of a goal satisfiability and deniability is propagated along the contribution link and degraded by ω in case of a (positive) contribution ($\omega > 0$). The second line states that the evidence for satisfiability of g_1 is propagated to the deniability value of g_2 and vice versa and degraded by ω in case of an interference (negative contribution; $\omega < 0$). Please note that according to these definitions, $Sat(g)$ and $Den(g)$ are always positive.

In addition to the propagation rules, Giorgini et al. provide an efficient algorithm to compute $Sat(g)$ and $Den(g)$ for an arbitrary goal model. This algorithm especially considers cycles and multi-edges in the graph. However, the algorithm can only work if $Sat(g)$ and $Den(g)$ is given for a set of goals before running the algorithm. These initial values correspond to an alternative, which should be evaluated.

To calculate the impact of a goal g, we initialise its satisfiability value $Sat(g) = 1$, its deniability value $Den(g) = 0$ and set the satisfiability and deniability values to zero for all remaining goals: $\forall_{g_i \in G \wedge g_i <> g}: Sat(g_i) = Den(g_i) = 0$. Based on these initial values we calculate the satisfiability and deniability values of all goals using Giorgini et al.'s algorithm. The positive (negative) impact $I^+(g)$ ($I^-(g)$) is then the sum of these satisfiability (deniability) values of each goal weighted by the importance of these goals.

For calculating the impact values we can write the following pseudo code:

1) Initialisation: $Sat(g) = 1, Den(g) = 0$ and $\forall_{g_i \in G \wedge g_i <> g}: Sat(g_i) = Den(g_i) = 0$.
2) Apply Giorgini et al.'s algorithm to compute $Sat()$ and $Den()$ for all goals.
3) Compute impact: $I^+(g) = \left(\sum_{g_i \in G} Sat(g_i) * P(g_i)\right)$ - $P(g)$ and $I^-(g) = \sum_{g_i \in G} Den(g_i) * P(g_i)$

The algorithm presented here works with quantitative measures for ω. However, design decision DD-5 prescribes a qualitative model for the contribution link strengths. Therefore, we have to define a mapping between the two systems. We use the following mapping here:

$$\omega_{quantitative} = \begin{cases} 1 \text{ if } \omega_{qualitative} = "++" \\ 0.5 \text{ if } \omega_{qualitative} = "+" \\ -1 \text{ if } \omega_{qualitative} = "—" \\ -0.5 \text{ if } \omega_{qualitative} = "-" \end{cases}$$

Similarly, we use the following quantification for the importance of goals:

$$P(g)_{quantitative} = \begin{cases} 1 \text{ if } P(g)_{qualitative} = \text{"very important"} \\ 0.5 \text{ if } P(g)_{qualitative} = \text{"important"} \\ 0.25 \text{ if } P(g)_{qualitative} = \text{"less important"} \end{cases}$$

Figure 2 provides an exemplary application of the algorithm to a simple goal model. The impact values for all goals are computed as follows:

- *Goal C* does neither contribute nor interfere with any other goal. Therefore, its positive and negative impacts are $I^+(C) = I^-(C) = 0$.
- *Goals B and E* support (interfere with) goal C. Therefore they have a positive (negative) impact of $I^+(B) = 0.5$ ($I^-(E) = 0.5$) and a negative (positive) impact of $I^-(B) = 0$ ($I^+(E) = 0$).
- *Goal D* contributes to E and has, therefore, a positive impact of $I^+(D) = 0.5$. In addition, it interferes with C transitively via E and has, therefore, a negative impact of $I^-(E) = 0.25$.
- *Goal A* contributes to goals B (strongly) and transitively to goal C (left hand side of Figure 2). The positive impact of this part is 1.5. In addition, goal A strongly interferes with goal D. Since goal D contributes to goal E, A also interferes transitively with goal E (see algorithm). Therefore, the negative impact of goal A from this part is 1.5. In addition, goal E interferes with goal C. Due to the interference of A with D the interference between E and C is calculated as positive impact. Therefore, the positive transitive impact from goal A on goal C is 0.25. Since $Sat(A)$ must be greater than 1.5 the 0.25 value from the sub-graph A-D-E-C is disregarded. Therefore, goal A has a positive and negative impact of $I^+(A) = I^-(A) = 1.5$.

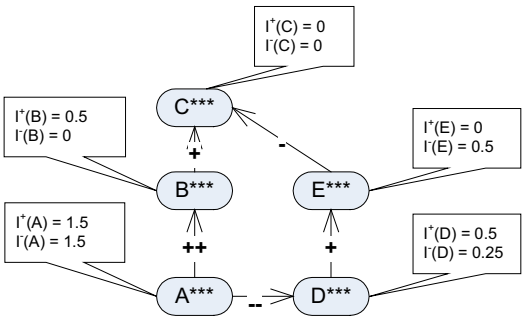


Figure 2: Exemplary Calculation of the Impacts in a Simple Goal Model

3.4 Lightweight Evaluation Using an Example

We have already shown that the technique is suitable for EA problems (fulfilment of Req-1; see Section 3.2) and that an algorithm can be constructed that calculates the impact of one goal on the entire goal model (fulfilment of requirements Req-2 and Req-3; see Section 3.3). Here, we are interested in demonstrating that the proposed technique and its corresponding algorithm are useful for answering important EA questions (substantiating the fulfilment of Req-1). We analyse this usefulness with the help of the example depicted in Figure 1.

For this proof of concept, we prototypically implemented the algorithm in Microsoft Visio (see Figure 3).

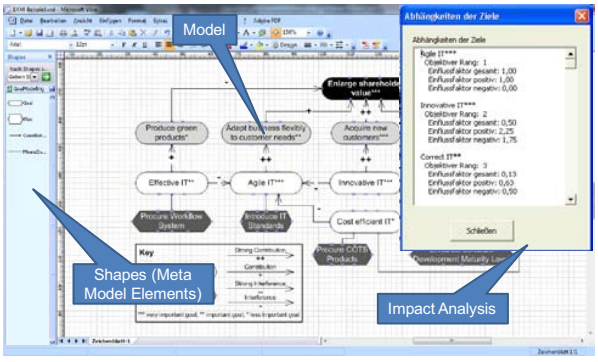


Figure 3: Screenshot of the Prototypical Tool Implementation

We used the introduced algorithm to calculate overall impact values $I(g)$ for all goals in Figure 1. This overall impact value respects positive (wanted) impacts as well as negative (unwanted) impacts respectively. It reflects our previous observation that the enterprise architect should concentrate on goals with high positive but low negative impact values [16, p. 95]. Figure 4 depicts the visualised impact values for all goals in the goal model of Figure 1. We have ordered the goals according to their impacts to improve the readability of the diagram.

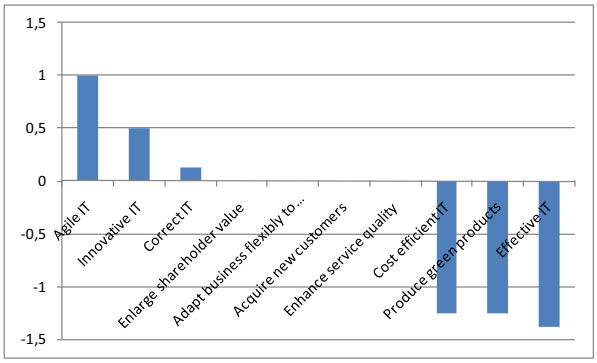


Figure 4: Impacts of all Goals for the Example in Figure 1

The typical structure of the goal model in Figure 1 results in a small number of goals with high impact values and a large number of goals with low impact values. Figure 4 clearly shows that the impact values are degrading rapidly in the example model. This property helps the enterprise architect to concentrate on few important goals rather than on may equally important ones. Having an agile and innovative IT are the two top goals in this example.

Another property of the algorithm can be clearly identified in Figure 4: The algorithm considers the impact values of all (transitively) connected goals. Therefore, it is most likely (depending on the concrete distribution of the contribution link strengths) that fine-grained goals have higher impacts values than coarse-grained ones. Therefore, fine-grained goals will be in the focus of the enterprise architect. Applying this finding to Figure 4, we clearly see that the top goal "Enlarge shareholder value" has no impact on

the goal model and that the business goals are distributed "around" this top-goal. Since these business goals are outside the scope of the enterprise architect, they can be disregarded for the prioritization of strategic IT projects.

Assumption A-1), furthermore, ensures that strategic IT projects (modelled as plans in Tropos) are assigned to exactly one goal. Since concrete strategic IT projects need to be assigned to concrete goals, this leads to the situation that the enterprise architect refines the goal model. Together with the previous observation, the algorithm ensures that fine-grained goals are preferred. Consequently, the enterprise architect gets advice on concrete strategic IT projects, which should be implemented in the future. The same holds true for strategic IT projects, which should be avoided due to the low impact value of their assigned goal.

Given our example the enterprise architect should concentrate on introducing IT standards assigned to the high impact goal "agile IT". In addition, the enterprise architect should not primarily strive for an effective and cost efficient IT and, therefore, should not implement the strategic IT projects "Procure Workflow System" and "Procure COTS products".

The example shows that there is no plan associated to the high impact goal "Innovative IT". The enterprise architect should consider either refining this goal or adding concrete strategic IT projects (plans) to it. In this way new and previously unrecognised activities are considered. Vice versa, our analysis demonstrates that one goal has a significant negative impact and should, therefore, not be supported by strategic IT projects (e. g. efficient IT in our example). The enterprise architect might – after a thorough analysis – decide to remove these activities from the model to avoid them entirely in the future.

We can summarise the findings from our lightweight evaluation of the proposed techniques with three guidelines for the enterprise architect:

1) Goals with high impact values should be achieved. Goals with low impact values should be disregarded.
2) Goals with high impact values and no associate plans should be refined and new plans should be associated to these goals.
3) Removing plans for goals with low impact values associated with these plans should be considered.

4. Discussion

Our approach is based on five design decisions and one assumption. The design decisions basically remove concepts from the Tropos goal modelling vocabulary, while the assumption restricts the ways Tropos is used. The impact of these decisions and assumptions are discussed in the following.

Design decision DD-1 removes actor models from the Tropos vocabulary. These actor models can potentially be used to analyse the enterprise architecture's stakeholders [12, pp. 281]. In combination with a formal reasoning algorithm, the impact of strategic IT projects could be traced back not only to business goals but also to the stakeholders. However, the current Tropos algorithm does not cover actor models and, therefore, the algorithm proposed here cannot be easily extended to actor models. To construct such an algorithm and to proof its usefulness in the EA domain is, therefore, subject to future research.

Design decision DD-2 disregards the hard goal concept. Hard goals are goals, which have clear-cut criteria to decide whether this goal is fulfilled. In terms of our algorithm it means that hard goals should have either a satisfiability value of 1 or a deniability value of 1 (but particularly no values <1). Giorgini et al.'s algorithm explicitly supports hard goals [10]. Therefore, relaxing this limitation does not have any impact on the formal part of the paper. However, it should be carefully investigated whether hard goals are useful in EA since they may increase the complexity of the goal model without having any other positive effect.

Design decision DD-3 introduced the meta-property importance for the goal concept with its possible values "very important", "important" and "less important". Although this extends the Tropos method, this extension was necessary to reflect the fact that not every goal is equally important for the enterprise architect. In addition, the extension has no impact on the existing Tropos algorithm so that existing formalisations can be used without modification.

Design decision DD-4 restricts the goal model to contribution links. Particularly decomposition links are not considered. Again, Giorgini et al.'s algorithm covers decomposition links so that our algorithm will work with decomposition links too [10]. The design decision is, therefore, not a restriction. However, decomposition links only add value in case of an And decomposition since Or decompositions are equivalent to contribution links with a strength of $\omega = 1$. With equivalent we mean that the algorithm treats Or decompositions and strong contribution links in the same way [see 8 for this argument]. From this respect, it does not add much value to the diagram and it should be carefully considered whether the decomposition concept is really needed.

Design decision DD-5 prescribes the use of qualitative strengths for contribution links. Since our algorithm is based on a quantitative calculation, the design decision could be removed without any effect on the algorithm. So, this decision does not restrict the application of the algorithm. However, from the practical view, the qualitative model for contribution link strengths and importance values should be preferred due to the easier elicitation of the contribution strengths and the enhanced readability of the diagram (requirement Req-4).

Assumption A-1) ensures that plans are only related to exactly one goal. This assignment means that the realisation of the plan fully satisfies the goal. There are two different situations, which seem to be impossible to model: 1) a plan might not completely satisfy the goal; 2) more than one plan might be necessary to satisfy the goal. In both cases, the initial goal assignment can be replaced by a goal model: in situation 1) a new goal is introduced and assigned to the plan – this goal contributes to the initial goal (with a contribution link strength "$\omega < ++$"); in situation 2) the initial goal is refined by more than one sub-goal and each sub-goal is assigned to exactly one plan. This assumption is, therefore, not a restriction.

We can conclude that our design decisions and assumptions do not have severe consequences for using Tropos and its formal reasoning techniques. Consequently, Tropos goal models could be used with all its concepts and the algorithms will still produce accurate results.

Furthermore, we decided how to translate qualitative contribution link strengths and goal importance values to quantitative number. The enterprise architect needs to assign appropriate mappings for the qualitative importance and contribution link strengths since there is no general guideline for such an assignment. However, the mapping for the contribution link strengths needs specific consideration since values smaller than one reduces the impact of transi-

tive goals on the overall impact value. Therefore, it seems reasonable to propose a contribution link strength of $\omega = 1$ for strong contribution links so that the impact of goals connected with these links does not degrade transitively.

A property of the proposed algorithm is the distinction between positive and negative impact values for each goal. Although we have only used the overall impact value in Section 3.4, the negative and positive impact values provide additional information for the enterprise architect. Consider for instance two goals with the same overall impact value. The enterprise architect should prefer the goal with the smallest negative impact. In addition, the situation $I^+(g) > 0 \wedge I^-(g) > 0$ indicates a situation of conflicting goals. These conflicting goals should be carefully analysed before making a prioritization decision. Supporting this analysis is subject to future research.

A criticism, which applies here as well, was brought forward by van Lamsweerde [23, p. 390]: He argues that the Sat and Den values and, hence, also the impact values derived from these Sat and Den values, have no clear meaning in terms of the (EA) domain. This problem remains and further research need to show how this problem can be avoided.

5. Conclusions

In this paper we have concentrated on one specific aspect of enterprise architecture: The enterprise architect may be interested in prioritizing strategic IT projects according to the business strategy and business goals. This prioritisation is important to achieve a good alignment between IT and business.

By analysing existing approaches to model business strategies we have found that techniques from the EA domain are not formal enough to apply formal reasoning techniques to the resulting goal models. We have shown that Tropos, a goal modelling technique from the RE discipline, could fill this gap. Its formal reasoning technique allows the enterprise architect to gain a quick overview of the prioritisation of strategic IT projects even if numerous business goals and many relations between them exist.

This prioritisation of strategic IT projects is realised by extending Tropos' algorithm to support the calculation of positive and negative impact values for each goal in the goal model. These impact values encode the relative importance of each goal as well as the different relations between the goals. A high overall impact of a goal signifies a high relevance of this goal to the enterprise architect; and a low overall impact value signifies a low relevance of this goal for the enterprise architect.

Based on the concept of impact value we have identified three rules for the enterprise architect: First, goals with high impact values should be achieved at first while goals with low impact values should be disregarded. Second, goals with high impact values and no associated strategic IT projects should be refined and new strategic IT projects should be assigned to them. Third, strategic IT projects associated with low impact goals should be reconsidered.

However, the outcome of the algorithm comes also with a warning: The algorithm was not yet tested empirically. This means that the algorithm's output should be treated as decision support not as final decision. Thus, the prioritisation of strategic IT projects should be analysed thoroughly before taking any action.

We see further research in three fields: 1) to analyse whether the algorithm can also be applied to actor models to analyse the dependencies between the various stakeholders of an Enterprise Architecture; 2) to extend the approach so that the enterprise architect can interpret the impact values using enterprise architecture phenomena without significantly increasing the effort of eliciting the required information; and 3) to support the enterprise architect when analysing conflicting goals.

6. REFERENCES

[1] Bleistein, S., Cox, K., Verner, J. and Phalp, K. B-SCP: A Requirements Analysis Framework for Validating Strategic Alignment of Organizational IT Based on Strategy, Context, and Process. *Information and Software Technology*, *48*, 9 (2006), 846-868.

[2] Bleistein, S.J., Cox, K. and Verner, J. Validating Strategic Alignment of Organizational IT Requirements Using Goal Modeling and Problem Diagrams. *Journal of Systems and Software*, *79*, 3 (2006), 362-378.

[3] Boer, F.d., Bonsangue, M., Buuren, R.v., Groenewegen, L., Hoppenbrouwers, S., Iacob, M.-E., Jonkers, H., Lankhorst, M., Proper, E., Stam, A., Torre, L.v.d. and Zanten, G.V.v. Concepts for Architectural Description, Centrum voor Wiskunde en Informatica, Radboud Universiteit Nijmegen, Leiden Institute of Advanced Computer Science, Ordina Public B.V., Telematica Instituut, 2007.

[4] Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F. and Mylopoulos, J. Tropos: An Agent-Oriented Software Development Methodology. *Autonomous Agents and Multi-Agent Systems*, *8*, 3 (2004), 203-236.

[5] Castro, J., Kolp, M. and Mylopoulos, J. Towards Requirements-Driven Information Systems Engineering: The Tropos Project. *Information Systems*, *27*, 6 (2002), 365-389.

[6] Dardenne, A., van Lamsweerde, A. and Fickas, S. Goal-Directed Requirements Acquisition. *Science of Computer Programming*, *20*, 1-2 (1993), 3-50.

[7] Engels, G., Hess, A., Humm, B., Juwig, O., Lohmann, M., Richter, J.-P., Voss, M. and Willkomm, J. *Quasar Enterprise - Anwendungslandschaften serviceorientiert gestalten*. dpunkt Verlag, Heidelberg, 2008.

[8] Gehlert, A., Bramsiepe, N. and Pohl, K., Proceedings of the 4th International Workshop on Service-Oriented Computing Consequences for Engineering Requirements (SOCCER 2008), September 8, 2008, Barcelona, Spain. in, (2008).

[9] Gehlert, A., Schermann, M., Pohl, K. and Krcmar, H., Towards a Research Method for Theory-Driven Design Research. in *Proceedings der 9. Internationale Tagung Wirtschaftsinformatik Business Services: Konzepte, Technologien, Anwendungen (WI 2009), 25. – 27. Februar 2009, Wien, Österreich*, (2009), Österreichische Computer Gesellschaft, 441-450.

[10] Giorgini, P., Mylopoulos, J., Nicchiarelli, E. and Sebastiani, R. Formal Reasoning Techniques for Goal Models. in *Journal on Data Semantics*, Springer, Berlin, Heidelberg, 2003, 1-20.

[11] Giorgini, P., Mylopoulos, J., Nicchiarelli, E. and Sebastiani, R., Reasoning with Goal Models. in *Proceedings of the 21st International Conference on Conceptual Modeling (ER 2002), October 7--11, 2002, Tampere, Finland*, (Berlin, Heidelberg, 2002), Springer, 167-181.

[12] Group, T.O. *TOGAF Version 9*. Van Haren Publishing, 2009.

[13] Lankhorst, M. *Enterprise Architecture at Work: Modelling, Communication and Analysis*. Springer, Berlin, 2009.

[14] Letier, E. and van Lamsweerde, A. Reasoning about Partial Goal Satisfaction for Requirements and Design Engineering. *SIGSOFT Software Engineering Notes*, *29*, 6 (2004), 53-62.

[15] Mylopoulos, J., Conceptual Modeling and Telos. in *Conceptual Modeling, Databases and CASE: An Integrated View of Information Systems Development*, (New York, 1992), Wiley, 49-68.

[16] Mylopoulos, J., Chung, L., Liao, S., Wang, H. and Yu, E. Exploring Alternatives During Requirements Analysis. *IEEE Software*, *18*, 1 (2001), 92-96.

[17] Scheer, A.-W. *ARIS -- Modellierungsmethoden, Metamodelle, Anwendungen*. Springer, Berlin, Heidelberg, 1998.

[18] Scheer, A.-W. *ARIS -- Vom Geschäftsprozeß zum Anwendungssystem*. Springer, Berlin, Heidelberg, 1998.

[19] Schermann, M., Gehlert, A., Krcmar, H. and Pohl, K., Justifying Design Decisions with Theory-Based Design Principles. in *Proceedings of the 17th European Conference on Information Systems (ECIS 2009), June 8-10, 2009, Verona, Italy*, (2009), 1065-1076.

[20] Sebastiani, R., Giorgini, P. and Mylopoulos, J., Simple and Minimum-Cost Satisfiability for Goal Models. in *Proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAiSE 2004), June 7-11, 2004, Riga, Latvia*, (2004), Springer, 20-35.

[21] Sowa, J.F. and Zachman, J.A. Extending and Formalizing the Framework for Information Systems Architecture. *IBM Systems Journal*, *31*, 3 (1992), 590-616.

[22] van Lamsweerde, A., Goal-Oriented Requirements Engineering: A Guided Tour. in, (2001), 249-262.

[23] van Lamsweerde, A., Reasoning about Alternative Requirements Options. in *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*, (Berlin, Heidelberg, 2009), Springer-Verlag, 380-397.

[24] van Lamsweerde, A., Requirements Engineering in the Year 00: A Research Perspective. in, (2000), 5-19.

[25] Yu, E. An Organisational Modelling Framework for Multiperspective Information System Design, Department of Computer Science, University of Toronto, Toronto, 1993.

[26] Zachman, J.A. A Framework for Information Systems Architecture. *IBM Systems Journal*, *26*, 3 (1987), 276-292.

A portal-based approach for user-centric legacy application integration in collaborative environments

Oliver Gmelch
Dept. of Information Systems
University of Regensburg
Germany
oliver.gmelch@wiwi.uni-regensburg.de

Günther Pernul
Dept. of Information Systems
University of Regensburg
Germany
guenther.pernul@wiwi.uni-regensburg.de

ABSTRACT

"Networked enterprises" are characterized by distributed teams of partner organizations, humans, computer applications, autonomous robots, and devices collaborating with each other in order to achieve higher productivity and to collaborate in joint projects or produce joint products that would have been impossible to develop without the contributions of multiple collaborators. In networked enterprises, special consideration must be paid to the IT systems which are in the position to integrate different applications across company boundaries as known from enterprise application integration. At the same time, high requirements are imposed on the employees within such alliances. The contribution of this paper is an architecture for legacy application integration in web-based portal systems, specifically tailored to the requirements of networked enterprises and focusing on a user-centric approach, allowing the user to customize his workspace to his own needs. Following its presentation, the proposed architecture is validated by a prototypical implementation.

1. INTRODUCTION

Globalization, shorter innovation cycles or increased competition are just a few of the challenges today's organizations are confronted with. The aggregation into virtual partnerships in order to create flexible and agile business networks between a number of partner organizations having complementary competencies appears as possible remedy to overcome this situation. As diagnosed by recent surveys performed for instance by AT&T [4] or the Gartner group [25], a significant increase in the number of business alliances can be expected in the near future.

Of major importance in virtual enterprises is the support by dedicated information and communication technology (ICT). Identified as a crucial characeristic of virtual partnerships by several authors (for instance, Katzy [21]), ICT systems are expected to provide support for different success factors such as integration on process level across company boundaries or information integration, implying provisioning the right information to the responsible user at the right time. As proposed by Shilakes et al. [28], one way of achieving this information integration is via the introduction of specifically-tailored enterprise portals, supporting the integration of different individual applications (even from different alliance partners or from dedicated application providers) in one common user interface consisting of a number of specifically-crafted portlets. It is the intention of these specific portlets to bridge the gap between application boundaries, which create the relations between individual applications, respective process steps and the users in a statical assignment.

Networked enterprises impose high demands on the parties involved in such collaborations, especially riquiring highly skilled personnel with special focus on their flexibility. This leads to the situation that companies are confronted with a "war for talent" as introduced by McKinsey employees Chambers et al. in 1998 [9]. The "war for talent" describes three qualitative challenges companies are confronted with, beginning with the statement that "[a] complex economy demands more sophisticated talent with global acumen, multicultural fluency, technological literacy, [and] entrepreneurial skills". Furthermore, according to Chambers et al., the emergence of small- to medium-sized companies has lead to a shortage of skilled personnel, as well as an increase in job mobility and thus an increase in job fluctuation.

As a consequence of the war for talent, we consider the employee in form of an IT user to be the origin of all value creation within an enterprise. Hence, this paper proposes a software architecture for a user-centric workplace specifically in the area of application integration with web-based portal systems. This architecture poses the chance to flexibly integrate distinct applications from both in- and outside a company's boundaries into a common web-based user interface and thus overcome the formerly static assignments between applications, tasks, and users.

The introduction of a user-centric workplace also bears the advantage that the user can adjust his workspace to his special needs himself. This is of special concern when talking about different application portfolios in use by different companies, which can make a difference in employees' productivity and the amount of training necessary to bring recently

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

employed people up to par. Likewise, the creation of a user-centric working environment can make a significant point to distinguish an employer from the competition merely by its IT systems and the resulting attitude towards their employees. Furthermore, flexibility with regard to the application portfolio in use can help to avoid the phenomenon of vendor lock-in, which is characterized by strong dependency of a company to a software vendor. Not limited to applications within a specific company, the proposed approach can also be applied across company boundaries as long as the goals for all individual tasks dealt with in a networked enterprise are well defined, especially in terms of data formats expected. This bears the advantage that individual competencies can be brought into networked enterprises even better.

One way of bringing this user-centric attitude to life in an enterprise environment could be via the introduction of an application store, where a user could select from a set of preconfigured applications per category. Even though this may not be appropriate nor desirable for all types of applications, it may make a big difference when used carefully in situations where interoperability across various application boundaries can be assured due to reliable import/-export filters or standards-compliant software components. For these settings, the user's preferences in the application store may be remembered and taken into account for his everyday working environment.

This paper is intended to present a software architecture for a user-centric application integration approach alongside a brief overview of a prototypical implementation as carried out within the SPIKE project[1], which is funded under the FP7 programme of the European Union and targets the creation of a secure collaboration platform for process integration of external application systems, specifically tailored for the usage in networked enterprise settings by small- to medium-sized enterprises. As such, it follows two main organizational objectives: (1) Outsourcing parts of the value chain to business partners (and vice versa, offering such parts in form of services) and (2) enabling collaboration between members of participating organizations through ad-hoc created as well as predefined business processes.

The remainder of this paper is structured as follows: following this introduction, section 2 gives an overview on related work. Section 3 then provides an introduction on different means of application integration in order to distinguish this approach from other pre-existing work. Section 4 describes the user-centric approach by providing an architecture for user-centric application integration in portal systems, which is then further explained regarding its prototypical implementation as part of the SPIKE project in section 5; section 6 finally concludes this paper and gives an outlook on future work.

2. RELATED WORK

Collaborative software as introduced in section 1 in most cases is built around a portal system. A rather technical definition of the term portal is given in the Java Portlet Specification JSR 168 [1]. According to this specification, "a portal is a web-based application that – commonly – provides personalization, single-sign-on, content aggregation from different sources and hosts on the presentation layer of information systems. Aggregation is the action of integrating content from different sources within a webpage. A portal may have sophisticated personalization features to provide customized content to users. Portal pages may have different sets of portlets creating content for different users."

From an application developer's perspective, portals consist of several independent web applications, called portlets, which are combined together into one uniform user interface, running under a Java application server. The Java portlet specification JSR 168 defines a standard for individual portlets, thus enabling platform independence of portlets, aiding usage across different application servers and thereby guaranteeing a high degree of interoperability. The nature of portals consisting of independent web applications has paved the way for integrating external applications into web-based portal systems. Its successor, JSR 286 [19], has extended the portlet standard amongst others with the ability to establish communication between individual portlets, which had been requested many times before.

Related to the portlet standards is the concept of web services for remote portlets (WSRP), which has been established by OASIS and initially released in 2003, with a second revision in 2008. WSRP defines ways for portlets to integrate external applications adhering to this standard into a compliant portal [30]. The WSRP standard follows the paradigm of service orientation, defining a SOAP-based interface for data exchange between an external application and the portal system. By returning the intended user interface in HTML format to the portal server where the request originated, a WSRP-capable application can itself define the expected user interface without the need for major transformations and hence be included easily into IT systems of external parties associated within a networked enterprise. Furthermore, the external application can employ the portal's event distribution mechanism to exchange information with other applications in the portal.

Of special interest in this area of portal-based application integration is the PADEM portal reference architecture as originally published by Gurzki and Hinderer [18]. Likewise, issues related to application integration via portals have been identified by Diaz et al. [13, 12], fostering the idea of a switch from traditional content syndication towards portal syndication and proposing an integrated framework for transformation of existing web applications into portal-aware applications. In their proposal, they introduce a model-driven approach to achieving this switch towards portal-aware applications. Bellas et al. [6, 5] have proposed approaches to displaying web applications as portlets, introducing a chain of user-configurable "transformers" in order to perform customized adaption regarding individual applications. They propose an annotation-based approach to allow for automatic transformations. In the same subject, Paz et al. [24] have extended the idea of portlet integration by the introduction of semantic integration based on annotations. In order to achieve this, an approach using annotator portlets, wrapping remote portlets, has been presented by the authors. With regard to security aspects introduced

[1]http://www.spike-project.eu

by inter-portlet communication, recent work has been published on integration of the XACML architecture into portal systems [17].

The idea of user-centric approaches has been examined in a variety of ways including but not limited to software engineering, grid computing or identity management, of which the subject of user-centric privacy considerations is one of the most noteworthy in recent years. User-centric privacy follows the idea to gain the users' trust by enabling the user to control which personal information to pass to which party. Even though the idea of user-centric approaches has been examined in a variety of subjects, little research has been carried out so far on the area of user-centric approaches in the area of enterprise application integration mechanisms, where a broad range of user capabilities can be employed by empowering the user to tailor his workspace to his special needs so that he feels most comfortable and hence productive with.

3. MEANS OF APPLICATION INTEGRATION

Ever since the introduction of data processing in the 1960s, an integrated view on the processed data has been requested. In the decades since, the number of applications holding and processing their single share of the data has increased massively, and so has the need for integration between these individual application silos. The introduction and widespread adoption of the personal computer in the 1980s and 1990s has put additional pressure on an integrated and consistent view on data and the applications processing this data, respectively. Due to this situation, a magnitude of different approaches to cope with this challenge has arisen.

On the level of enterprise IT architecture design, Winter [32] has proposed an application architecture with recommendations for the design and implementation of applications (i.e. by bundling of functionality or by introduction of responsibilities and data usage), ultimately resulting in a reduction of complexity and the introduction of well-defined interfaces and thus an improved ratio of costs for application systems and the respective interfaces. Also on the level of application infrastructure, a pattern-based integration approach has been initiated [2], which is, however, limited to IBM products and hence only applicable to limited extent.

Also on the level of IT architectures, the idea of enterprise application integration (EAI) has been examined for a long time since the 1990s, resulting in a plethora of different definitions and differentiations between various types of integration identified [22]. In a nutshell, the idea of EAI is to introduce a central middleware component, governing all communication between individual applications and hence further reducing the number of interfaces required for all applications to communicate with each other properly. Linthicum [23] and Ruh [27] argue that EAI should allow for unrestricted usage of information and application functionality between all applications within an enterprise. Linthicum, for instance, has identified the following types of EAI approaches [23]:

- *Data-Level EAI*, resulting in integration on the level of different databases,
- *Application-Level EAI*, containing integration on the level of individual application programming interfaces,
- *Method-Level EAI*, meaning the sharing of business logic or methods, and
- *User Interface-Level EAI*, resulting in the integration of various applications on the level of their corresponding user interfaces.

As can be seen, the level of abstraction varies greatly between individual approaches in the context of EAI; a majority of methods, however, is focused on the technical side of EAI. Another method to enable enterprise integration on the technical level is the pattern-based approach as presented by Hohpe and Woolf [20], which provides a system of integration relations in the context of EAI. Including different design patterns and the respective areas of applicability, their approach shows potential solutions to individual challenges on a rather technical yet not product-oriented level. Dealing with the subject on a technical level and limited to individual patterns, however, it remains rather vague when it comes to an overall architecture, supporting networked enterprises and the introduction of a user-centric approach.

Also coming from a technical perspective with special focus on the area of portal systems, Daniel et al. have performed extensive work on the aspects of application integration [10, 11], resulting in the discovery and definition of problems associated with certain technologies and opportunities to overcome their respective weaknesses. Of major interest for Daniel et al. were desktop UI components, components realized via browser plug-ins, web mashups as well as web portals and portlets. One of the major findings in their research is the awareness about the diversity of different UI presentation techniques. For this reason, the authors suggest a standardization of service interfaces for user interface integration to succeed. Furthermore, the authors mention a lack of abstraction in order to establish composition in the context of user interface integration. Even though their works have shown the need for unification on the level of user interface representation, it does not provide new strong points on an architectural integration.

As stated by Lee et al., "enterprise integration means both technical and behavioral integration" [22]. In the field of behavioral integration, we consider the user as one crucial part. None of these approaches, however, have evaluated the user as the center of their respective application integration approaches. With their focus on the technical perspective of application integration, all these concepts fall short in considering the user-implied aspects of EAI, which may enable the user to customize his workplace up to his own needs. For this reason, this paper presents a user-centric approach towards portal-based application integration on user interface level, which is demonstrated in a software architecture overview in the next section. As such, the presented software architecture can be seen in line with the ideas as introduced by EAI, most notably the introduction of a middleware component for inclusion of and brokerage between external applications. Even though the concepts and techniques as followed by EAI are considered by some as a hyped topic which has exceeded its climax [3], the results and general require-

ments presented in this paper can also be adopted to other integrative techniques such as service-oriented architectures.

4. AN ARCHITECTURE FOR APPLICATION INTEGRATION IN PORTAL SYSTEMS

Based on the requirements of networked enterprises as evaluated by the SPIKE project in [7], this paper presents the approach of user-centric application integration. This section focuses on the description of the related architecture. The presentation is split up into two parts: section 4.1 gives an overview on the overall architecture and its elements, whereas section 4.2 presents the followed approach with regard to individual protocol types in more detail. The proposed architecture and its support for a broad range of protocols is considered fundamental for the introduction of a user-centric integration approach in collaborative environments, which allows users to customize application usage to their own needs. Due to the sheer amount of different applications and involved protocols available in the market, the proposed architecture offers support for a multitude of different protocols typically found in enterprise use as outlined in the next sections without, however, making any application-specific adjustments with regard to integration. The presentation of the architecture is implementation-agnostic; for implementation details, please see section 5.

4.1 Overall architecture

As the architecture is intended as collaborative platform for networked enterprises, it is embedded into a surrounding system architecture providing supporting components for networked enterprises, especially in the field of process orientation and user management (which are not presented in detail within this paper due to space limitations). In the following paragraph, however, a brief overview is given on the respective components and their interaction with outside components of the architecture. This architecture can serve as a blueprint for application integration in a collaboration platform such as the SPIKE platform as presented in [15].

The core of the portal architecture as depicted in figure 1 is represented by three components, namely *Intra Portlet Communication Management*, *Session Management*, and *Portal Display Management*. *Inter Portlet Communication Management* is responsible for collecting events from available portlet sources and delivering them to other portlet destinations, including Notification Management for external delivery in other user sessions. Within the portal session, this is performed using the established method of Inter Portlet Communication. *Session Management*, on the other hand, is in charge of user session context handling as well as storage and retrieval of user sessions. To do so, it makes use of the central storage repository as maintained by Content Management. The third component in the portal architecture in figure 1 is *Portal Display Management*. It is the task of this component to provide users with a visual representation of applications, thus carrying out the actual work of user interface integration of external applications. All three components inside the Portal Instance are in contact with *Interface Management*, which forms the interface to the external applications as connected via the integrative platform.

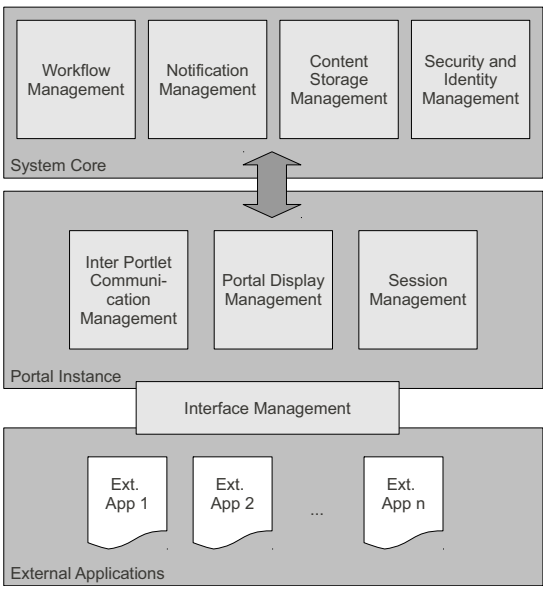


Figure 1: Portal Architecture

Surrounding this portal architecture, a number of other components can be found, which are aggregated in the System Core. First of all, the aforementioned *Content Management* provides a central storage repository, allowing for retrieval, update, and storage of all data present and brokered via the platform.

It is the duty of the *Security and Identity Management* component to provide access decisions based on the user's identity, containing information about the user's home company as well as personal information (i.e., department, email address, etc.) about a specific user. This information is retrieved from the external identity management system associated with a specific user, i.e. provided by his employer and selected and queried during login. Likewise, it has to be assured that only trustworthy users may enter the platform who have been granted the necessary privileges by their employer beforehand. Even more important, it also has to be guaranteed that only associated members of a networked enterprise may enter the system and perform actions with it as this platform is working on potentially highly sensitive information.

Notification Management, as briefly introduced before, is necessary for sending event notifications to the rest of the architecture and to receive events from other parts of the architecture, which are then further processed within the portal. These events can consist of various types, ranging from inter portlet communication between different application instances in the portal to exchange of status messages, containing i.e. information about a new user session or a logout event.

Finally, *Workflow Management* keeps track of all workflows, their respective instances and all associated tasks deployed within an alliance and hence has to maintain communication with the individual user sessions, represented via Session

Management.

Likewise, the correlation between Workflow Management and Session Management allows for inclusion of a user-centric application integration approach, where a user for certain tasks can customize the set of applications being used for task execution up to his own preferences. This can be performed following the widespread idea of an application store, where users can select from a range of applications, differentiated by categories. This way, the user can stick to the type of application he is used to and feels most comfortable with, thus reducing the need for his employer to provide further training in one specific application. On the level of process definition, this requires a concise definition of the goals to achieve for the individual tasks associated in a workflow as well as definition of the applications or data formats that these results can be produced with.

Even though this approach may not be desirable or feasible in all circumstances, enterprises can gain a broader range of options to provide their users with. Likewise, it can be expected that training costs can be reduced when users have to make little or no changes at all to their accustomed working environment whilst ensuring user satisfaction and creating a feeling of affiliation of the employees with their respective employer. This in turn can help to keep personnel fluctuation low and thus ensure customer satisfaction, which successively can help to strengthen the company's position in the market.

4.2 Detailed UI architecture

The major component for UI integration of external applications – as was already described in the previous section – is the Portal Display Management component, which is intended to provide support for a broad range of different application types. As different applications usually are provided by different protocols, the Portal Display Management component is designed to offer a flexible plugin mechanism to allow for integration of different protocol handlers. For an overview on the different types of applications currently supported by the architecture, see table 1. Differentiated by their nature as request- or communication-oriented protocols, the support ranges from HTTP and the SOAP-based WSRP protocol for the case of requested-oriented protocols to support for connection-oriented protocols, whereas further differentiation between text-based (currently, Telnet and SSH are considered here) and GUI-based protocols (where the current state considers RDP and VNC protocols) is made. At this point, it is important to note that this listing is neither complete nor exhaustive, meaning that protocol support is not limited to these protocols on an architectural level, but these protocols have been introduced by the SPIKE project based on the user requirements analysis performed by the project. Moreover, the architecture strives to provide a general, application-independent view at the supported protocols. For the case of HTML content presented later on, due to the large amount of different usage scenarios, three approaches were introduced to support these different scenarios appropriately.

The architecture regarding those individual protocol handlers was designed to follow the paradigm of service-oriented architectures. The services are generally stateless and atomic

Table 1: Currently Considered Protocol Types

	Connection-Oriented Protocols		Request-Oriented Protocols
	Text-based	GUI-based	
Protocols	SSH, Telnet	RDP, VNC	HTTP, WSRP

and grouped by functionality. Common to all protocol handlers is the idea of a proxy mechanism, which allows the protocol handlers to act as an intermediary between the target application and the targeted end user. This proxy mechanism can be implemented in two distinct manners: One approach is that the portal merely is responsible for application connection intialization and hands over a session token representing this connection to the service consumer, which then performs all interaction with the application independently from the portal instance.

Secondly, the portal can be acting as a proxy for the whole user session, meaning that all communication between service provider and -consumer will be routed through the portal. This bears a number of advantages: First of all, firewall policies will have to be adapted by the service provider only once. As all requests from the service provider's point of view will stem from the portal system, access on IP level will have to be granted to a target application only once for an unlimited number of users for this application type. Secondly, acting as a proxy bears the capability to perform transformations on the presentation layer, for instance on CSS/JavaScript level in case of HTML content. Moreover, based on the type of application, the proxy approach can allow for flexible and dynamic extraction of content elements adhering to previously defined rules as implemented during workflow specification phase. This can for instance be performed for the case of Telnet- or HTML-based user interfaces. Furthermore, as the platform provider is aware about all parties associated with its platform, it can make more educated access decisions about whether or not certain kinds of access attempts adhere to the access policy for the specific user or the user's employee.

The major drawback of this approach, however, is that the platform provider acting as the intermediary between service requester and service provider can potentially carry out man-in-the-middle attacks, thus putting all contents exchanged via said platform at risk. This risk therefore requires the platform provider to be a trustworthy third party, relied upon by all participants and entrenched respectively, i.e. by legal means. Furthermore, the aspect of runtime performance must be paid attention to. For the case of full-featured application proxying, a significant increase in the workload of the IT infrastructure is to expected, potentially leading to an increase in costs for the platform provisioning. Likewise, since all accesses are carried out via the portal platform, this imposes higher requirements on the availability of said platform.

As can be seen from figure 2, both types of communication are supported by the architecture. In the case of connection-

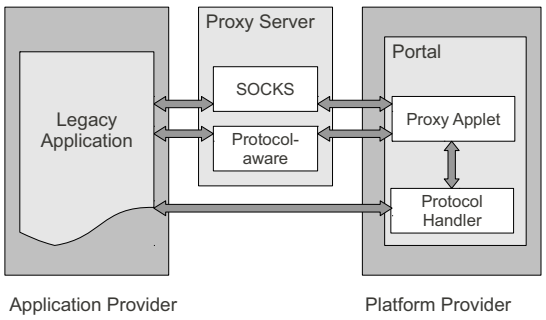


Figure 2: Asynchronous Applications: Architecture Overview

oriented protocols, communication is performed via a tailored applet running inside the portlet of this specific application instance, which can either communicate directly with the connected external application or create the connection via a dedicated proxy server. The latter case can – for future developments – be used to allow for session reliability as the proxy server could be used to maintain connectivity to the respective application even in the case that the browser session crashes or the applet is reloaded unintentionally by the user's browser, allowing the user to seamlessly continue his session at the point it was interrupted. With protocol-agnostic support being propagated by a SOCKS-compliant proxy module, further functionality can be gained by protocol-aware proxy modules. Some details about possible enhancements by such protocol-aware proxy modules are given in section 4.2.2. Also, figure 2 shows that the situation for request-oriented protocols is similar to the approach taken for connection-oriented protocols except for one difference: as markup-based content currently is the only content type supported here, the dedicated applet can be omitted since all web browsers can render markup natively.

The design of the architecture follows a number of software design patterns as originally presented by [16]. First and foremost, the model-2 pattern has widely influenced the design as it presents itself as the foundation for merely all web-based UI interaction, especially in the context of Java applications as is the case for SPIKE. Adopted from the well-known MVC pattern which is considered to be a set of strongly related patterns itself [14], model-2 is specifically tailored for web applications. The MVC pattern separates an application into three distinct parts: (1) the *controller*, connecting the latter two parts of the application with each other, (2) the *model*, responsible for the actual application logic, and (3) the *view*, which is in charge of rendering a visual representation of the application, the so-called user interface which can therefore be considered a visual representation of the model.

Moreover, the *facade pattern* has found major consideration in the design as it enables the support for a pluggable modules system as the architecture defines an interface on an abstract level, the facade, which is then implemented by every protocol plugin. All protocol-specific functionality is hidden by this service facade afterwards, thus making all accesses transparent to the portal system as the portal system merely hands over control to the corresponding plugin, which helps to keep the portal code clean and flexible.

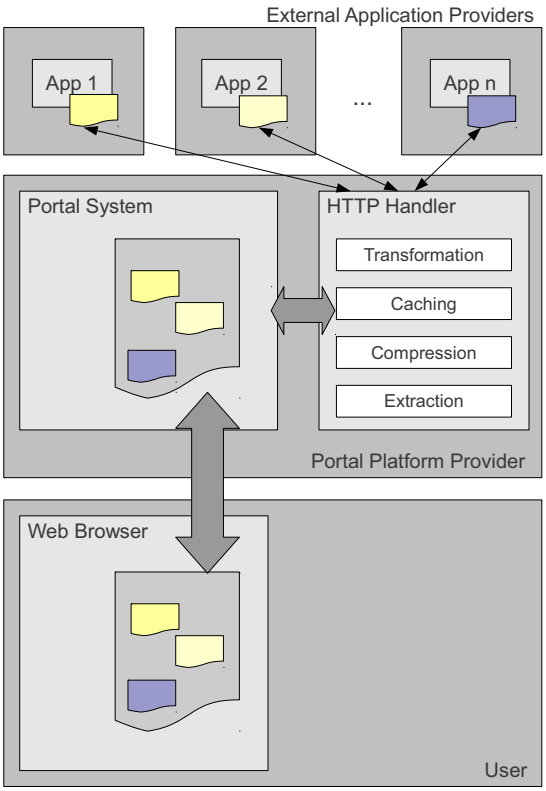


Figure 3: HTTP Handler: Architecture

4.2.1 Support for request-oriented protocols

Based on user requirements analysis, the highest ranked type of protocol to support in the presented architecture is the hypertext transfer protocol (HTTP), which forms the basis for the world wide web in its current form. HTTP Communication generally follows the request-/response paradigm and is not connection-oriented and stateless. Another important protocol taken into account by the software architecture is the WSRP protocol, enabling the integration of markup-based remote applications into a portal system. Due to the flexible nature of the architecture, however, support could be extended for other protocol types appropriately.

HTTP support in the presented software architecture focuses on integration of content provided in the hypertext markup language (HTML) and considers three distinct ways of integration in order to establish flexible yet efficient ways for application integration which are briefly outlined below. In this enumeration, the level of complexity and hence the possibilities of the respective methods increase significantly from top to bottom. Due to the wide nature of predominant HTML-based applications, the best-fitting integration mechanism must be determined on a case-by-case basis.

First of all, integration of HTML content can be carried out in a straight-forward manner using the HTML tag <*iframe*>,

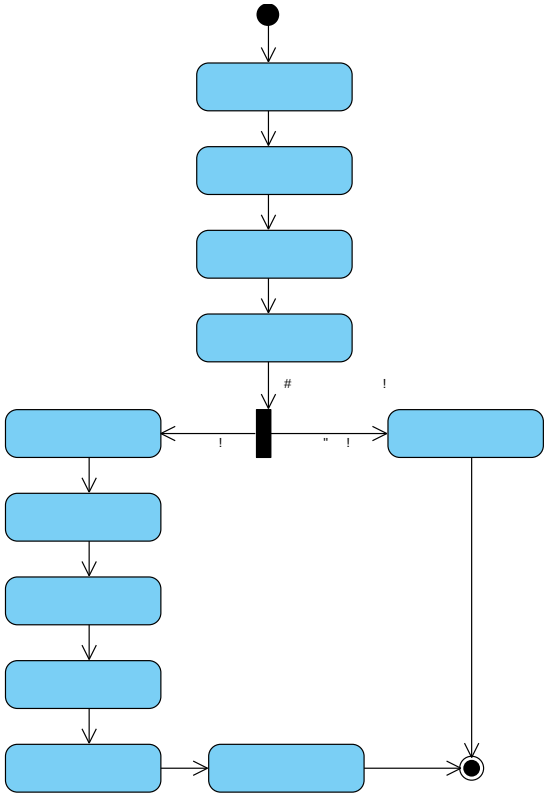


Figure 4: HTTP Handler: State Diagram

which is the most non-intrusive way of integrating an application as it merely passes all content through to the web browser, which renders the application as part of the portal website. Since HTML frames exhibit some limitations in terms of user experience (for instance, the risk of losing the portal session due to external links), this is also the least powerful integration mechanism and should be considered ultima ratio.

Secondly, integration can be carried out by a dedicated applet, which renders content described by markup in a separate applet independent from the end user's web browser and thus can achieve a significantly higher level of unity in application appearance and platform independence. This, however, comes at the price of increased system requirements for the associated client systems due to the separate rendering process and should preferably be considered for applications under direct influence of the application provider.

The third option is to perform integration via a dedicated HTTP handler acting as a reverse proxy and as such performing content rewriting, which allows for direct integration of HTML contents into the portal whilst posing higher browser compatibility requirements on the target application to serve all end users reliably. Due to the heterogeneous nature of HTML documents, this approach presents itself the most complex. Figure 3 shows the underlying architecture: Content from external applications is integrated into the platform using the HTTP handler, altering the content in such way that it properly fits into the portal, where all contents from individual applications are merged and presented to the user in a unified user interface. The HTTP handler acting as a reverse proxy, all content is modified as part of the transformation phase so that all requests in return to content delivered by the HTTP handler module will again invoke the HTTP handler, avoiding unintended side channel data flows. The execution of content rewriting also aids to avoid situations where users of the portal could be leaving the platform unintentionally due to popup windows or other external links. Moreover, the HTTP handler is to provide caching mechanisms in order to speed up connections by storing temporary copies of frequently accessed files as well as image compression capabilities in order to reduce the footprint of images and other large resources, which may be desirable for mobile devices with low-bandwidth network connectivity. Finally, the extraction stage serves to extract information retrieved from an external application based on predefined goals. Figure 4 shows a state diagram of HTTP handler, giving an overview on its individual states when processing a HTTP request. Beginning with a client request, the resource is first looked up in the cache, maintained by the HTTP handler for performance reasons. In case a resource is found, it is directly sent to the client; otherwise, the external resource is retrieved and applied a number of transformations in order to perform the tasks of information extraction and compression as outlined above and finally stored in the cache and sent to the client.

Another type of request-based application integration is gained by supporting the service-oriented paradigm, and more specifically the WSRP standard, allowing portals to include external applications into portal systems both on the data level as well as on the level of the user interface representation. This is achieved by providing WSRP consumer support in the architecture, which allows external application providers to integrate with the proposed solution without any changes to their existing WSRP-aware application.

Figure 5 shows the general flow of information for the case of WSRP application integration. An external application provider makes available an arbitrary number of WSRP-capable applications, each running in a portlet container. The WSRP producer is responsible for embedding the individual markup generated by every application into corresponding SOAP messages, which is then extracted by the respective consumer on the portal side, where the overall portal is rendered, consisting of the combined markup from the external applications plus any portal-specific markup, i.e. for user management, status message display, etc.

4.2.2 Support for connection-oriented protocols

As is the case for request-oriented protocols, the support for connection-oriented protocols in the presented architecture is currently limited to four protocols to demonstrate the feasibility of the approach. These protocols can be grouped into protocols supporting graphical user interfaces (GUI) such as RDP or VNC and text-oriented protocols such as SSH or Telnet. However, the extensibility gained by grouping functionality into separate plugins allows to fulfill later requirements in case of need.

Generally, in the application integration architecture, two

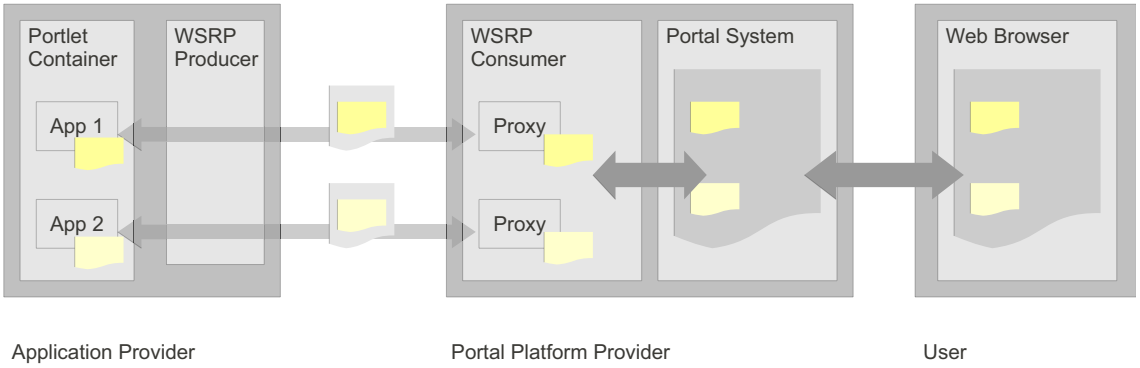


Figure 5: WSRP Support

modes of operation are supported as previously shown in figure 2: with and without intermediary proxy server, whereas operation with a proxy server between application and user can again be differentiated into usage of a SOCKS-compliant proxy server and an application- or protocol-specific proxy server, which can be thought to provide protocol-specific functionality for later requirements. Whilst additional functionality such as information extraction from the user interface can be implemented easily for the cluster of text-based protocols, the situation becomes more difficult when it comes to graphical user interfaces as is the case for instance with RDP and VNC. Depending on the protocol specification, however, additional features of the protocol can be employed. For instance, the RDP protocol foresees capabilities to not only transfer information about the GUI but also offers features such as file system redirection, enabling the integration platform to provide a file system which the external application can store task results on. The level of integration which can be achieved by the architecture is therefore limited by the specifics of the protocol in use.

5. IMPLEMENTATION AND EVALUATION

The software architecture as presented in the previous section has been implemented in order to prove its feasibility. The implementation has demonstrated the general applicability of the proposed solution in the environment of user-centric legacy application integration. While a number of issues has shown up during the implementation, these issues are not related to the overall architecture and hence do not diminish its value.

One major pillar during the implementation was the route of embracing and extending existing open source components to achieve the goals. The rationale behind using the source code of publicly available open source projects for the implementation is manifold: First of all, open source projects have – at least for major players on the market – reached a satisfactory degree of maturity. For instance, analysis of different operating system kernels from both open and closed source products has shown that despite greatly verying processes by which the respective products were developed, the metrics applied to the corresponding source code showed a comparable performance [29]. Likewise, commercial support is available for most major open source projects, meaning that no special knowledge has to be created inside an enterprise prior to using and potentially extending a specific open source project. Beyond that, a number of soft facts have been identified, ranging from customer relationship enhancement to earlier feedback from the community and voluntary actions by the community like localization activities [8]. Finally, available code provided by various open source projects presents a huge wealth of knowledge which has been tested in numerous installations around the world and which has been of great use for the prototypical implementation. For instance, the Liferay portal server, which forms the basis for all portal-related software development within the implementation, has been evaluated to consist of more than 1.7 million lines of code, which has been estimated to result in 502 person years on effort[2]. With the open source movement existing for a few years only, little expertise is available on the long-term performance and development of the open source paradigm. However, recent research seems to indicate that a commercial open source business model, where commercial open source firms build their business around an open source project that they fully control, can prove successful [26]. Even though this approach seems to constrain customers to the vendor of a specific product, there is still the availability of the product's source code that can allow them to continue usage of that product even after the case that the original vendor should no longer exist.

Although implementation of the proposed software architecture has been performed employing open source components, this does not limit the applicability of the underlying software architecture to open source components. Instead, the general architecture is implementation-agnostic, whereas the implementation only serves to prove its validity and has been carried out employing as many existing components as possible for complexity reasons.

For the reasons given above, for all portal developments during the prototypical implementation, the Liferay portal server[3] in its community edition has been chosen to form the basis. Liferay Portal is freely available under an open source license and has been selected mainly for its widespread acceptance, its user community and its extensibility. It is especially the extensibility provided by a so-called extension environment which allowed to create extensions to the por-

[2] `http://www.ohloh.net/p/liferay`
[3] `http://www.liferay.org`

tal's core while not touching the portal server's source code directly and which has therefore proven very useful for future version updates of the portal server underneath.

During the implementation, special emphasis was put on support for existing HTML-based web applications. For the integration of web applications, the three approaches as laid out in section 4.2 were implemented, of which the reverse proxy method has received the most attention due to its complexity and its possibilities. Evaluation has shown afterwards that while it works well for basic standards-compliant websites, the situation changes when non-conforming HTML documents come into play. As the reformatting and information extraction methods require the HTML code to be transformed into XHTML, standards conformance plays a crucial role. Despite not limiting the proposed architecture's value as a prototype, work is ongoing with regard to a more robust parsing of HTML pages.

During the implementation of WSRP support, the separation of duties in producer and consumer has shown to provide valuable support for application developers. Implementation itself has been carried out using the WSRP extension for Liferay portal and has shown that existing portlets can be easily provided as external applications, thus enabling the introduction of external application providers. This also undermindes the idea of networked enterprises where all partners can focus on their key competencies and hence improve the overall competitiveness and outcome of such alliances.

Besides the aforementioned adversities, the implementation has proven the proposed architecture's general applicability in the context of user interface integration, especially following the presented user-centric approach, which allows a user to employ functionalities commonly known from the nowadays prevalent application stores, where the user can choose from a set of available applications to adopt his workspace to his special needs.

6. CONCLUSION AND FUTURE WORK

This paper has presented reasons to perform application integration on a user interface level and has demonstrated an architecture for the inclusion of applications into portal systems. The architecture is designed to allow for a high level of flexibility in order to support future enhancements and to achieve a high degree of maintainability. At the same time, this paper has briefly introduced the prototypical implementation of this architecture, which has demonstrated the feasibility of the approach in an application prototype.

The combination of this architecture with the special requirements of networked enterprises can pose a significant advantage over traditional companies. Networked enterprises generally are characterized by a high degree of flexibility, and so are the requirements on their employees. The introduction of the user-centric attitude into the context of application integration can provide a way for companies to get advantage over their competition in the war for talent by providing their users with the ability to specifically tailor their work environment to their own needs whilst also providing ways to ensure or improve their competitiveness.

The implementation has proved the general applicability of the proposed architecture for the problem setting of interoperability in networked enterprises. At the same time, a number of weak points could be identified by the implementation, first and foremost the lack of standards compliance of many existing web applications. Even though a significant increase in the percentage of standards compliant websites has been detected recently [31], their share still remains at mere 4.1%. Therefore, implementations will have to be made more robust to also support non-standards compliant websites as well. Adding to that the sheer number of different techniques for web content presentation currently prevalent on the market besides HTML-based content like Adobe Flash, Microsoft Silverlight, or Oracle JavaFX-based content, the situation gets even more difficult. This also supports the point of Daniel et al. [11] who pledge for standardization on the UI level. A first step towards a more homogeneous presentation layer may occur due to an increasing acceptance of the HTML standard in its fifth release, which is expected to introduce – amongst others – capabilities for video playback and drag-and-drop facilities but has not been officially released by W3C and WHATWG yet. Also, the HTML 5 standard is likely to include recommendations regarding a user agent's behaviour in case of erroneous documents, which may help to increase the number of standards compliant websites. At the same time, the advent of HTML 5 will further ease the introduction of specific so-called widgets, small applications targeting at a limited set of functionality. In combination with the presented user-centric approach, this is expected to provide valuable extensions whilst reducing complexity by breaking up formerly strict application boundaries at the same time.

Besides improving their implementation to cope with the identified lack of standards compliance, future work of the authors is going to focus primarily on carrying out detailed analysis of the proposed architecture. The architecture will be evaluated based on the prototypical implementation alongside a number of axes. These axes will first of all comprise reliability aspects of individual types of applications in order to assess the degree of maturity of the proposed solution. Furthermore, implementation evaluation will also include detailed analysis of the runtime performance of the proposed solution as well as assessment of further protocol candidates for future implementation. Another field for future work is measuring the appropriateness of the proposed architecture in terms of the currently prevalent topic of cloud computing. With this architecture and networked enterprises as potential users of said applications, company boundaries are converging. As a result, provided that legal issues are considered, the location of an application provider no longer is of importance and hence opens the door for the proposed architecture which allows to include distinct applications into a uniform user interface for the networked enterprise. Consequently, aspects like availability, access times and provisioning times for individual applications become more and more important. Also, this paper primarily focuses on the technical implications, whilst business dimension considerations are subject to future research.

Acknowledgements

The research leading to these results is receiving funding from the European Community's Seventh Framework Programme under grant agreement no. 217098 and from the

European Regional Development Funds (ERDF). The content of this publication is the sole responsibility of the authors and in no way represents the view of the European Commission or its services.

7. REFERENCES

[1] A. Abramski and T. Schaeck. *Java Portlet Specification Version 1.0*. Java Community Process, 2004.

[2] J. Adams and S. Koushik. *Patterns for E-Business – A Strategy for Reuse*. IBM Press, 2001.

[3] S. Aier and J. Schelp. EAI und SOA–Was bleibt nach dem Hype? In *Proceedings der Multikonferenz Wirtschaftsinformatik*, pages 1469–1480, 2008.

[4] AT&T Corp. Collaboration across borders. `http://www.corp.att.com/emea/docs/s5_collaboration_eng.pdf`, 2008.

[5] F. Bellas, I. Paz, A. Pan, and O. Díaz. *Handbook of Research on Web Information Systems Quality*, chapter New Approaches to Portletization of Web Applications, pages 270–285. Idea Group Inc, 2008.

[6] F. Bellas, I. Paz, A. Pan, Óscar Díaz, V. Carneiro, and F. Cacheda. An automatic approach to displaying web applications as portlets. In *Distributed Computing and Internet Technology*, volume 4317/2006, pages 264–277. Springer Berlin / Heidelberg, 2006.

[7] C. Broser, C. Fritsch, O. Gmelch, G. Pernul, R. Schillinger, and S. Wiesbeck. Analysing Requirements for Virtual Business Alliances – The Case of SPIKE. In *Digital Business*, volume 21 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 35–44. Springer Berlin Heidelberg, 2010.

[8] E. Capra, C. Francalanci, and F. Merlo. An empirical study on the relationship between software design quality, development effort and governance in open source projects. *IEEE Transactions on Software Engineering*, 34:765–782, 2008.

[9] E. Chambers, M. Foulon, H. Handfield-Jones, S. Hankin, and E. Michaels III. The war for talent. *The McKinsey Quarterly*, 1(3), 1998.

[10] E. Daniel and J. Ward. Enterprise portals: Adressing the organisational and individual perspectives of information systems. In *Proceedings of the Thirteenth European Conference on Information Systems*, 2005.

[11] F. Daniel, J. Yu, B. Benatallah, F. Casati, M. Matera, and R. Saint-Paul. Understanding UI Integration: A Survey of Problems, Technologies, and Opportunities. *IEEE Internet Computing*, 11:59–66, 2007.

[12] O. Díaz, A. Irastorza, J. S. Cuadrado, and L. M. Alonso. From page-centric to portlet-centric Web development: Easing the transition using MDD. *Information and Software Technology*, 50(12):1210 – 1231, 2008.

[13] O. Díaz and I. Paz. Turning web applications into portlets: Raising the issues. *Symposium on Applications and the Internet (SAINT'05)*, pages 31–37, 2005.

[14] E. Freeman, E. Freeman, B. Bates, K. Sierra, and M. Loukides. *Head First Design Patterns*. O'Reilly Media, Inc., 2004.

[15] K. Furdík. Secure Process-oriented Infrastructure for Networked Enterprises. In *Workshop on Data Analysis WDA'2009*, volume 2, pages 98–105. 2009.

[16] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns. Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional Computing Series, 1995.

[17] O. Gmelch and G. Pernul. Preventing malicious Portlets from Communicating and Intercepting in Collaboration Portals. In *Proceedings of the International Conference on Security and Cryptography (SECRYPT)*, 2010.

[18] T. Gurzki and H. Hinderer. Eine Referenzarchitektur für Software zur Realisierung von Unternehmensportalen. In U. Reimer, A. Abecker, S. Staab, and G. Stumme, editors, *Professionelles Wissensmanagement - Erfahrungen und Visionen*, GI-Edition - Lecture Notes in Informations (LNI). Bonner Köllen Verlag, 2003.

[19] S. Hepper. *JSR 286: Java Portlet Specification Version 2.0*. Java Community Process, 2008.

[20] G. Hohpe and B. Woolf. *Enterprise Integration Patterns*. Pearson Education, 2004.

[21] B. R. Katzy. Design and implementation of virtual organizations. In *Proceedings of the Thirty-First Hawaii International Conference on System Sciences (HICSS)*, volume 4, pages 142–151, Los Alamitos, CA, USA, 1998. IEEE Computer Society.

[22] J. Lee, K. Siau, and S. Hong. Enterprise Integration with ERP and EAI. *Communications of the ACM*, 46(2):54–60, 2003.

[23] D. S. Linthicum. *Enterprise Application Integration*. Addison-Wesley Information Technology Series. Addison-Wesley, 2000.

[24] I. Paz, O. Díaz, R. Baumgartner, and S. F. Anzuola. Semantically integrating portlets in portals through annotation. In *Web Information Systems - WISE 2006*, pages 436–447. Springer Berlin / Heidelberg, 2006.

[25] C. Pettey. Gartner Says Worldwide Web Conference and Team Collaboration Software Markets Will Reach $2.8 Billion in 2010. `https://www.gartner.com/it/page.jsp?id=507717`, June 2007.

[26] D. Riehle. The commercial open source business model. In *AMCIS 2009 Proceedings*, 2009.

[27] W. A. Ruh, F. X. Maginnis, and W. J. Brown. *Enterprise Application Integration*. John Wiley & Sons, Inc. New York, 2001.

[28] C. C. Shilakes and J. Tylman. Enterprise information portals. *Merril Lynch*, November 1998.

[29] D. Spinellis, G. Gousios, V. Karakoidas, P. Louridas, P. J. Adams, I. Samoladas, and I. Stamelos. Evaluating the quality of open source software. *Electronic Notes in Theoretical Computer Science*, 233:5 – 28, 2009.

[30] R. Thompson. *Web Services for Remote Portlets Specification v2.0*. OASIS, 2008.

[31] B. Wilson. MAMA: What is the Web made of? online publication at `http://dev.opera.com/articles/view/mama/`,

October 2008.

[32] R. Winter. Ein Modell zur Visualisierung der Anwendungslandschaft als Grundlage der Informationssystem-Architekturplanung. In J. Schelp and R. Winter, editors, *Integrationsmanagement*, pages 1–29. Springer Berlin / Heidelberg, 2006.

The BabelNEG System - A prototype Infrastructure for protocol-generic SLA Negotiations

Sebastian Hudert
Department of Information Systems Management, University of Bayreuth
Universitaetsstr. 30
95447 Bayreuth, Germany
sebastian.hudert@uni-bayreuth.de

Torsten Eymann
Department of Information Systems Management, University of Bayreuth
Universitaetsstr. 30
95447 Bayreuth, Germany
torsten.eymann@uni-bayreuth.de

ABSTRACT

Visions of the next-generation Internet of Services are driven by digital resources traded on a global scope. For the resulting economic setting, automated on-line techniques for handling services and resources themselves, for advertising and discovering as well as for the on-the-fly negotiation of proper terms for their use are needed. Hence, a flexible infrastructure for the respective management of services and associated service level agreements is mandatory.

In this paper we present a novel approach for such an infrastructure, building on software-agent technology and an expressive but still machine manageable protocol description language, capable of specifying a multitude of different negotiation protocols. It supports the discovery of services with appropriate SLA negotiation styles as well as the actual SLA negotiations based on the chosen protocol description documents.

Keywords

service level agreements, internet of services, electronic negotiations, software agents, quality of service

1. INTRODUCTION AND MOTIVATION

Current developments in the area of Information Systems show a tendency towards massively distributed infrastructures, consisting of highly specialized digital resources. Today's Internet of mainly human interactions will evolve towards a socio-technical and global information infrastructure, where humans as well as software agents, acting on their behalf, continuously interact to exchange data and computational resources. This vision can be observed in both research and industry alike and is commonly referred to as the Internet of Services (IoS) [27, 25].

Building on currently applied computing paradigms, such as Service-oriented [11], Grid [12] or Cloud Computing [6], the IoS vision defines highly dynamic networks of composable services, offered and consumed on demand and on a global scope. Taking such ideas one step further, it rigorously focuses on the goal of an Internet-based service economy, similar to the real-world service sector. Digital services will be offered over electronic service markets, purchased by respective customers and then combined with internal or other external services to business workflows of varying complexity.

Hence, the IoS will primarily focus on new business models and the commercial application of distributed computing, concerning trading processes down to the level of an individual service, and the subsequent charging based on its usage and delivered quality-of-service (QoS). In such a setting even very small and specialized companies can find a niche in the digital economy where they can compete with the ubiquitous international enterprises, which in turn have to face a much higher competition on the global market [9].

Summarizing the IoS scenario thus results in the following characteristics [27]:

- The IoS focuses on a (potentially huge) set of electronic services of varying complexity.
- These services will be employed in potentially mission-critical business processes and thus have to fulfill a (pre-negotiated) set of QoS guarantees as stated in a Service Level Agreement (SLA)[1].
- New business models will cope with the possibility of trading even very fine-granular services and charging them based on their actual usage.
- It will consist of a global set of SPs and SCs, negotiating over digital services as well as some mediating nodes, such as service brokers.

Two of the main challenges for the IoS, from a commercial perspective, are reliability of the services traded and the technical infrastructure underlying the service economy.

[1]A SLA is a structured document, describing a bilateral agreement between a service provider (SP) and a service consumer (SC) on the terms and conditions of the invocation(s) of an (electronic) service.

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

The need for guaranteed reliability and service quality becomes more prominent, as no longer the question of "who provides the service?" matters but only whether he is able to achieve the requested result.

Since such scenarios inherently lack the applicability of centralized QoS management, service guarantees must be obtained in the form of bi-lateral SLAs assuring service quality across individual sites [18]. These SLAs subsequently act as a signed contract governing the actual service invocation [6], enabling the structured monitoring and assessment of the service's compliance[2].

A very crucial part of the SLA-based service life cycle can be seen in the discovery and, above all, the negotiation phase. All subsequent steps (binding, execution and monitoring, post-processing etc.) depend on the SLA documents which were agreed-upon in this phase. On that account, we focus on a the negotiation and prior discovery of SLAs for our work presented in this paper.

Aiming at the negotiation phase, economic reserach claims that differences in system configuration, or the services actually traded, demand different negotiation protocols in order to reach the highest-possible efficiency of the overall system (see for example [19]). Based on these findings and the global context of the envisioned scenario it is not likely, or even efficient, that only one central marketplace for electronic services will emerge, offering a single, known protocol. Instead a system of marketplaces offering different protocols will probably emerge, each of which is best be suited for a given context.

Fortifying this, we argue that restricting SCs in that they are only able to interact with one distinct service market they were implemented for (and are therefore only technically compatible with the applied negotiation protocol), unnecessarily decreases the potential flexibility and efficiency of the whole system. SCs should be able to buy, and therefore negotiate about, any fitting service, regardless of the market it is offered in, and thus regardless of the protocol with which it is offered. SPs, on the other hand, should be able to offer their services with the protocol best fitting to the current market situation instead of being restricted to a given protocol by the market infrastructure.

Also, given the dynamic nature of distributed workflow executions and the increased complexity of global service selection manual negotiations of the human users are by far not efficient enough. This process should be automated by electronic software agents that negotiate on the users' behalf [17].

The research goal of our work is thus to develop a service-oriented infrastructure supporting software agents to discover and negotiate about electronic SLAs and not restricting them to a pre-defined negotiation protocol.

[2]In our work we focus on electronic SLAs as needed in the anticipated fully automated setting. Such documents are mainly used for resource management and scheduling in various research projects [22] and mark a promising approach for representing real-world, legally binding, contracts in an electronic way.

The remainder of this paper is structured as follows: In section 2 we will present a short overview on related research projects in the area of electronic SLA management systems and the IoS. Next, the main contribution of this paper is presented in section 3, the design of our SLA negotiation infrastructure. Section 4 will provide information on the developed proof-of-concept prototype system and the evaluation steps already conducted. Finally, we conclude our paper with a short summary and the identification of future work.

2. RELATED WORK

A significant amount of research projects exist, dealing with distributed QoS management based on SLAs. Such efforts have risen after traditional distributed systems came to maturity and reliability came into focus. Some of the most notable approaches were presented by Ludwig et al. [18], Yarmolenko and Sakellariou [26] or Tosic et al. [30]. These works mainly address the internal structure of SLAs, the relevant service metrics and SLA-based resource management and scheduling mechanisms.

Building on such theoretical works, an ever growing amount of research projects, such as CoreGRID[3] or SLA@SOI[4] employ SLAs for resource management and designed respective SLA management systems. Also the ongoing Web Services Agreement [1] standardization effort at the Open Grid Forum[5] shows the growing interest in SLA-based QoS management from both research and industry.

On the other hand a variety of different negotiation settings and respective protocols for both electronic and real-world markets have been introduced (see for example [24] for very fundamental work on negotiations). As a next step, these findings were ported to the digital world, forming the new research discipline of Electronic Negotiations [2]. This lead to the definition of formal descriptions and characterizations of given negotiation protocols [28] as well as first attempts to software infrastructures for (electronic) negotiations (see for example [32]).

After software agent technologies [31] reached maturity it was only a logical next step to employ the mechanisms developed therein for the implementation of electronic markets [23].

Surprisingly, there is little research done in combining the economic considerations on the one hand and the technical developments on the other hand. No more than a couple of research groups address the agent-based, electronic negotiation of SLAs; resulting infrastructures were presented for example in [7] or [21]. Even those projects mainly focus on static and centralized architectures within which only one particular, and fixed, negotiation protocol is implemented. Hence, they allow for the definition of individual service markets, but still lack the possibility for SCs to migrate from one market, and thus one negotiation protocol, to another (in analogy to a real-world economy).

[3]http://www.coregrid.net
[4]http://sla-at-soi.eu
[5]http://www.ogf.org

Although the need for protocol-generic systems is widely agreed-upon (see for example [19]), only a few projects incorporate the mere possibility of different protocols within one infrastructure; [17] and [3] being two of the most prominent examples. However, both still lack important flexibility by restricting the negotiation protocols to a small and fixed set and by building on static, centralized architectures without appropriate discovery mechanisms.

3. INFRASTRUCTURE DESIGN

Before describing the actual system design, we will now present the results of a literature-based requirements analysis, underlining the need for identified protocol-generic SLA negotiation systems.

3.1 Requirements Analysis

Among the abovementioned researchers, a common agreement exists on which requirements are posed on electronic SLA management infrastructures for the IoS. The most significant requirements concerning the discovery and negotiation phases include:

R1 After the discovery phase all parties must have a common understanding of the protocol to be executed in the negotiation phase [17].

R2 This common understanding must be generated dynamically at runtime [4].

R3 Services (and thus SLAs) of different complexity must be negotiable [19].

R4 Different marketplaces and protocols (fixed-price catalogues, bargaining, auctions etc.) even within one infrastructure are needed for different services to be traded [19]. In order to cope with the highly dynamic IoS environment the available set of protocols should not be restricted a priori [15].

R5 Software agents should act as negotiators [17].

R6 Intermediaries, such as auctioneers or brokers, should be present [6].

Although most of these requirements were even identified in the context of either one of the abovementioned research projects, no infrastructure to date fulfills all of them. Especially the claimed protocol-flexibility of SLA negotiations, in particular at run time, is not reached or even not addressed at all. In the following we will present the design of our system, aiming at closing that gap.

3.2 Abstract Design Idea

The basic design idea underlying this work is to offer a given product (SLA for an electronic service) independently from the way an agreement concerning this product can be attained (negotiation protocol). This allows for flexible combinations of product and protocol to be chosen for each market situation individually.

Such an approach has many analogies in real-world settings. For example a TV set, offered at an electronic retailer implicitly states that the only way to negotiate about it is to accept the stated price. This protocol thus corresponds to a classic cataloque pricing model. Then again, the same TV set, sold on an online auction platform such as Ebay[6], implies that the consumers have to outbid each other until a certain deadline occurs. This in turn corresponds to an English Auction protocol. Although the product sold in both cases is exactly the same (a new TV set of a given type) the negotiation protocols applied are quite different. Transferred into the electonic SLA management scenario, this means that a given SLA (template[7]) under negotiation could be offered with a whole set of different negotiation protocols at different points in time, best fitting the current market situation.

Since software agents will be employed for the service management within our systen (R5), the negotiation protocol applied for a given service is not only decoupled from the actual SLA but must also be made explicit in terms of its communication rules. This allows for run time adaption of the SC agents to the respective protocol.

To this end, the designed infrastructure will build on a conceptual architecture of machine-readable description documents, as described in the following.

3.3 Service Description Documents

As just introduced, a set of service description documents is needed enabling a) the discovery of an appropriate service or respective SLA (template) and b) the description of the negotiation protocol used to reach an agreement. For that purpose three different data structures were designed:

- Service Type (ST): definition of the functional and non-functional aspects, a given class of services can offer.
- Extended SLA Template (EST): definition of initial QoS guarantees (building on the non-functional aspects given in the respective ST) as an input for the subsequent negotiation as well as the applied negotiation protocol.
- Service Identificator (SI): identification of an individual service instance along with links to the associated ST and EST documents.

Figure 1 gives a short overview on these documents and their relations.

[6]http://www.ebay.com/

[7]Most of the present SLA (management) systems incorporate the possibility to define SLA templates (see for example [1]). A SLA template is basically a partially filled out SLA document, that is offered from the SP to potential cunsumers. It regularly marks the starting point for an actual negotiation process, during which the already stated service guarantees (also called service level objectives (SLOs)) are altered according to the preferences of the negotiators until the final SLA is reached. The internal structure of a SLA template is equivalent to an actual SLA document. The only difference is, that a set of rules is optionally defined, helping the SC to identify (combinations of) SLOs that are a) valid and/or b) acceptable by the SP (see for example [1, pp. 30-33]).

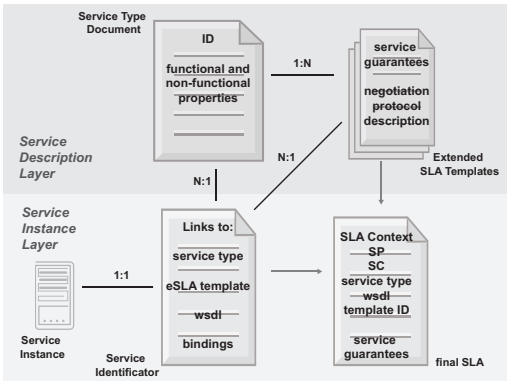


Figure 1: Document-based Architecture

Note: All data structures were defined on the basis of XML Schema [29] due to the massive adoption of XML [5] for communications in service-oriented systems.

3.3.1 Service Type

The basic goal of the ST document is to describe a class of service instances in terms of their functional capabilities and the non-functional parameters, such as throughput or price.

The functional aspects are primarily used to search for a service, taking into account the search criteria as received from the user / service requestor. In contrast, the non-functional parameters, or rather their actual values, are mostly the very object of any SLA negotiation. These parameters' actual values vary during run-time, depending on the current workload. The incorporation of respective monitoring and enforcement components at design-time however, offers the management agents the possibility to negotiate about such metrics and subsequently enforce the resulting guarantee, even at run-time.

3.3.2 Extended SLA Template

In our work, the SLA template concept was not only also used to define initial SLOs (R3), but additionally extended to also include a description of the applied negotiation protocol[8].

The respective part of the EST is created, building on a negotiation protocol description language developed by the authors (a previous version of which can be found in [14]). It builds on simple parameters where possible and, if more complex restrictions are needed, allows the application of any desired (external) rule language within some of the language elements.

A respective protocol description provides the querying SC agent with information about the:

- *negotiation context* (permitted agents and their permissions and obligations),
- *object* (actually negotiable SLOs),
- *offer restrictions* (constraints posed on the negotiable SLOs, such as upper or lower bounds of acceptable values),
- *allocation* (matching function applied for winner determination),
- *information policy* (information access rules, as needed for example to define sealed-bid or outcry negotiations)
- *negotiation process* (definition of the allowed behavior of the involved agents, employing the event-condition-action paradigm).

[8]Since this aspect is very closely linked to the SLOs already stated in the template it was considered appropriate to integrate it into the same data structure.

3.3.3 Service Identificator

Each individual service instance is finally defined within a respective SI. Such a document describes where exactly this service can be found (important for the actual binding process), what its type is and which EST is offered for it. It thus consists of the following elements:

- *serviceID*: identifier for this particular SI. This element should be of the type URI in order to ensure uniqueness in a global setting.
- *serviceTypeID*: link to the description of this service's type
- *slaTemplateID*: link to the EST offered for this service
- *wsdlFile*: reference to the WSDL [8] file, describing the actual service interface in terms of operations with input and output parameters as well as error types.
- *negotiationCoordinator (NC)* and *serviceProvider*: these two elements represent role bindings for this service[9].

3.4 Protocol Design

In order to support both, the discovery and the negotiation of SLAs, a set of simple protocol primitives have been developed, building on the abovementioned data structures.

3.4.1 Discovery Phase

The discovery phase basically represents the set of activities ultimately leading to a situation where potential business transaction partners (SCs and SPs) know one another and can start a negotiation process. This means the discovery phase is supposed to support a given SC to find one or more SI documents fitting the search criteria it received from the user (R1 and R2). To this end, the SPs should be able to publish the service it offers (in terms of the respective description documents), in a way that it can be found by potentially interested SCs.

For our proof-of-concept implementation, a very simple registry node, which can be found via a broadcast-based discovery protocol, was chosen (see figure 2 for more details). Such a simple architecture was considered sufficient for investigating the research question at hand. Future versions

[9]For more details on the role architecture, see subsection 3.5.

however, will probably favor more robust mechanisms, such as Peer-to-Peer discovery approaches.

For either the registry discovery, the publication of one or more service documents or the discovery of those a set of message types was defined. These messages strictly follow the internal structure of SOAP [13] messages to allow a seamless integration into present Web Service infrastructures.

At first the SP, upon receiving a request to sell a given service, publishes the respective service description documents at the registry node. Subsequently it waits for a SC to request admission to the actual negotiation. The SC process is on the other hand triggered by the reception of a user request for a particular service. In a first step, the SC requests all SIs fitting the search criteria and retrieves the (until then unknown) ST and EST documents from the registry. Given a list of adequate SIs were found, the SC chooses one of them and tries to start / join a negotiation via an explicit admission step at the NC agent (for each negotiation protocol exactly one agent adopts this role; see subsection 3.5 for more details). In case of a successful admission the SC and SP now engage in the actual negotiation process.

3.4.2 Negotiation Phase

Since the main goal of our work is to define an infrastructure for protocol-generic negotiations, no single negotiation protocol can be identified for this phase[10]. Rather, a set of negotiation message types along with their respective contents was defined. These messages can then be used in a given negotiation process, orchestrated according to the protocol description in the EST document (R1, R2 and R4).

After a thorough literature review on currently applied negotiation protocols (as listed for example in online libraries such as [10]), a minimally necessary set of such messages could be identified, consisting of *offer(toSell)*, *accept*, *reject*, *callForBids* and *stillInterested* messages. Using these message types a variety of different protocols can be described in an EST document and subsequently processed by the service management agents[11].

3.5 Role-based Architecture

The developed prototype infrastructure builds on a defined set of roles, the service management agents can adopt (see figure 3):

The two basic roles present in this system are the SC and the SP, representing buyer and seller agents.

Additionally a set of registry services / agents (RA) are needed for supporting the publication and discovery of service description documents. Finally, the NC role represents the agents mediating negotiation processes as a broker agent (R6).

[10]Therefore no distinct process model for this phase (analogue to figure 2) can be given. Individual Negotiation phases always differ, depending on the applied protocols.

[11]A set of mutually very different negotiation protocols, designed only building on these message types, can be found at `https://sourceforge.net/apps/mediawiki/simis/index.php?title=BabelNeg`

Both SP and SC agents mainly offer an interface to one another, allowing them to send and receive messages related to the discovery and negotiation phases. Additionally, each of theses roles offers one routine to external users of the infrastructure: A SP agent offers a method for publishing and selling and a SC one for discovering and purchasing a service on a user's behalf, respectively.

A RA only accepts discovery related messages as it does not take part in any other phases of the life cycle. Finally, NC agents are responsible for admission of SCs or SPs to and potentially mediation of a given negotiation. Hence, they again offer methods for exchanging respective messages.

In the following the internal routines of each role, as implemented in the proof-of-concept prototype, will be further detailed.

3.5.1 Service Provider

A SP agent's basic purpose is to publish a given service to potential customers and sell it to them subsequently. On an abstract level a SP agent first receives the request for publishing a service and the respective description documents (triggering a state change from IDLE to PUBLICATION). Now it stores all respective documents at a RA node (which potentially has to be discovered first). If successful, it transits to the BUSY state, in which the service is continuously offered, negotiated about and executed. The SP only exits this state when the service is taken off-line or is re-deployed.

3.5.2 Service Consumer

The default process for a SC is to receive a service request and move to the discovery phase, during which a set of fitting SI documents are retrieved. After choosing an appropriate SI the SC agent requests admission to the respective negotiation, joins the negotiation process if admission was successful and finally invokes the service in the EXECUTION state if it could win the negotiation.

A fundamental principle was implemented all along the process of a SC agent: the reluctant escalation backtracking in terms of internal states. In case of a failed attempt to move to the next state (e.g. from the DISCOVERY to the INITIATION_NEGOTIATION state) it checks its options for another such attempt. In this case it checks whether or not any other SI was found, for which it could request admission to the respective negotiation. Only if no such options are available it moves back to the predecessor state and checks whether it can start over from there or if even there no other options are available and so on. This principle ensures the agent to fully exploit all possibilities it could discover for reaching an agreement.

3.5.3 Protocol-generic SC Strategy Component

During the actual negotiation phase a protocol-generic strategy component is employed by the SC agents. All routines needed for adaption to a new negotiation protocol are implemented herein.

When admitted to a negotiation process, the SC instantiates such a generic negotiator (GN) component and passes over the ST, EST and SI documents, as applied for the respective

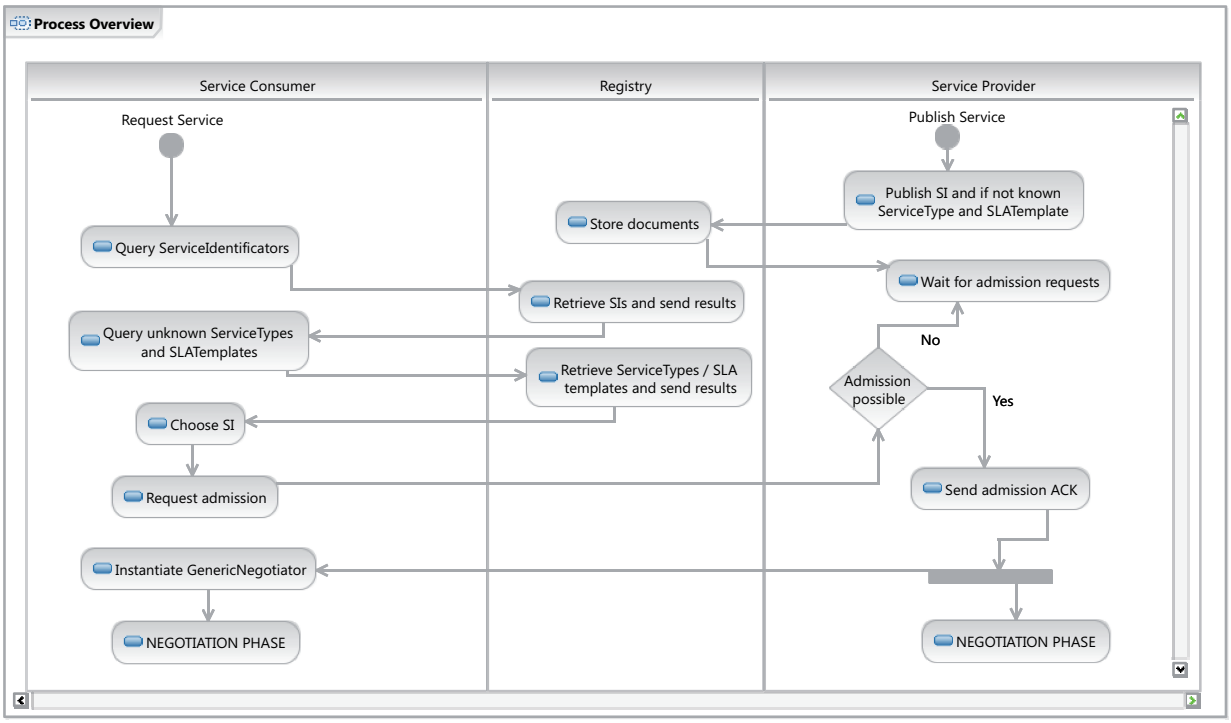


Figure 2: Discovery Phase Sketch

negotiation, as well as all internal constraints received from the user (such as reservation values for certain SLOs).

Once all this information was extracted the GN checks whether it has to start the negotiation (pro-active protocol). If so, two possible actions can be defined: this agent can be allowed to post an offer or to simply accept all the the values stated in the EST (corresponding to a catalogue pricing model).

Here a basic principle of the GN becomes obvious: it always seeks to maximize its utility and thus will always choose some actions over others (if both are allowed). In this case it will first check, whether an offer is possible (such checks are always done by inspecting the respective event-condition-action tuples stated in the EST), which could further improve the currently offered agreement. If so, an offer will be sent; if not it will simply accept the current values, as no negotiation on them is allowed (if they are acceptable).

The same principle applies throughout the whole negotiation process. Whenever a negotiation message is dispatched to the GN it checks its options: In case of a reject or accept message the negotiation is over. It simply processes the result in that it passes the respective information / reached agreement to the SC agent. In case of a callForBids message it creates an offer and sends it to the SP agent[12]. All these possibilities are straightforward, as they don't give the GN any option to choose among a set of possible actions.

When receiving an offer, this could potentially change. If the offer is not completely rejectable (this is checked by iterating over all involved slo constraints) and a counter offer is possible, the GN will always do so. A counter offer is always the best option, as it potentially increases value of the agreement for the user. If this is not possible, the GN will check whether a stillInterested-message is allowed, providing it with the possibility of an ongoing negotiation, even if it cannot actively influence the changing of the negotiated values. If even this is not possible it will finally check whether the received offer can be accepted or in the end rejected completely and do so. This routine gives the GN the possibility to react on incoming messages in a way that maximizes its further options during the negotiation and in the end potentially its utility in terms of the reached agreement.

3.5.4 Registry

The RA node is basically just a placeholder element for any discovery mechanism used in future versions. It was designed to simply receive register messages and store the corresponding data into an internal data structure. This is for example used when a SP publishes a new SI document. Supporting the discovery of service documents it also offers the possibility to query the stored SI, EST and ST documents. No complex internal states are maintained and changed throughout its life time. It simply processes requests to register or query service documents or answers registry discovery messages.

[12]Offer messages are simply created by iterating over all not yet fulfilled user constraints and creating a counter offer value for each. These values are then combined to one offer message.

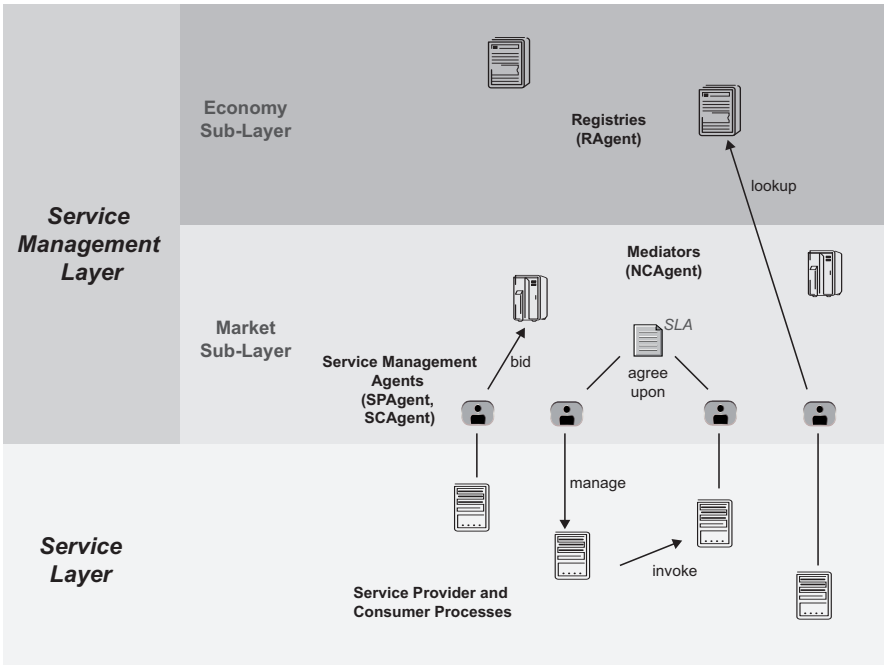


Figure 3: System Architecture

3.5.5 Negotiation Coordinator

The main task of a NC agent is to handle the admission of SCs to a given negotiation. For this it offers the possibility to submit respective messages along with a set of credentials, as needed for the admission decision. Additionally, if the negotiation phase is also assigned to the NC (mediated negotiation), it must also be able to process incoming negotiation messages. Just as described for the SP and SC agents, these are simply forwarded to a negotiation strategy component, which in turn provides the protocol-specific functionality.

Similarly to the RA nodes, a NC thus does not expose complex internal states and state changes. Admission requests are evaluated and answered based on the received data and the implicit service availability information, and negotiation related messages are simply forwarded to the strategy component.

4. PROTOTYPE IMPLEMENTATION AND EVALUATION

The presented mechanisms were implemented in a Java-based proof-of-concept prototype infrastructure, building on the agent-based IoS simulation toolkit SimIS[13]. The simulation experiments, presented in the following, represent one evaluation step for our work, allowing us to model different environmental settings (i.e. market configurations) and demonstrate the feasibility of our approach therein. In a second step, future versions of the developed components will be ported to a productive IoS platform dealing with the whole service life cycle.

[13]http://sourceforge.net/projects/simis

SimIS was co-developed by the authors and aims at providing researchers with a comprehensive framework for investigating distributed algorithms or protocols within the context of the IoS vision. Building on the generic Recursive Porous Agent Simulation Toolkit (REPAST) [20], it proposes a two-tiered architecture dividing the overall system into an *Application Layer* and an *Infrastructure Layer* (see figure 4).

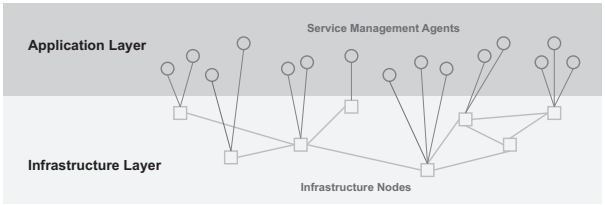


Figure 4: SimIS Toolkit

The Infrastructure Layer models topological settings of the IoS. The basic idea is that all Application Layer Agents / Services are linked to a single Infrastructure Agent each, which is representing their server platform, mainly performing message handling tasks.

Within the Application Layer the actual services of the IoS vision are modelled, communicating via the offered message objects and routing functionality. Each service, i.e. Application Layer Agent, is to be implemented as a plain Java class and can therefore exploit the full potential this programming language offers. A more detailed description of the SimIS toolkit can be found in [16].

The agent types developed for our infrastructure prototype were consequently implemented as specific Application Layer agents within SimIS. Each of the management agents (SCs and SPs) is accompanied by a strategy object, the SCs with the protocol-generic negotiator and the SPs with a strategy distinctly fitting to the offered protocol (as parameterized during startup).

In a first step the document-based architecture and the respective discovery mechanism were tested by running a variety of different configurations within the simulated IoS setting (varying numbers of SCs and respective service requests, SPs with a given service / protocol combination, and RAs supporting the discovery process). Both the protocol steps applied and the RA implementation were able to prove effectiveness very well.

Secondly an initial set of negotiation protocols was chosen to be a) described with the developed language and b) applied within SimIS to show whether or not the GN is able to correctly extract the relevant information and take part in the subsequent negotiation.

For the subsequent evaluation runs the Alternate Offers, an English Auction and Double Auction Protocol were instantiated. This selection covers 1:1 (Alternate Offers), 1:N (English Auction) and M:N (Double Auction) settings and can thus be regarded as a representative subset of the most commonly used negotiation protocols [14].

In the Alternate Offers Protocol, both parties basically exchange offer messages. Upon receving such a message the agent has three options: accepting the offered values, rejecting them ultimately or creating a counter-offer message. In the English Auction, the SP offers a given value for the negotiated metric (in the shown experiments, the price) to all SCs, which in turn can answer with stillInterested messages, given they are still interested in the negotiated service under the offered conditions. In the next round the SP increases the value of the negotiated SLO and offers it to all remaining SC in the same way. This process is repeated until only one SC is left that accepts the offered value. During a Double Auction both, SP and SC, post messages indicating their service offers or demands. The broker (NC) subsequently matches these according to a defined matching function and passes the results to back to the SPs and SCs.

For demonstration purposes, we configured the simulator with ten infrastructure nodes, upon which 40 SPs and 50 SCs are placed. Each SP offers the same service, however 8 of those with the Alternate Offers, 12 with the English Auction and the remaining 20 with the Double Auction protocol, as described above. Additionally one registry node for storing and retrieving the service and protocol description documents as well as one broker for the Double Auction are present.

The SCs generate service requests at random points in time (mean duration between two request arrivals: 50 ticks), starting at tick 60; this way start-up effects of the simulation can be avoided. Each of these agents starts with the same valuation for the offered service but adopts its valuation over time (in case of a successful negotiation it decreases its valuation by 0,05, in case of an unsuccessful negotiation vice versa). Additionally, a set of timeouts were employed for timing of new auction rounds or in case of not answering opponents. The experiments were delimited to a maximum tick count of 20000.

[14]The used protocols, described in terms of UML Sequence Diagrams as well as a respective EST documents, can be found at: `https://sourceforge.net/apps/mediawiki/simis/index.php?title=BabelNeg`

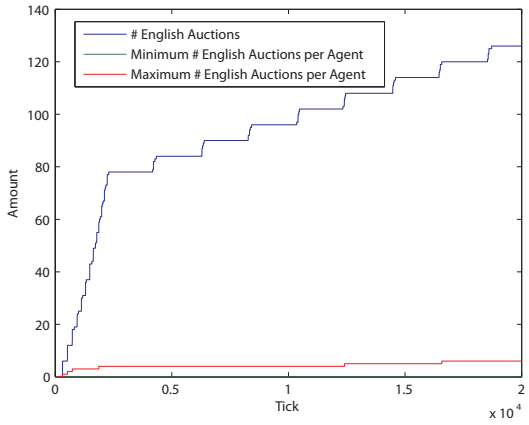


Figure 5: Successful English Auctions

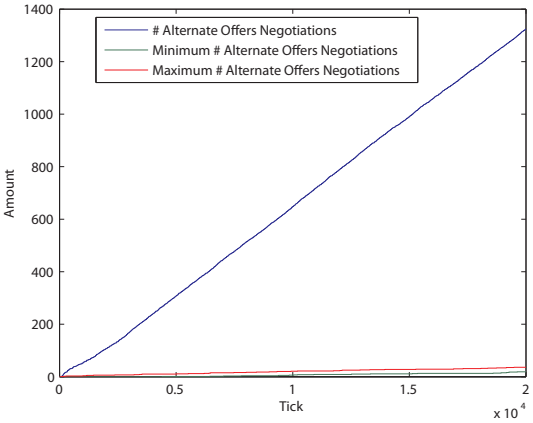


Figure 6: Successful Alternate Offers Negotiations

The main statement to be proven with these experiments is that the GN node is at all able to adapt to different protocols, only based on their description in the EST documents (proof-of-concept). Given the results shown in figures 6, 5 and 7, this assertion can be approved. Each of the introduced protocols was successfully integrated in the actual market behaviour; each protocol type present was executed by a significant number of agents.

This fact is underlined by figure 8 which shows that each

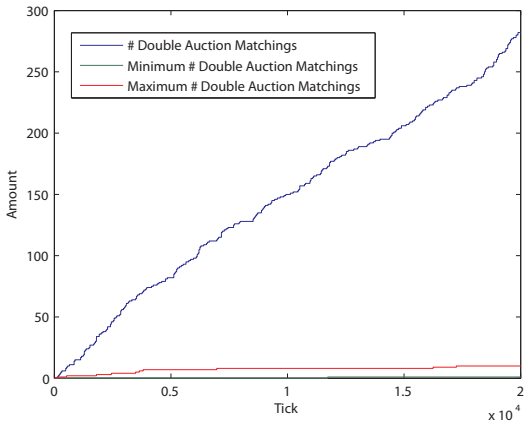


Figure 7: Successful Double Auctions

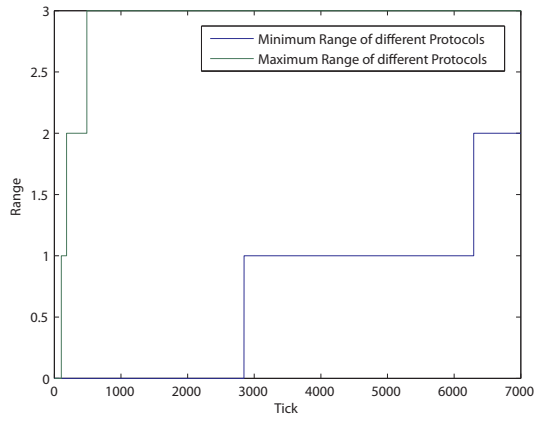


Figure 8: Range of used Protocols

agent could successfully take part in at least two different negotiation protocol types. On the other hand, there exist SC agents that have successfully taken part in all available protocols.

We will now shortly relate the developed system to the requirements stated in section 3.1.

R1 and R2: Dynamically generated understanding of negotiation protocol, during discovery phase

Each SC enters the discovery phase with no prior knowledge about the protocol that is executed in the subsequent negotiation phase. All information needed for a successfull participation in a respective negotiation is coded in the service description documents, which are queried during discovery. Therefore the requirements R1 and R2 are fulfilled by our system.

R3: Support for SLAs of different complexity

The ST, EST and SLA documents provide a very generic service (SLA) description structure. By offering free-form text elements along with the pre-defined and typed elements for quantitatively measurable service aspects, a comprehensive service description can be created. The defined document structures also enable the usage of external, standard languages for describing service characteristics (for example WSDL when describing the service interface).

R4: Support for different negotiation protocols

By employing a structured protocols description language in the EST document, each service (and SLA template respectively) can be offered over a different protocol. The GN strategy component, being able to adapt to the described protocols, supports all protocols that can be described in that language, thus fulfilling R4.

R5: Software agents acting as negotiators

The system architecture, as implemented in the proof-of-concept prototype, heavily builds on sofware agents as instantiations of a particular role. Such components are the basic actors in our system.

R6: Need for intermediaries

Our system incorporates an explicit intermediary role, used for the definition of market brokers, the NC. Therefore R6 is fulfilled by our infrastructure proposal.

5. CONCLUSION AND FUTURE WORK

In this paper we presented a novel infrastructure for the discovery and protocol-generic negotiation of SLAs. For this we employed a set of structured document types and software-agent technology for managing the incoming requests to buy and sell electronic services and negotiate about SLAs respectively. We evaluated our system subsequently, with a set of different negotiation protocols and system configurations, within the context of an IoS simulation toolkit.

Future work will comprise the simulative investigation of more complex scenarios in which a multitude of different brokers and other negotiation protocols are present at the same time. Such evaluation runs will especially focus on the utility gain a protocol-generic infrastructure can provide over one offering only one single protocol. Also, a more thorough investigation of negotiation strategies, able to efficiently cope with different protocols, has to be done.

6. REFERENCES

[1] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu. Web services agreement specification, version 03/2007. 2007.

[2] M. Bichler, G. Kersten, and S. Strecker. Towards a structured design of electronic negotiations. *Group Decision and Negotiation*, 12(4):311–335, 2003.

[3] I. Brandic, S. Venugopal, M. Mattess, and R. Buyya. Towards a meta-negotiation architecture for sla-aware grid services. In *Workshop on Service-Oriented*

Engineering and Optimizations 2008. In conjunction with International Conference on High Performance Computing 2008 (HiPC2008), Bangalore, India, December 17 - 20, 2008.

[4] I. Brandic, S. Venugopal, M. Mattess, and R. Buyya. Towards a meta-negotiation architecture for sla-aware grid services. Techreport, University of Melbourne, August 2008.

[5] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Yergeau. Extensible markup language (xml) 1.0 (fourth edition). *W3C*, August 2006.

[6] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*, 25(6):599 – 616, 2009.

[7] M. B. Chhetri, J. Lin, S. Goh, J. Y. Zhang, R. Kowalczyk, and J. Yan. A coordinated architecture for the agent-based service level agreement negotiation of web service composition. In *ASWEC '06: Proceedings of the Australian Software Engineering Conference (ASWEC'06)*, pages 90–99, Washington, DC, USA, 2006. IEEE Computer Society.

[8] R. Chinnici, J.-J. Moreau, A. Ryma, and S. Weerawarana. Web services description language (wsdl) version 2.0 part 1: Core language. *W3C*, March 2006.

[9] T. P. Consortium. Texo: Business webs im internet der dienste (german). http://theseus-programm.de/anwendungsszenarien/texo/default.aspx, 2009. last checked: 2010.01.13.

[10] T. F. for Intelligent Physical Agents. |"www.fipa.org. last checked: 17. 07. 08.

[11] I. Foster. Service-oriented science. *Science*, 308(5723):814–817, 2005.

[12] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 15:2001, 2001.

[13] M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, and H. F. Nielsen. Soap version 1.2 part 1: Messaging framework. *W3C*, June 2003.

[14] S. Hudert, T. Eymann, H. Ludwig, and G. Wirtz. A negotiation protocol description language for automated service level agreement negotiations. In *Proceedings of the 11th IEEE Conference on Commerce and Enterprise Computing (CEC 09), Vienna, Austria*, 2009.

[15] S. Hudert, H. Ludwig, and G. Wirtz. Negotiating slas - an approach for a generic negotiation framework for ws-agreement. *Journal of Grid Computing*, 7(2):225–246, June 2009. ISSN: 1570-7873 (Print) 1572-9814 (Online).

[16] S. König, S. Hudert, and T. Eymann. Socio-economic mechanisms to coordinate the internet of services Ű the simulation environment simis. *Journal of Artificial Societies and Social Simulation (JASSS)*, 13(2), 2010.

[17] A. Ludwig, P. Braun, R. Kowalczyk, and B. Franczyk. A framework for automated negotiation of service level agreements in services grids. In *Lecture Notes in Computer Science, Proceedings of the Workshop on Web Service Choreography and Orchestration for Business Process Management, 2006*, volume 3812/2006, 2006.

[18] H. Ludwig, A. Keller, A. Dan, R. King, and R. Franck. A service level agreement language for dynamic electronic services. *Journal of Electronic Commerce Research*, 3:43–59, 2003.

[19] D. Neumann, J. Stoesser, C. Weinhardt, and J. Nimis. A framework for commercial grids - economic and technical challenges. *Journal of Grid Computing*, 6(3):325–347, September 2008. ISSN: 1570-7873.

[20] M. J. North, N. T. Collier, and J. R. Vos. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Trans. Model. Comput. Simul.*, 16(1):1–25, 2006.

[21] M. A. Oey, R. J. Timmer, D. G. A. Mobach, B. J. Overeinder, and F. M. T. Brazier. Ws-agreement based resource negotiation in agentscape. In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–2, New York, NY, USA, 2007. ACM.

[22] M. Parkin, R. M. Badia, and J. Martrat. A comparison of sla use in six of the european commissions fp6 projects. Technical Report TR-0129, CoreGRID, 2008.

[23] S. Paurobally, V. Tamma, and M. Wooldridge. A framework for web service negotiation. *ACM Trans. Auton. Adapt. Syst.*, 2(4):14, 2007.

[24] H. Raiffa. *The Art and Science of Negotiation*. Harvard University Press, Cambridge, Mass., 1982.

[25] R. Ruggaber. Internet of services sap research vision. In *WETICE '07: Proceedings of the 16th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, page 3, Washington, DC, USA, 2007. IEEE Computer Society.

[26] R. Sakellariou and V. Yarmolenko. On the flexibility of ws-agreement for job submission. In *MGC '05: Proceedings of the 3rd international workshop on Middleware for grid computing*, pages 1–6, New York, NY, USA, 2005. ACM.

[27] C. Schroth and T. Janner. Web 2.0 and SOA: Converging concepts enabling the internet of services. *IT Professional*, 9(3):36–41, 2007.

[28] M. Stroebel and C. Weinhardt. The Montreal Taxonomy for electronic negotiations. *Journal of Group Decision and Negotiation*, 12:143–164, 2003.

[29] H. S. Thompson, D. Beech, M. Maloney, and N. Mendelsohn. Xml schema part 1: Structures, second edition. *W3C*, October 2004.

[30] V. Tosic, K. Patel, and B. Pagurek. Wsol - web service offerings language. *Lecture Notes in Computer Science*, 2512/2002:57–67, 2002. ISSN 0302-9743.

[31] M. Wooldridge. Agent-based software engineering. *IEEE Proceedings Software Engineering*, 144(1):26–37, 1997.

[32] P. R. Wurman, M. P. Wellman, and W. E. Walsh. The Michigan Internet Auctionbot: A configurable auction server for human and software agents. *Second International Conference on Autonomous Agents*, May 1998.

Goal-oriented requirements modeling as a means to address stakeholder-related issues in EA

Carsten Lucke
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg, Germany
carsten.lucke@unibw.de

Ulrike Lechner
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg, Germany
ulrike.lechner@unibw.de

ABSTRACT

In this paper we explore goal-oriented requirements engineering (GORE) as a means to address stakeholder-related issues in the enterprise architecting process. We elaborate on a recent literature analysis on enterprise architecting issues. We refine this literature analysis results by identifying problem areas that we consider solvable by increasing the focus on the stakeholders in EA. We develop a conceptual model, which we use to provide reasoning about means to foster stakeholder orientation and thereby to address stakeholder-related issues. We argue that a stronger focus on the stakeholders' benefits EA and that this increased stakeholder orientation can be reached by leveraging intentional modeling used in software engineering.

Keywords

enterprise architecture, stakeholders, goal-oriented requirements engineering, GORE, intentional modeling, goal modeling.

1. INTRODUCTION

A recent literature review on critical issues of the enterprise architecting process indicates a number of problem areas [1]. One particular problem area that has been identified is concerned with the stakeholder topic. Stakeholder focus is a relatively new topic in EA literature. Until 2008 EA literature was rather focused on overview on EA, best practices, EA frameworks and enterprise modeling in general [2]. Recent publications show that the attention being paid to stakeholders in EA increases [3-10].

In this paper we elaborate on the results of the aforementioned literature review on issues in enterprise architecting [1] and argue that stakeholder orientation is crucial in EA since a significant number of problem areas in the EA process are related to stakeholders, stakeholders' goals and requirements. We develop a conceptual model to capture critical issues, the relations between critical issues and the role of stakeholder management. We argue that the goal-oriented approach [11-13] to requirements modeling applied in requirements engineering is a means to an increased stakeholder focus in EA and allows to address stakeholder-related issues in the EA process.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

This paper is structured as follows. Section 2 provides the theoretical foundations for this paper and section 3 describes the research method. In section 4, we describe the predominant focus areas of issues occurring during the EA process. Section 5 presents our conceptual model of stakeholder orientation and section 6 provides reasoning for our proposition that GORE is a means to address stakeholder-related EA issues. Section 7 concludes this paper with a brief discussion of our argumentation.

2. THEORY

2.1 Stakeholders in Enterprise Architecture

Stakeholder theory is a concept originating from strategic management addressed in disciplines like business ethics [14], project management [15] and also information systems [16]. An important proposition of stakeholder theory is, that financial benefit of its shareholders should not be the only obligation of a company. Stakeholder theory recognizes, that organizations are as well dependent on a number of constituency groups and have moral and ethical obligations over these groups [14, 17]. These constituencies are referred to as stakeholders. Mitchell et al. present a chronology of the stakeholder term [18]. According to this chronology, the term can be traced back to 1963 where it appeared in a Stanford memo describing it as "those groups without whose support the organization would cease to exist" [18]. In Freeman's seminal publication a stakeholder is defined as "any group or individual who can affect or is affected by the achievement of the organization's objectives" [19].

The stakeholder notion is also discussed in information systems and more specifically in enterprise architecture literature (cf. [3-5, 7, 9, 10, 20]). ISO/IEC 42010 defines the stakeholder of a system as "individual, team, organization, or classes thereof, having concerns with respect to a system" [3]. "EA stakeholders are individual or grouped representatives of the organization who are affected by EA products, either by providing input to EA decision making or having to conform to the EA products" [8]. Closely connected to the notion of EA stakeholders is the recognition of their concerns [3, 21] as well as the selection of viewpoints, framing their concerns [3, 7, 9, 10]. A viewpoint is defined as a "work product establishing the conventions for the construction, interpretation and use of architecture views and associated architecture models" [3] and a concern marks an "area of interest in a system pertaining to developmental, technological, business, operational, organizational, political, regulatory, social, or other influences important to one or more of its stakeholders" [3].

Stakeholder Management in EA comprises stakeholder identification, stakeholder classification, communication of architectural information and tailoring of architectural work

products [7]. Stakeholder Management is concerned with managing the relationships between various stakeholders and their concerns [8]. These management activities are supposed to take care of stakeholders impacted by an EA effort as well as these sponsoring an EA effort [8]. Thus, an EA stakeholder is not only an individual or group that is affected by EA (cf. [8]), but also something or someone that can affect an EA effort. This bi-directional relationship conforms to the aforementioned stakeholder definition of Freeman [19].

2.2 The enterprise architecting process

The ISO/IEC 42010 standard defines architecture as "fundamental conception of a system in its environment embodied in elements, their relationships to each other and to the environment, and principles guiding system design and evolution" [3]. This definition generally refers to a system's architecture. More specifically, "enterprise architecture" may be defined as "a coherent whole of principles, methods and models that are used in the design and realization of an enterprise's organizational structure, business processes, information systems, and infrastructure" [22]. An enterprise in this context is any kind of organization or part thereof (e.g., a company or an agency) [23].

In the ISO/IEC 42010 standard, architecting is defined as a "set of interrelated activities of conceiving, defining, describing, documenting, maintaining, improving, and certifying proper implementation of, an architecture throughout a system's lifecycle"[3]. Armour and colleagues describe enterprise architecting as "the process of developing enterprise Information Technology architecture – both its description and its implementation" [24]. Op't Land et al. provide a similar description: "Enterprise architecting is a continuous process involving the creation, modification, enforcement, application, and dissemination of different results. This process should be in sync with developments in the environment of the enterprise as well as developments internal to the enterprise, including both its strategy and its operational processes" [20].

Careful consideration of EA stakeholders and their needs is of critical importance to the success of any EA endeavor [4, 7, 9, 10]. Different analyses highlight challenges occurring during the enterprise architecting process [1, 25, 26]. A recent literature review [1] identifies critical issues, related to stakeholders. Requirements Engineering is a means to understand stakeholders and their needs [27, 28]. The importance of requirements engineering for EA is acknowledged in a number of publications [7, 29] and management of requirements is a central aspect in the Open Group Architecture Framework (TOGAF) architecture development method (ADM) [7], which is a widely adopted process model for enterprise architecting. Lankhorst et al. present a model (cf. Figure 1) that points out the relationship between requirements and architecture models [29], indicating the importance of understanding architecture requirements for the architecting process.

According to them, "the first step is to analyze the problem and elicit goals and requirements that address the problem". A requirements model represents these goals and requirements. Should the baseline enterprise architecture not sufficiently fulfill these requirements, a to-be architecture model has to be conceived in a second step, which defines a composition of products, services, processes and applications fulfilling the defined requirements. "Both steps can again be repeated for (the problem of) realizing the elements of the architecture" [29].

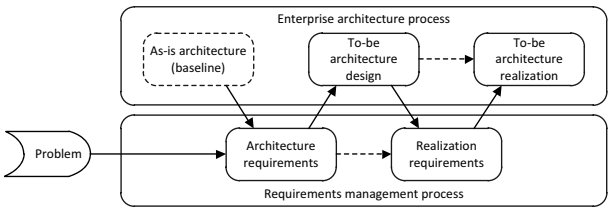


Figure 1: Relation between requirements and architecture models (cf. [29])

Figure 1 visualizes that requirements engineering and the enterprise architecting process are tightly coupled and requirements models play an important role in the architecture design and realization.

2.3 Requirements Engineering and EA

"Requirements engineering is the branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behavior, and to their evolution over time and across software families" [30]. The critical importance of requirements engineering (RE) in software engineering is reflected by the statement of Brooks, who writes: "No other part of the work so cripples the resulting system if done wrong" [31]. RE tasks are requirements elicitation, modeling, analysis, validation & verification and requirements management [27, 28]. More specific areas of research focus on requirements technologies (i.e. notations, methodologies and techniques) to accomplish these tasks (cf. [27]).

Stakeholders are of critical importance in RE as they are the main source of requirements [32, 33]. RE aims to find solutions for stakeholder problems. Requirements elicitation and modeling offers two approaches to describe a solution – a problem-oriented or a solution-oriented view [30, 34]. Problem-oriented RE has its origin in systems engineering, emphasizing the analysis of a problem domain whereas solution-oriented RE represents a classic software engineering view on RE [35]. The problem- and the solution-oriented view are also referred to as early and late RE phases in [36]. A common solution-oriented approach is object-oriented analysis (OOA) [37]. OOA models typically utilize the Unified Modeling Language (UML) to create object models, behavioral models or domain descriptions (cf. [27]). A popular problem-oriented RE approach is goal-oriented requirements engineering (GORE) [11-13]. "Goals capture, at different levels of abstraction, the various objectives the system under consideration should achieve. Goal-oriented requirements engineering is concerned with the use of goals for eliciting, elaborating, structuring, specifying, analyzing, negotiating, documenting, and modifying requirements" [12]. Like problems, goals are closely related to stakeholders. The main GORE approaches are the NFR framework [38], i*[36], KAOS [39] and the Goal-based Requirements Analysis Method (GBRAM) [11, 40].

We argue that goal-oriented modeling is a promising approach to a better understanding and documentation of the motivation for EA undertakings (i.e., the WHY or intentions behind an EA effort). It can be a means to provide a sound reasoning and justification for EA endeavors. Efficient collaboration between

architects and EA stakeholders is seen as one of the main critical success factors for EA [1] and we consider GORE approaches as aid in this respect.

The remainder of this paper explores the impact and importance of goal-oriented requirements modeling on the enterprise architecting process and how increased attention to stakeholder goals and requirements can help to address stakeholder-related problem areas that are predominant in enterprise architecting.

3. RESEARCH METHOD

We develop a model to capture predominant and stakeholder-related enterprise architecting problem areas and relations between them. This model is subject to and aid of our argumentative reasoning on the support of goal-oriented requirements modeling to address the depicted EA issues.

We elaborate on the results of a recently conducted literature review on critical issues in Enterprise Architecting [1]. The research method is a database-driven literature review [41, 42] using the AIS Electronic Library (AISel) and IEEE Xplore. The search was conducted on November 17th, 2009 and double-checked on December 8th, 2009. The two literature databases were chosen as they provide access to a noteworthy number of publications with a high rating in the ranking lists (German IS lists for conference proceedings and journals 2008) published by the IS chapter of the "Gesellschaft für Informatik" [43] – see Table 1 for an aperture. Both databases provide access to journals and conference proceedings. Thus, publications with presumably higher quality (i.e., journal publications according to [42]) as well as content that is more likely up-to-date (i.e., conference proceedings) are covered. Furthermore, AISel and IEEE Xplore provide a good coverage of both scholarly and practice-oriented publications with AISel's focus mainly in scholarly publications and the IEEE Xplore contents being more focused on practice.

Table 1: Journals and conference proceedings accessed by AISel and IEEE Xplore

AISel	IEEE Xplore
Journals	
Information Systems Journal	IEEE Software
Journal of the Association of Information Systems	IEEE Transactions journals
MIS Quarterly	IEEE Computer
Communications of the AIS	IEEE Intelligent Systems
MIS Quarterly Executive	IEEE Internet Computing
	IEEE Pervasive Computing
Conference proceedings	
European Conference on Information Systems	EDOC Conference
Int'l Conference on Information Systems	Hawaii Int'l Conference on System Sciences
Americas Conference on Information Systems	IEEE Conference on E-Commerce Technology
Pacific Asia Conference on Information Systems	IEEE Conference on Enterprise Computing, E-Commerce and E-Services
	IEEE Int'l Conference on Data Mining
	IEEE Security and Privacy

The search term "enterprise architecting" is used, since we consider it a well-accepted term in the EA community (cf. section 2.2). A full text search was conducted for peer-reviewed contents in AISel and a search without any other filters limiting the search request in IEEE Xplore. The AISel search yielded 40 publications dated from 1996 to 2009, with 18 articles dated from 2005 or earlier. The IEEE Xplore search yielded a number of 46 publications dated from 1999 to 2009, with 43 articles dated from 2005 or earlier. The database search yielded a total of 86 publications. 13+2 articles contained just a table of contents (TOC) of proceedings or were duplicate papers – these were not reviewed. The remaining 71 articles were read, identifying 27 referring to EA issues and 44 dealing with other topics. A content analysis approach analogous to grounded theory literature [44] was used.

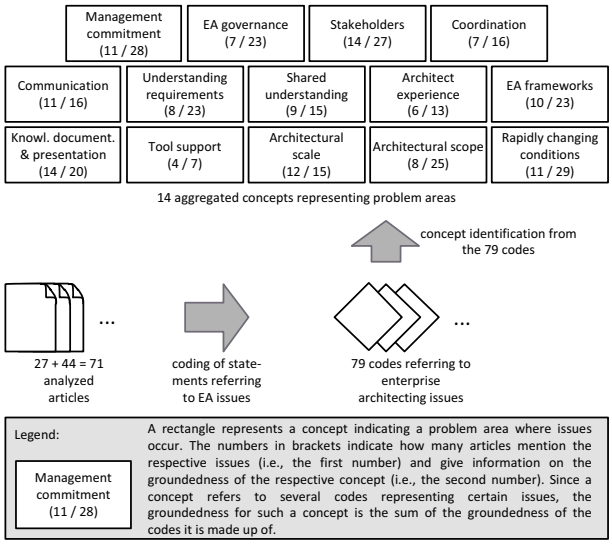


Figure 2: Content analysis approach

Our content analysis approach is depicted in Figure 2. In comparison to [1] the content analysis was completely repeated, utilizing the qualitative data analysis tool Atlas.TI [45] instead of spreadsheets for the detailed analysis of the 27 articles describing enterprise architecting issues. Statements indicating EA issues (i.e., describing an obstacle or gap between a current unsatisfying and a desired more satisfying situation) are identified using open coding in a bottom-up comparative process. Identified issues were considered relevant and assigned one or more codes, when they could clearly be related to a step of the enterprise architecting process. The widely accepted TOGAF ADM [7] was used as reference model for the architecting process. Thus, an issue was considered an architecting issue, when being relatable to a step in the TOGAF ADM. Similar collections of codes were grouped by inductive reasoning to identify underlying concepts. Numbers in brackets behind concept names (cf. Figure 2) denote the number of articles (i.e., literature references) referring to the codes making up a concept and the groundedness (cf. [44]) belonging to the respective concept. The detailed coding and concept identification was conducted by two researchers and discussed with fellow practitioners to reach better intersubjectivity and to agree upon reasonable concepts and categories. The use of the aforementioned qualitative data analysis tool for content analysis instead of spreadsheets and the formulation of concise definitions for these concepts resulted in a refined categorization model compared to [1]. Section 4 will present definitions for the 14

aggregated concepts. Since the 14 problem areas were developed by inductive reasoning, the definitions are developed from coded quotations rather than a theoretical body of knowledge.

We develop a model that depicts those EA problem areas (i.e., concepts) we consider closely related to a weak stakeholder orientation in EA. The relationships between problem areas depicted in our model are explained by causal connections that were identified during the content analysis of reviewed articles. We use this model as an aid for our argumentative reasoning on the expected positive effects of the adoption of goal-oriented requirements modeling in EA. Understandably; this reasoning might to a certain degree be based on the discretion of the authors of the paper at hand.

4. CRITICAL ISSUES IN ENTERPRISE ARCHITECTING

In this section we describe the 14 concepts derived from coded quotations in literature analysis. Each of these concepts represents a focus area of issues occurring during the enterprise architecting process. We use the terms concept and problem area synonymously. Note that these concepts are not in all cases disjunctive or on the same level of abstraction. E.g., *Architect experience* (cf. section 4.8) might for instance be seen as a subset of the *Stakeholders* problem area (cf. section 4.3) because the enterprise architect is an EA stakeholder. The subsequent subsections will provide inductively derived definitions and brief examples of the 14 concepts. More detailed examples and further references can be found in [1].

4.1 Management commitment

This concept is defined as the lack of support for an EA effort from management representatives who are in charge of monetary and organizational resources. Findings are made in two main areas: (1) dimensions and shapes of such insufficiency and (2) influencing factors leading to insufficient management support.

Examples: Lack of meaningful metrics [25] makes it hard to provide justification for EA efforts to management representatives and to develop meaningful value propositions [25, 46, 47]. This is a weakness because return on investment is often expected within a too short amount of time [25, 46-48]. Precipitate expectations for return on investment also seems to result from misunderstanding EA as a project instead of a process [48]. "The reality is that architectural thinking is needed continuously in enterprises because enterprises are 'living things' and in SoS enterprises this need is even greater" [49]. Armour et al. describe that they have seen EA efforts succeed or fail on the basis of this issue (i.e., lack of senior management commitment). "Architecture building often crosses organizational boundaries. The team must be able to capture the information they need. In a large, distributed enterprise, this is a tall order. Your team will need cooperation on many levels, which means they need a strong champion. If the enterprise's senior management doesn't support the effort, don't start it" [48].

4.2 EA governance

This concept stands for lack of authoritative steering, control and process operationalization of EA endeavors. The problem is twofold. Core aspects are the lack of a clearly defined EA process with uncertain goals and the less than optimal organizational structures enforcing EA governance rules.

Examples: Lam [46] describes a lack of governance structures in many EA projects. This is caused by insufficiently defined roles, responsibilities, processes and procedures. There is a need for EA governance "because architectural decisions must be made, coordinated and overseen on several interrelated levels" [50]. Often there exists no common agreement on principles or guidelines for the EA development process [51-54] and although EA frameworks try to address this issue the EA approach is often not rigid enough. Another reason why EA projects sometimes fail is because they do not focus on the right objectives [46, 55] – "one has to first define the key objectives and this would require the inputs of the top management for both, IT and business" [55].

4.3 Stakeholders

The "Stakeholders" concept focuses on the fact that in enterprise architecting there are a plethora of stakeholders that affect or are affected by EA. Dimensions of this problem are identification and classification of stakeholders, management of relevant perspectives suitable to stakeholder needs, a distributed decision making process and connected to this, involvement of relevant stakeholders.

Examples: The plethora of stakeholders is an issue mentioned by several authors [49, 56, 57]. It leads to a number of related challenges like incomplete stakeholder involvement or buy-in [46, 48-50, 54, 55]. Missing relevant stakeholders may lead to the undermining of stakeholder consensus [48]. The large number of stakeholders results in different or even conflicting stakeholder needs and perspectives [4, 25, 46, 47, 53, 55]. A further stakeholder-related issues is distributed decision making [58]. Decision-makers may make local design decisions where they should have incorporated other stakeholders [25, 54, 58, 59].

4.4 Coordination

The concept of "Coordination" describes the challenge of coordinating all parties involved in an EA endeavor, which are highly interdependent due to the multi-layered and multi-faceted nature of EA (cf. for example the rows and columns in the Zachman framework [60]). According to our findings, coordination is needed between people, projects and IT systems. Request for coordination is raised between activities, decisions and roles of people as well as budgets, decisions, priorities and schedules of projects or IT systems in a company or organization. Mediating variables in terms of intensity of coordination issues are time and geographical separation.

Examples: Since enterprise architecting often involves multiple organizational units or even whole branches of an organization, coordination is a major issue [50, 55, 59, 61]. Coordination is directly influenced by two important boundaries: (a) geographic distance and separation and (b) time separation [51, 55, 57, 61]. "[…] systems management is essential in creating timelines for developing component systems and synchronizing them in order to ensure interoperability in a timely manner […] challenge is to balance schedules, while also considering appropriate development lifecycles, risks, configurations, and budgetary issues" [59].

4.5 Communication

The "Communication" concept is concerned with the exchange of information between the different stakeholders in an EA undertaking and the fact of ineffective or mismatched

communication. The establishment of effective communication mechanisms is a central aspect of this problem area.

Example: In EA diverse groups of interest have to avoid mismatched communication in collaboration [46, 50, 54, 55, 62]. "Although each group depends on each other, their levels of specialization have led to group specific languages that thwart effective communication" [58]. "In one large organization […] different groups were running the EITA development effort […], and the groups did not talk to each other. This is one way to guarantee that the target architecture will be out of sync with any new business requirements from the start" [48].

4.6 Understanding requirements

This concept describes the problem of an insufficient understanding of a business's requirements. Important aspects are disregard of the EA vision and business requirements being ignored or misunderstood because of insufficient domain knowledge at the side of the EA team creating architectural descriptions.

Example: Builders and users of architectural descriptions are frequently not the same people. "This division complicates the process of understanding what the application requirements are" [58]. In a similar way, disregard of the EA vision and objectives is an issue, because "you may develop a great architecture for the wrong business" [51]. Further issues described in literature are a lacking understanding of business requirements [46, 51, 62] and ignoring the business requirements [48, 50, 51, 62, 63].

4.7 Shared understanding

"Shared understanding" is a concept which indicates that different stakeholders perform actions and make decisions at a differing level of awareness of the interrelationships of architectural elements. These stakeholders act with a cleft awareness of EA goals, visions and requirements. Often they also share no common vocabulary and have no perception of the implications of their actions.

Example: Literature shows that the EA process suffers from the lack of a shared vision [46, 48] and a shared/common vocabulary [48, 51, 64]. A related issue is the lack of a distributed cognition [25, 49, 50, 58]. "Individual project managers may understand the impact of such changes on local platforms, but often do not understand the impact of changes on other platforms" [25]. Dreyfus refers to this as "local optimization with global ramifications", where these global ramifications are badly understood. Thus, decision-makers in the EA process often operate with imperfect knowledge and understanding [50, 58, 59].

4.8 Architect experience

This concept describes a lack of experienced architects. Enterprise architects are either insufficiently educated or inexperienced – skilled and experienced architects are considered a scarce resource. The complexity and steep learning curve of EA (e.g., EA frameworks) acts as a mediating factor.

Example: A serious issue is the lack of experienced enterprise architects [46, 50, 51] – "competent architects are on high demand" [50]. The field of EA is very complex [46, 51, 58] and so are the EA frameworks that are adopted [65]. Thus, the learning curve in the EA context is very steep – a "critical problem for EA implementation is the short timeframe for learning and getting acquainted with the frameworks and governance model" [50]. Skilled architects are an insufficient resource [48].

4.9 EA frameworks

The "EA frameworks" concept is defined as the challenge of selection, utilization and applicability of enterprise architecture frameworks (EAF).

Example: "The efforts to characterize enterprises in general has led to a plethora of enterprise architecture frameworks" [49], which complicates the selection of an appropriate framework [51, 66]. Furthermore, several shortcomings of EAF are identified. EAF are often overly complex [49, 50] and provide no sufficiently described process for generating the postulated products [51, 67, 68]. Moreover stakeholder-related and a number of other concerns in EAF are bemoaned to be too abstract [4]. EAF are often not capable of taking organizational concerns adequately into account [4, 50, 54, 68, 69]. Literature also shows that there is a disagreement on essential EA layers and segments [49, 51, 54, 56]. EAF adaptability is another key challenge "to make sure the framework guides overall architectural design but is still broad enough to withstand all the modifications from different groups within the enterprise who will need more specific support" [51].

4.10 Knowledge documentation & presentation

This problem area is concerned with the capture, interpretation, representation, prioritization and presentation of architectural information and knowledge. An important part of representation is the question about which notations and modeling techniques should be adopted. Presentation is to the main extent concerned with the communication of architectural knowledge, typically from the architectural team towards users of EA.

Example: A serious knowledge management related EA issue is poor documentation [46]. Architecture rationale is often poorly documented, making it difficult to track "what decisions were made and why" [48]. "There is no single repository (human or otherwise) containing knowledge of the purpose, functionality, or implementation detail of all the applications and their interdependencies [58]". Documenting and retrieving architectural knowledge is far from ideal conditions [46, 48, 54, 57, 70, 71]. The absence of commonly understandable representations, which facilitate cross-group discussions, limits the ability to achieve a well-aligned and agreed architecture [56, 58].

4.11 Tool support

The concept of "tool support" describes issues in the offering of EA tools. This is a twofold problem. First, tool features are described as insufficient in meeting the requirements of practitioners. Second, the standardization of the tool landscape is considered inadequate, leading to ambiguity in documentation of EA information.

Example: A general issue described in literature is unsatisfying tool support [25, 54, 57, 70]. "There is minimum tool support to track and maintain this diverse collection of entities" like strategic goals & objectives on different hierarchy levels, stakeholders, business process descriptions, applications, data and so on [25]. Additionally, the multitude of available tools is described as an issue. "People use different tools to produce different models, resulting in an ambiguous documentation of the architecture" [54].

4.12 Architectural scale

"Architectural scale" is comprised of two aspects which cause a major problem of integration. The first aspect which defines the problem area of "Architectural scale" is the typically large scale of the organization to be modeled, often having an application landscape of hundreds to thousands of applications. The second aspect is that a system of interest is modeled from a number of different perspectives which are highly interdependent and thus need to be integrated to allow for a traceability of elements from one perspective to another.

Example: Often the immense complexity of EA endeavors is underestimated [46, 47, 49, 53, 56-59, 62, 70] as it "applies to very large-scale, complex open systems which are technologically enabled and have extensive social implications" [49]. The large number of applications in today's organizations [58, 72] and the dependencies that exist between the different layers and segments described in architectural descriptions [25, 54-56] are resulting in the challenge to maintain inter-view consistency [51, 73] and traceability [51, 54, 74].

4.13 Architectural scope

The concept of "Architectural scope" is centered on the challenge to determine what is in and what is out of EA (i.e., the determination of architectural boundaries). This concept is comprised of two challenges. First, the commitment on activities to be undertaken and second, the determination of a suitable information model (i.e., what information needs to be captured and modeled).

Example: The scope of architectural descriptions (ADs) has to be defined in breadth and depth [25, 47, 49, 56, 59, 70]. Insufficient scoping efforts can lead to overscoping [47, 48, 51] and/or overmodeling [48, 49, 56]. Overscoping means to choose a too broad scope – "when architects are at high levels, they see more things – and everything they see they model" [48]. Overmodeling refers to the "overuse of detail" [48] in architectural descriptions. Not knowing the scope of the architecting effort may lead to "analysis paralysis" – the architect gets "caught in a never-ending series of analyses" [48]. A related issue is the scheduling of architectural activities. "The team's morale suffers if you don't show results early on. Set schedules such that deliverables arrive within weeks, not months" [51].

4.14 Rapidly changing conditions

This problem area is best described as engineering under uncertainty due to changing conditions. Changes might be triggered either reactively or providently. These changes occur predominantly in the IT landscape caused by different lifecycle phases of systems or applications. The other main source of uncertainty is changing stakeholder objectives and needs. It is a problem of keeping track with operational changes.

Example: Rapidly changing conditions in technology and business are an important issue in EA [51, 54]. "It's impossible to specify an enterprise-wide architecture in a single effort. Technology and business conditions change so rapidly that the architecture would be out of date before it's complete" [51]. Architects have to deal with dynamics and constraints that are caused by different (and shortened) lifecycle phases of systems and applications [25, 46, 47, 58, 59, 70]. There is "a tension between the continuing operations and the introduction of enhanced or new systems" [25].

5. A MODEL OF STAKEHOLDER ORIENTATION IN EA

The previous sections define and describe the critical issues of enterprise architecting. We consider 5 of these 14 issues to be caused to a large extent by weak stakeholder orientation in EA:

- Stakeholders;
- Understanding requirements;
- Architectural scope;
- Knowledge documentation & presentation;
- Communication.

The conceptual model in Figure 3 depicts these problem areas and the relationships between them.

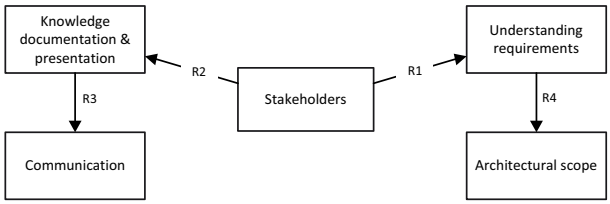


Figure 3: Conceptual model of stakeholder orientation in EA

Subsequently we explain the model. The *Stakeholders* problem area is caused by the plethora of stakeholders relevant in EA endeavors and therefore inherently stakeholder-related. Special dimensions of this problem are stakeholder identification, classification and perspectives suitable to certain stakeholders. According to our findings, the *Stakeholders* problem area is related to two other problem areas: *Understanding requirements* and *Knowledge documentation & presentation*. The problem area of *Understanding requirements* mainly deals with an insufficient understanding of business requirements. Considering EA as a means to Business/IT alignment, business and IT are two very important stakeholder groups involved in the EA process. The large amount of stakeholders involved in EA undertakings, makes it very hard to collect, understand and find a compromise between conflicting requirements (cf. [50, 58, 64]). The problem area of *Understanding requirements* is therefore aggravated by the Stakeholders problem area, which is indicated by relation *R1* (cf. Figure 3).

Knowledge documentation & presentation as a problem area describes issues in capturing and communicating architectural information. Thus, information is captured from information providers and presented and communicated to information users – both being stakeholders involved in the EA process. Since, stakeholders have different, sometimes conflicting needs and perspectives [4, 25, 54], the more stakeholders are involved the more perspectives need to be considered. A stakeholder can have multiple roles connected to different needs, which adds to the problem of the many perspectives [4]. Thus, the *Stakeholders* problem area complicates *Knowledge documentation & presentation* because important stakeholders are easily missed and left unconsidered (cf. relation *R2*).

The *Communication* problem area is to a large extent concerned with ineffective information exchange and communication between stakeholders in the EA process. Mismatched communication among stakeholders, and the lack of tools and

artifacts that can enable improved communication is seen as an important issue [46]. The absence of notations and representations that facilitate cross-group discussions [54, 56, 58, 66, 71] and allow for a stakeholder- and role-specific communication of architectural aspects [56, 59, 71] are seen as an important reason for existing communication issues. Therefore, *Knowledge documentation & presentation* is one reason for *Communication* problems (cf. relation *R3*).

The problem area *Architectural scope* describes difficulties in the proper determination of architectural boundaries. One of the main problems of defining architectural scope is the decision what information is inside and what is outside of an EA effort [56]. This information is typically defined in an information model. A comprehensive understanding of the requirements of an EA effort is seen as important influence factor for successful scoping [51, 59]. Capturing unnecessary information is cost-intensive and may lead to information overload. Therefore only required information should be captured. Thus, *Architectural scope* is influenced by *Understanding requirements* (cf. relation *R4*). Unclear requirements lead to difficulties in scoping of an EA effort.

Summarizing, stakeholder management issues in EA comprise – according to our literature analysis – keeping track of the stakeholders, understanding their requirements, knowledge documentation and presentation, communication and architectural scoping.

6. ADDRESSING STAKEHOLDER ORIENTATION IN EA

In software engineering, goal-orientation is an established way of understanding stakeholders and their objectives in early phases of a project. In the following subsections we explore how goal-orientation is currently implemented in the EA process and how it can help to overcome stakeholder-related problem areas elaborated in section 5.

6.1 Stakeholder orientation in EA

Requirements engineering literature identifies early- and late-phase requirements engineering [27, 28, 30]. Early-phase requirements engineering (i.e., requirements elicitation) comprises activities enabling the understanding of goals, objectives and motives for building a system [27]. Different models are used during elicitation phase (e.g., use-cases, scenarios, goal models) to facilitate early feedback from stakeholders [27]. In late-phase requirements engineering (i.e., requirements modeling) a project specification is expressed in one or more models which compared to requirements elicitation tend to be more precise and unambiguous [27]. Common notations are object models, behavioral models or domain descriptions. Current EA frame–works mainly deal with "structure, function, and behavior, neglecting the intentional dimension of motivations, rationales, and goals" [75]. Their focus is on late-phase requirements engineering.

The ISO/IEC 42010 standard [3] provides a conceptual model for architectural description which many current EA frameworks use as an orientation (e.g., [7, 76, 77]). Architecture rationale is an element in this conceptual model and is defined as "the reasoning about the architecture decisions made" [3]. "The rationale for a decision can include the basis for a decision, the alternatives and trade-offs that were considered, the potential impact of the decision including its pros and cons on other decisions, and citations to further sources of information" [3]. Yu et al. argue that generally architectural descriptions predominantly identify non-intentional elements [75]. Architecture rationale, if provided, therefore typically relates to these non-intentional elements. The documentation of WHY knowledge (i.e., the intentions and motivations behind an EA endeavor) is poorly supported by current EA frameworks [35, 75]. Therefore, this knowledge is typically "embedded in documents, meeting minutes, or held in the minds of individuals involved. Intentional knowledge is therefore often implicit, hard to get at, not systematically managed, and easily lost." [75]. Goal-oriented requirements models can add to the architecture rationale by providing information on the motivation WHY an architectural description is created in the first place [75] and help to better understand a given problem.

So far, early-phase requirements engineering or more specifically goal-oriented requirements engineering has not been widely adopted by current EA frameworks and practitioners on a broad basis. However, a few publications exist, which indicate that goal-orientation can provide important benefits for EA [35, 75]. Yu et al. assess the potentials of the Business Motivation Model (BMM) [78] and i* [36, 79] intentional modeling languages in the context of EA. They identify three main strengths of integrating intentional modeling (IM) with EA: (1) IM can introduce rationality to the EA construction process and justify decision making; (2) IM can provide traceability between high-level business objectives and low-level EA elements; (3) IM stimulates explicit, structured thinking about business transformation and the underlying drivers [75]. The ARMOR language [35] is another recently presented approach which represents an adoption of some traditional GORE approaches to the field of EA, allowing the modeling of goals and requirements in EA. "The ARMOR language is based on existing requirements modeling languages and is aligned with the standard enterprise modeling language ArchiMate" [35]. It supports the goal concept and further concepts like goal refinement, goal conflicts or assumptions (i.e., a refinement of some goal being based on certain assumptions). Applications of ARMOR are the traceability of stakeholder concerns; evaluation of alternative architectures; or the detection of conflicting interests and solutions.

6.2 Goal-oriented requirements modeling to address stakeholder-related EA issues

We see an increased stakeholder orientation as a way to address the stakeholder-related problem areas (cf. section 5). Goal-oriented modeling approaches allow to capture functional as well as non-functional requirements [38] by providing descriptions of stakeholders and their goals. A goal under responsibility of a stakeholder becomes a requirement [12].

The i* notation [36, 79] offers a so-called Strategic Dependency (SD) model, which describes the dependencies among stakeholders (i.e., actors in the i* context) in a given context and provides information on the type of relationship between these stakeholders (e.g., a dependency or a task relationship). Dependencies may involve goals for example and provide criteria for its fulfillment. Thus, i* emphasizes the WHY that underlies a system's requirements [35]. Goal-oriented modeling helps to (a) depict who the stakeholders are, (b) understand the relationships between stakeholders, (c) depict what functional and non-functional requirements exist and (d) understand which stakeholders have those requirements. Therefore we deem these

goal-oriented approaches as a possible means to address the problem areas of *Stakeholders* and *Understanding requirements*.

Typically, a goal-oriented modeling approach comprises a specific notation and provides one or more different model types to capture information about requirements ([11, 36, 38-40]). Thus, these approaches add to the problem space of *Knowledge documentation & presentation* by providing models, which offer help to foster communication about stakeholders and their goals.

Note that some of these notations are formal and it might be difficult especially for business stakeholders to understand them [75]. Nevertheless they provide a proven way of representing information about stakeholders and associated goals. Therefore, we consider them as a means to address the problem area of *Communication*. First, by documenting the information to be shared and second, by disclosing who should communicate with whom in the case of common or conflicting goals or similar.

As mentioned, a main issue in terms of architectural scoping is "clearly identifying what is in, and what is outside, the enterprise" [56]. We see goal-oriented modeling approaches as a solution to this problem (i.e., *Architectural scope*) since they clearly identify functional and non-functional requirements by documenting stakeholders and their goals [12, 38]. Aspects and information that cannot be related to a goal or requirement should not be part of the information model of an EA effort. Thereby, goal-modeling helps to restrict the information model of an EA to the necessary elements. This will help to define a reasonable architectural scope.

7. CONCLUSION AND OUTLOOK

In this paper we elaborate on the results of a recent literature analysis [1] and argue that a number of problem areas identified are to be seen in relationship with stakeholders of EA efforts. We introduce a model to capture these problem areas and the relationships between them and elaborate that an increased stakeholder orientation is a means to address these issues. We provide argumentative reasoning, why we consider the goal-oriented requirements modeling approach adopted in software engineering as a means to address these issues by fostering an increased stakeholder focus. Additionally, we mention two contributions [35, 75] which explore the integration of intentional modeling in EA. These articles show goal-modeling as a promising approach to a better understanding of stakeholders and their objectives in EA. A number of benefits have been mentioned (cf. section 6.1).

Therefore, our message is that EA research should add more emphasis to the stakeholder topic by adopting intentional modeling. Our introduction shows that recently the interest in the stakeholder topic has begun to increase. The popular TOGAF framework can be seen as an example for this trend. The 2007 version 8.1.1 [80] already described requirements management as a central aspect in its ADM but had no chapter dedicated to stakeholder management, whereas the 2009 version 9 [7] offers such a chapter providing guidance on stakeholder management. This latest TOGAF version mentions important tasks like stakeholder identification and classification and provides a template stakeholder map (i.e., a plain table) that helps to identify and capture stakeholder concerns as well as associated viewpoints framing those concerns. However, no modeling techniques or notations to document stakeholders and their goals and requirements are presented in this latest TOGAF version. Scholarly literature argues, that goal-oriented requirements modeling contributes further value to stakeholder management by documenting stakeholders, their relationships and their specific requirements [35, 75]. These approaches model high-level goals in early phases of an EA undertaking in terms of stakeholders and their concerns [35]. High-level objectives can be related to low-level architecture elements like products, services, processes or applications. This means a synthesis of the current EA approach and its models with the GORE approach adopted in the software engineering field. Further research will be needed to determine how to leverage and integrate these requirements engineering principles and approaches from software engineering into EA and how to connect current EA modeling with intentional modeling.

We conclude that goal-oriented models can be considered a very reasonable addendum to existing modeling approaches in EA. They provide a way to capture the goals and intentions of stakeholders [35, 75] and provide rationale for an EA effort as a whole. As we show in this paper, goal-oriented requirements modeling is furthermore an approach to increase stakeholder orientation in EA and can as such help to overcome a significant number of predominant stakeholder-related enterprise architecting problem areas in EA.

8. REFERENCES

[1] Lucke, C., Krell, S. and Lechner, U. Critical Issues in Enterprise Architecting – A Literature Review. In *Proc. of the 16th Americas Conf. on Information Systems (AMCIS)* (2010), Paper 305.

[2] Schöenherr, M. *Towards a Common Terminology in the Discipline of Enterprise Architecture*. Springer, Berlin, 2009.

[3] Int'l Organization for Standardization. *ISO/IEC 42010 (WD4)*. 2007.

[4] Niemi, E. Enterprise Architecture Stakeholders - a Holistic View. In *Proc. of the 13th Americas Conf. on Information Systems* (2007), Paper 41.

[5] Lankhorst, M. *Enterprise Architecture at Work: Modelling, Communication and Analysis* Springer-Verlag, Berlin, 2009.

[6] Bender, G. *Designing a Stakeholder-Specific Enterprise Architecture Management based on Patterns*. M.Sc. thesis, TU München, 2009.

[7] The Open Group. *TOGAF Version 9 - The Open Group Architecture Framework (TOGAF)*. The Open Group, 2009.

[8] Raadt, B., Schouten, S. and Vliet, H. Stakeholder Perception of Enterprise Architecture. In *Proc. of the 2nd European Conf. on Software Architecture* (2008), 19-34.

[9] Aier, S., Kurpjuweit, S., Riege, C. and Saat, J. Stakeholderorientierte Dokumentation und Analyse der Unternehmensarchitektur. In *Proc. of the INFORMATIK 2008: Beherrschbare Systeme – dank Informatik* (2008), 559-565.

[10] Kurpjuweit, S. *Stakeholder-orientierte Modellierung und Analyse der Unternehmensarchitektur unter besonderer Berücksichtigung der Geschäfts- und IT-Architektur*. Ph.D. thesis, Universität St. Gallen, Hochschule für Wirtschafts-, Rechts- und Sozialwissenschaften (HSG), St. Gallen, 2009.

[11] Anton, A. I. Goal-based requirements analysis. In *Proc. of the 2nd Int'l Conf. on Requirements Engineering* (1996), 136-144.

[12] Lamsweerde, A. v. Goal-Oriented Requirements Engineering: A Guided Tour. In *Proc. of the IEEE Int'l Conf. on Requirements Engineering* (2001), 249-262.

[13] Leveson, N. G. Intent specifications: an approach to building human-centered specifications. In *Proc. of the 3rd Int'l Conf. on Requirements Engineering* (1998), 204-213.

[14] Phillips, R. and Freeman, R. E. *Stakeholder Theory and Organizational Ethics*. Berrett-Koehler Publishers, 2003.

[15] Achterkamp, M. C. and Vos, J. F. J. Investigating the use of the stakeholder notion in project management literature, a meta-analysis. *Int'l Journal of Project Management*, 26, 7 (2008), 749-757.

[16] Pouloudi, A. Aspects of the Stakeholder Concept and their Implications for Information Systems Development. In *Proc. of the 32nd Annual Hawaii Int'l Conf. on System Sciences* (1999), 7030.

[17] Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L. and Colle, S. d. *Stakeholder Theory: The State of the Art*. Cambridge University Press, 2010.

[18] Mitchell, R. K., Agle, B. R. and Wood, D. J. Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review*, 22, 4 (1997), 853-886.

[19] Freeman, R. E. *Strategic Management - A Stakeholder Approach*. Pitman Publishing, Marshfield, MA, 1984.

[20] Op't Land, M., Proper, E., Waage, M., Cloo, J. and Steghuis, C. *Enterprise Architecture: Creating Value by Informed Governance*. Springer, Berlin, 2009.

[21] Buckl, S., Ernst, A., Lankes, J. and Matthes, F. *Enterprise Architecture Management Pattern Catalog (Version 1.0, February 2008)*. Technical Report TB0801, Chair for Informatics 19 (sebis), TU München, 2008.

[22] Lankhorst, M. *Enterprise Architecture at Work. Modelling, Communication and Analysis* Springer-Verlag, Berlin, 2005.

[23] Aier, S., Riege, C. and Winter, R. Unternehmensarchitektur - Literaturüberblick und Stand der Praxis. *Wirtschaftsinformatik*, 50, 4 (2008), 292-304.

[24] Armour, F., Kaisler, S. and Bitner, J. Introduction to Enterprise Architecture: Challenges [Minitrack Introduction]. In *Proc. of the 41st Annual Hawaii Int'l Conf. on System Sciences (HICCS '08)* (2008), 395-395.

[25] Kaisler, S., Armour, F. and Valivullah, M. Enterprise Architecting: Critical Problems. In *Proc. of the 38th Annual Hawaii Int'l Conf. on System Sciences* (2005), 224b-224b.

[26] van der Raadt, B., Soetendal, J., Perdeck, M. and van Vliet, H. Polyphony in architecture. In *Proc. of the 26th Int'l Conf. on Software Engineering* (2004), 533-542.

[27] Cheng, B. H. C. and Atlee, J. M. Research Directions in Requirements Engineering. In *Proc. of the 2007 Future of Software Engineering* (2007), 285-303.

[28] Nuseibeh, B. and Easterbrook, S. Requirements engineering: a roadmap. In *Proc. of the Conf. on The Future of Software Engineering* (2000), 35-46.

[29] Lankhorst, M. and Quartel, D. Architecture-Based IT Valuation - Supporting portfolio management and investment decisions. *VIA NOVA ARCHITECTURA* (2010).

[30] Zave, P. Classification of research efforts in requirements engineering. *ACM Comput. Surv.*, 29, 4 (1997), 315-321.

[31] Brooks, F. P., Jr. No Silver Bullet Essence and Accidents of Software Engineering. *Computer*, 20, 4 (1987), 10-19.

[32] Balzert, H. *Lehrbuch der Softwaretechnik: Basiskonzepte und Requirements Engineering* Spektrum Akademischer Verlag, 2009.

[33] Rupp, C. *Requirements-Engineering und -Management: Professionelle, iterative Anforderungsanalyse für die Praxis*. Hanser Fachbuch, 2009.

[34] Wieringa, R. J. Requirements Engineering: Problem Analysis and Solution Specification (Extended Abstract) In *Proc. of the Lecture Notes in Computer Science (LNCS) - Web Engineering* (2004), 13-16.

[35] Quartel, D., Engelsman, W., Jonkers, H. and van Sinderen, M. A Goal-Oriented Requirements Modelling Language for Enterprise Architecture. In *Proc. of the IEEE Int'l Enterprise Distributed Object Computing Conf.* (2009), 3-13.

[36] Yu, E. S. K. Towards modelling and reasoning support for early-phase requirements engineering. In *Proc. of the 3rd IEEE Int'l Symposium on Requirements Engineering* (1997), 226-235.

[37] Balzert, H. *Lehrbuch der Objektmodellierung: Analyse und Entwurf mit der UML 2*. Spektrum Akademischer Verlag, 2004.

[38] Mylopoulos, J., Chung, L. and Nixon, B. Representing and using nonfunctional requirements: a process-oriented approach. *Software Engineering, IEEE Transactions on*, 18, 6 (1992), 483-497.

[39] Lamsweerde, A. v. and Letier, E. *From Object Orientation to Goal Orientation: A Paradigm Shift for Requirements Engineering*. Springer, Berlin, 2004.

[40] Anton, A. I. *Goal identification and refinement in the specification of software-based information systems*. Ph.D. thesis, Georgia Institute of Technology, 1997.

[41] Webster, J. and Watson, R. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26, 2 (2002), XIII-XXIII.

[42] vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R. and Cleven, A. Reconstructing the giant: On the importance of rigour in documenting the literature search process. In *Proc. of the 17th European Conf. on Information Systems* (2009).

[43] Sprecher der WKWI und GI-FB WI WI-Orientierungslisten - WI-Journalliste 2008 sowie WI-Liste der Konferenzen, Proceedings und Lecture Notes 2008. *WIRTSCHAFTSINFORMATIK*, 50, 2 (2008), 155-163.

[44] Glaser, B. G. and Strauss, A. L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Pub, 1999.

[45] Mayring, P. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz Verlag, Weinheim, Basel, 2008.

[46] Lam, W. Technical Risk Management on Enterprise Integration Projects. *Communications of the Association for Information Systems*, 13(2004), 290-315.

[47] Namba, Y. and Iljima, J. City Planning Approach for Enterprise Information System. In *Proc. of the 8th Pacific Asia Conf. on Information Systems* (2004), Paper 14.

[48] Armour, F., Kaisler, S. and Liu, S. Building an enterprise architecture step by step. *IT Professional*, 1, 4 (1999), 31-39.

[49] Rhodes, D., Ross, A. and Nightingale, D. Architecting the system of systems enterprise: Enabling constructs and methods from the field of engineering systems. In *Proc. of the 3rd Annual IEEE Systems Conf.* (2009), 190-195.

[50] Seppanen, V., Heikkila, J. and Liimatainen, K. Key Issues in EA-Implementation: Case Study of Two Finnish Government Agencies. In *Proc. of the IEEE Conf. on Commerce and Enterprise Computing* (2009), 114-120.

[51] Armour, F., Kaisler, S. and Liu, S. A big-picture look at enterprise architectures. *IT Professional*, 1, 1 (1999), 35-42.

[52] Avgeriou, P., Kruchten, P., Lago, P., Grisham, P. and Perry, D. Architectural knowledge and rationale: issues, trends, challenges. *SIGSOFT Softw. Eng. Notes*, 32, 4 (2007), 41-46.

[53] Delic, K. A., Riley, J. A. and Faihe, Y. Architecting Principles for Self-Managing Enterprise IT Systems. In *Proc. of the 3rd Int'l Conf. on Autonomic and Autonomous Systems* (2007), 60.

[54] Shah, H. and Kourdi, M. E. Frameworks for Enterprise Architecture. *IT Professional*, 9, 5 (2007), 36-41.

[55] Espinosa, J. A. and Boh, W. F. Coordination and Governance in Geographically Distributed Enterprise Architecting: An Empirical Research Design. In *Proc. of the 42nd Annual Hawaii Int'l Conf. on System Sciences (HICCS '09)* (2009).

[56] Armour, F., Kaisler, S., Getter, J. and Pippin, D. A UML-Driven Enterprise Architecture Case Study. In *Proc. of the 36th Annual Hawaii Int'l Conf. on System Sciences (HICSS'03) - Track 3 - Volume 3* (2003).

[57] Henttonen, K. and Matinlassi, M. Open source based tools for sharing and reuse of software architectural knowledge. In *Proc. of the Joint Working IEEE/IFIP Conf. on Software Architecture & Europ. Conf. on Softw. Archit.* (2009), 41-50.

[58] Dreyfus, D. Information System Architecture: Toward a Distributed Cognition Perspective. In *Proc. of the Int'l Conf. on Information Systems* (2007), Paper 131.

[59] Bubak, O. Composing a course book for system and enterprise architecture education. In *Proc. of the IEEE/SMC Int'l Conf. on System of Systems Eng.* (2006), 230-235.

[60] Zachman, J. A framework for information systems architecture. *IBM systems journal*, 26, 3 (1987).

[61] Espinosa, J. A. and Armour, F. Geographically Distributed Enterprise Architecting: Towards a Theoretical Framework. In *Proc. of the 41st Annual Hawaii Int'l Conf. on System Sciences (HICCS '08)* (2008).

[62] Wang, X., Ma, F. and Zhou, X. Aligning Business and IT Using Enterprise Architecture. In *Proc. of the 4th Int'l Conf. on Wireless Communications, Networking and Mobile Computing* (2008), 1-5.

[63] Melchert, F., Winter, R. and Klesse, M. Aligning Process Automation and Business Intelligence to Support Corporate Performance Management. In *Proc. of the Americas Conf. on Information Systems* (2004), Paper 507.

[64] da Cunha, P. and de Figueiredo, A. Quality Management Systems and Information Systems: Getting More than the Sum of the Parts. In *Proc. of the Americas Conf. on Information Systems* (2005), Paper 236.

[65] Schekkerman, J. *How to survive in the jungle of enterprise architecture frameworks: creating or choosing an Enterprise Architecture Framework*. Trafford Publishing, 2004.

[66] Goel, A., Schmidt, H. and Gilbert, D. Towards formalizing Virtual Enterprise Architecture. In *Proc. of the 13th Enterprise Distributed Object Computing Conf. Workshops* (2009), 238-242.

[67] Moghaddam, M. R. S., Sharifi, A. and Merati, E. Using Axiomatic Design in the Process of Enterprise Architecting. In *Proc. of the 3rd Int'l Conf. on Convergence and Hybrid Information Technology* (2008), 279-284.

[68] Wilton, D. R. The relationship between IS strategic planning and enterprise architectural practice: case studies in New Zealand enterprises. In *Proc. of the Pacific Asia Conf. on Information Systems* (2008), Paper 19.

[69] Janssen, M. and Hjort-Madsen, K. Analyzing Enterprise Architecture in National Governments: The Cases of Denmark and the Netherlands. In *Proc. of the 40th Annual Hawaii Int'l Conf. on System Sciences* (2007), 218a-218a.

[70] Meilich, A. System of systems (SoS) engineering & architecture challenges in a net centric environment. In *Proc. of the IEEE/SMC Int'l Conf. on System of Systems Engineering* (2006), 1-5.

[71] Templeton, G. F., Lee, C. and Snyder, C. Validation of a Content Analysis System Using an Iterative Prototyping Approach to Action Research. *Communications of the Association for Information Systems*, 17, 1 (2006), 539-561.

[72] Mocker, M. What Is Complex About 273 Applications? Untangling Application Architecture Complexity in a Case of European Investment Banking. In *Proc. of the 42nd Annual Hawaii Int'l Conf. on System Sciences* (2009), 1-14.

[73] Nordstrom, L. and Cegrell, T. Analyzing utility information system architectures using the common information model. In *Proc. of the CIGRE/IEEE PES Int'l Symposium* (2005), 274-281.

[74] Buuren, R., Gordijn, J. and Janssen, W. Business Case Modelling for E-Services. In *Proc. of the Bled Conf.* (2005), Paper 8.

[75] Yu, E., Strohmaier, M. and Xiaoxue, D. Exploring Intentional Modeling and Analysis for Enterprise Architecture. In *Proc. of the 10th IEEE Int'l Enterprise Distributed Object Computing Conf. Workshops* (2006).

[76] Department of Defense. *Department of Defense Architecture Framework v2, from http://cio-nii.defense.gov/sites/dodaf20/*.

[77] North Atlantic Council. *NATO Architecture Framework (NAF) Ver 3*. North Atlantic Council, 2007.

[78] Business Rules Group. *The Business Motivation Model – Business Governance in a Volatile World. Version 1.3 from http://www.businessrulesgroup.org/bmm.shtml*.

[79] i* Wiki. *http://istar.rwth-aachen.de/*.

[80] The Open Group. *TOGAF Version 8.1.1 - The Open Group Architecture Framework (TOGAF)*. The Open Group, 2007.

A Longitudinal, Multi-Project Study of Bug Tracking Productivity and Learning in Open Source Software Development

Janina Matz
Darmstadt University of Technology
Chair of Information Systems
Hochschulstr. 1,
64289 Darmstadt, Germany
matz@is.tu-darmstadt.de

Arne Beckhaus
SAP Research
Bleichstr. 8,
64283 Darmstadt, Germany
arne.beckhaus@sap.com

Dierk Erdmann
Darmstadt University of Technology
Chair of Information Systems
Hochschulstr. 1,
64289 Darmstadt, Germany
erdmann@is.tu-darmstadt.de

Peter Buxmann
Darmstadt University of Technology
Chair of Information Systems
Hochschulstr. 1,
64289 Darmstadt, Germany
buxmann@is.tu-darmstadt.de

ABSTRACT

The dynamics of Open Source Software development have recently received a lot of attention from an organizational learning perspective. Following a network theoretic approach, we study the temporal development of communication network structure and productivity in order to detect associations among these constructs. Thereby, we identify a research gap in that related work either focuses on too few projects or utilizes insufficient numbers of analyzed timeframes. Our study is both multi-project and longitudinal in order to detect holistic influencing factors of successful Open Source Software development. First, we find that learning effects are present since productivity increases over time. Next, we observe that growth in team size impedes productivity whereas the continuous concentration on central nodes coincides with increasing productivity. Against our expectation, we also find that increasing centralization does not yield decreasing network density, possibly due to Open Source Software developers deliberately avoiding the dependence on bottleneck nodes.

Keywords

Open Source Software, software development, organizational learning, productivity, network analysis

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

1. INTRODUCTION

Research on Open Source Software communities has flourished as a consequence of scholars pointing to the vast potential of information that is coded in these platforms, (e.g., [57]). Thereby, various research streams can be differentiated. Researchers have been interested in questions regarding participation in Open Source Software communities and the motivational incentives for contributions therein [6, 30, 50, 62]. Closely aligned with this view is the question of how new members are attracted to a community [29]. It frequently takes a network perspective that is also addressed in the research streams of socialization [23] and social networks [8, 65]. In addition to the different explanatory goals, studies on Open Source Software are also conducted with differing focus. While a large stream of research rather takes a focused view by observing the bug fixing process [17, 19, 66], other works take a market perspective by analyzing the competition with proprietary software [24, 32, 52].

Within this broad field of research on Open Source Software development, this study's objective is to explain productivity from a network theoretic perspective. Thereby, our research intents to detect temporal effects in order to address the dynamics of many Open Source Software projects. It is hence in line with related work on organizational learning in the field of software engineering in general [11, 47] and Open Source Software development in particular [31]. As will be detailed in Section 4, we see a research gap in the design of this analysis. Related work has primarily chosen an unbalanced relation between the number of Open Source projects studied and the amount of temporal data available therein. We address this research gap by conducting a study that is both multi-project and longitudinal. With this research design, we intent to answer the research question whether

network theoretic constructs and productivity show corresponding temporal developments. In other words, we analyze, e.g., whether increasing communication network centrality coincides with increasing productivity. Due to studying these associations for different projects at the same time, we are able to holistically identify the dynamics of successful Open Source Software development.

The remainder of this paper is structured as follows. In Section 2, we present the theoretical foundation of our work. From this foundation, research hypotheses are derived in Section 3. Section 4 introduces our research design with a special focus on differentiation from related work. Section 5 explains the data source analyzed in this study. Suitable variables for the measurement of the postulated effects are presented in Section 6 and analyzed in Section 7. Potential implications of our results are discussed in the conclusions in Section 8.

2. THEORETICAL FOUNDATION

Research on Open Source Software communities is frequently based on two alternative theories. Crowston's coordination theory [16, 38, 39] is a suitable foundation for studies that analyze process flows. It builds on a rather strict formalization of the underlying process and hence does not optimally reflect the 'creative chaos' types of coordination in communities. This view is rather supported by the second frequently chosen theoretical basis being Granovetter's embeddedness theory [25–27]. Studies on Open Source Software communities that take this network embeddedness perspective usually analyze network structure and its association with performance (e.g., [28]). Our work follows this research stream and uses network embeddedness as its theoretical foundation. We hence believe that informal communication network structure is more important than formal coordination processes. This view is in line with theories of self-coordinating virtual communities that have appeared in various research disciplines such as virtual online groups [49], knowledge management [60], and information systems [40].

In addition to network embeddedness being the theoretical foundation of this article, it also significantly draws from organizational learning literature. This can primarily be attributed to the longitudinal research design that facilitates the analysis of learning effects over time. While researchers have stated that the term 'learning' represents different concepts in different research streams [61], organizational learning on an abstract level represents the need for firms and their employees to continuously learn and transfer knowledge [4]. Learning effects have consequently been studied and identified in various industries including software engineering (e.g., [11, 47]). Thereby, Open Source Software development has been said to be "a particularly rich environment for studying organizational learning" [31, p. 485].

3. RESEARCH HYPOTHESES

It is common to many studies in the field of information systems research to analyze variances in order to map input factors to some kind of output measure. The definition of output is thereby discussed controversially. While for information systems success from a customer perspective, DeLone and McLean's [21] model has become the de-facto standard, software development is an information work process whose productivity is difficult to define [3, 22]. Despite the challenges to define a productivity centric outcome measure, many studies in related work use this factor as a dependent variable. In the context of organizational learning, it plays an even more crucial role. The assumption that productivity increases over time is the essence of organizational learning theory (cf. [5]). Hence, we hypothesize that in Open Source Software projects, learning effects are present and that productivity and time are positively associated.

Hypothesis 1. *Productivity increases over time.*

Studies on organizational learning often only consider the phenomenon from a productivity and time perspective [5]. However, many other factors are of interest and can reveal interesting learning effects. The remaining hypotheses thereby do not directly focus the time construct but rather discuss the associations of other pairs of factors that are later on related to their temporal dynamics.

Mixed effects have been reported to exist in the association of productivity and team size. According to [5], this can be attributed to the opposed effects of large teams' coordination overheads and small teams' resource utilization challenges. The debate goes back to Brook's [14] seminal writing on the mythical man-month, in which he argues why doubling the man-power of a project might not increase but even decrease productivity due to the quadratical rise in potential communicative links. This effect has been found to also exist in the context of Open Source Software development [1].

From an economic perspective, the association of productivity and team size can be related to the debate on economies of scale in software engineering. Economies of scale are defined to exist in scenarios of increasing average productivity with rising volume [53] and have been frequently analyzed in the application context of software engineering (e.g., [7, 34, 45]). However, these studies lack a consensus on findings with studies reporting the full range of positive, negative, and not existing associations of productivity and size [34, 45]. As already outlined in the previous section on the research design, many of these studies conduct a cross-sectional analysis of size and productivity levels across projects. The advantage of our study's approach to analyze multiple Open Source Software projects longitudinally is also evident in the analysis of economies of scale. In accordance with our theoretical foundation on network embeddedness theory, we postulate that the communication overhead of large teams rather impedes productivity.

Hypothesis 2. *Productivity and team size are negatively associated.*

The analysis of network structure and especially centrality has gained popularity over the last decades in information systems research and the social sciences [12, 44]. Network centrality has also been identified to be an influential concept in the research domain of Open Source Software development [8, 37]. In fact, the onion-like, core/periphery communication structures of a small, highly central group of developers and a large, decentral group of users has been

named an identifying characteristic of the Open Source Software development model [20, 48, 51]. While many studies find centrality to be beneficial for performance (e.g., [8]), partly due to efficiency in information flow, others report examples where too much centrality that ultimately takes the form of separation also impedes performance in Open Source Software development [17]. Due to our study's focus on individual project, we hypothesize that centrality is beneficial for development productivity since it resembles the Open Source Software movements core/periphery characteristic. If the focus would have been on larger collaboration networks of active developers (as is the case in many studies analyzing large parts of the Sourceforge.net community, e.g., [29,55]), the opposite effect might have been more predominant.

Hypothesis 3. *Productivity and communication network centrality are positively associated.*

The fourth research hypothesis regards typical effects of network theory and consists of two parts. At first, we postulate that communication becomes more difficult when many nodes are present in the communication network. As Wu and Goh state, "a higher project density makes the dissemination of knowledge more time-consuming, as information and knowledge needs to travel through the extended hierarchies of the project team."[63, p. 4] Since density reflects the number of present pairwise communicative links in relation to all theoretically possible links in a network of the same size, increases in size require quadratic increases in new communication links in order for constant density. Hence, we hypothesize that communication network density and team size are negatively associated as it is very difficult for large teams to be as connected as small ones.

Hypothesis 4a. *Team size and communication network density are negatively associated.*

The second part of Hypothesis 4 also regards network density. We hypothesize that networks with high centrality require less communication links among peers due to the existence of key players. Information is efficiently routed over these central nodes and information paths become shorter [8]. As a consequence, we assume that network density can be low for these highly central networks since less communicative links are required when a lot of pairwise communication is routed over central nodes.

Hypothesis 4b. *Communication network density and centrality are negatively associated.*

4. RESEARCH DESIGN

The research designs of related work primarily differ in two dimensions. First, different data sources are chosen. Some studies make use of surveys among developers (e.g., [36]). Data acquisition by means of questionnaires has the advantage to retrieve diverse information ranging from age, gender, cultural background to experience. It is, however, also very limited in size. The alternative data acquisition strategy is to access existing archival data sources (e.g., [43]). Software engineering is a digitized process with large amounts of information available in databases. While less information can be retrieved per individual in comparison with surveys, archival data sources have the advantage that they are nearly unlimited in size and structural properties of networks are often only visible at this scale [35]. We hence conduct our research on the basis of archival data such as bug tracking systems and mailing list archives.

The second important dimension of research design regards the temporal nature of the study. Many works analyze static phenomena in cross-sectional studies (e.g., [62]). While this design has the advantage to thoroughly acquire data and measure effects, it neglects the dynamics of the fast moving software engineering domain. Hence, a second stream of research conducts longitudinal studies that explicitly analyze temporal development of effects (e.g., [50]). We follow this stream and analyze the dynamics of Open Source Software development teams.

Several studies in related work can be identified that share our research design. However, our work differs from these studies in the implementation of the longitudinal, multi-project study. Some studies shift the focus towards rich information regarding the time factor. However, this comes at the expense of analyzing a reasonably large number of different projects. The works of [43, 50, 56] fall into this category. Other studies shift their focus in the opposite direction and analyze many projects at the expense of very limited data on timing. Often, factors are measured for a large number of projects at two points in time and a delta-study of these two measurements is conducted in order to reveal temporal effects (e.g., [54]). Our research design aims at providing a sufficient number of measurements for both projects and time slots. This well-balanced selection of projects and time slots comes at the expense of applicable research methodology. Just like the analysis of [18], our sample does not allow for standard time-series analysis. As will be elaborated on in Section 7, it is, however, suitable for correlation analysis and can reveal other and potentially more interesting findings than one of the two other research designs discussed above.

5. DATA SOURCE

For our analysis of communication patterns and Open Source Software project's productivity on a reasonable number of projects over time, there are several constraints on data acquisition. The biggest challenge is to identify projects that fulfill the criteria of consistent data over time. The analysis of communication network structure requires rather big projects in order to be confronted with interesting network effects that can reasonably be analyzed by means of standard social network analysis techniques. While large parts of related work on Open Source Software development use a large number of Sourceforge.net projects, we could not solely rely on this vast data source since it does not host very many big projects that are suited to our analysis of communication structures over time. We hence aggregated candidates from various Open Source Software directories and manually identified projects that provide sufficient amounts of data. Thereby, automated approaches such as that of [2] are not feasible in our setup since not all information is publically available. Instead, in some cases, we need to contact the projects' administrators and ask them for mailing

list archives. We do not believe that our data acquisition strategy introduces a serious bias since small Open Source projects, which were not covered by our study, do not often represent truly collaborative and already successful projects.

Thereby, we use the archived mailing lists of eleven different Open Source projects for analyzing communication patterns and structures. For a given period of time, a network graph is constructed, where each node represents a person. The graph's edges linking the persons represent the e-mails that have been exchanged. The edges' weights indicate the number of e-mails that have been exchanged between the corresponding two participants. Formally, a graph or network is given by $G = (\mathcal{V}_G, \mathcal{E}_G)$, where $\mathcal{V}_G$ is the set of vertices and $\mathcal{E}_G$ is the set of edges that constitute the graph. We do not consider directed networks that show the direction of the sent e-mails, as this is not relevant to our analysis.

For measuring productivity, we use the Open Source Software projects' bug tracking databases, which is in line with related work [33]. In the case of the analyzed projects, these additional data sources are freely accessible. They represent a type of workflow management system and help the developers to organize the bug fixing process. For each listed bug, the system keeps track of several attributes, e.g.,

- the initial reporting date,
- a description of the bug,
- its current status (e.g., 'fixed', 'open', ...), and
- the dates of subsequent actions, such as the resolution of the bug.

Using these attributes, we can define a measure of productivity as given in Section 6.1. Of course, the software development process and its productivity consists of much more than bug tracking. However, as elaborated on in the following, productivity of software engineering is very difficult to measure. We hence follow the argument of [17] that bug tracking is the one sub-process of software engineering that involves everybody of the community. Quality management, which bug tracking belongs to, is also a very important aspect of the entire software engineering process from a financial point of view that accounts for large parts of total software budgets [42].

6. VARIABLE MEASUREMENT

6.1 Productivity

While in traditional research settings, productivity has often been defined as 'output per input', in knowledge work, a definition of productivity as a combination of effectiveness and efficiency is becoming more common [46]. Effectiveness means 'doing the right things' and is in our setting of less relevance than efficiency ('doing things the right way') since software engineering tasks are rather fixed and well defined. Efficiency is the more interesting factor here as it can be seen as the main outcome measure influencing productivity.

An important argument for the higher importance of output than input in knowledge work productivity is data availability and measurement. While in conveyor-belt-type of work, input can be easily measured via operation hours, knowledge work is much more complex and interweaved over tasks. It is hence hardly feasible to precisely measure input for a given task. A knowledge worker usually executes diverse tasks in parallel and does not keep time on a fine grained level, such as minutes or even seconds. This is one of the reasons for the dominance of efficiency-based over output/input-based productivity definitions in knowledge work contexts.

In line with this knowledge work perspective on productivity, we follow Kidane and Gloor's [33] productivity definition in a software engineering and bug tracking context for a given period of time as

$$\text{productivity} = \frac{\text{Number of bugs reported } \textit{and} \text{ fixed}}{\text{Number of reported bugs}}.$$

This definition deviates from [33, p. 20] in that we only count bugs that have been reported *and* fixed in the same period of time. This is due to a correction of an apparent data inconsistency in the analyzed Open Source Software projects' bug tracking databases. These list large amounts of bugs as fixed on the exact same point of time with an accuracy of a second. Obviously, the "date of last change" does not represent the true bug resolution time in these cases but is rather the result of an automatic data correction procedure. However, the initial reporting date of the bugs did not exhibit any inconsistencies. If we included all these bugs with wrong timestamps, the observed productivity values would be much too high for certain periods of time. Our way of measuring productivity avoids this problem and is more robust against bug database entries with erroneous date entries.

6.2 Network Centralization

The second measure of network topology is centralization. We use betweenness centrality for measuring how centralized the communication networks are. While there exist other concepts of measuring centrality, the betweenness centrality is one of the most commonly used ones [9, 13, 59]. The betweenness centrality for a *single node* v is defined as

$$\mathrm{BC}(v) = \underbrace{\left(\sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \right)}_{\text{centrality}} \cdot \underbrace{\left(\frac{2}{(|\mathcal{V}_G| - 1)(|\mathcal{V}_G| - 2)} \right)}_{\text{normalizing factor}}, \quad (1)$$

where σ_{st} is the number of shortest paths from node s to t, and $\sigma_{st}(v)$ is the number of shortest paths between s and t that pass through node v. The "normalizing" factor in Equation (1) rescales the betweenness centrality to the interval $[0, 1]$ [9].

The centralization for a *network* G with n nodes is defined as

$$\mathrm{BC}(G) = \frac{\sum_{i=1}^{n} (\mathrm{BC}(n^*) - \mathrm{BC}(v_i))}{n - 1}, \quad (2)$$

where $\mathrm{BC}(n^*)$ is the maximum betweenness centrality value of a node in the whole network. The denominator $(n - 1)$ normalizes the entire network's centralization to a value between zero and one [59].

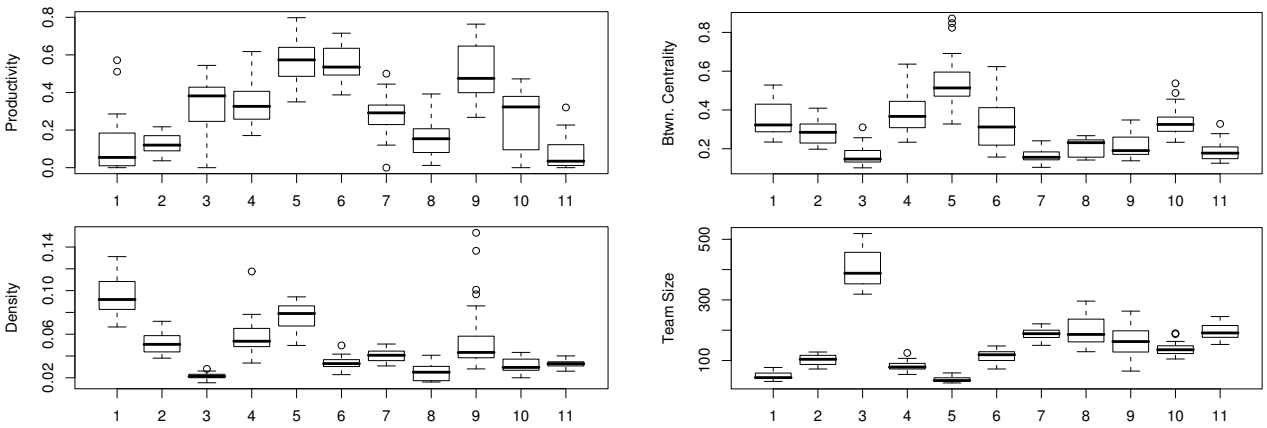


Figure 1: Overview of Intra-Project Variances.

6.3 Network Density

The third measure of network topology is density. It is defined as the number of edges present in the graph, divided by the maximum number of edges that could possibly exist. Therefore the network's density $\Delta(G)$ is given by [59, p. 129]

$$\Delta(G) = \frac{2 \cdot |\mathcal{E}_G|}{|\mathcal{V}_G| \cdot (|\mathcal{V}_G| - 1)}. \quad (3)$$

The density $\Delta(G)$ measures how connected different persons (nodes) of a network are. The networks we analyzed show rather small density values, usually below 0.1, which means that 10% or less of the possible connections between the nodes are actually established by mail exchange. However, small values for the density are not uncommon, as sociology regards values above 0.4 as high [41]. Studies have shown that networks with higher densities display a higher degree of mutual control and solidarity [41, p. 84]. Furthermore dense networks are said to represent communities that a more capable of finding and pursuing common goals than less dense ones. Anyhow, one should be careful to conclude that low density networks – like the ones we analyze here – exhibit a slow distribution of information (c.f. [15]), because findings for non-electronic social networks are only partly applicable to virtual communities [58]. Actually, researchers have shown that Open Source Software communities show rather fast patterns of information exchange [10].

6.4 Team Size

Our hypotheses of Section 3 include a size construct which we measure by the "team size" variable. In order to measure how many people have been participating in the communication during a certain period of time, we simply interpret the number of nodes in the communication network G as the size of the team. As defined in Section 5, every participant of the mailing list communication is represented by a single node in the network. Formally, the team size is given by the set of nodes' cardinal number:

$$\text{team size} = |\mathcal{V}_G|.$$

7. DATA ANALYSIS

Building on the variable definitions of the previous section, we now present the results of our data analysis. As stated in Section 4, our research design explicitly balances the amount of projects and timeframes available therein. The advantage of this holistic view on the dynamics of Open Source Software development come at the expense of applicable research methodology. In line with [18], we cannot apply standard time-series analysis in this setup but introduce an alternate methodological approach to answering our research question before presenting analysis results.

In this section, we present our approach for analyzing and interpreting the data sources described above. We first outline the aggregation of variables and measures and then explicitly address the problem of different variances across the OSS projects. This problem has been ignored by studies conducted before [63, 64].

7.1 Methodology

Our research methodology consists of three steps:

1. Analyze differences in variances across projects in order to demonstrate the necessity to analyze timeframes of different projects separately.
2. Per project, analyze the correlations of the variables with time.
3. Compare the results of the previous step across projects and legitimize the finding's statistical validity with an additional correlation and Pearson product momentum test.

In step 1, we explicitly address a shortcoming of parts of related work (e.g., [63, 64]) in that we analyze whether two randomly chosen samples of a single project are more alike than two randomly chosen samples of different projects. [63, 64] do not analyze these variances but choose a research design that inflates the number of samples by analyzing all timeframes of all projects in a single analysis. However, our methodology checks for the assumption that these samples are drawn from a unique distribution.

In step 2, we assume (and confirm in Section 7.2) that the variances among samples within a project are less than those

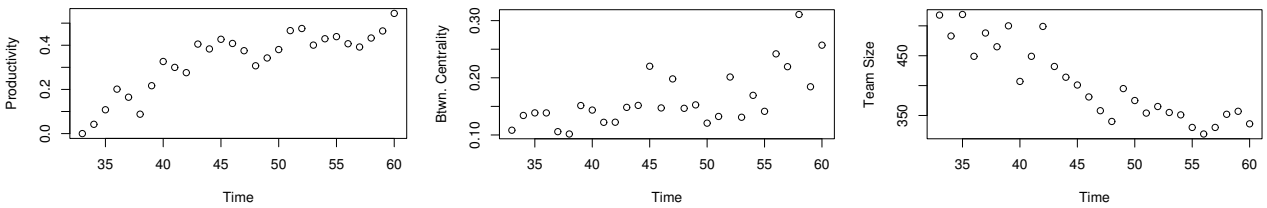


Figure 2: Project 'gcc'.

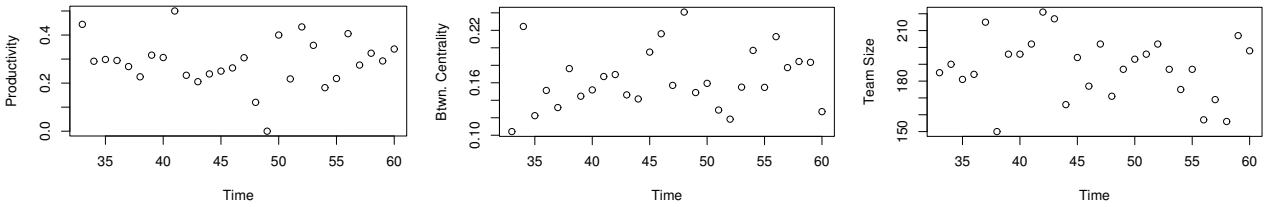


Figure 3: Project 'kde'.

spread over all projects and that a joint analysis of all available timeframes of all projects is therefore not methodologically sound. Instead, we rather analyze each project separately according to the following formal scheme:

For each project $p \in P$ and each variable v, let V_p be the set of corresponding measurements of v for the time slots analyzed for p. Further let

$$a_v = \frac{1}{|P|} \sum_{p \in P} \left(\frac{1}{|V_p|} \sum_{v \in V_p} v \right).$$

Then, for each project p, the first aggregated measure per variable v is

$$d_{v,p} = -a_v + \frac{1}{|V_p|} \sum_{x \in V_p} x.$$

It represents the difference of a project's average variable value from the cross-project average of the same variable. Hence, this measure does not yet reflect the longitudinal nature of our study but rather serves as a comparison to related work on a cross-sectional basis.

The time-dependent aggregation builds on significant correlations of a variable v with the time variable. For each project $p \in P$ and each variable v with measurements V_p and the time slots T_p, let $c_{v,p}$ be the Pearson correlation coefficient of the correlation between V_p and T_p if the Pearson product momentum test of this correlation is significant at the 5% level and 0 otherwise.

In step 3 of our research methodology, we compare the values of $d_{v,p}$ and $c_{v,p}$ across all analyzed projects. This comparison can reveal interesting findings on a qualitative level already. It can be observed which variables' temporal developments coincide. In addition to answering our research hypotheses for each project separately, this final step enables a holistic answer. In order to demonstrate the statistical significance of these findings, the associations of postulated effects are analyzed by correlating the $d_{v,p}$ and $c_{v,p}$ values of all eleven projects. Despite the small number of samples in this additional meta-analysis' correlation, the hypotheses' significance can be confirmed by means of the Pearson product momentum test.

7.2 Analysis and Results

The first step of analysis according to our research methodology as described in the previous section is to compare the variances of our variables across different projects. In line with related work, we choose intra-project variance boxplots for this analysis (e.g., [17]). Figure 1 shows four different charts, one for each variable. All x-axes show the eleven different projects and the y-axis shows the corresponding variable measure in each case. The boxplots in Figure 1 reveal that all four variables take very different levels across projects. There is little overlap between different boxplots of a subchart, which indicates that the measurements per timeslots of different projects must be seen to be drawn from different distributions. This result legitimizes our research design to analyze projects and timeslots separately. It must also be seen as a critique of part of related work that ignores the different variations (e.g., [63, 64]).

Next, we take a more detailed look at the temporal development of our variables in two example projects. In Figure 2, three variables are plotted on the y-axes against time on the x-axis. The three scatterplots in this figure all represent the 'gcc' project. All three plots take a more or less linear form. Accordingly, the correlations of all three variables with time are significant at the 5% level. The opposite case is present in the project 'kde' which is depicted in Figure 3. The seemingly random plots yield insignificant correlations of the three variables and time. Hence, various of the effects postulated in our research hypotheses can be confirmed to exist in 'gcc' but cannot be observed in 'kde'.

The comparison of these two projects shows that the dynamics of Open Source Software development are complex and may vary a lot across different projects. Due to our longitudinal, multi-project research design, we are able to go beyond merely discussing the effects present in individual projects. We can rather derive holistic answers to our research hypotheses that consider all analyzed projects at

Project	Temporal Correlation				Delta from Avg. Level			
	Produc.	Centr.	Density	Size	Produc.	Centr.	Density	Size
ant	0.692	–	–	–	-0.173	0.070	0.047	-104
apache	–	–	0.480	-0.638	-0.181	-0.005	0.005	-50
gcc	0.865	0.670	0.555	-0.904	0.026	-0.128	-0.026	252
gimp	0.426	–	-0.630	0.680	0.034	0.094	0.011	-70
gnumeric	0.667	–	0.555	-0.683	0.260	0.266	0.029	-116
gtk	–	-0.606	–	–	0.260	0.042	-0.014	-37
kde	–	–	–	–	-0.017	-0.126	-0.007	35
maemo	0.771	–	0.887	-0.669	-0.141	-0.085	-0.022	51
python	-0.715	-0.402	-0.741	0.781	0.205	-0.076	0.007	9
samba	0.727	–	0.680	-0.708	-0.039	0.051	-0.016	-14
wine	0.747	–	–	–	-0.235	-0.103	-0.014	43

Table 1: Aggregated Correlation Results and Descriptive Statistics per Project.

the same time. For this purpose, the third step of our proposed research methodology consists of a meta-analysis of the temporal correlation results.

Table 1 serves as a basis for this analysis. It can be seen as an outcome of the project-centric analysis up to the second step of our research methodology. The projects are listed in the left column. For each project and each variable, two further values are given. The 2^{nd} to 5^{th} columns represent the Pearson correlation coefficients of the corresponding variables and time, i.e., $d_{v,p}$ as defined in Section 7.1. Where this correlation is not significant at the 5% level, a dash is displayed. For example, the three scatterplots in Figure 2 correspond with row 'gcc' and the columns 'Produc.', 'Size', and 'Centr.' in left part of Table 1.

In addition to the analysis of Open Source Software dynamics by means of temporal correlations, we also double-check our results by analyzing the static level of a variable in comparison to all projects. Table 1's four rightmost columns serve as a basis for this analysis. These columns represent the difference of the corresponding project's arithmetic mean and the cross-project arithmetic mean of the same variable, i.e., $c_{v,p}$ of Section 7.1. Analyzing these static levels provides valuable additional insights. For example, the size-delta of 252 for project 'gcc' reveals that despite this project's decreasing team size, 'gcc' is still much larger than the average of all analyzed projects (each averaged over the entire time covered by this study). Additionally, the right part of Table 1 representing the static levels of variables can be compared to the left part representing dynamics. For example, in project 'gnumeric', productivity and size show temporal correlations in opposite direction as postulated in Hypothesis 2. This effect can also be observed statically, since 'gnumeric' has an above average level of productivity but its size is below average. However, this comparison does not hold in many cases and does not yield statistical significance in the meta-analysis conducted below. It must hence be seen as an add-on to our analysis only.

Regarding Hypothesis 1, the second column (temporal correlation of productivity) reveals that seven projects show the expected association (embodied in the positive correlation coefficient), that project *python* exhibits the opposite direction, and that three projects do not have a significant association between productivity and time. The first hypothesis can hence be said to be confirmed and is statistically significant at the 5% level according to a t-test (the alternative hypothesis that the true mean productivity coefficient is negative can be rejected, $t = 2.543, p = 0.015$).

Hypothesis 2 postulates a negative association of productivity and team size. An additional correlation of the temporal variables for the corresponding two variables (productivity and size) yields a coefficient of -0.605 and is significant at the 5% level. Hence, projects whose size increases over time tend to show decreasing productivity and vice versa. This dynamic effect is also supported by the static average levels (right hand side of Table 1) which also yield a negative correlation coefficient, despite not being statistically significant. Those projects with comparably high levels of productivity are rather small in terms of team size and vice versa.

The third hypothesis is the first to regard network structure. It postulates a positive association of communication network centrality and productivity. The additional correlation of both corresponding dynamic variables yields a coefficient of 0.645 and is significant at the 5% level. Those projects that show increasing productivity rates also become more central over time and vice versa. Again, the effect is also visible when correlating the static variables, though lacking statistical significance.

The next hypothesis of network structure, Hypothesis 4a, regards a potential negative association of team size and communication network density. Its intuition is the larger teams get, the more difficult it is to communicate with everyone. The additional correlation of the temporal 'Size' and 'Density' variables yields a coefficient of -0.970 and is significant at the 1‰ level. The static variant confirms this drastic association and is also significant at the 1% level. Hence, both variables must be seen to be highly correlated both dynamically and statically.

The last hypothesis, Hypothesis 4b, also regards network density and postulates a negative association with network centrality. However, against our expectations, the opposite effect can be observed in our data. Despite not being statistically significant, the additional correlation reveals that centrality and density dynamics are coherent. Increasing

centrality seems to not necessarily lead to decreasing density and vice versa. A possible explanation is that projects want to benefit from shortened information paths through central nodes but also want to keep the less efficient communicative links as a backup and bottleneck prevention. When analyzing the static variant of this hypothesis, the same effect is present (coefficient 0.649) and is even significant at the 5% level.

8. CONCLUSIONS

In this study, we analyze the temporal development of factors that have previously been found to influence Open Source Software development success. Thereby, we identify a weakness of related longitudinal studies in this field. They either analyze very few projects for many timeslots or conduct a delta-study of two points in time for many projects. A third group of related work fails to account for significant variations between projects by analyzing timeslots of all projects jointly.

We address these shortcomings of related work by analyzing multiple projects over a reasonable number of timeslots. This research design comes at the expense of the unapplicability of standard time-series analysis techniques [18]. We apply a two-stage Pearson correlation-based research methodology which enables us to draw statistically significant findings.

In addition to the distinction from related work by our longitudinal, multi-project research design, we make an important contribution to the literature on collaboration in Open Source Software development with our findings. Our results show that the theory of organizational learning is applicable in the context of Open Source Software development. Most of the analyzed projects exhibit increasing levels of productivity over time. Our work hence confirms that the findings of learning theory in commercial software development (e.g., [11]) are also valid in the context of Open Source Software. The implications for project coordinators are manifold. For example, they should attempt to foster participants' long-term commitment to the project in order to fully exploit knowledge accumulation and learning benefits.

Our findings regarding diseconomies of scale and the negative temporal association of productivity and team size imply that projects should not become too big. Open Source Software project administrators should react to growth in size with strict modularization and splitting too big projects in parts in order to keep productivity high. In line with related work on collaboration network structure, we find that centrality is beneficial for productivity. It is thus desirable to identify and develop key contributors. Strengthening their position in the collaboration network yields rises in productivity.

A rather surprising finding of our work is the positive temporal association of collaboration network density and centrality. It implies that Open Source Software projects do not rely on central nodes too much in order to avoid bottlenecks. This collaboration network evolution suggests that the advantages of centrality – such as shortened paths of information flow – can be utilized, but that the risks associated with centralization – such as reliance on central nodes and the creation of potential bottlenecks – are prevented at the same time by maintaining backup collaborative ties. Should the central nodes become unavailable, work can quickly be routed via the less efficient, but nevertheless maintained backup connections. We are not aware of other works that have detected this efficient and failsafe communication strategy in the dynamics of Open Source Software development.

Future work should focus on transferring our results to an even larger number of Open Source Software projects. Additionally, more constructs than the analyzed ones can be transferred from the organizational learning and communication network literature. For example, researchers could transfer the results from related works studying membership dynamics to the context analyzed in this study.

9. REFERENCES

[1] P. Adams, A. Capiluppi, and C. Boldyreff. Coordination and productivity issues in free software: The role of Brooks' law. In *IEEE International Conference on Software Maintenance (ICSM 2009)*, pages 319–328, 2009.

[2] P. Anbalagan and M. Vouk. On mining data across software repositories. In *Proc. of the International Workshop on Mining Software Repositories (MSR 2009)*, pages 171–174, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[3] S. Aral, E. Brynjolfsson, and M. van Alstyne. Information, technology and information worker productivity: Task level evidence. In *Proc. of the International Conference on Information Systems (ICIS)*, Milwaukee, WI, 2006.

[4] L. Argote. *Organizational learning: Creating, retaining and transferring knowledge*. Kluwer Academic Publishers, Norwell, MA, 1999.

[5] Y. A. Au, D. Carpenter, X. Chen, and J. G. Clark. Virtual organizational learning in open source software development projects. *Information & Management*, 46(1):9–15, 2009.

[6] R. P. Bagozzi and U. M. Dholakia. Open source software user communities: A study of participation in Linux user groups. *Management Science*, 52(7):1099–1115, 2006.

[7] R. D. Banker and S. A. Slaughter. A field study of scale economies in software maintenance. *Management Science*, 43(12):1709–1725, 1997.

[8] D. Barbagallo, C. Francalenei, and F. Merlo. The impact of social networking on software design quality and development effort in open source projects. In *Proc. of the International Conference on Information Systems (ICIS)*, Paris, France, 2008.

[9] M. Barthélemy. Betweenness centrality in large complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):163–168, 3 2004.

[10] J. Bitzer and P. Schröder. The economics of open source software development: An introduction. In J. Bitzer and P. Schröder, editors, *The Economics of Open Source Software Development*, pages 1–13. Elsevier Science, 1 edition, 10 2006.

[11] W. F. Boh, S. A. Slaughter, and J. A. Espinosa. Learning from experience in software development: A multilevel analysis. *Management Science*, 53(8):1315–1331, 2007.

[12] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, Feb 2009.

[13] U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136 – 145, 2008.

[14] F. P. Brooks. *The Mythical Man-Month*. Addison Wesley, New York, 1975.

[15] V. Buskens. *Social Networks and Trust (Theory and Decision Library)*. Springer Netherlands, 1 edition, 3 2002.

[16] K. Crowston. A coordination theory approach to organizational process design. *Organization Science*, 8(2):157–175, Mar. - Apr. 1997.

[17] K. Crowston and J. Howison. Hierarchy and centralization in free and open source software team communications. *Knowledge, Technology and Policy*, 18(4):65–85, December 2006.

[18] K. Crowston, J. Howison, and H. Annabi. Information systems success in free and open source software development: Theory and measures. *Software Process–Improvement and Practice*, 11(2):123–148, 2006.

[19] K. Crowston and B. Scozzi. Bug fixing practices within free/libre open source software development teams. *Journal of Database Management*, 19(2):1–30, April-June 2008.

[20] K. Crowston, K. Wei, Q. Li, and J. Howison. Core and periphery in free/libre and open source software team communications. In *Proc. of the 39th Annual Hawaii International Conference on System Sciences (HICSS)*, Washington, DC, USA, 2006. IEEE Computer Society.

[21] W. H. DeLone and E. R. McLean. Information systems success: The quest for the dependent variable. *Information Systems Research*, 3(1):60–95, 1992.

[22] P. F. Drucker. Knowledge-worker productivity: The biggest challenge. *California Management Review*, 41(2):79–94, 1999.

[23] N. Ducheneaut. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work*, 14(4):323–368, 2005.

[24] N. Economides and E. Katsamakas. Two-sided competition of proprietary vs. open source technology platforms and the implications for the software industry. *Management Science*, 52(7):1057–1071, 2006.

[25] M. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, Nov 1985.

[26] M. Granovetter. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1):33–50, 2005.

[27] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

[28] R. Grewal, G. L. Lilien, and G. Mallapragada. Location, location, location: How network embeddedness affects project success in open source systems. *Management Science*, 52(7):1043–1056, July 2006.

[29] J. Hahn, J. Y. Moon, and C. Zhang. Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties. *Information Systems Research*, 19(3):369–391, 2008.

[30] D. Hu and J. L. Zhao. Discovering determinants of project participation in an open source social network. In *Proc. of the International Conference on Information Systems (ICIS)*, Phoenix, AZ, 2009.

[31] C. L. Huntley. Organizational learning in open-source software projects: An analysis of debugging data. *IEEE Transactions on Engineering Management*, 50(4):485–493, Nov 2003.

[32] J. Jaisingh, E. W. K. See-To, and K. Y. Tam. The impact of open source software on the strategic choices of firms developing proprietary software. *Journal of Management Information Systems*, 25(3):241–275, 2008.

[33] Y. Kidane and P. A. Gloor. Correlating temporal communication patterns of the eclipse open source community with performance and creativity. *Computational & Mathematical Organization Theory*, 13(1):17–27, 3 2007.

[34] B. A. Kitchenham. The question of scale economies in software—why cannot researchers agree? *Information and Software Technology*, 44(1):13–24, January 2002.

[35] J. Kleinberg. The convergence of social and technological networks. *Communications of the ACM*, 51(11):66–72, 2008.

[36] A. G. Koru and J. Tian. Defect handling in medium and large open source projects. *IEEE Software*, 21(4):54–61, July-Aug. 2004.

[37] Y. Long and K. Siau. Impacts of social network structure on knowledge sharing in open source software development teams. In *AMCIS 2008 Proceedings*, 2008.

[38] T. W. Malone and K. Crowston. What is coordination theory and how can it help design cooperative work systems? In *Proc. of the ACM Conference on Computer-Supported Cooperative Work (CSCW)*, pages 357–370, Los Angeles, CA, 1990.

[39] T. W. Malone and K. Crowston. The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1):87–120, 1994.

[40] M. McLure-Wasko and S. Faraj. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29(1):35–57, 2005.

[41] J. Mewes. *Ungleiche Netzwerke? Vernetzte Ungleichheit: Persönliche Beziehungen im Kontext von Bildung und Status*. Vs Verlag, 1 edition, 3 2010.

[42] P. Middleton and J. Sutton. *Lean Software Strategies: Proven Techniques for Managers and Developers*. Productivity Press, New York, May 2005.

[43] W. Oh and S. Jeon. Membership herding and network stability in the open source community: The Ising perspective. *Management Science*, 53(7):1086–1101, 2007.

[44] E. Otte and R. Rousseau. Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6):441–453, 2002.

[45] P. C. Pendharkar. Scale economies and production function estimation for object-oriented software component and source code documentation size. *European Journal of Operational Research*, 172(3):1040–1050, August 2006.

[46] R. D. Pritchard. *Productivity Measurement and Improvement: Organizational Case Studies*. Greenwood Publishing Group, 1995.

[47] N. Ramasubbu, S. Mithas, M. S. Krishnan, and C. F. Kemerer. Work dispersion, process-based learning, and offshore software development performance. *MIS Quarterly*, 32(2):437–458, 2008.

[48] E. S. Raymond. *The Cathedral and the Bazaar*. O'Reilly, Cambridge, MA, 1999.

[49] H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley, Reading, MA, 1993.

[50] J. A. Roberts, I.-H. Hann, and S. A. Slaughter. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science*, 52(7):984–999, 2006.

[51] G. Robles, J. M. Gonzalez-Barahona, and I. Herraiz. Evolution of the core team of developers in libre software projects. In *Proc. of the International Workshop on Mining Software Repositories (MSR 2009)*, volume 0, pages 167–170, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[52] R. Sen. A strategic analysis of competition between open source and proprietary software. *Journal of Management Information Systems*, 24(1):233–257, 2007.

[53] A. Silberston. Economies of scale in theory and practice. *Economic Journal*, 82(325):369–391, 1972.

[54] K. J. Stewart, A. P. Ammeter, and L. M. Maruping. Impacts of license choice and organizational sponsorship on user interest and development activity in open source software projects. *Information Systems Research*, 17(2):126–144, 2006.

[55] Y. Tan, V. Mookerjee, and P. Singh. Social capital, structural holes and team composition: Collaborative networks of the open source software community. In *Proc. of the International Conference on Information Systems (ICIS)*, Montreal, Canada, 2007.

[56] G. von Krogh, S. Spaeth, K. R. Lakhani, and R. Policy. Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, 32(7):1217–1241, July 2003.

[57] G. von Krogh and E. von Hippel. The promise of research on open source software. *Management Science*, 52(7):975–983, 2006.

[58] M. Wasko, R. Teigland, and S. Faraj. The provision of online public goods: Examining social structure in an electronic network of practice. *Decision Support Systems*, 47(3):254 – 265, 2009. Online Communities and Social Network.

[59] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 11 1994.

[60] E. Wenger. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge, UK, 1998.

[61] J. M. Wilson, P. S. Goodman, and M. A. Cronin. Group learning. *Academy of Management Review*, 32(4):1041–1059, 2007.

[62] C.-G. Wu, J. H. Gerlach, and C. E. Young. An empirical analysis of open source software developers' motivations and continuance intentions. *Information & Management*, 44(3):253–262, 2007.

[63] J. Wu and K. Goh. Evaluating longitudinal success of open source software projects: A social network perspective. In *Proc. of the 42nd Annual Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, HI, 2009.

[64] J. Wu, K.-Y. Goh, and Q. Tang. Investigating success of open source software projects: A social network perspective. In *Proc. of the International Conference on Information Systems (ICIS)*, 2007.

[65] J. Xu, S. Christley, and G. Madey. Application of social network analysis to the study of open source software. In J. Bitzer and P. J. H. Schröder, editors, *The Economics of Open Source Software Development*, chapter 11, pages 205–224. Elsevier Press, Amsterdam, The Netherlands, 2006.

[66] L. Zhao and S. Elbaum. Quality assurance under the open source development model. *Journal of Systems and Software*, 66(1):65–75, April 2003.

Understanding Acceptance of Information System Development and Management Methodologies by actual Users: A Review and Assessment of Existing Literature

Kunal Mohan
EBS Business School
Söhnleinstraße 8D, 65201 Wiesbaden
kunal.mohan@ebs.edu

Frederik Ahlemann
EBS Business School
Söhnleinstraße 8D, 65201 Wiesbaden
frederik.ahlemann@ebs.edu

ABSTRACT
Despite the advantages of using IS methodologies, they are often rejected by actual users. Consequently, researchers have repeatedly attempted to understand why individuals accept certain methodologies while rejecting others. In order to differentiate what has been done from what needs to be done in research, a systematic review of academic studies that examine the acceptance of IS methodologies by actual users was conducted. This review revealed 19 articles. We found that the studies were either: a) descriptive, b) focused on specific determinants, or c) applied a holistic approach, examining methodology acceptance from a number of dimensions. Furthermore, while cognitive aspects have received considerable attention, none of the publications studied the effect of habits, emotions and the personal characteristics of individuals. We also examined the studies with respect to the reported research practices, and thereby identified areas of improvement. Based upon our findings, we developed a research agenda to guide future studies on this crucial subject.

General Terms
Management, Human Factors, Theory

Keywords
Methodology acceptance, literature review

1. INTRODUCTION
Early systems development projects often applied unsystematic methods. As these systems, embedded in a dynamic environment, became increasingly complex, the use of more disciplined methodological approaches proved to be indispensable [1]. Some of the most fundamental concepts that justify the use of structured methodologies, as identified by Fitzgerald, [1] are: i) they reduce complexity by subdividing the development and management process into plausible and coherent steps, ii) they increase transparency and therefore control of the development process, thus reducing risk and uncertainty, iii) they provide a goal-oriented framework that helps to direct the application of techniques and resources at appropriate times during the development and management process, and iv) they enable the standardisation of the development and management process. This facilitates the application of lessons learned from past experiences and also increases productivity and quality, because resource requirements can be predicted and made available as and when necessary [1].

As such, in search of ways to achieve predictable results, organisations either adopt, or customise and adaptively apply, information systems development (ISD) (e.g. object oriented systems development, agile system development etc.) and information system management (ISM) methodologies (e.g. IT project management (PM), enterprise architecture management (EAM), project portfolio management (PPM), IT benefits management (BM) etc.). These consist of tested bodies of methods, rules, and assumptions that fit the organisation [2,3]. Even though a methodological approach to solving complex tasks holds certain benefits, a methodology in itself is no silver bullet [3]. Despite the overwhelming advantages of using an IS methodology, only a handful of organisations are able to develop and implement one that is useful to the individuals that actually have to use it. Furthermore, only about 50% of organisations are able to motivate their staff to use such IS methodologies [4]. In a survey conducted by Russo et al. [5], they report that only 6% of organisations claim that their methodologies are always used as specified. Eva and Guilford [6] conducted a survey of 152 organisations, and found that only 17% of respondents use a methodology as a whole. As a result of this inconsistent use, despite the high investment in the development of IS methodologies and the pressure to use them, their practical usefulness is still a controversial issue [7]. The root of these problems lie, among others, in the failure to understand the needs of actual methodology users, which ultimately leads to the development and implementation of an IS Methodology that does not suit the user's needs and skills, and which they consequently reject.

The reason why user acceptance of IS methodologies is so much more important for consistent use, than acceptance at an organisational level is because, although an IS methodology is "adopted" by an organisation, the extent of its use (i.e. breadth and depth) is usually decided by the actual users of the methodology [8,9]. The importance to distinguish between the intentions of individuals to use a methodology from those of an organisation, is also suggested by Fichman [10]: "the relative lack of attention to individual adoption of technologies is unfortunate because, while the organisation as a whole makes the initial adoption decision for such technologies, the actions of individual adopters (e.g., how enthusiastically they embrace the innovation) can be expected to have a large impact on the implementation process". Thus, overcoming resistance to IS

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

methodology acceptance at an individual level is a critical area of concern in IS research [12].

Another issue, critical to understanding acceptance of methodologies, is recognising that comprehensive methodologies are not similar to individual methods (e.g. stakeholder analysis, use cases, entity relationship diagrams etc) and tools (e.g. ARIS, CASE tools, project management information systems, etc.). Although literature exists on the use of methods and tools, there is not sufficient justification for assuming, without empirical validation, that the results from the method/tool acceptance domain would be applicable to the *methodology* usage context [8]. Reasons why the adoption and success of new methodologies might be so different and so much more challenging than the adoption and success of specific methods and tools lies partly in the tacit organisational and individual problems that are caused by the introduction of new methodologies (which still remain insufficiently explored) [13]. Radical changes that are accompanied by new methodologies justify the need for exclusive research on their impact, instead of simply considering research on the adoption of methods and tools (which represent minor changes), to be directly applicable in the context of methodologies [14]. For example, the stress associated with the learning of a new methodology, the fear, and the impact on self-esteem and identity that is associated with the organisational restructuring or re-engineering can be grave. Little consideration is given to the emotional costs of role conflict and ambiguity, organizational conflict or workplace transformation, which recognizes the communication practices, personal relationships and co-ordination within the organization [13]. Consequently, the magnitude of behavioural change entailed by the adoption of a methodology is greater than that of a method or a tool [8]. All this warrants considering the antecedents of adoption and success of methodologies and the interrelationships between them to be different than that of individual methods and tools.

In order to better understand the domain of methodology acceptance, there is a need to conduct a critical review of the extant literature with the aim of: a) distinguishing what has been done from what needs to be done, b) synthesising and gaining a new perspective, c) discovering avenues for future research on methodology adoption at an individual level [15], and d) developing a research agenda for future studies. Consequently, our critical analysis is organised along the following research questions: 1) What findings have been reported in studies up to now?, and 2) what are the potential fruitful avenues for future research regarding an *individual's* acceptance of IS *development* and *management* (ISDM) *methodologies*? As to our knowledge, no systematic review of ISDM methodologies has been published before. Existing reviews focus on very specific types of ISD methodologies, especially agile and object-oriented system development (for e.g. [16,17]. Furthermore, the previous reviews, except that of [17], generally do not include any examination of the research design and methodology of the published studies, as in this systematic review. We feel that this overview will be important for researchers who wish to identify areas that have been researched or in which research is lacking, as well for practitioners who want to stay up to date on the current state of research in the general domain of ISDM methodologies.

The remainder of the paper is organized as follows: section 2 defines IS methodology as it is used in this paper, differentiating it from methods and tools. In this section, we also provide justification for the necessity of this differentiation. Section 3 explains the literature review and research practice assessment methodology. In section 4, we present the results, critically examine the literature review, and discuss future research opportunities. In section 5, we discuss the limitation of the literature review and provide an overview of the next steps in our research that i) aims at providing a solution to the limitations and ii) elaborates on how we plan to build upon our literature review results in order to develop a better understanding of the research topic at hand. We conclude with section 6, highlighting the contributions of the current research.

2. BACKGROUND – WHAT IS A METHODOLOGY?

One of the most fundamental problem in the literature – as identified by Iivari et. al. [3] – is the debate on the use of method as opposed to methodology, or vise versa. They find that some authors use method and methodology interchangeably; that some think methods encompass methodologies; that some think methodologies encompass methods; and that some believe that there are no methodologies, only techniques. Results of our literature review revealed that this debate can be extended to include tools, since some studies regard methods to be tools (e.g. [18]). Therefore, we consider that the discussion on ISDM methodologies can be updated to methodologies vs. methods vs. tools. The four-tiered conceptual structure, developed by Iivari et al. [3], makes it relatively easy to classify the large number of existing methodologies as a result of its abstract and parsimonious construction. Using this structure, we propose the following definitions, which help us to better understand methodologies, their parts and the interrelationships between them (see Figure 1).

Methodology. An ISDM methodology is a collection of goal-oriented, problem solving methods/techniques governed by a set of normative principles [19], beliefs, and a multi-step procedure that prescribes what to do and how to do things [20,21].
Methods/Techniques. An ISDM method/technique consists of a well-defined sequence of elementary operations for conducting a portion of a phase of a methodology (consult [22] for a detailed overview of existing IS methods).
Tools. An ISDM tool is an artifact, (usually software programs) that individuals may or may not use to support and facilitate the execution of a method/technique [12,14].

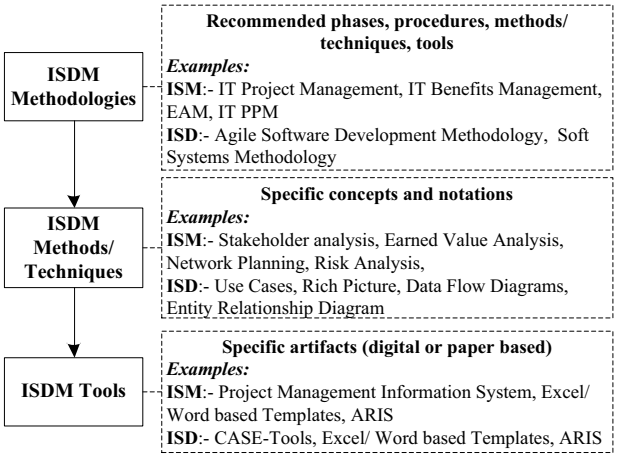


Figure 1. Overview of ISDM Methodology, Methods, & Tools

An ISDM methodology implies a holistic goal-oriented approach, with cultural, educational, ideological and/or strategic implications, that guides the work and cooperation of the various parties (stakeholders) involved in the *development* and *management* of IS/IT [23]. Methods and tools are only a subset of methodologies meant to support them (a means to an end) [23]. It is important that we distinguish between the use of tools, techniques/methods, and the use of an entire methodology, since tools and techniques can be used in the absence of a formal methodology. Furthermore, the use of a methodology represents a much more radical change than the use of tools and methods [12,8]. This distinction was demonstrated by Orlikowski [14], and mentioned by Hardgrave et al. [12]. Orlikowski [14] examined two CASE tool adoption environments: in the first environment, a methodology was present and CASE tools were adopted to support the existing methodology (a minor change for the stakeholders); the second environment had no methodology in place. Therefore, an ISD methodology and a CASE tool, specifically designed for that methodology, was adopted (a radical change for the stakeholders). Comparing the two adoption scenarios, Orlikowski found that the reactions of the stakeholders significantly differed. He concluded that this was because stakeholders in the first environment did not have to undergo radical change, compared to those in the second environment who had to adopt a complete new methodology with the CASE tool. This particular example illustrates clearly the need for a more holistic approach when studying methodology acceptance.

3. RESEARCH METHODOLOGY

We conducted an extensive review of existing literature between July 2009 – Nov 2009, as recommended by Webster and Watson [24]. Contrary to the more popular review approach of studying only selected top journals, we also included conference proceedings, working papers, editorials, book chapters, and dissertations. We felt that a complete review should not be confined to one methodology, one set of journals, or one geographic location [24].

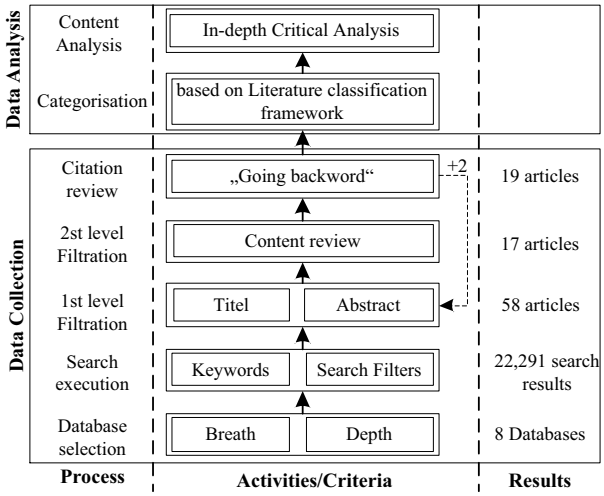


Figure 2: Data Collection and Analysis Methodology

Additionally, because MIS is an interdisciplinary field, we extended our review to include related fields such as marketing, psychology, sociology, and operations research. Figure 2 provides an overview of our literature review methodology as it is explained in the subsequent sections.

3.1 Data collection

For purposes of data collection, we again relied on the advice of Webster and Watson [24], and applied a structured approach. We searched a number of online databases, using a combination of keywords, for example methodology, adoption, use etc. The use of multiple databases and keywords allowed us to cover a large number of different publications, preventing the review from being too narrow or shallow [25]. Since our goal was to potentially investigate all published academic articles in the area of interest, we did not confine the search to certain time periods. The search resulted in a total number of 22,291 results (see Table 1).

Table 1. Search Criteria and Results

Database	Search Filters	Hits
EBSCO (covers 4 sub-databases)	case study, editorial, interview, proceeding, report, scholarly (peer reviewed) journals, collective volume, working paper, dissertation, journal article, periodicals, conference proceedings, book chapter, report	6,719
Science Direct	article, editorial, report, discussion, short survey, publisher's note	6,055
ACM Portal	journal, proceeding , thesis, report	4,932
Emerald Insight	conceptual paper, general review, case study, literature review, research paper, technical paper, viewpoint	351
SpringerLink	journal articles, book chapters	4,234

Although the number might seem overwhelming at first, a large number of search hits had little in common with the specific research area and were consequently dismissed quickly. The initial filtering through the search results was done by examining the title and the abstract. However, in a large number of cases, the abstract was not examined, because the title was found to provide sufficient evidence that the article did not address the

research topic at hand. We identified 58 publications that were related to the research at hand and were consequently selected for further examination. The full text of each research paper was further reviewed to eliminate those that were not actually related to ISDM methodology usage behaviour of individuals. The review yielded 17 articles related to our very specific research interest. This sharp reduction in the final list of research papers can be explained by the fact that many papers studied methods and tools but classified them as a study of methodologies (for e.g. [26,27]) or focused on organisational adoption decisions instead of individual acceptance [3]. Such papers appeared in the search results only because their title contained the keywords we used, but were discarded after we recognised that they did not address the research topic at hand. To conclude the data-gathering phase, we "went backward" [24] by reviewing the citations of the pool of 17 articles to find relevant articles that we might have overlooked and that should be considered. This revealed another 2 articles, increasing the final number of publications to 19. A work-log revealed that a total of 74 hours were spent on data collection (this does not include data analysis) and that the majority of the work i.e. 71% was done on weekends and holidays.

3.2 Data analysis

We subjected the final pool of 19 papers to a classification, to systematically categorise and describe the selected literature. The classification framework (see Figure 3) was constructed after examining the classification scheme of similar studies (for e.g. [28], [29], [30]), which present the most comprehensive classification of MIS topics. We also adapted by added further categories and items to cover all the important aspects of the research objectives at hand. The full text of each of the papers was studied to classify the entire literature, based on a number of dimensions of our classification framework such as object of analysis, unit of analysis, independent and dependent variables, theories used, sample source, sample size, data collection method, data analysis method and research type [29].

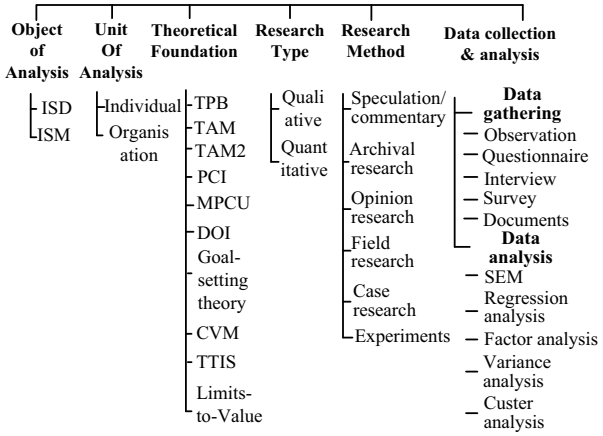


Figure 3. Literature Classification framework

3.3 Examination of Research Practices

Each of the final 19 studies was examined according to 10 criteria (C1-C10 in table 2). These criteria were based on principles of good practice for qualitative (for e.g. [31,32]) and quantitative research (for e.g. [33,34]), in a process similar to that applied by Dybå and Dingsøyr [17]. The 10 criteria, adopted from Dybå and Dingsøyr [17], covered four main issues pertaining to quality, which need to be considered when evaluating studies: a) *Academic* - Is it an academic Article?, b) *Rigour* - Has a thorough and appropriate approach been applied to key research methods in the study?, c) *Credibility* - Are the findings well-presented and meaningful?, and d) *Relevance* - Are the findings useful to the industry and/or the research community?. These 10 criteria provide a deeper understanding of the *"reported"* research practices/methodology of the studies. Each of the 10 criteria was examined, using an evaluation form that consisted of 34 questions, developed and validated by Dybå and Dingsøyr [17] (in some cases, wording of the original questions were changed to suit our research.) (Please contact the authors for the evaluation form.)

Table 2. Evaluation Criteria adopted from Dybå and Dingsøyr [17]

Academic	**C1. Research**: Is the paper based on research (or is it merely a ''lessons learned" report, based on expert opinion, without a concrete methodology)?
	C2. Aim: Is there a clear statement of the aims of the research?
	C3. Context: Is there an adequate description of the context in which the research was carried out?
Rigour	**C4. Research Design:** Was the research design described sufficiently and was it appropriate to address the aims of the research?
	C5. Sampling: Was the sampling strategy described sufficiently with regard to the aims of the research?
	C6. Data Collection: Was the data collected in a way that addressed the research issue?
	C7. Data Analysis: Was the data analysis described sufficiently with regard to the aims of the research?
Credibility	**C8. Reflexivity**: Has the relationship between researcher and participants been considered to an adequate degree?
	C9. Findings: Is there a clear statement of findings?
Relevance	**C10. Value:** Is the study of value for research and/or practice?

4. RESULTS AND FUTURE RESEARCH OPPORTUNITIES

4.1 Critical Review

In general, while development of methodologies has been widely researched, there has been little research on the determinants of individual intentions to use methodologies in the more general context of ISDM methodologies. A number of studies suggest that the use of methodologies is limited in practice, and that – even when they are used – are not literally applied. This signals a fundamental flaw in methodology engineering. Other authors go so far as to suggest that methodologies are useful to beginners, rather than to experienced individuals [3]. A number of studies have attempted to understand the adoption of methodologies by organisations, using organisations as their unit of analysis (for e.g. [35-38]). While these studies shed light on the important organisational-level decision to adopt software development innovations, they do not focus on the individual-level determinants of intentions. Others study the effects of using a methodology on project success or task performance (for e.g. [17,39,40]). On the whole, while there is abundant software engineering research on development of particular methodologies

(for e.g. [41]), studies that examine the determinants of methodology use and success at an individual level, considering not only ISD but also ISM methodologies, are scarce [42]. A plethora of research projects address the use of certain tools and techniques/methods [43-47] that may form part of a methodology. Some studies regard adoption merely as *intention to use* and do not study the *actual use* of the methodology (for e.g. [8], [48], [12]). In the following subsections we provide an overview of the specific studies related to the research topic at hand.

4.1.1 Descriptive results

Our literature review revealed, as shown in Figure 4, that research on ISDM methodology acceptance and usage at an individual level started as early as 1993 and peaked in 2002 when 3 articles were published. There was a significant gap in research in the 1990s. No articles were published between 1994 and 1996. After 2000, the number stabilised, with regular publications. Furthermore, 17 articles were published in academic journals, and 2 appeared in conference proceedings. The contribution and innovativeness of these publications needs to be examined further, since almost all studies were published in second tier or lower outlets (we analyse this situation in the "Discussion and Research Agenda" section after critically examining the content of the studies).

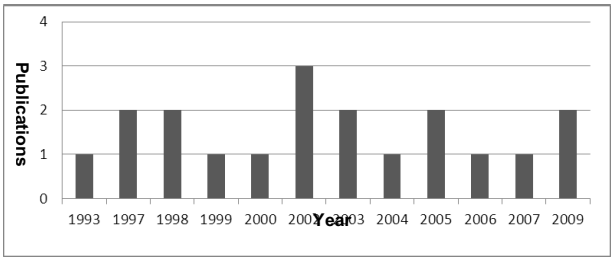


Figure 4. Overview of the studies by publication year

We also found that previous studies in the area of qualitative research consisted mainly of case research (CR) (2 publications), opinion research (OR) (1 publication), archival research (AR) including literature reviews (3 publications), and field research (FR) (2 publications). Surprisingly, all 19 studies focused on ISD methodologies and none analyzed ISM methodologies.

Table 3. Overview of Research Approach of Previous Studies

	Qualitative Research				Quantitative Research		Total
	AR	OR	CR	FR	Descriptive/ Exploratory	Confirmatory/ Positivistic	
ISD	3	1	2	2	8	3	**19**
ISM	-	-	-	-	-	-	**0**
Total	3	1	2	2	8	3	**19**
	8				**11**		

Table 3 summarises the research approach of prior studies on the adoption of ISDM methodologies, based on the categorisation scheme of Stone [28] and exposes areas in which research is severely lacking. Studies based on quantitative research comprise mostly descriptive and exploratory research (8 publications) with few studies of confirmatory positivist nature based on valid measures and extensive statistical analysis [8]. The lack of quantitative confirmatory studies that conform to the generally accepted validity criteria [33,34,49-51] highlights the need to develop a conceptual theoretical framework, which can then form a basis for future confirmatory positivist research. In the next section, we discuss and critically examine a subset of these publications, which we consider to have significantly contributed to the existing body of knowledge on this very specific topic.

4.1.2 Assessment of Research Practices

The goal of the assessment was not to criticise the commendable efforts of the researchers, but to point out aspects/issues of academic research that future studies should clarify and demonstrate with appropriate depth, so that the readers can better understand the context, method, and limitations of the research, thereby increasing comparability of findings with other similar studies.

We found that almost all the articles had clearly formulated research questions, and an adequate description of the research context. Only one article failed to state the research objectives clearly. For three of the 19 studies, the research design was found to be described insufficiently, and three did not apply a sampling strategy suitable for their research design. In these studies either a) data characteristics and origin was not mentioned, b) sample was not random, c) participants were chosen subjectively and therefore potentially affected by researcher bias, or d) the sampling did not fully cover the various segments of the target population. Two studies did not mention how data was collected and six articles did not describe their data analysis procedures sufficiently. For example, in some studies authors failed to address aspects of researcher triangulation, or did not mention analysis methods applied or tools used. In three studies we found the possibility of researcher bias was mentioned. Only three studies were found to have reported in a manner as to meet the 10 criteria. In general, we found that a) methods were not described sufficiently, b) biases in qualitative as well as quantitative studies were not addressed adequately, and c) data collection and analysis methods were not always described well. This is similar to the findings reported by Dybå and Dingsøyr [17], which suggests that studies in the domain of methodology acceptance are also plagued by common shortcomings found in other domains.

4.1.3 Content overview

As illustrated in the previous section, earlier studies on ISDM methodologies are largely descriptive and do not explain acceptance at an individual level [52]. It is only very recently that more explanatory studies that actually study the research problems at hand [52], have appeared. Westrup [53] conducted longitudinal case studies of the development and implementation of IS to explore how ISD methodologies are acquired by developers, and describe some of the ways in which methodologies are used in practice. An important conclusion of the study is that users reinterpreted the methodologies in each situation. Therefore, they did not follow the methodology rigorously. They also observe that developers used methodologies to complete deliverables and as insurance, to deny responsibility in case of project failure. Based upon the use of methodology manuals, Hidding [54] comes to a similar conclusion. He finds that even though practitioners seldom read methodology material, they are still able to produce the deliverables. Based on his research, Hidding [54] suggests that people assume different

roles when they use methodologies. Based on their roles, they have different information needs, which, when not satisfied, may lead to a rejection of a particular methodology. Roberts et al. [55,56] identified a number of factors, based on an exploratory factor analysis (EFA) of 88 survey items, that might affect the acceptance of methodologies among software developers. However, researchers (e.g. [57]) have pointed out that the study lacked a theoretical basis, compromising its internal validity; used measures that were not rigorously validated, and did not analyse the relationship between the identified factors and developers' usage intentions. Fitzgerald [1] found, after conducting a survey, that project size, -type, -client, user experience, and contingency might affect the decision of individuals to use an ISD methodology.

Some studies focus on very specific determinants of methodology adoption and suggest (directly or indirectly) determinants that could impact the behavioral decision. Kautz and Heje [58] conducted explorative studies to understand the role of formal university education on the adoption of systems development methodologies by means of grounded theory. They found a positive effect. However, the authors mentioned themselves that only a simple statistical analysis was carried out in their research. Huisman and Iivari [48] studied the perceptions of IS managers and developers, and found that managers had more positive views on the use of ISD methodologies than the developers. They concluded that ISD methodologies reflect management's agenda, implying different expectations, assumptions and norms. This provides further evidence for the widespread understanding that while developing and implementing IS methodologies, organisations often do not consider the values, beliefs and needs of the actual users. This might be the cause of the methodologies eventually being rejected. In a related study, Iivari and Huisman [52] found that organisational culture orientations, especially hierarchical and rational organisational cultures, affect the use of ISD methodologies. Most of the factors that have been reported have been studied separately and, for this reason, the relationships among them have not been explored sufficiently [9]. Although these studies identify some factors to have a significant effect on usage, when grouped together with other factors, they might become insignificant. Therefore, we suggest, along with Khalifa and Verner [9], that in order to determine what factors really drive the extent of use of ISDM methodologies, the combined effects (instead of isolated study of the effects) of these factors need to be examined.

In a plethora of research, the use of ISDM methodologies per se has not been studied, but rather the adoption of certain methods/techniques (such as object-oriented programming) and tools (such as CASE tools) [8]. Some of these studies have contributed in a major way to understanding the antecedents of an individual's decision to use ISDM methodologies. For example, Leonard-Barton [46] studied innovation acceptance, based on the adoption or rejection of structured systems analysis (SSA), and suggested that social pressure and training positively influence the use of methodologies. Although the author herself mentions that SSA is a *method* used only in the first phase of systems development and not a comprehensive methodology, we still consider her findings worthy of acknowledgement since they were published at a time when research on the adoption of IS methodologies was in its infancy. Khalifa and Verner [9] studied several determinants of software developers' use of two specific approaches, namely prototyping and waterfall. Although, similar to Palvia and Nosek [22], they regard prototyping to be a methodology, we consider it to be a method/technique since it is very often used within the waterfall methodology, as the authors point out themselves. Johnson et al. [59] applied TPB to examine the beliefs that underlie attitudes, social norms and behavioural control constructs, to examine IS developers' beliefs underlying intentions to use object oriented (OO) *methods*; however, they did not empirically test the relationships between the constructs. Research discussed in this paragraph represents an important step towards examining the underlying topic, but since the use of ISDM methodologies involves radical change compared to using simple methods and tools, we consider human behaviour, in the context of using complete methodologies, to be more complex. This requires a deeper examination that should take into consideration not only cognitive but also automatic user behaviour, such as emotions and habits.

Riemenschneider et al. [8] attempted to remedy some of the shortcomings of the research of Johnson et al. [59] by using five theoretical models to study the intentions of software developers to use methodologies. He found that the usefulness of methodologies plays the most crucial role in the adoption process. Being the first study to empirically test these five models in a methodology context, Riemenschneider et al. [8] provided thoughtful insights on the determinants of intention to adopt methodologies. In a related study (based on the quantitative data of Riemenschneider et. al. [8]), Hardgrave et al. [12] investigated the determinants of the individual developer's *intentions* to follow methodologies, based on TAM and DOI. In their study, perceived usefulness, social pressure, perceived compatibility and organisational mandate were found to have a direct influence on individual developers' intentions to follow methodologies, whereas social pressure, complexity and perceived compatibility were found to be significant determinants of perceived usefulness. In a similar study, Hardgrave and Johnson [60] found that organisational usefulness (OU), subjective norm and perceived behavioural control-internal significantly influenced intentions of software developers to use OO-SD processes. Although they propose that personal usefulness (PU) might also affect the intention to use, they could not psychometrically separate it from organisational usefulness. They suggest that the cause of this might lie in that "…developers do not view their personal benefits separately from organisational benefits" [60]. We suggest differentiating OU and PU based on other dimensions, influenced by purely personal interests (independent of organisational usefulness) such as materialism and enjoyment. Kacmar et al. [61] conducted a field study of ISD methodologies, applying theories of social exchange, task-technology fit, and technology acceptance. They found that perceptions of the outputs and deliverables from a methodology, and perceptions of challenges and obstacles to using and applying a methodology, to significantly and positively influence perceived usefulness. They found that these factors also negatively influence ease of use of a methodology, respectively, within a developer's organisation. Although Riemenschneider et al.[8], Hardgrave et al. [12] and Hardgrave and Johnson [60]

contribute significantly in understanding the topic at hand, they consider adoption to be merely the *intention to use* and do not study the *actual use* of the methodology. In our view, the mere *intention* to use a methodology, even though it plays a major role in determining actual use, does not imply that the individual will actually use the methodology. As such, future research could focus on studying the *actual use* of a methodology, rather than the mere intention to do so.

4.2 Discussion and Research Agenda

Existing research has attempted to examine usage behavior of individuals regarding IS methodologies from a technology adoption perspective. Some of these studies view software development methodologies as technology innovations and make use of technology adoption theories and models, such as Diffusion of Innovations Theory (DOI) and Technology Acceptance Model (TAM) (for e.g. [8,12,62,61]). Others apply sociological models such as the Theory of Planned Behaviour (TPB) and Triandis' Theory of Interpersonal Behavior (TTIB) to examine the development of the intention of individuals to use methodologies (for e.g. [63,9]. While previous studies, based on the technological and behavioural models, have been found to be suitable for examining the acceptance of IS methodologies, they focus mainly on *technology characteristics*, such as perceived usefulness, perceived ease of use, perceived complexity, and adaptability [16]. Moreover, these technical characteristics examine the decision-making process of individuals to adopt a methodology based upon the (potential) *benefits* that the particular methodology provides. *Costs* of adopting and using new methodologies have, up until now, not been studied, and depicts a potential gap in this research topic. In order to remedy this, future studies might be able to use the extensive switching costs topology proposed by Burnham et al. [64]. In the context of methodology acceptance at an individual level, following switching costs in particular might inhibit a person's desire to use new methodologies [64]: a) *Economic risk costs* are the costs of accepting uncertainty with the potential for a negative outcome when switching to a new methodology about which the user has insufficient information, b) *Evaluation costs* are the time and effort costs, associated with the search, and analysis is needed to make a decision to switch to a new methodology, c) *Learning costs* are the time and effort costs of acquiring new skills or know-how in order to use a methodology effectively, and d) *Personal relationship loss costs* are the affective losses associated with breaking the bonds of identification that have been formed with the people with whom the individual user used to interact before, when using old methodologies/processes (e.g. new reporting processes/roles change the way users interact with whom).

In order to fully understand the effect of costs, researchers should, in addition to switching costs, also examine *sunk costs* (i.e. irretrievable expenditures). Numerous empirical studies (for an overview, consult [65]) have shown that sunk costs cause a decision-making bias known as sunk-cost fallacy (or also escalation of commitment) that reflects the tendency in individuals to invest more future resources in a situation in which a prior investment has been made, compared to a similar situation in which a prior investment has not been made. Based on this research stream, sunk costs might hinder individuals from adopting and using new methodologies since these people have already invested considerable time and effort in learning their present methodology/way of doing things (some might even have costly certifications such as PRINCE2 or PMI, which might be of use in the context of a new methodology).

There has been a significant movement in the psychology discipline, in recent decades, in which the affective or emotional aspect is moving towards mainstream psychology, [66] based on the realisation that a realistic human being has more than just the physical and cognitive aspects. However, strikingly, none of the studies conducted in the past have attempted to examine the effect of *non-technological characteristics* such as a) traits of individuals, b) habits and emotions, c) self-beliefs such as self-concept, and d) organisational and national culture in the context of methodology usage. Research, in particular, has not attempted to understand the effect of deep-rooted personal characteristics and traits of individual users, such as their needs, as examined by needs theories, such as Maslow's hierarchy of needs [67] and Murray's theory of psychogenic needs [68], expectancies, age and gender. Needs theories, specifically, have become widely accepted in research studies because they are considered to be the most enduring ways to understand the motivation of an individual to act in a particular way [69]. According to the needs theories, individuals are motivated to use a particular methodology by their individual desire to satisfy certain needs. Many definitions of basic needs have been proposed. The one presented by Ryan and Deci [70] is most consistent with the scope of methodology acceptance. They indicate that "a basic need, whether it be a physiological need or a psychological need, is an energising state that, if satisfied, conduces toward health and well-being but, if not satisfied, contributes to pathology and ill-being" [70]. This implies that if a methodology fails to satisfy an individual's basic needs, this might result in serious discomfort, and this dissatisfaction might be visible in the individual's rejection of the particular methodology.

On another note, a lack of significant innovativeness and originality can be observed in the field of quantitative research (e.g. [8,9,12,59,63]). Studies based upon this research type, in particular, have generally resorted to simply applying highly validated theoretical models from other fields, for example TPB, TAM, DOI, TTIB etc. without modifications on the domain of methodology acceptance. Such research is not without merit. However, it leads to conclusions that are at best already known and well established. This might help explain why almost no studies have been published in top tier journals (see descriptive results section). Researchers are therefore advised to conceptually analyse the problem at hand from different angles, rather than only from what is already known, in order to develop new theoretical concepts and a deeper understanding of human behaviour, specific to the methodology acceptance domain.

Future research could also focus on understanding the actual use of methodologies (measured via documented usage) and not just self-reported intention of using methodologies, since intention might not always lead to actual use. Another crucial area that was neglected in past studies is post-adoption use, i.e. reuse of methodologies past the initial adoption and usage. This is important because, while seeking to examine why individuals

accept particular methodologies, the goal is not just to understand "one-time" use, but rather the repeated continued long-term use of a methodology. Furthermore, past studies have been conducted almost exclusively in the field of software development (e.g. agile and object-oriented software development methodologies), neglecting IS management methodologies. We feel that the discussion on IS methodologies should be more general, taking into consideration not only IS development but also IS management practices, since both ISD and ISM methodologies usually tend to address tasks and processes, consist of phases and procedures that are to be followed strictly [13]. Both types of methodologies are "…concerned with exploring and understanding information technology as a corporate resource that determines both the strategic and operational capabilities of the firm in designing and developing products and services for maximum customer satisfaction, corporate productivity, profitability and competitiveness" [71]. As such, results of research on ISD methodologies might not be different from ISM methodologies. This calls for research on methodology adoption to be of higher generalizability by taking into consideration not only ISD but also ISM methodologies. From a research design perspective, while previous quantitative and qualitative studies are largely cross-sectional research, longitudinal studies involving repeated observations of the same individuals over long periods of time might be better suited to observe the development of behaviour, since time is one of the most important explanations of change. Therefore, longitudinal studies can give answers to questions concerning behavioural change/intention to change that cross-sectional studies cannot.

5. LIMITATIONS OF THIS REVIEW AND NEXT STEPS

The main limitation of this review is potential bias in the selection, classification and assessing of the literature, which might be caused by subjective opinions of the researchers. In order to reduce this bias and, as part of the next steps, another researcher will independently analyse and classify the final pool of 19 articles. Subsequently, in discussion with the researcher, we will develop a common understanding of the results by comparing his evaluation with ours and critically reflecting on it. In case of unresolvable differences, we will call upon another independent researcher to provide further feedback. Another limitation pertains to the data collection that might be hindered by the keywords we used. Considering that there were more than 1,000 ISDM methodologies and that most of them are commercial products named creatively and not standardised [36,1,3], our choice of keywords and search strings might have failed to address "buzz words" and unique names of methodologies. Concerning data extraction, we found that some studies did not describe their methods and samples adequately. There is therefore a possibility that the extraction process might have resulted in some inaccuracy in data. Furthermore, our categorisation might have suffered, and could not always be conducted to a very satisfactory degree because some articles lacked sufficient details about the design and findings. Owing to this, we might have differed in what we actually extracted. There is therefore a possibility that the extraction process may have resulted in some inaccuracy in the data.

What follows in our research program will build upon our literature review findings. After uncovering what has been done and what needs to be done, the next step is *how* to do it. Our long-term goal is to discover new variables and relationships, beyond what is already known. For this, two researchers will catalogue and classify existing validated and tested theories and models that might be useful in examining methodology acceptance by individuals, especially in the areas lacking research. Following the cataloguing and classification, the researchers will extract, from these theories, relevant factors and constructs that might help explain methodology acceptance as per guidelines of good qualitative analysis (for e.g. [31,32]), with the help of the software Atlas.ti. The research community might be able to use our work as a rich source to develop a better understanding of the theoretical fieldwork of methodology usage and success. We hope that such a "database" might prove to be a useful source of guidance to researches when looking at the problem at hand from different perspectives. It might help them by sparking new ideas and developing exciting concepts. Regarding the current status of our categorisation and classification project, we have to date identified and conducted an initial classification of 46 theories. We acknowledge that, as a result of subjectivity, limited resources and information processing capabilities, we are sure to have missed out some potential theoretical concepts. We also advise researchers to be critical when they use a theory, because theories are subjectively measured and as such one must make a judgment about which theories are most helpful.

6. CONCLUSIONS

The present study attempts to further the research on individual acceptance and use of ISDM methodologies by providing an overview of research conducted in this area, and by discussing what needs to be done. Our assessment of research practices of the extant literature is a rigorous approach to identify areas of improvement. While such a thorough assessment might not be practically possible for large-scale literature reviews (because of time constraints), we feel that reviews focusing on specific topics of interest, analysing a relatively small number of studies, should not fear going the extra mile to enrich the research community with deeper insights.

The 19 identified studies fell into three broad categories: a) those that are mainly descriptive, providing a snapshot of current state of methodology acceptance, b) those that focus on very specific determinants of methodology use, such as education and training, and c) those that apply a holistic approach in examining a methodology acceptance from a number of dimensions, including usefulness, social pressure, ease of use and organisational support. Our research has implications for practitioners as well as researchers. The various areas reveal a different aspect of human behaviour and personality, and each can serve as a point of attack for organisations in attempts to steer it in the desired direction [72]. A better understanding of these topics would enable organisations to design interventions that would increase the use of ISDM methodologies in order to improve productivity and quality, and to reduce effort.

A clear finding of this review is that non-technical, or "soft factors", such as culture, needs of individuals, habits and

emotions have not been addressed. We also do not know much about post-adoption use of methodologies. Another very promising field of focus is how culture influences the decision of individuals to adopt methodologies. Although the understanding of cultural influences has been repeatedly emphasised by top journal editors – e.g., Straub [73] – it is seldom incorporated in research, generally because of the difficulty of data collection. From a research design perspective, we found that previous studies were mostly of qualitative nature. Even quantitative research is mainly descriptive. This calls for building conceptual models and testing them in a confirmatory fashion, to discover causal relations that might aid a better understanding and predict ion of methodology usage. Furthermore, studies could adopt a longitudinal approach in order to better understand change in behaviour of individuals over time, since time provides one of the most important explanations of change. In conclusion, user acceptance of ISDM methodologies remains a complex and elusive, yet important, phenomenon. Past research has made progress in unravelling some of its mysteries, but we see that there is a backlog of research issues, which still need to be addressed.

7. REFERENCES

[1] Fitzgerald B., 1998, "An empirical investigation into the adoption of systems development methodologies," Inf. Manage., **34**(6), pp. 317-328.

[2] Wynekoop J. L., and Russo N. L., 1995, "Systems development methodologies: unanswered questions," Journal of Information Technology (Routledge, Ltd.), **10**(2), p. 65.

[3] Iivari J., Hirschheim R., and Klein H. K., 2000, "A Dynamic Framework for Classifying Information Systems Development Methodologies and Approaches.," Journal of Management Information Systems, **17**(3), pp. 179-218.

[4] Glass R. L., 1999, "A Snapshot of Systems Development Practice," IEEE Softw., **16**(3), pp. 112-111.

[5] Russo N. L., Hightower R., and Pearson J. M., 1996, "The Failure of Methodologies to Meet the Needs of Current Development Environments," Proceedings of the British Computer Society's Annual Conference on Information System Methodologies, pp. 387-393.

[6] Eva M., and Guilford S., 1996, "Committed to a Radical approach? A survey of systems development methods in practice," Proceedings of the Fourth Conference of the British Computer Society Information Systems Methodologies Specialist Group, pp. 87-96.

[7] Huisman M., and Iivari J., 2002, "The Individual Deployment of Systems Development Methodologies," Advanced Information Systems Engineering, pp. 134-150.

[8] Riemenschneider C. K., Hardgrave B. C., and Davis F. D., 2002, "Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models," IEEE Transactions on Software Engineering, **28**(12), pp. 1135-1145.

[9] Khalifa M., and Verner J. M., 2000, "Drivers for Software Development Method Usage.," IEEE Transactions on Engineering Management, **47**(3), p. 360.

[10] Fichman R. G., 1992, "Information Technology Diffusion: A Review of Empirical Research," Proceedings of the thirteenth international conference on Information systems, pp. 195--206.

[11] Rogers E. M., 2003, Diffusion of Innovations, 5th Edition, Free Press.

[12] Hardgrave B. C., Davis F. D., and Riemenschneider C. K., 2003, "Investigating Determinants of Software Developers' Intentions to Follow Methodologies.," Journal of Management Information Systems, **20**(1), pp. 123-151.

[13] Vickers M. H., 1999, "Information technology development methodologies.," Journal of Management Development, **18**(3), p. 255.

[14] Orlikowski W. J., 1993, "CASE Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development.," MIS Quarterly, **17**(3), pp. 309-340.

[15] Hart D. C., 1998, Doing a Literature Review: Releasing the Social Science Research Imagination, Sage Publications Ltd.

[16] Chan F. K., and Thong J. Y., 2009, "Acceptance of agile methodologies: A critical review and conceptual framework," Decision Support Systems, **46**(4), pp. 803-814.

[17] Dybå T., and Dingsøyr T., 2008, "Empirical studies of agile software development: A systematic review," Inf. Softw. Technol., **50**(9-10), pp. 833-859.

[18] Besner C., and Hobbs B., 2006, "The Perceived Value and Potential Contribution of Project Management Practices to Project Success," Project Management Journal, **37**(3), pp. 37-48.

[19] Lyytinen K., 1987, "A taxonomic perspective of information systems development: theoretical constructs and recommendations," Critical issues in information systems research, John Wiley \& Sons, Inc., pp. 3-41.

[20] Checkland P., 1999, Systems Thinking, Systems Practice: Includes a 30-Year Retrospective, Wiley.

[21] Avison D. B., and Fitzgerald G., 2003, "Where Now for Development Methodologies?," Communications of the ACM, **46**(1), pp. 78-82.

[22] Palvia P., and Nosek J. T., 1993, "A field examination of system life cycle techniques and methodologies," Inf. Manage., **25**(2), pp. 73-84.

[23] Iivari J., Hirschheim R., and Klein H. K., 1998, "A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies," Information Systems Research, **9**(2), pp. 164-193.

[24] Webster J., and Watson R., 2002, "Analyzing the Past to Prepare for the Future: Writing a Literature Review," MIS Quarterly, **26**(2), pp. 23, 13.

[25] Levy Y., and Ellis T. J., 2006, "A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research.," Informing Science, **9**, pp. 181-212.

[26] Johnson R. A., Hardgrave B. C., and Doke E. R., 1999, "An industry analysis of developer beliefs about object-oriented systems development," SIGMIS Database, **30**(1), pp. 47-64.

[27] Leonard-Barton D., 1987, "Implementing Structured

Software Methodologies: A Case of Innovation in Process Technology.," Interfaces, **17**(3), pp. 6-17.
[28] Stone E. F., 1981, Research Methods in Organizational Behavior, Scott Foresman & Co.
[29] Palvia P., En Mao P., Salam A. F., and Soliman K. S., 2003, "Management Information Systems Research: What's There in a Methodology," Communications of AIS, **2003**(11), pp. 289-308.
[30] Palvia P., Leary D., En Mao, Midha V., Pinjani P., and Salam A. F., 2004, "Research Methodologies in MIS: An Update," Communications of AIS, **2004**(14), pp. 526-542.
[31] Eisenhardt K. M., 1989, "Building Theories from Case Study Research.," Academy of Management Review, **14**(4), pp. 532-550.
[32] Gerring J., 2006, Case Study Research: Principles and Practices, Cambridge University Press.
[33] Straub D., Boudreau M., and Gefen D., 2004, "Validation Guidelines for IS Positivist Research," Communications of AIS, **2004**(13), pp. 380-427.
[34] Straub D. W., 1989, "Validating Instruments in MIS Research.," MIS Quarterly, **13**(2), pp. 147-169.
[35] Chad Lin, Yu-An Huang, Man-Shin Cheng, and Wo-Chung Lin, 2007, "Effects of Information Technology Maturity on the Adoption of Investment Evaluation Methodologies: A Survey of Large Australian Organizations.," International Journal of Management, **24**(4), pp. 697-711.
[36] Fitzgerald B., 1997, "The use of systems development methodologies in practice: a field study," Information Systems Journal, **7**, pp. 201-212.
[37] Sultan F., and Chan L., 2000, "The Adoption of New Technology: The Case of Object-Oriented Computing in Software Companies.," IEEE Transactions on Engineering Management, **47**(1), p. 106.
[38] Zmud R. W., 1982, "Diffusion of Modern Software Practices: Influence of Centralization and Formalization," Management Science, **28**(12), pp. 1421-1431.
[39] Liu L., Grandon E. E., and Ash S. R., 2009, "Trainee reactions and task performance: a study of open training in object-oriented systems development.," Information Systems & e-Business Management, **7**(1), pp. 21-37.
[40] Maruping L. M., Venkatesh V., and Agarwal R., 2009, "A Control Theory Perspective on Agile Methodology Use and Changing User Requirements.," Information Systems Research, **20**(3), pp. 377-399.
[41] Vavpotic D., and Bajec M., 2009, "An approach for concurrent evaluation of technical and social aspects of software development methodologies," Information and Software Technology, **51**(2), pp. 528-545.
[42] Pfleeger S. L., 1999, "Understanding and improving technology transfer in software engineering.," Journal of Systems & Software, **47**(2/3), p. 111.
[43] Agarwal R., and Prasad J., 2000, "A Field Study of the Adoption of Software Process Innovations by Information Systems Professionals.," IEEE Transactions on Engineering Management, **47**(3), p. 295.
[44] Chau P. Y. K., 1996, "An empirical investigation on factors affecting the acceptance of CASE by systems developers," Inf. Manage., **30**(6), pp. 269-280.
[45] Kozar K. A., 1989, "Adopting Systems Development Methods: An Exploratory Study.," Journal of Management Information Systems, **5**(4), pp. 73-86.
[46] Leonard-Barton D., 1987, "Implementing Structured Software Methodologies: A Case of Innovation in Process Technology.," Interfaces, **17**(3), pp. 6-17.
[47] Sheetz S. D., Irwin G., Tegarden D. P., Nelson H. J., and Monarchi D. E., 1997, "Exploring the Difficulties of Learning Object-Oriented Techniques.," Journal of Management Information Systems, **14**(2), pp. 103-131.
[48] Huisman M., and Iivari J., 2006, "Deployment of systems development methodologies: perceptual congruence between IS managers and systems developers," Inf. Manage., **43**(1), pp. 29-49.
[49] Boudreau M., Gefen D., and Straub D. W., 2001, "Validation in Information Systems Research: A State-of-the-Art Assessment," MIS Quarterly, **25**(1), pp. 1-16.
[50] Lee A., 2001, "Validation in Information Systems Research: A State-of-the-Art Assessment," MIS Quarterly, **25**(1), p. 1.
[51] Podsakoff P. M., MacKenzie S. B., Jeong-Yeon Lee, and Podsakoff N. P., 2003, "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies.," Journal of Applied Psychology, **88**(5), p. 879.
[52] Iivari J., and Huisman M., 2007, "The Relationship Between Organisational Culture and the Deployment of Systems Development Methodologies," MIS Quarterly, **31**(1), pp. 35-58.
[53] Westrup C., 1993, "Information systems methodologies in use.," Journal of Information Technology (Routledge, Ltd.), **8**(4), p. 267.
[54] Hidding G. J., 1997, "Reinventing methodology: who reads it and why?," Communications of the ACM, **40**(11), pp. 102-109.
[55] Gibson T. L., and Gibson M. L., 1999, "System Development Methodology Implementation: Perceived Aspects of Importance.," Information Resources Management Journal, **12**(3), p. 27.
[56] Tom L. Roberts J., Gibson M. L., Fields K. T., and R. Kelly Rainer J., 1998, "Factors that Impact Implementing a System Development Methodology," IEEE Trans. Softw. Eng., **24**(8), pp. 640-649.
[57] Yadav S. B., Shaw N. G., Webb L., and Sutcu C., 2001, "Comments on 'Factors that Impact Implementing a System Development Methodology'," IEEE Trans. Softw. Eng., **27**(3), pp. 279-281.
[58] Kautz K., and Pries-Heje J., 1999, "Systems development education and methodology adoption," SIGCPR Comput. Pers., **20**(3), pp. 6-26.
[59] Johnson R. A., Hardgrave B. C., and Doke E. R., 1999, "An industry analysis of developer beliefs about object-oriented systems development," SIGMIS Database, **30**(1), pp. 47-64.
[60] Hardgrave B., and Johnson R., 2003, "Toward an information systems development acceptance model: the case of object-oriented systems development," Engineering Management, IEEE Transactions on, **50**(3), pp. 322-336.

[61] Kacmar C. J., McManus D. J., Duggan E. W., Hale J. E., and Hale D. P., 2009, "Software Development Methodologies in Organizations: Field Investigation of Use, Acceptance, and Application.," Information Resources Management Journal, **22**(3), pp. 16-39.

[62] 2007, "Effects of Information Technology Maturity on the Adoption of Investment Evaluation Methodologies: A Survey of Large Australian Organizations."

[63] Hardgrave B., and Johnson R., 2003, "Toward an information systems development acceptance model: the case of object-oriented systems development," Engineering Management, IEEE Transactions on, **50**(3), pp. 322-336.

[64] Burnham T. A., Frels J. K., and Mahajan V., 2003, "Consumer Switching Costs: A Typology, Antecedents, and Consequences," Journal of the Academy of Marketing Science, **31**(2), pp. 109 -126.

[65] Singer M. S., and Singer A. E., 1986, "Individual Differences and the Escalation of Commitment Paradigm.," Journal of Social Psychology, **126**(2), p. 197.

[66] Forgas J., 1995, "Mood and judgment: The affect infusion model (AIM)," Psychol. Bull., **117**(1), pp. 66, 39.

[67] Maslow A. H., 1954, Motivation and Personality, Harper & Brothers.

[68] Murray H. A., 1938, Explorations in Personality, John Wiley & Sons Inc.

[69] Arnolds C. A., and Boshoff C., 2000, "Does higher remuneration equal higher job performance?: an empirical assessment of the need-progression proposition in selected need theories.," South African Journal of Business Management, **31**(2), p. 53.

[70] Ryan R. M., and Deci E. L., 2000, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.," American Psychologist, **55**(1), p. 68.

[71] Badawy M. K., 1998, "Technology Management Education: ALTERNATIVE MODELS.," California Management Review, **40**(4), pp. 94-116.

[72] Ajzen I., 1991, "The theory of planned behavior," Organizational Behavior and Human Decision Processes, **50**(2), pp. 179-211.

[73] Straub D. W., 2009, "Creating Blue Oceans of Thought Via Highly Citable Articles.," MIS Quarterly, **33**(4), pp. iii-vii.

Sind Ereignisgesteuerte Prozessketten besser für Fachanwender geeignet als UML Aktivitätsdiagramme? Eine empirische Untersuchung

Sven Overhage
Forschungsgruppe Component & Service Engineering
Universität Augsburg, 86159 Augsburg, Deutschland
sven.overhage@wiwi.uni-augsburg.de

Sebastian Schlauderer
Forschungsgruppe Component & Service Engineering
Universität Augsburg, 86159 Augsburg, Deutschland
sebastian.schlauderer@wiwi.uni-augsburg.de

Dominik Birkmeier
Forschungsgruppe Component & Service Engineering
Universität Augsburg, 86159 Augsburg, Deutschland
dominik.birkmeier@wiwi.uni-augsburg.de

ABSTRACT

Die Modellierung und Verbesserung von Geschäftsprozessen gehört zu den zentralen Aufgaben in Unternehmen. Um diese Aufgabe effizient umzusetzen, ist es erforderlich, dass die eingesetzte Modellierungssprache neben Analysten und Entwicklern auch Fachanwendern eine aktive Beteiligung erlaubt. Im Rahmen des Beitrags werden zwei weitverbreitete Sprachen, Ereignisgesteuerte Prozessketten und UML Aktivitätsdiagramme, empirisch untersucht. Im Mittelpunkt steht dabei die Frage, welche Notation für Fachanwender besser zur Abbildung von Geschäftsprozessen geeignet ist und durch welche sprachlichen Eigenschaften Unterschiede entstehen. Grundlage der Untersuchung ist eine Studie, in der 73 Teilnehmer Geschäftsprozesse unter kontrollierten Bedingungen mit beiden Sprachen modellierten. Die Ergebnisse der Studie werden explorativ ausgewertet, um wissenschaftlich gesicherte Erkenntnisse über die Eignung beider Sprachen zu gewinnen und zur Theoriebildung beizutragen.

1. EINLEITUNG

Die Modellierung von Geschäftsprozessen mit dem Ziel einer Verbesserung der Abläufe gehört zu den zentralen Zielen in Unternehmen [15]. Eine wesentliche Voraussetzung für die Erreichung dieses Ziels ist die Nutzung einer Geschäftsprozessmodellierungssprache, die es Managern, Fachanwendern, Analysten und Entwicklern gleichermaßen erlaubt, sich an der Abbildung und Verbesserung von Unternehmensabläufen zu beteiligen [1, 30]. Dem Bedarf an entsprechend geeigneten Sprachen steht eine Vielzahl an Notationen gegenüber, darunter formale Notationen (z.B. Petri-Netze), technische Notationen der Software-Entwicklung (z.B. UML Aktivitätsdiagramme) und betriebswirtschaftliche Notationen, die auf eine Verwendung durch Fachanwender ausgelegt sind (z.B. Ereignisgesteuerte Prozessketten). Hierzulande haben sich einer Studie zufolge vor allem Ereignisgesteuerte Prozessketten (EPK) sowie – die mit einigem Abstand folgenden – UML Aktivitätsdiagramme (AD) durchgesetzt [26].

Um die Auswahl eines Modellierungsansatzes methodisch zu unterstützen, ist es notwendig, vorhandene Sprachen systematisch in Bezug auf ihre Anwendbarkeit für die unterschiedlichen Parteien eines Unternehmens zu vergleichen und zu bewerten [16]. Jedoch finden sich in der wissenschaftlichen Literatur nur wenige Arbeiten, die Aufschluss über die Eignung der verschiedenen Sprachen zur Abbildung bzw. Interpretation von Geschäftsprozessen geben [24]. Häufig basiert die Einschätzung der Sprachen deshalb auf Expertenaussagen. So wird der Erfolg der EPK u.a. darauf zurückgeführt, dass sie „für den Fachanwender leicht zu erlernen sind und die abgebildeten Prozesse strukturiert und übersichtlich darstellen“ [13]. Durch ihre verglichen mit technischen und formalen Ansätzen „höhere Anschaulichkeit“ unterstützen EPK „nicht nur den Informationssystem-, sondern auch den Organisationsgestalter“ [2]. AD werden dagegen trotz einer gewissen Ähnlichkeit als zu technikorientiert [31] oder gar als fremdartig für Fachanwender [22] eingestuft. Jedoch bleibt unklar, wieso EPK im Vergleich zu AD über eine vermeintlich höhere Anschaulichkeit bzw. bessere Eignung für Fachanwender verfügen.

Im Folgenden wird die Eignung von EPK und AD für Fachanwender deshalb systematisch untersucht. Das Ziel des Beitrags ist es, wissenschaftliche Erkenntnisse über deren Verwendbarkeit zur Modellierung und die ursächlichen sprachlichen Eigenschaften zu gewinnen. Insbesondere werden folgende Forschungsfragen untersucht: *Sind EPK besser für Fachanwender zur Abbildung von Geschäftsprozessen geeignet als AD? Durch welche Spracheigenschaften werden Unterschiede in der Eignung verursacht?* Grundlage der Untersuchung ist eine von den Autoren im Jahr 2010 durchgeführte empirische Studie, in deren Rahmen 73 Teilnehmer unter kontrollierten Bedingungen Geschäftsprozesse mit beiden Sprachen modellierten. Die Ergebnisse der Studie werden explorativ ausgewertet, um bestehende Unterschiede im Datenmaterial zu erkennen und auf die untersuchten Sprachen zurückzuführen. Kapitel 2 beschreibt zunächst verwandte Ansätze, um den Kontext und die bestehende Forschungslücke herauszuarbeiten. Danach werden in Kapitel 3 die für die weitere Untersuchung notwendigen theoretischen

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

Grundlagen gelegt. Kapitel 4 beschreibt die Forschungsmethode und das Studienkonzept. In Kapitel 5 werden die Ergebnisse der Studie ausgewertet und analysiert. Am Ende des Beitrags werden Auswirkungen auf Wissenschaft und Praxis sowie der weitere Forschungsbedarf diskutiert.

2. VERWANDTE ANSÄTZE

In der Literatur werden folgende Untersuchungsansätze für den Vergleich von Modellierungssprachen unterschieden: der direkte Vergleich sprachlicher Eigenschaften, vergleichende Bewertungen von Sprachen auf Basis theoretischer Bezugsrahmen und empirische Studien [27]. Um zu ermitteln, inwiefern Sprachen geeignet sind, einen relevanten Wirklichkeitsausschnitt abzubilden, zu kommunizieren und dessen Verständnis zu ermöglichen, wird zum einen die Durchführung theoretischer Untersuchungen empfohlen, in denen die Ausdruckskraft der Sprachen analysiert wird [7]. Zum anderen wird vorgeschlagen, die Anwendung der Sprachen durch die verschiedenen Nutzer insbesondere bei der Modellerstellung und der Interpretation von Modellen empirisch zu vergleichen [3, 7, 8]. Mit der wachsenden Bedeutung der Geschäftsprozessmodellierung entstehen auch Arbeiten, die die Eignung vorhandener Sprachen zur Abbildung und Kommunikation von Geschäftsprozessen im Allgemeinen und die Eignung von EPK und AD im Speziellen untersuchen.

Diese Arbeiten basieren jedoch meist auf theoretischen Untersuchungsansätzen. Sie verwenden *theoretische Rahmenwerke* (z.B. die Workflow Patterns [31, 32] und kognitive Gestaltungsprinzipien [6]) oder *ontologische Gegenstandseinteilungen* der Realwelt (z.B. die Bunge-Wand-Weber Ontologie [4, 24]), um die Ausdruckskraft von Modellierungssprachen und ihre Eignung zur Abbildung und Kommunikation relevanter Sachverhalte zu beurteilen. Insbesondere die Bunge-Wand-Weber Ontologie wurde wiederholt herangezogen, um mögliche Schwächen von EPK zu diskutieren [5, 9] oder Defizite im Vergleich zu anderen Modellierungssprachen herauszustellen [24]. EPK werden dabei einerseits wegen ihrer kompakten Notation kritisiert, die dazu führen soll, dass einige Sachverhalte der Realwelt nur umständlich bzw. nicht angemessen im Modell dargestellt werden können [24]. Andererseits werden die expliziten Verknüpfungsoperatoren (Konnektoren) der EPK bemängelt, da sie keine Entsprechung in der Realwelt haben und vom Anwender deshalb nicht intuitiv verwendet werden können [24]. Da die Verknüpfungsoperatoren der EPK außerdem eine hohe syntaktische Ähnlichkeit aufweisen, könnten sie zudem leichter verwechselt werden als die graphisch eigenständigeren Verknüpfungsoperatoren des AD [6]. Weitere Vergleiche mit den als eher technisch eingestuften AD lassen sich allerdings kaum ziehen, da AD in den genannten Arbeiten häufig nicht berücksichtigt sind. Kritisch anzumerken ist ferner, dass die Relevanz solcher theoretisch begründeten Erkenntnisse noch nicht ausreichend empirisch überprüft wurde. Die wenigen vorhandenen Arbeiten zur Überprüfung von theoretischen Schwächen der EPK (vgl. [4, 10]) beschränken sich lediglich auf qualitative Befragungen von Nutzern und sind nur eingeschränkt aussagekräftig.

Empirische Studien, in denen die Anwendung von EPK und AD beobachtet wird, um daraus Rückschlüsse über die Eignung der Sprachen für bestimmte Nutzer zu ziehen, existieren bislang kaum. Vorhandene Arbeiten nutzen vor allem *Laborexperimente*, um den Einfluss der Modellierungssprache auf die Erstellung bzw. Interpretation von Geschäftsprozessmodellen zu untersuchen. Unter diesen Arbeiten konnten bei einem Vergleich von EPK und Petri-Netzen bspw. Hinweise darauf gefunden werden, dass gerade die kritisierten expliziten Verknüpfungsoperatoren der EPK einen positiven Einfluss auf die Interpretation von Geschäftsprozessmodellen durch Fachanwender haben [25]. Auch die Eignung von EPK und AD zur Erstellung und Interpretation von Geschäftsprozessen wurde bereits empirisch verglichen [11]. Dabei wurde vor allem festgestellt, dass die Teilnehmer bei der Erstellung von Prozessmodellen mit AD weniger Fehler begingen als mit EPK. Bei der Interpretation konnten dagegen keine Unterschiede zwischen den Sprachen gefunden werden. Die Aussagekraft der Studie ist allerdings schon aufgrund der geringen Anzahl an Teilnehmern beschränkt. Da bei der Auswertung nicht zwischen verschiedenen Fehlerarten unterschieden wurde, ist außerdem keine differenzierte Rückführung auf die Eigenschaften der verwendeten Notationen möglich. Darüber hinaus handelte es sich bei den im Rahmen der Modellierung beobachteten Teilnehmern um informationstechnisch ausgebildete Analysten. Somit kann keine Aussage darüber getroffen werden, inwiefern es die Sprachen auch Fachanwendern erlauben, sich an der Abbildung und Verbesserung von Geschäftsprozessen aktiv zu beteiligen. Um diese Forschungslücke zu schließen, wird nachfolgend die Eignung beider Sprachen für die Modellerstellung durch Fachanwender untersucht. Dabei wird insbesondere auf Auswirkungen eingegangen, die sich durch die unterschiedliche graphische Darstellung von Sachverhalten in beiden Sprachen ergeben [17].

3. THEORETISCHE GRUNDLAGEN

Die Erstellung konzeptioneller Modelle wird durch mehrere Faktoren beeinflusst, die zu Variationen im erstellten Modell führen können (Abb. 1). Neben der Komplexität des abzubildenden Wirklichkeitsausschnitts werden in der Literatur vor allem der Modellierer, die Modellierungssprache und der Modellierungsprozess als Einflussfaktoren genannt [28, 12].

Der *Modellierer* beeinflusst das zu erstellende Geschäftsprozessmodell auf verschiedene Weise. Sein subjektives Qualitätsempfinden hat einen unmittelbaren Einfluss auf das erstellte Modell (z.B. den Detaillierungsgrad) [12]. Mittelbar wirken sich ferner unterschiedliche Vorkenntnisse über den Anwendungsbereich und die individuelle Wahrnehmung der Realwelt aus, die den Modellierungsprozess beeinflussen [28]. Ebenso wirkt sich das Verständnis der Modellierungssprache aus, das mit vorhandenen persönlichen Erfahrungen variieren kann [12]. Soll der Einfluss von Modellierungssprachen bei der Erstellung von Modellen untersucht werden, sind die durch den Modellierer bedingten Einflüsse zu kontrollieren.

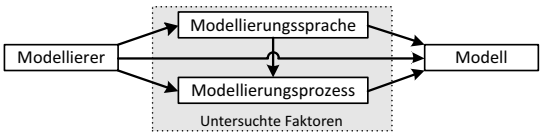


Figure 1: Einflussfaktoren der Modellierung [12]

Die *Modellierungssprache* hat einen direkten Einfluss auf das zu erstellende Geschäftsprozessmodell, da sich ihre Grammatik auf die Darstellung von Sachverhalten auswirkt. Unterschiede in der graphischen Syntax von Sprachkonstrukten und der Ausdruckskraft von Modellierungssprachen führen dabei zu Variationen im Geschäftsprozessmodell [12, 17].

Die im Beitrag verglichenen Sprachen nutzen unterschiedliche graphische Symbole und weisen kategoriale Unterschiede in ihren Grammatiken auf, die zu Variationen führen können. Um den Kontrollfluss präzise festzulegen, nutzen AD die Token-Semantik von Petri-Netzen [20].

Hierdurch wird auch die implizite Aufspaltung und Synchronisation von Kontrollflüssen, die ohne die Verwendung expliziter Konnektoren erfolgt, geregelt [20]. Die einzig zentralen Einheiten eines AD sind Aktionen, die zu einem Kontrollfluss verknüpft sind. Im Gegensatz zu Petri-Netzen, bei denen sich Stellen und Transitionen abwechseln, gibt es bei AD also keine alternierenden Modellkonstrukte. Die von den Aktionen verarbeiteten Informationsobjekte werden üblicherweise in den Kontrollfluss integriert, was deren Weiterleitung entsprechend der Flussrichtung impliziert. Verzweigungen des Kontrollflusses wirken sich dabei auch auf die Weiterleitung der Informationsobjekte aus. EPK stellen Geschäftsprozesse als Graphen dar [19]. Sie basieren zwar auf einer speziellen Form von Petri-Netzen, den sog. Bedingungs-Ereignis-Netzen, übernehmen aber nicht die formale Token-Semantik [25]. Zusätzlich zu Funktionen und Ereignissen, die die zentralen Einheiten eines Prozessmodells bilden und sich im Kontrollfluss stets abwechseln, müssen in EPK deshalb explizite Verknüpfungsoperatoren verwendet werden, um Verzweigungen und Synchronisationen darzustellen [19, 2]. Zwischen EPK und AD bestehen weitere Unterschiede (vgl. Abb. 4 und 5). Bspw. müssen Informationsobjekte in EPK vom Kontrollfluss getrennt modelliert und für jede Funktion dargestellt werden. Zur ausführlicheren Diskussion beider Sprachen sei hier auf die Literatur verwiesen.

Zusammen mit der Modellierungssprache hat auch der *Modellierungsprozess* unmittelbaren Einfluss auf das Geschäftsprozessmodell. Während dem Modellierungsprozess wird die relevante Semantik eines Geschäftsprozesses unter Anwendung einer Modellierungssprache in ein konzeptionelles Modell abgebildet. Sind die Konstrukte einer Sprache für den Anwender nicht intuitiv zu verwenden oder nicht direkt einsetzbar, um einen bestimmten Sachverhalt abzubilden, entstehen dabei zwangsläufig Variationen [12, 17]. Um die Eignung von Modellierungssprachen zu beurteilen, sind deshalb sowohl die Modellierungssprache selbst als auch der Modellierungsprozess zu untersuchen (Abb. 1).

Die Qualität der resultierenden Modelle ist ein wesentlicher Indikator, um die Eignung der eingesetzten Modellierungssprache bei der Modellerstellung zu beurteilen [8]. Zur Bewertung der Qualität von konzeptionellen Modellen im Allgemeinen bzw. von Geschäftsprozessmodellen im Speziellen existieren mehrere Ansätze in der Literatur [8, 14, 2]. Die meisten bleiben jedoch abstrakt und beinhalten keine konkreten Metriken zur Messung von Qualitätskategorien [18]. In diesem Beitrag wurde die Bewertung der erstellten Prozessmodelle anhand des 3QM-Frameworks [21] durchgeführt, dessen Qualitätskategorien auf den o.g. Ansätzen aufbauen und um Metriken zur Messung ergänzt wurden (Abb. 2). Analog zur Sprachwissenschaft unterscheidet es zwischen der Syntaktik, der Semantik und der Pragmatik als Qualitätskategorien von Geschäftsprozessmodellen. Diese Kategorien wurden in empirischen Studien als zur Beurteilung notwendig, hinreichend und voneinander unabhängig bestätigt [18]. Die *Syntaktik* bezieht sich auf die formale Ordnung der Zeichen in einem Modell, die durch die Grammatik der Sprache vorgegeben wird. Sie umfasst die formale Ordnung der einzelnen Zeichen (Wortsyntax), der unmittelbaren Verknüpfung von Zeichen (Satzsyntax) und der komplexen Sachverhalte (Textsyntax). In EPK ist bspw. vorgegeben, dass Ereignisse mit Sechsecken zu repräsentieren sind (Wortsyntax) und nicht vor exklusiven Verzweigungen stehen dürfen (Satzsyntax). Verzweigungen dürfen nicht durch Synchronisationen zusammengeführt werden (Textsyntax).

Die *Semantik* steht für die Bedeutung der Zeichen im Modell, die während der Modellierung festgelegt wird. Relevante Bedeutungsträger sind bei der Prozessmodellierung die (je nach Sprache) darzustellenden Aktionen, Ereignisse, Kontrollflüsse, Ablaufbedingungen, Informationsobjekte, Informationsobjektzustände, Datenflüsse und Organisationseinheiten [30, 2]. Die für die Abbildung eines Realitätsausschnitts notwendigen Bedeutungsträger dürfen im Modell weder fehlen (Vollständigkeit) noch inhaltlich falsch dargestellt sein (Korrektheit). Ferner sollten keine unnötigen Bedeutungsträger im Modell aufgeführt werden, die zur Abbildung nicht gebraucht werden (Relevanz). Schließlich sind voneinander unabhängige Abläufe im Modell als parallel darzustellen, um den Kontrollfluss nicht unnötig einzuschränken (Flexibilität). Die *Pragmatik* thematisiert die Interpretation der Zeichen im Modell durch Dritte, die durch die Modellierung mit beeinflusst wird. Im Sinne einer effizienten Interpretation sollte die Darstellung frei von Redundanzen und Widersprüchen sein (Eindeutigkeit) sowie einheitliche Bezeichnungen verwenden, die den Konventionen der Sprache entsprechen (Verständlichkeit). Eine gängige Konvention ist z.B., dass Funktionen durch das bearbeitete Objekt und ein Verb im Infinitiv bezeichnet werden [19].

Die zur Zusammenführung der Qualitätskategorien im 3QM-Framework genutzte Gewichtung (Abb. 2) wurde im Rahmen einer Expertenbefragung unter Nutzung des Analytic Hierarchy Process (AHP) ermittelt [21]. Bei der nachfolgenden Auswertung wird diese Gewichtung übernommen. Im Mittelpunkt steht allerdings die Betrachtung der Einzelfehlerklassen, die von der Gewichtung unabhängig ist. Anzumerken ist, dass die genannten Qualitätskategorien lediglich die Effektivität der Modellierung thematisieren. In die Auswertung sollten deshalb auch die Bearbeitungszeit und die Zufriedenheit der Anwender mit der Sprache als weitere relevante Aspekte einbezogen werden [7, 8].

4. FORSCHUNGSMETHODE

Um die Eignung beider Notationen für Fachanwender zu vergleichen wurde ein Laborexperiment durchgeführt, bei dem die Modellierungssprache und das Training die einzigen gezielt variierenden Einflussfaktoren darstellten. Mögliche Störfaktoren wurden durch Stichprobengröße und Auswahl der Teilnehmer kontrolliert. Mit der Einführung unterschiedlicher Trainingslevel wurden realitätsnahe Bedingungen für die Studie geschaffen, da in der Praxis meist eine Vielzahl unterschiedlicher Kenntnisstände über Modellierungsansätze vorzufinden ist [23]. Eine Auswertung von Trainingseffekten bleibt in diesem Beitrag aber eine Nebenbeobachtung.

Bei der Planung des Experimentverlaufs (Abb. 3) fanden die Erfahrungen einer 30 Teilnehmer umfassenden Vorstudie sowie die aus anderen relevanten Studien abgeleiteten Empfehlungen Berücksichtigung [3]. Um persönliche Merkmale der Modellierer als Störeinfluss zu minimieren, wurde eine hinreichend große, bezüglich der Vorkenntnisse homogene Stichprobe ausgewählt. Als Studienteilnehmer wurden Studenten

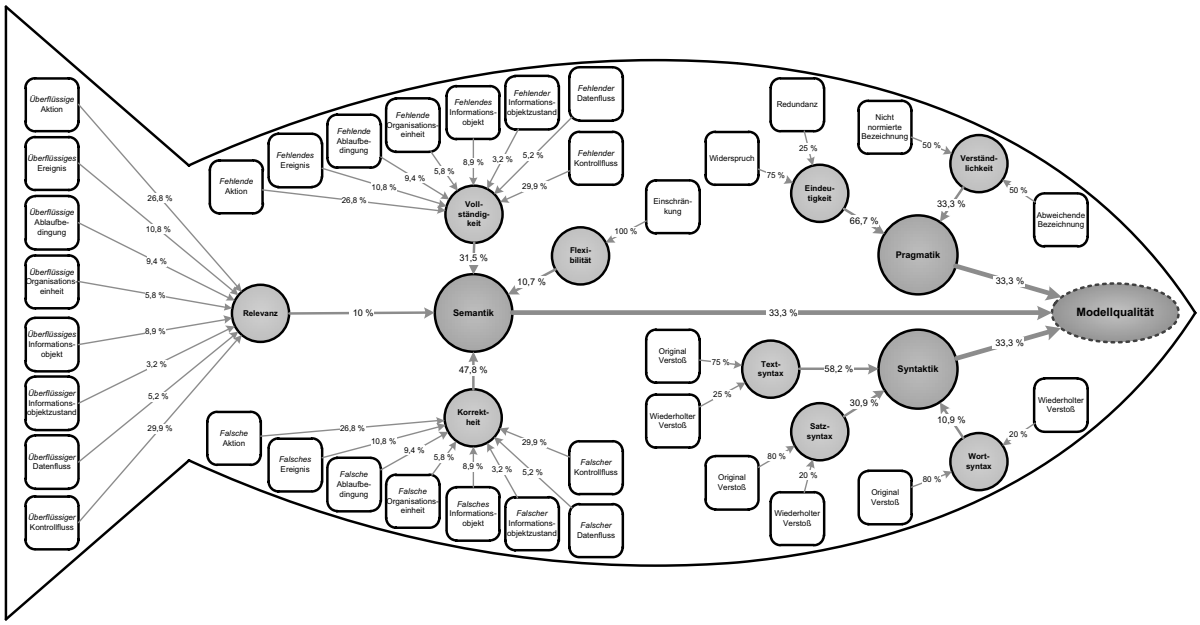


Figure 2: Qualitätsmodell des 3QM-Frameworks mit Kategorien (grau) und Metriken (weiß) [21]

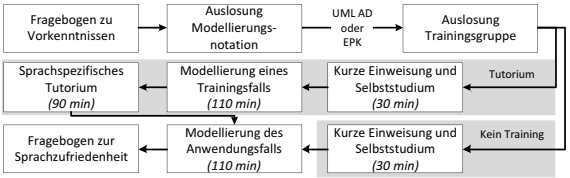


Figure 3: Ablauf des Experiments

der BWL/VWL im Hauptstudium – und somit zukünftige Fachanwender – herangezogen. Diese stellen in Bezug auf ihre Ausbildung einen gleichwertigen Ersatz für Fachanwender dar und bringen entscheidende Vorteile mit sich, um die Eignung von Modellierungssprachen zur Abbildung von Geschäftsprozessen unter kontrollierten Bedingungen zu beurteilen. So ist zu erwarten, dass die Ergebnisse nicht durch ein stark variierendes Wissen über den Anwendungsbereich und die Modellierungssprache verzerrt werden [8, 3, 27]. Ferner lassen sich Einsteiger eher auf neue Methoden ein als Experten, bei denen sich ein bestimmtes Vorgehen bereits tief eingeprägt hat [29]. Alle Teilnehmer wurden zufällig aus einer größeren Menge an Freiwilligen ausgewählt. Zu Beginn des Experiments wurden sie nach ihren bisherigen Erfahrungen mit Geschäftsprozessmodellen und den eingesetzten Sprachen befragt. Die darauf folgende Versuchsgruppeneinteilung wurde vollständig randomisiert vorgenommen. Alle Teilnehmer bekamen anfangs eine kurze Einweisung in die jeweilige Notation und konnten sich zusätzlich im Selbststudium mit der Sprache vertraut machen. Die Gruppen mit höherem Trainingslevel bekamen anschließend einen Testfall zur Bearbeitung, welcher in Länge und Komplexität dem späteren Anwendungsfall entsprach. Weiterhin wurde für diese Gruppen ein sprachspezifisches, 90-minütiges Tutorium durchgeführt, in dem neben den Grundlagen auch Hintergründe und Anwendungsdetails besprochen wurden.

Als Anwendungsfall erhielten alle Teilnehmer eine 641 Wörter umfassende, natürlichsprachliche Beschreibung eines Geschäftsprozesses zur Produkteinführung in einem Joint-Venture. Dieser Prozess ist im Wesentlichen an die realen Geschäftsprozesse eines führenden europäischen Großhändlers im Elektrotechnikbereich angelehnt und somit von praxisrelevanter Komplexität (siehe Abb. 4 und 5). Aufgabe der Probanden war es, den beschriebenen Prozess mit Stift und Zettel in der ihnen zugewiesenen Notation abzubilden. Erlaubte Hilfsmittel waren eine bereitgestellte Kurzbeschreibung der Sprache, sowie ggf. die Schulungsunterlagen aus dem Tutorium. In der abschließenden Umfrage wurden alle Probanden zur Sprach- und Prozessverständlichkeit, sowie zur Zufriedenheit mit der jeweiligen Notation befragt.

Primäre Zielgröße im Experiment ist die Qualität der erstellten Modelle. Die Bewertung erfolgte mithilfe von Referenzlösungen, die für jede Sprache erstellt wurden (Abb. 4, 5). Abweichungen der von den Probanden modellierten Prozesse wurden von drei Korrektoren beurteilt. Da bei jeder Abbildung der Realwelt auf ein Modell gewisse subjektive Freiheitsgrade bestehen, wurden nur Abweichungen, die über einen festgelegten Rahmen hinausgehen, als Fehler gezählt und jeweils genau einer Fehlerklasse (Abb. 2) zugeordnet. Insgesamt wurden 35 Einzelfehlerklassen erhoben und zu 13 Qualitätskategorien auf unterschiedlichen Ebenen aggregiert. Inklusive Ex-ante- und Ex-post-Befragung und unter Berücksichtigung der Bearbeitungszeit ergibt sich je Proband ein Datensatz mit 103 Variablen als Auswertungsgrundlage.

5. AUSWERTUNG

Die im Experiment erhobenen Daten wurden mit statistischen Methoden und Tools (*SPSS*, *R* und *GGobi*) ausgewertet. Hypothesentests im Rahmen einer ANOVA wurden

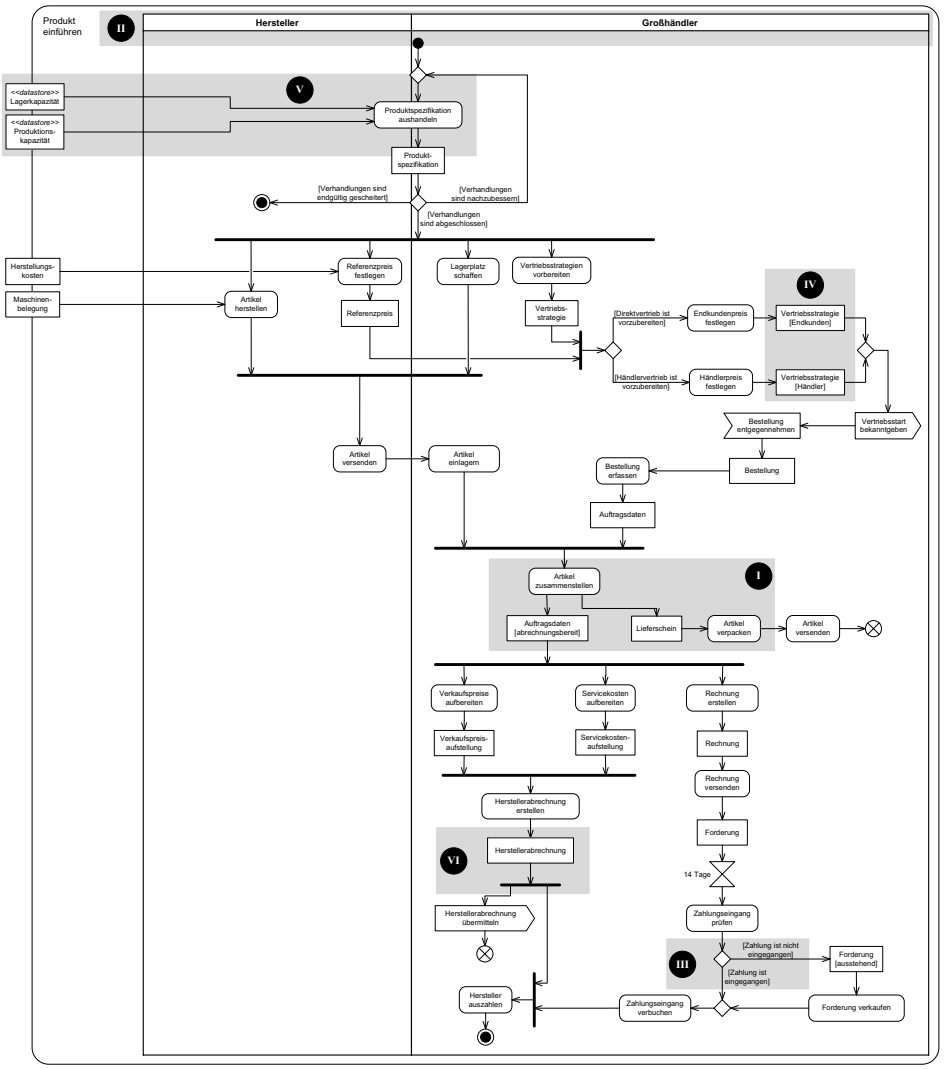


Figure 4: Referenzlösung für den Prozessablauf als AD

ergänzt um den t-Test für unabhängige Stichproben, sowie ggf. alternativ um den nichtparametrischen Mann-Whitney-U-Test. Zur Überprüfung von Voraussetzungen des t-Tests fanden der Kolmogorow-Smirnow-Test (Normalverteilungsannahme) und der Levene-Test (Varianzgleichheit) Anwendung. Tabelle 1 fasst die wichtigsten Ergebnisse sowie deskriptive Statistiken zu den Einzelfehlerklassen zusammen. Die folgende Analyse konzentriert sich auf die signifikanten Unterschiede zwischen den beiden Notationen und ihre Auswirkungen für den Fachanwender.

Bei der Auswertung des Ex-ante-Fragebogens ergeben sich keine signifikanten Unterschiede zwischen den Gruppen. Alle teilnehmenden Studenten hatten ähnliche Vorkenntnisse in konzeptioneller Modellierung, welche hauptsächlich im universitären Umfeld erworben wurden. Berufserfahrung hatten 40 der 73 Teilnehmer, die in der Mehrheit BWL im siebten Semester studierten und durchschnittlich 25 Jahre alt waren. Die ursprünglich gleichen Gruppengrößen nach der zufälligen Einteilung der Probanden haben sich im Verlauf des Experiments verschoben, so dass schlussendlich 36 AD und 37 EPK erstellt wurden, wobei je 19 Teilnehmer ohne ausführliches Training modellierten.

Mit den Elementen der beiden Sprachen konnten grundsätzlich alle Sachverhalte der Prozessbeschreibung abgebildet werden. Auffällig ist jedoch, dass der Fertigstellungsgrad bei den AD deutlich höher ist als bei den EPK. Während bei ersteren alle Prozessmodelle vollständig waren, haben 13,5% der EPK-Anwender ein unvollständiges Modell abgegeben. Dies ist zum einen auf die bipartite Darstellung von EPK-Graphen zurückzuführen, da sowohl die konsequente Modellierung von Ereignissen als auch von nicht prozessrelevanten Aktionen (sog. Hilfsfunktionen, Abb. 5, IV) nötig ist. So sind in der UML Referenzlösung (Abb. 4) 22 Aktionen und 4 Ereignisse vorgesehen, bei EPK dagegen 26 Funktionen und 30 Ereignisse (Abb. 5). Durch die Integration der Informationsobjekte in den Kontrollfluss kommt das AD mit 19

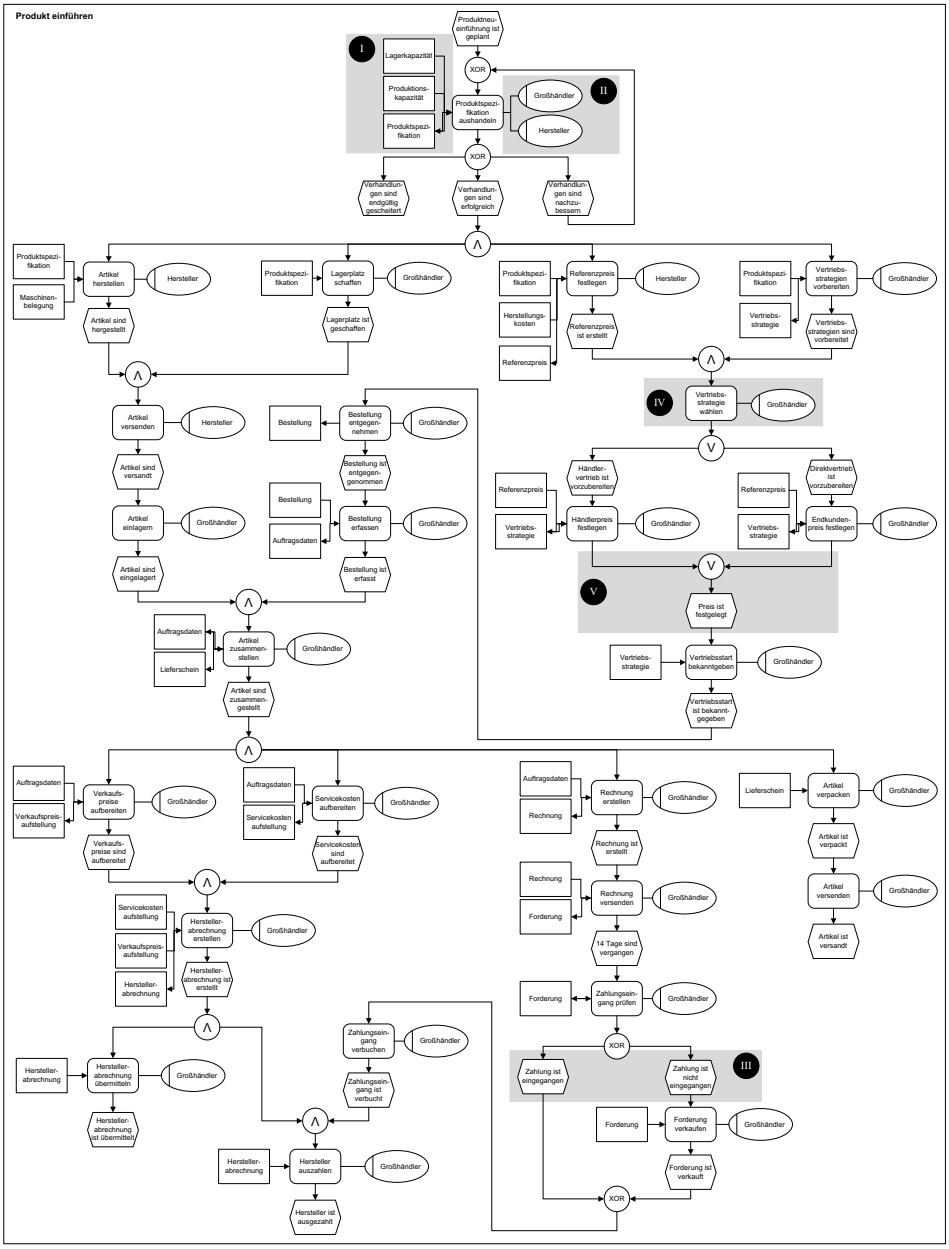


Figure 5: Referenzlösung für den Prozessablauf in EPK Notation

Modellelementen aus, wohingegen in der EPK 37 Informationsobjekte benötigt werden. Letztere ordnet weiterhin jeder Aktion eine eigene Organisationseinheit zu, während im AD allen Aktionen eines bestimmten Bereichs (sog. Swimlanes) eine Organisationseinheit zugewiesen wird (Abb. 4, II und 5, II). Entsprechend sind im Referenzprozess beim AD zwei Elemente und in der EPK 27 Elemente darzustellen.

5.1 Syntaktik

Im Vergleich zu EPK stellt das AD dem Nutzer eine höhere Anzahl an Elementen zur Verfügung. Im Referenzprozess wurden beim AD 16 verschiedene Notationselemente verwendet, in der EPK hingegen lediglich neun. Dies erklärt die signifikanten Unterschiede der Qualitätskategorie Wortsyntax, in der die falsche Darstellung existierender Elemente gemessen wird. Die höhere Anzahl an zur Verfügung stehenden Elementen bedingt auch mehr Fehlerquellen und spiegelt sich in den Originalverstößen wider. Weiterhin fordert die höhere Sprachkomplexität des AD, also die Handhabung von mehr Elementen, ein höheres Maß an Konzentration. Oftmals wurde deshalb bspw. vergessen, Bedingungen oder Informationsobjektzustände in eckigen Klammern zu modellieren (Abb. 4, III, IV). Die konsequente Missachtung dieser Vorschriften wird durch den signifikanten Unterschied der

Table 1: Testergebnisse und deskriptive Statistiken zu den Einzelfehlerklassen

7bdibbb7				B7di. du. d		B7d.			D7dkdp.iv7d.p.id.ik7.				
				NV [1]	HH) [2]	Teststatistik	Signifikanz	oornoil	Mi. .[Mr]	Ml x.[Mx]	Mi.[5M5]	M4d.d[Md]	Md.dAb.[Md]
*ybrakmk	Wdxmyb.	Oddi. pp	EPK	×	×	y ubb-Wxy-bey-W-Wu-yl-ykä	*,***[***]	Wyd	0	6	6,78	6	1,22
			UML	×		WyddW			U	U	LMU	UMI	UMU
		Wi7d7dh7lh. d	UM5	×	✓	y ubb-Wxy-bey-W-Wu-yl-ykä	*,***[***]	EPK	U	UU	UM9	U	3M3
			3 M4	×		Wydxd			3	34	4M4	4	3M3
	Wi-.l yb.	Wddi. hh	3M4	×	×	y ubb-Wxy-bey-W-Wu-yl-ykä	*,***[***]	Wy d	3	4	3M4	4	3M3
			3 M4	×		WyWxd			3	3	3M4	3	3M4
		Nicht signifikant:	Wyedexxdlube										
	Tex-l yb.	Wddi. hh	Ni c	×	✓	E PLLgWhgLLygUgStPtgotgk:	*,***[***]	EPK	N	c	NicN	N	NiNN
			Ni c	×		3P344			N	N	NiNh	N	Nicc
		Wi7d7dh7lh. d	Ni c	×	×	y ubb-T xy-bey-T-T-u-yl-ykä	*,**2[**]	EPK	N	N	NiNN	N	NiNN
			Ni c	×		TyTdd			N	c	NiNN	N	NiNN
Te- ub-yk	ydlll-äbdyDkDy	mk.i7.	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRdd[R]	Ry d	N	Nh	NiNh	c	NihN
			Ni c	×		dydxx			N	c	NiNN	N	Nicc
		mhlhhmh7di. dh. d	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RRR]	Ryd	N	c	NiNc	N	Nicc
			Ni c	×		RydRR			N	c	NihN	Nic	ciNc
		Eh.7. mhdd	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[R]	Ry d	N	N	NicN	N	NiNc
			Ni c	×		dydRd			N	N	NiNh	N	NicN
		E m7dmh.i7. d7hj7k.	Ni c	✓	×	--T-u-yl-ykä	0,000[***]	U0 L	N	NN	NNihN	NN	hicN
			Ni c	✓		Tyxdd			N	cN	ciNh	N	ciNc
		If foamatiof oobjektzuotaf d	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	*,***[***]	Ryd	N	N	NiNN	N	NiNN
			Ni c	×		RyRRR			N	N	NiNh	N	NiNh
		Nicht signifikant:	ErLgLgr Orgr Lgr tgLoLgLhLgr KoLtromfruoo										
	ddxxDk-xDy	K7. .d7lhmhdd	Ni c	×	×	y ubb-T xy-bey-T-T-u-yl-ykä	*,***[***]	Ryd	N	c	NihN	N	NiNN
			Ni c	✓		TyTdd			N	c	ciNN	c	Nihc
		Eh.7. mhdd	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RRR]	Ryd	N	N	NiNN	N	NiNN
			Ni c	×		dyddd			N	N	ciNN	c	NiNN
		R m7wmh.i7. w7hj7k.	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	0,089′	Ryd	N	N	NiNc	N	Nicc
			Ni c	✓		RydRR			N	c	Nihh	N	ciNh
		Nicht signifikant:	Rk-ydbyRxDyDbyl yOxDubyl u-ydbl DybxDy-yRbluufbDdybDubDyIbfdx- u-ydbl dbjDk-. ul -ubd										
	RDID. ub.	Ed7id. id	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RR]	Ry d	N	Nh	Nich	N	NicN
			Ni c	×		dyddR			N	NN	NiNc	N	Nich
		mhlhhmh7di. dh. d	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRx[R]	Ryd	N	N	NiNN	N	NiNN
			Ni c	×		dyRdR			N	N	NiNh	N	NicN
		Rddh. idh.i7. d7i. h7i.	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRd[RR]	*V d	N	N	NiNc	N	Nicc
			Ni c	×		RyRRd			N	N	NiNN	N	NiNN
		I. m7dmb.i7. d7bj7k.	Ni c	✓	×	--T-u-yl-ykä	RyRdR[R]	Ryd	N	Nc	cihN	c	NicN
			Ni c	✓		dydxT			N	cN	ciNN	Nic	cich
		K m7dmh.i7. d7hj7k.zhd.h. d	Ni c	×	×	y ubb-T xy-bey-T-T-u-yl-ykä	RyRRd[RR]	Rdy	N	N	NiNN	N	NiNN
			Ni c	×		dydxd			N	c	NicN	N	NicN
		Nicht signifikant:	Ak-ydbyDu-Dbflull yddb-xdllflull										
	Flexi.	Einschränkung	Ni c	×	✓	Mann-Wh-tnny-U-Stat-st-k:	RyRRR[RRR]	U5 L	N	NN	ciNh	c	ciNc
			Ni c	×		5,711			N	N	Nihc	N	NihN
yxuD- u-yk	Rybd.	E id7ddhdhch	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RRR]	Ryd	N	N	NiNN	N	NiNN
			Ni c	×		dyddd			N	N	ciNh	c	NicN
		R7dh. dh. z	Ni c	×	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RRR]	Ry d	N	N	ciNN	c	NiNc
			Ni c	×		dydRR			N	N	NicN	N	NihN
	y Dxl -.	hhw.vR7z7ich. h. w	Ni c	✓	×	y ubb-R xy-bDy-R-R-u-yl-ykä	RyRRR[RR]	Ry d	N	NN	NiNc	N	NiNh
			Ni c	×		dyxRd			N	c	Nihc	N	NicN
		Nicht signifikant:	Rydx- bdx- yDx-D BD. Dydx bubD										

[(()] (ormulvermeilunei [(oo]VurVunzr omoeenVrämV[((o]V VhVnumV[(oo]V uxVnumV[(oo]V Vmelwexm[((o]V edVunV[(dd]SmubduxdubweVdx ubeV[***] xddx l VebVfVkubm[**] l exx l VebVfVkubm[*] l VebVfVkubm[u] l dx wudx l VebVfVkubm

Kategorie Wiederholung dokumentiert.

Bei der Satzsyntax zeigt sich ein signifikanter Unterschied zu Gunsten des AD. So wurden bei EPK häufig bei der Aufteilung in bzw. der Zusammenführung von mehreren Kontrollflüssen zwei aufeinanderfolgende Funktionen oder Ereignisse modelliert. Ebenso wurde teilweise nicht berücksichtigt, dass bei EPK lediglich Funktionen Entscheidungskompetenz besitzen und folglich Verzweigungen auf Basis eines Ereignisses unzulässig sind (Abb. 5, IV). Zudem bereitete die EPK-Prozessschnittstelle den Teilnehmern deutlich mehr Probleme als das Konzept der Sprungmarke im AD. Insbesondere die Anbindung ersterer an ein Ereignis bzw. die Berücksichtigung des letzten Ereignisses bei der Fortführung nach der Prozessschnittstelle wurde häufig nicht bedacht.

In der Kategorie Textsyntax wurde vor allem die Zusammenführung mehrerer Kontrollflüsse mit einem nicht der Ablauflogik entsprechenden Verknüpfungsoperator festgestellt (bspw. mit einem Oder- statt einem Und-Konnektor nach vorausgegangener Parallelisierung). Hierbei wurden in beiden Einzelfehlerklassen beim AD signifikant mehr Verstöße gefunden. In der Literatur finden sich Hinweise darauf, dass die Verwendung expliziter Operatoren das Prozessverständnis fördert [25]. Zwar werden auch im AD explizite Verknüpfungsoperatoren verwendet, allerdings besteht zudem die Möglichkeit, den Kontrollfluss implizit zu parallelisieren (Abb. 4, I) oder zu synchronisieren. Gerade diese Optionen haben sich bei der Textsyntax negativ ausgewirkt. Die verkürzte Darstellungsform wurde von den Teilnehmern zwar angewandt, jedoch mitunter ohne sich der Token-Semantik bewusst zu sein. Weiterhin wurden auch bei Verwendung expliziter Verknüpfungsoperatoren mit dem AD mehr Textsyntaxfehler modelliert, weshalb die Verständlichkeit der Operatoren der EPK als höher anzusehen ist.

Insgesamt ist bei der Wort- und Textsyntax ein signifikanter Unterschied zu Gunsten der EPK festzustellen, bei der jedoch deutlich mehr Satzsyntaxfehler begangen wurden (Tabelle 2). Dennoch ergibt sich bei der Aggregation der Ergebnisse kein signifikanter Unterschied in der Syntaktik.

5.2 Semantik

Die Analyse der Semantikfehler identifiziert Informationsobjekte und Datenflüsse als Ursache für die größten Differenzen bei der Anwendung der Notationen. In AD fehlen signifikant weniger Informationsobjekte, dafür treten gleichzeitig signifikant mehr falsche und überflüssige auf. Diese Beobachtung kann auf die unterschiedliche Einbindung von Daten zurückgeführt werden. Während EPK die explizite Modellierung aller ein- und ausgehenden Informationsobjekte jeder Aktion (Abb. 5, I) erzwingen, ist dies zwar auch bei AD der Fall, allerdings werden hier die Elemente nur einmal aufgeführt. Generierte Informationsobjekte, welche nicht in der darauffolgenden Aktion benötigt werden, sind explizit aus dem Kontrollfluss auszugliedern (Abb. 4, I). Als Konsequenz werden in AD zwar seltener Daten vergessen, da in der Regel die Ausgabe einer Aktion gleichzeitig die Eingabe der nächsten ist, jedoch bedingt dies ebenso mehr falsche und überflüssige Informationsobjekte. Ein gleiches Bild ergibt sich für fehlende Datenflüsse, welche verursacht werden indem eine Aktion ein Informationsobjekt als Eingabe verwendet, welches zwar in der vorhergehenden Aktion generiert werden sollte, jedoch nicht modelliert wurde.

Informationsobjekte können durch Aktionen ihren Zustand verändern, wobei AD im Gegensatz zur EPK eine Möglichkeit bieten, dies explizit darzustellen (Abb. 4, IV). Folglich schneiden EPK bei den fehlenden und überflüssigen Statusinformationen signifikant besser ab. Auf der anderen Seite empfinden es EPK-Anwender jedoch scheinbar als notwendig, den Zustand in den Informationsobjekten oder auch Ereignissen zu notieren. Dies führte wiederum, neben der erhöhten Anzahl an Elementen insgesamt, zu signifikant mehr abweichenden Bezeichnungen für gleiche Objekte innerhalb eines Prozesses. Die Einführung eines expliziten Informationsobjektzustands in EPK erscheint demzufolge sinnvoll.

Weitere, in den Modellen teilweise fehlende Elemente sind Aktionen und Ablaufbedingungen. Bei ersteren schneiden AD signifikant besser ab, was insbesondere daran liegt, dass Einzelaktionen in EPK oftmals zu komplexen Funktionen aggregiert wurden. Die Ursache hierfür kann nicht eindeutig geklärt werden, da beide Gruppen die gleichen Ausgangsbedingungen hatten. Somit bleibt zu vermuten, dass die insgesamt höhere Anzahl zu modellierender Elemente EPK-Anwender zu Zusammenfassungen verleitete. Eine Untersuchung dieser Beobachtung in weiteren Studien steht noch aus. Auf der anderen Seite wurden Ablaufbedingungen, die in EPK als Ereignisse modelliert werden, dort signifikant seltener weggelassen. Durch die Bipartitheit wird der Anwender faktisch zu deren Modellierung gezwungen (Abb. 5, III), wohingegen Ablaufbedingungen in AD explizit an den Kontrollflusskanten ergänzt werden müssen (Abb. 4, III). Ein Nachteil der Bipartitheit kommt bei der Betrachtung überflüssiger Ereignisse zum Tragen. Obwohl beide Gruppen gleich häufig unnötig feingranulare Funktionen modellieren, wirkt sich dies bei EPK zusätzlich negativ auf die Anzahl der sprachlich notwendigen, jedoch überflüssigen Ereignisse aus. Weiterhin tendieren EPK-Anwender zu signifikant mehr überflüssigen Organisationseinheiten. Wie bereits gezeigt, werden diese in AD durch Swimlanes für mehrere Aktionen und in EPK durch Knoten an jeder Einzelaktion dargestellt. Zusätzliche Organisationseinheiten lassen sich somit in EPK einfacher ergänzen als komplette Swimlanes, die von den Anwendern restriktiver eingesetzt werden.

In Bezug auf die Eindeutigkeit der Modelle fallen hoch signifikante Unterschiede zwischen den Notationen auf. EPK-Modelle enthalten deutlich mehr Redundanzen als AD, was bei genauerer Betrachtung insbesondere auf redundante Ereignisse zurückzuführen ist, wohingegen gleiche Aktionen kaum mehrfach modelliert werden. In EPK-Modellen ist es hierbei entscheidend, wo genau Ereignisse platziert werden. Äquivalente Ereignisse vor der Zusammenführung von Kontrollflüssen werden bspw. als redundant angesehen, da stattdessen auch eine einmalige Platzierung dahinter eine Option ist (Abb. 5, V). Gleiches gilt für redundante Ereignisse nach Parallelisierungen, wohingegen vor Entscheidungsknoten kein Ereignis stehen darf. Diese Vielzahl zu beachtender Regelungen führt insgesamt zu einer erhöhten Zahl an Redundanzen bei den EPK Anwendern. Andererseits enthalten die UML-Modelle im Mittel erheblich mehr Widersprüche, was auf die angesprochene Token-Semantik zurückzuführen ist. Insbesondere die Tatsache, dass Informationsobjekte implizit mit genau einem Token belegt sind, welches bei konkurrierenden Aktionen von der zuerst ausgeführten konsumiert wird, verschafft den Studienteilnehmern in der UML-Gruppe Probleme. So muss bei Schleifen oftmals auf einen *datastore* zurückgegriffen werden (Abb. 4, V) und bei Verzweigungen ist auf die genaue Platzierung des Informationsobjekts zu achten, je nachdem ob es einer oder mehreren Aktionen als Eingabe dienen soll (Abb. 4, VI).

Bereits in der Vorstudie zeichnete sich eine unerwartete Tendenz ab, die in der vorliegenden Hauptstudie bestätigt werden konnte: Prozesse in UML-Notation sind signifikant flexibler modelliert als mit EPK. Flexibel bedeutet in diesem Zusammenhang, dass Nebenläufigkeiten von Prozesssträngen so weit wie möglich ausgenutzt werden. Für Unternehmen kann es ein entscheidender Vorteil sein, dass Prozesse nicht in starrer, sequentieller Folge ablaufen, sondern parallele Bearbeitungsmöglichkeiten so weit wie sinnvoll in Anspruch genommen werden. Die genaue Ursache für die Überlegenheit von AD in dieser Hinsicht kann nicht abschließend geklärt werden. Eine mögliche Ursache wären die vielfältigeren Möglichkeiten zur Parallelisierung und Sequentialisierung, welche sowohl über die expliziten Konnektoren, als auch über Aktionen selbst modelliert werden können. Auf der anderen Seite führt die erhöhte Flexibilität von UML-Prozessen gleichzeitig zu einer signifikant erhöhten Anzahl falscher Kontrollflüsse gegenüber EPK. Kontrollflüsse gelten als falsch, wenn diese zu früh auseinander laufen oder zu spät zusammengeführt werden, da hierdurch gravierende Fehler im Ablauf entstehen. Der Trend zur flexibleren Modellierung ist demzufolge Vor- und Nachteil zugleich, wegen der erhöhten Flexibilität einerseits und der Vermehrung an falschen Abläufen andererseits.

Tabelle 2 fasst die Testergebnisse für die aggregierten Qualitätskategorien zusammen. Neben den bereits besprochenen Signifikanzen in der Syntaktik finden sich auch hoch signifikante Unterschiede in den Unterkategorien der Semantik. AD führen zu vollständigeren Prozessen und damit allgemein zu weniger fehlenden Elementen. Andererseits zeigen EPK Vorzüge bei der Korrektheit und Einfachheit der modellierten Prozesse. Insgesamt muss jedoch beachtet werden, dass aggregierte Kategorien nur grundsätzliche Tendenzen aufzeigen und die Einzelergebnisse in beide Richtungen abweichen können. Aussagen auf den höheren Aggregationsstufen sind nicht mehr sinnvoll möglich.

Table 2: Testergebnisse zu den Qualitätskategorien

ol ril blo		Bodinr unr		Rorn		
		NV	HH	Teststatistik	Signifikanz	oornoil
Goruv n	EPK	✓	✓	t-ntathstley	0,383	t - n
	UML	✓		t -9n9		
Gynnukrik	UML	✓	✓	t-ntathstley	0,201	t - n
	UML	✓		a-n9n		
Wornrynrl x	UML	×	×	- ann-Whltney-t -ntathstley	0,000***	(Vo
	UML	✓		n-n9n		
Wncrynrl x	UML	✓	×	V wooWl l(oef V(Vo(w(lz(lef	*,**6**	(V o
	UML	×		oVoo(		
Eoxnrynnux	UML	×	×	V wooWl l(oef V(Vo(w(lz(lef	*,******	(Vo
	UML	×		oVcow		
Eov unrik	UML	✓	✓	(Vo(w(lz(lef	t -a9n	(V o
	UML	✓		wVooo		
oollrnändir koin	UML	✓	×	t-7tatistik:	t -t nn[t]	t - 7
	UML	✓		7-a77		
Korrokrhoin	UML	✓	×	t-7tatistik:	t -t t t[ttt]	t - 7
	UML	✓		7-977		
Rolovl nc	UML	✓	✓	t-7tatistik:	t -n9n	t - n
	UML	✓		t -555		
Fuoxiyiünyn	UML	×	×	- ann-Whltney-t -ntathstley	t -t t t[ttt]	t - 7
	UML	×		n-naa		
Prl r ml nik	UML	✓	✓	t-Statistik:	t -a44	t - 7
	UML	✓		a-48		
Eindounir koin	UML	×	×	- ann-Whltney-t -ntathstley	t -t t t[ttt]	t - 4
	UML	✓		n-nnn		
oorrnändlichkoin	UML	✓	×	t-4tatistik:	t -t 9t[a]	t - 4
	UML	✓		a-77		

[(a)] Wormmlver(elluogNWWmrlmozl omogeol(ä(WWocl Wgolflemo(WWel cWgolflemo(WWzlgolflemo(WWcl wwcl Wgolflewo(

5.3 Training und Zufriedenheit

Wie eine Kovarianzanalyse bestätigte, wirkte sich der Lerneffekt, der durch die Modellierung des Testfalls und den Besuch des Tutoriums entstanden ist, bei beiden Sprachen positiv aus. Dabei konnte festgestellt werden, dass unter den Teilnehmern mit Training die Unterschiede zwischen den Sprachen geringer sind. Insbesondere bei den Qualitätskategorien Syntaktik und Vollständigkeit sind bei beiden Sprachen deutliche Verbesserungen zu beobachten. Dies ist zum einen darauf zurückzuführen, dass die Teilnehmer mit Training Hinweise zum Abbildungsprozess erhielten und sich bereits in die Verwendung der Sprache eingefunden hatten. Zum anderen konnte im Rahmen des Tutoriums vor allem auf sprachgrammatische Besonderheiten eingegangen werden, wohingegen die Perzeption und damit das Prozessverständnis stets subjektiv und schwerer zu trainieren ist.

Eine Analyse der Zufriedenheit ergab mehrere signifikante Unterschiede. Vor allem bei den Fragen nach der Komplexität und der Übersichtlichkeit des Prozesses sowie zur benötigten Zeit konnten Vorteile zu Gunsten des AD festgestellt werden. Diese Erkenntnisse sind primär darauf zurückzuführen, dass mit der EPK wesentlich mehr Elemente benötigt werden, und bestätigen damit die bisherigen Erkenntnisse. Weitere Indizien hierfür sind, dass die Teilnehmer das AD als signifikant geeigneter empfanden, um den Prozess übersichtlich darzustellen und Sachverhalte sauber auszudrücken.

5.4 Rezeption der Ergebnisse

Die vorausgegangene Analyse der im Experiment erstellten Prozessmodelle zeigt keine Überlegenheit einer der beiden Sprachen in Bezug auf deren Eignung für Fachanwender. Vielmehr müssen die Ergebnisse entsprechend verschiedener Kriterien differenziert betrachtet werden. Die *Bipartitheit der EPK* stellte sich hierbei als ebenso nützlich wie problematisch heraus. Einerseits hilft sie dem Anwender dabei, wichtige Details wie etwa die korrekten Ablaufbedingungen zu berücksichtigen. Andererseits erhöht sich jedoch auch die Komplexität des Prozesses durch eine Vielzahl zusätzlich notwendiger Elemente. Letzteres erschwert es dem Anwender, den Überblick über das Modell zu behalten und resultiert häufig in Satzsyntaxfehlern durch Verletzung wichtiger Grammatikregeln der Sprache. Intuitiv würden viele Anwender auf künstlich formulierte Zwischenereignisse verzichten und diese stattdessen nur an prozessrelevanten Stellen modellieren. Ein großer Vorteil der EPK ist wiederum die *kompakte Notation*, die es den Teilnehmern ermöglichte, alle Sachverhalte mit nur neun Elementen abzubilden, was zu weniger wortsyntaktischen Fehlern führt. Lediglich die Einführung von expliziten Informationsobjektzuständen würde sich als hilfreich für den Anwender erweisen.

In AD ist die *Token-Semantik* und damit verbunden die implizite Konnektoreigenschaft von Aktionen ein Problem, da diese für den Fachanwender nur schwer zu verstehen ist und u.a. mehr Textsyntaxfehler verursacht. Indessen scheint sie jedoch auch der Grund für deutlich flexiblere Prozesse einerseits, und die erhöhte Zahl an Kontrollflussfehlern andererseits zu sein. Lässt sich Letzteres in den Griff bekommen, bieten sich für Unternehmen erhebliche Vorteile durch Identifikation und Nutzung von Parallelisierungspotenzialen in Geschäftsprozessen. Ein weiterer deutlicher Unterschied zwischen den Notationen zeichnet sich im *Umgang mit Informationsobjekten* ab, wobei weder AD noch EPK eine optimale Lösung bieten. Für den Fachanwender hat die Auswertung gezeigt, dass eine Trennung von Daten- und Kontrollfluss sinnvoll ist, ohne jedoch die Informationsobjekte mehrfach modellieren zu müssen. Eine mögliche Lösung könnte hier das Datenflusskonzept der Business Process Modeling Notation (BPMN) bieten.

Bei der Bewertung der vorgestellten Ergebnisse sind jedoch auch einige Einschränkungen zu berücksichtigen. Anzumerken ist zunächst, dass sich die Modellqualität durch den Einsatz von Werkzeugen insbesondere hinsichtlich der Syntaktik verbessern lässt. Da in diesem Beitrag die prinzipielle Eignung der Sprachen im Fokus stand, wurde eine Werkzeugunterstützung nicht betrachtet. Einschränkend ist ferner hervorzuheben, dass zur besseren Kontrolle von Störfaktoren auf Studenten als Studienteilnehmer zurückgegriffen wurde. Die gewonnenen Erkenntnisse sind deshalb stets in diesem Kontext zu würdigen und nicht automatisch auf die Realwelt übertragbar [27]. Inhaltlich ist hervorzuheben, dass zunächst lediglich die Erstellung von Modellen untersucht wurde. Um ein umfassenderes Bild zu gewinnen, sind weitere Einsatzgebiete und Faktoren zu analysieren, bspw. Trainingseffekte oder die Fähigkeit von Fachanwendern, Geschäftsprozessmodelle zu interpretieren [8]. Schließlich bleibt zu überprüfen, ob sich die vorgestellten Ergebnisse auch auf andere Anwendergruppen übertragen lassen.

6. SCHLUSSBETRACHTUNG

In diesem Beitrag wurde die Eignung von EPK und AD zur Abbildung von Geschäftsprozessen für Fachanwender untersucht. Grundlage der Untersuchung war eine empirische Studie, in deren Rahmen die praktische Anwendung beider Sprachen bei der Modellerstellung unter kontrollierten Bedingungen beobachtet wurde. Bei der anschließenden Untersuchung der erstellten Modelle konnten zahlreiche qualitative Abweichungen auf sprachliche Unterschiede zwischen EPK und AD zurückgeführt werden. Die Ergebnisse der Auswertung zeichnen insofern ein detailliertes Bild der Eignung beider Sprachen, Geschäftsprozesse in konzeptionelle Modelle abzubilden. Die vorgestellten Ergebnisse sind für Wissenschaft und Praxis gleichermaßen bedeutsam.

Aus praktischer Sicht tragen sie dazu bei, die Auswahl geeigneter Modellierungssprachen besser zu unterstützen. Zwar konnte keine prinzipielle Überlegenheit einer der Sprachen in Bezug auf die Modellerstellung ausgemacht werden. Vielmehr ergab sich eine differenzierte Übersicht der Stärken und Schwächen beider Sprachen. Gerade diese ermöglicht es jedoch, die Auswahl einer geeigneten Sprache situativ im jeweiligen Unternehmenskontext besser zu treffen und die Anwendung der Sprache wirksam zu unterstützen: ist die Flexibilität von Prozessen bspw. ein wichtiger Faktor bei der Modellierung, sind AD tendenziell gegenüber EPK zu bevorzugen. Da Fachanwender die Wortsyntax der wenigen EPK-Konstrukte jedoch leichter beherrschen, ist bei der Modellierung mit AD eine intensivere Unterstützung durch Analysten und Entwickler nötig.

Aus wissenschaftlicher Sicht wird mit den Erkenntnissen ein Beitrag zur Bildung von Theorien über die Eignung von Geschäftsprozessmodellierungssprachen geleistet. Einerseits können auf Basis der empirischen Ergebnisse neue Theorien abgeleitet werden. Andererseits können die Ergebnisse zur Überprüfung theoretischer Aussagen beitragen. So konnte übereinstimmend mit anderen empirischen Untersuchungen [25] bspw. festgestellt werden, dass die expliziten Konnektoren der EPK von Fachanwendern besser ge-

handhabt werden können als implizite Verzweigungen, die von AD und Petri-Netzen unterstützt werden. Die Tatsache, dass explizite Konnektoren keine direkte Entsprechung in der Realwelt haben, führt also nicht zu der in theoretischen Arbeiten vorhergesagten Beeinträchtigung der Anwendbarkeit (vgl. [24]). Entgegen den Erwartungen konnten die Verknüpfungsoperatoren der EPK von den Anwendern insgesamt besser gehandhabt werden als diejenigen der AD (vgl. [6]). Deshalb verbleibt zu untersuchen, wann und wie stark sich theoretisch begründete Schwächen in der praktischen Anwendung auswirken. Andere Vorhersagen konnten dagegen bestätigt werden. Bspw. konnte entsprechend dem Grundsatz der Sparsamkeit [17] beobachtet werden, dass die kleinere Zahl an Konstrukten bei der Modellierung mit EPK tatsächlich zu weniger Fehlern in der Wortsyntax führte.

Im Zuge der weiteren Forschung ist zunächst geplant, die externe Validität der vorgestellten Ergebnisse zu stärken, u.a. durch eine Variation der abzubildenden Prozesse sowie der teilnehmenden Probanden. Ferner soll das Design der Studie genutzt werden, um weitere Sprachen zu vergleichen. Im Sinne einer Langzeitforschung kann so insbesondere überprüft werden, ob neue Sprachen wie die BPMN Fortschritte für einzelne Anwendergruppen bringen.

7. REFERENCES

[1] J. Becker, M. Rosemann, and C. von Uthmann. Guidelines of Business Process Modeling. In *Proceedings of the BPM Conference*, pages 30–49, Berlin, 2000. Springer.

[2] J. Becker and R. Schütte. *Handelsinformationssysteme*. Redline, Frankfurt, 2004.

[3] A. Burton-Jones, Y. Wand, and R. Weber. Guidelines for Empirical Evaluations of Conceptual Modeling Grammars. *Journal of the AIS*, 10(6):495–532, 2009.

[4] I. Davies, M. Rosemann, and P. Green. Exploring Proposed Ontological Issues of ARIS with Four Different Types of Modellers. In *Proceedings of the ACIS*, 2004.

[5] P. Fettke and P. Loos. Ontologische Evaluierung von Ereignisgesteuerten Prozessketten. In *Proceedings of the EPK Workshop*, pages 61–78. Ges. f. Inform., 2003.

[6] K. Figl, J. Mendling, M. Strembeck, and J. Recker. On the Cognitive Effectiveness of Routing Symbols in Process Modeling Languages. In *Proceedings of the BIS Conference*, pages 230–241, Berlin, 2010. Springer.

[7] A. Gemino and Y. Wand. Evaluating Modeling Techniques based on Models of Learning. *Commun. ACM*, 46 (10):79–84, 2003.

[8] A. Gemino and Y. Wand. A Framework for Empirical Evaluation of Conceptual Modeling Techniques. *Requirements Engineering*, 9(4):248–260, 2004.

[9] P. Green and M. Rosemann. Integrated Process Modeling: An Ontological Evaluation. *Inform. Syst.*, 25(2):73 – 87, 2000.

[10] P. Green and M. Rosemann. Perceived Ontological Weaknesses Of Process Modeling Techniques: Further Evidence. In *Proceedings of the ECIS*, 2002.

[11] A. Gross and J. Dörr. EPC vs. UML Activity Diagrams - Two Experiments Examining their Usefulness for Requirements Engineering. In *Proceedings of the RE Conference*, pages 47–56. IEEE Computer Society, 2009.

[12] I. Hadar and P. Soffer. Variations in Conceptual Modeling: Classification and Ontological Analysis. *Journal of the AIS*, 7(8):569–593, 2006.

[13] S. Hillner, H. Kern, and S. Kühne. Berechnung von Modelldifferenzen als Basis für die Evolution von Prozessmodellen. In *Proceedings of the SE Conference*, pages 375–382, Bonn, 2009. Köllen.

[14] O. Lindland, G. Sindre, and A. Sølvberg. Understanding Quality in Conceptual Modeling. *IEEE Softw.*, 11:42–49, March 1994.

[15] M. McDonald, J. Begin, and S. Fortino. Meeting the Challenge: The 2009 CIO Agenda. Gartner Executive Programs, Gartner, Inc., 2009.

[16] D. Moody. Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models: Current State and Future Directions. *Data Knowl. Eng.*, 15(3):243–276, 2005.

[17] D. Moody. The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Trans. Softw. Eng.*, 35(6):756–779, 2009.

[18] D. Moody, G. Sindre, T. Brasethvik, and A. Sølvberg. Evaluating the Quality of Information Models: Empirical Testing of a Conceptual Model Quality Framework. In *Proceedings of the ICSE*, pages 295–305. IEEE Computer Society, 2003.

[19] M. Nüttgens. Ereignisgesteuerte Prozesskette. In K. Kurbel, J. Becker, N. Gronau, E. Sinz, and L. Suhl, editors, *Enzyklopädie der Wirtschaftsinformatik - Online-Lexikon*. Oldenbourg, 2010.

[20] OMG. Unified Modeling Language Specification: Version 2. Revised Final Adopted Specification ptc/05-07-04, Object Management Group, 2005.

[21] S. Overhage, D. Birkmeier, and S. Schlauderer. Qualitätsmerkmale, -metriken und -maßnahmen für Geschäftsprozessmodelle: Das 3QM-Framework. Forschungsbericht, Universität Augsburg, 2010.

[22] M. Owen and J. Raj. BPMN and Business Process Management - Introduction to the New Business Process Modeling Standard. Popkin Software, 2003.

[23] J. Recker. BPMN Modeling - Who, Where, How and Why. *BPTrends*, 5(3):1–8, 2008.

[24] J. Recker, M. Rosemann, M. Indulska, and P. Green. Business Process Modeling - A Comparative Analysis. *Journal of the AIS*, 10(4):333–363, 2009.

[25] K. Sarshar and P. Loos. Comparing the Control-Flow of EPC and Petri Net from the End-User Perspective. In *Proceedings of the BPM Conference*, pages 434–439, Berlin, 2005. Springer.

[26] S. Schnägelberger. Umfrage Status Quo Prozessmanagement 2008/2009. Analyst Report, BPM&O Architects GmbH, 2009.

[27] K. Siau and M. Rossi. Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis. *Info. Syst. J.*, 2007.

[28] H. Topi and V. Ramesh. Human Factors Research on Data Modeling: A Review of Prior Research, and Extended Framework and Future Research Directions. *Journal of Database Management*, 13(2):3–15, 2002.

[29] I. Vessey and S. A. Conger. Requirements Specification: Learning Object, Process, and Data

Methodologies. *Commun. ACM*, 37(5):102–113, 1994.

[30] M. Weske. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin, 2007.

[31] S. A. White. Process Modeling Notations and Workflow Patterns. In *Workflow Handbook 2004*, pages 265–294. Future Strategies Inc., 2004.

[32] P. Wohed, W. van der Aalst, M. Dumas, A. ter Hofstede, and N. Russel. On the Suitability of BPMN for Business Process Modelling. In *Proceedings of the BPM Conference*, pages 161–176. Springer, 2006.

Requirements of Process Modeling Languages – Results from an Empirical Investigation

Susanne Patig
University of Bern, IWI
Engehaldenstr. 8
CH-3012 Bern
susanne.patig@iwi.unibe.ch

Vanessa Casanova-Brito
University of Bern, IWI
Engehaldenstr. 8
CH-3012 Bern
vanessa.casanova-brito@iwi.unibe.ch

ABSTRACT
The majority of large and mid-sized companies are active in Business Process Management (BPM). Documenting business processes is a key task of BPM, but the variety of process modeling languages makes it difficult to determine 'the best' one. Basically, the suitability of a process modeling language depends on the companies' requirements. In this paper we adopt a bird's eye view on the issue: By an empirical investigation of 130 public companies from all over the world and any sector, we gather the common requirements of process modeling languages and use them to assess the most popular ones (i.e., BPMN, UML Activity Diagrams, Event-driven Process Chains). Our results show that these languages are (1) equally expressive and (2) presumably equally understandable concerning the common core notion of 'business process'; thus, they can be used interchangeably. However, the BPMN is the most complex process modeling language.

Categories and Subject Descriptors
D.2.1 [**Requirements**]: Languages; BPMN, UML, EPC

General Terms
Languages, Measurement

Keywords
Business process management, Requirements, process modeling languages, empirical investigation

1. INTRODUCTION
According to the BP Trends Report 2010, the majority of large and mid-sized companies in the world are active in business process management [31]. *Business process management (BPM)* comprises the design, administration, configuration, enactment and analysis of business processes [27]. All of these BPM activities require that the business processes are documented, which can be achieved by (business) *process modeling languages*; for brevity, we omit the term 'business' in the following. But, the large number of these languages confronts the companies with the problem of selecting 'the best one'. It is the aim of our research to contribute to the solution of this problem.

Several evaluations and comparisons of process modeling languages have been proposed; they are summarized in Section 2. The main weakness of the existing work is that the requirements, which are used to evaluate particular process modeling languages, stem from literature or BPM tools, but not from the companies' reality. Especially focusing on BPM tools is dangerous because this market is highly competitive, and some functionalities of BPM tools and features of process modeling languages rather satisfy the tool vendor's marketing department (and its attempts to diversify from other software solutions) than the customers' needs.

The research presented here overcomes this weakness: We have conducted an empirical investigation to gather the requirements concerning process modeling languages (see Section 3) and use them to assess the most popular process modeling languages (see Section 4). Our results (Section 5) are relieving for those trying to find 'the best' process modeling language.

2. RESEARCH BACKGROUND
Evaluations of process modeling languages either account for language pluralism or concentrate on the Business Process Modeling Notation (BPMN) [14] as the current de facto standard [31] of process modeling.

The pluralism-driven research addresses the question of 'the best' process modeling language as measured by expressiveness, understandability or complexity. *Expressiveness* means the capability to represent any meaning intended for some purpose, and it is usually assessed based on a reference [15]. In the field of process modeling languages, two types of *references* can be identified (see also Table 11 in Section 4.2):

1) *Widespread evaluation frameworks*[1] such as the Bunge-Wand-Weber (BWW) representation model (e.g., used in [5]) or different types of workflow patterns (e.g., used in [22]), and
2) *New evaluation frameworks* that were derived from the process modeling literature, e.g., [11], [24], [9].

The outcomes of such evaluations are either lists containing the equivalent, incomplete, overloaded, redundant or excess constructs of the analyzed process modeling languages [5] or evaluation vectors whose symbols indicate to which degree the elements of the reference are supported by the respective process modeling languages (e.g., [22], [9], [11]). Usually, both types of outcome convey the impression that neither process modeling language is 'good enough' in the sense of expressiveness. However,

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

[1] For a comparison between the ontology- and the pattern-based evaluation of process modeling languages see [20].

this impression is misleading as the evaluation references do not reflect the companies' requirements, but theoretical literature [5], [11], [24], [9] or functionality provided by BPM tools [22] - and provided functionality is neither necessarily needed nor used [17].

Research on the 'best' process modeling language in terms of *understandability* has just started out [12], [18]. Though no definite answer can be given yet, first findings indicate that process modeling languages are equally expressive as far as the common notion of 'business process' is concerned [18]. In contrast, if the 'best' process modeling language is the one with the smallest *complexity*, then UML activity diagrams are superior to the BPMN [21]. The complexity was calculated from the numbers of objects, relationships and properties in the meta models of the process modeling languages.

The standard-oriented evaluation research concentrates on the BPMN as the most widespread process modeling language [31] and tries to find out how good it is in real-life conditions. Once more, an important quality criterion is expressiveness: So, a series of interviews with nineteen participants from Australian organizations was conducted to test hypotheses about missing, redundant, superfluous and overloaded BPMN constructs [19]. The hypotheses were derived from the BWW representation model, and most of them had no or limited support in practice. On the one hand, these findings challenge the BWW representation model as a basis to assess process modeling languages; on the other hand they are first hints at the constructs needed in practice to model processes. We will discuss the results in more detail in Section 4.1.

Other research efforts analyze the use of the BPMN in real-life projects to identify unused constructs as well as the most frequently used ones [32]. It turned out that only 20% of the BPMN vocabulary appears in process models (see Section 4.1 for details). However, this analysis was conducted ex post facto and relied on given process models. Thus, it does not provide information on the constructs that would have been needed to adequately represent the companies' processes.

Altogether, the existing research does not help companies in selecting the process modeling language to be used, mainly for the following reasons: First, the academic way of assessing expressiveness does not start from the companies' requirements. Secondly, many of the process modeling languages evaluated in [9], [11], and [24] are not usable in large or mid-sized companies because appropriate tool support is missing. Though this argument does not apply to the BPMN, it is just one process modeling language among so many others. Recommendations, thirdly, should be given language-neutral, starting from the characteristics of the companies' processes and the resulting requirements. To remove these deficiencies, we have conducted an empirical investigation that is described in the next section.

3. EMPIRICAL INVESTIGATION ON THE REQUIREMENTS OF BUSINESS PROCESS MODELING

3.1 Research Goal

Our investigation adopts a bird's eye view on the common requirements of process modeling (i.e., irrespective of the individual companies' specifics) to enable a general assessment whether or not the current process modeling languages are sufficient and which one is possibly superior.

Since recent research findings indicate that there may be no significant difference in the understandability of process modeling languages if a common notion of 'business process' is kept (see Section 2), we concentrate on this common notion and try to define it more precisely. In particular, we want to find out what must be represented by process modeling languages to describe real-life business processes. By *business process* we mean a set of linked activities that collectively realize a business objective or policy goal [29]. Section 3.2 describes how we proceeded to reach this goal.

3.2 Method

Participants: The basic population for our investigation consisted of the companies from the 'Forbes Global 2000' list [3]. This list is an annual ranking of the top 2,000 public companies in the world based on sales, profit, assets and market value. As automation of process execution is the fourth most important reason to model processes [7], we require 'e-readiness'. *E-Readiness* describes the ability of a country and its businesses to use information and communication technology to their benefits [2]. The 'e-readiness' of 70 countries is assessed yearly on a scale between 1 (lowest) and 10 [2]. Since our investigation was conducted in 2009, we used the e-readiness rating of the year 2008, for which the average e-readiness score amounted to 6.4. All companies that are headquartered in countries with below-average e-readiness were discounted. Altogether, we draw a sample of 1,172 companies (by random numbers), which were contacted as described below.

Table 1. Numbers of responses per country

Countries	Responses	
	Per country	Total
United States	17	17
Switzerland	15	15
Germany; Japan	11	22
UK; Canada	9	18
Australia; France; Spain	5	15
Hong Kong; Italy; Portugal; Sweden; Taiwan	4	20
Austria; Belgium	3	6
Netherlands; Singapore; Greece; Denmark; Ireland; Korea; Norway	2	14
New Zealand; Luxemburg; Iceland	1	3
Sum N		130

In total, N = 130 companies responded (see Table 1); so, the response rate was 11%. Some questions were not answered by all companies; the resulting *missing values* are represented by '*na*' in the tables of this paper. Companies whose headquarters are located in Europe account for the majority (58%) of responses, followed by companies from North America (20%), Asia (18%) and Oceania (5%). All sectors are represented. Most responses came from the banking sector (19%), followed by utilities (8%), transportation and insurance (7% in each case) as well as oil & gas, and technology hardware (6% in each case). Some of the companies operate in more than one sector.

Half (53%) the companies' representatives who answered the questionnaire work in IT departments; other affiliations include departments for BPM (22.9%) or for company organization (13.3%), functional areas (9.6%) and product divisions (1.2%). In addition to knowledge about the processes of their own departments (29.4%), the participants stated that they also knew the processes of other departments (23.9%) or even have a company-wide picture (46.7%) of the processes.

Materials: The questionnaire consisted of 42 questions that were grouped in several sections dealing with BPM (current status, tools), processes (characteristics, change and statistics), process modeling (procedure, languages) and socio-demographic information. In this paper we concentrate on process modeling.

All questions of the questionnaire were partially open-ended, i.e., they provided a list with alternatives as well as an alternative 'other' to enter free text for unanticipated answers. The provided alternatives were derived from the process modeling literature (e.g. [1], [25], [6], [27]) and existing standards [28], [29]. Some questions were optional. The data was collected on nominal scales (participants were asked to select *all* alternatives that applied to their case) or on ordinal scale (participants were asked to rate some alternative or to make a ranking). All rating scales had four levels (to avoid neutral answers) and an additional level (e.g., 'Don't know', 'Not applicable', 'Not needed', 'None') to avoid forced ratings [8]. The resulting total number of five rating levels is generally reckoned optimal since the ability to differentiate between ratings decreases with an increasing number of rating levels [8]. In detail, we have used the following *rating scale types*:

- *Scale 1*: 1=Essential, 2=Frequently needed, 3=Occasionally needed, 4=Rarely needed, 5=Not at all needed
- *Scale 2*: 1=Always, 2=Very Often, 3=Sometimes, 4=Rarely, 5=Never
- *Scale 3*: 1=All processes, 2=Most processes, 3=Some processes, 4=A few processes, 5=No processes
- *Scale 4*: 1=Very important, 2=Important, 3=Not so important, 4=Not at all important, 5=Don't know

The questionnaire, including definitions of key terms used in the questions, was written in English and implemented as an online form with the 'LimeSurvey' tool [10]. To contact the companies, we sent letters by surface mail containing the link to the online form and explaining the goals and importance of our investigation. These letters were written in English, German, Spanish and Japanese – depending on the location of the company's headquarter – to make the goals and importance of our research understandable for the addressees.

Procedure: We conducted a pretest of the questionnaire with 10 BPM experts from practice and academia to check the questions for understandability and unambiguity of responses; afterwards, the questionnaire was revised. As little is known about the internal organization of BPM in a company, we sent the letters with the link to the revised questionnaire to the CIOs or CEOs of the companies in the sample and asked them to forward the letters to the persons responsible for BPM. After four to six weeks, we phoned the offices of the CIOs or CEOs to inquire after the status of our information request. After six to eight weeks, we sent reminder letters by surface mail. The survey was conducted from January to December 2009. No incentives were given; the companies were only offered the opportunity to obtain the results of the investigation free of charge.

3.3 Results

The wording of the questions (Q1) to (Q17) discussed here is given in the Appendix. As our questions on *nominal scale* ('choose *all* alternatives that apply …') allow more than one answer, the sum of counts c_i for an alternative (answer) can exceed the number N of responding companies. The answer alternative with the highest count (absolute frequency) represents the *mode* [4]. In the text, we give the count c_i for each alternative and the *percentage of responses* (c_i / N), which relates the count of an alternative to the number N of responses ($N \leq 130$, depending on the number *na* of *missing values*). If questions involved entering *open text*, this was optional, and we refer to the total number of received answers (Σc_o). For the open text questions, we state the *percentage of answers* ($c_i / \Sigma c_o$). Questions on *ordinal scale* can be recognized by the rating scale type from Section 3.2. According to this rating scale type, a numerical value is associated with each rating level to calculate the mean rating (μ) and the standard deviation (σ) of ratings; the numerical values 1 and 5 correspond to the highest and the lowest rating, respectively.

Naturally, the requirements concerning process modeling languages depend on the characteristics of the processes to be described. For that reason we first gathered information on the nature of the companies' processes (Questions (Q1) to (Q6)).

Concerning their scope **(Q1)**, the processes in the companies of our survey are related to (N=130, na=0) products (78/60.9%), administration (71/55.5%), customer contact (67/52.3%), system integration (66/51.6%), system development (65/50.8%), emergency procedures (33/25.8%) and other things (12/9.4%). To check whether the understanding of what constitutes a business process agrees among the companies, we asked the participants (as an open text question) to freely list at least one title or short description of a typical process **(Q2)**. Having obtained $\Sigma c_o = 265$ short descriptions, we clustered similar answers. The clusters of the most frequently mentioned titles can be summarized as 'Order-to-pay' (18/7%) and 'Purchase-to-Pay' (17/6%). The next smaller clusters are related to project management (9/3%), human resources issues (9/3%) and incident management (8/3%). Other process titles mentioned characterize development processes (7/3%) and change management processes (7/3%). Across all clusters, 7 titles (3%) emphasize that the processes involve approval. Altogether, the entered answers to Question (Q2) basically confirm our results of Question (Q1).

Table 2. Nature of processes according to modes

Question	Rating			N/na
	Most	Some	A few	
Distribution of processes (Q3)	1 Company & >1 Department (83)	1 Department & >1 Person (51); >1 Company (55)	1 Department & 1 Person (60)	130/0
Measure of average process run time (Q4)	In days (54)	In weeks (49)	In month (34); in years (30)	87/43
Execution frequencies of processes (Q5)	Several times a day (59)	Several times a week (58) / a month (56)	Several times a year (50)	102/ 128

In addition to the process scope, we inquired after the distribution of the processes **(Q3)**, the average run time **(Q4)** and execution frequency **(Q5)**, see Table 2, as well as process statistics (Q6). The cells of Table 2 show all alternatives of the questions and their most frequent rating (i.e., the mode). The rating scale (Type

3) corresponds to the columns; the particular counts are given in brackets. Obviously, most processes stay within one company, but span several departments. The processes are executed several times a day and have a run time that is measured in days.

Table 3 gives the statistics for the companies' average processes **(Q6)**. The numbers confirm Table 2 in that the processes involve few other companies (μ = 1.60), but usually several departments (μ = 3.70).

Table 3. Statistics for an average process (N=128, na=2)

(Q6) Please estimate the…	Min	Max	μ	σ	N/na
Number of involved persons from the same department	1	55	7.31	11.59	70/60
Number of involved departments	1	18	3.70	2.56	73/57
Number of other companies involved	0	8	1.60	1.39	58/72
Number of applications involved	1	24	4.27	4.22	69/61
Number of tasks	3	120	19.09	20.82	60/70

The aim of the Questions (Q1) to (Q6) was to get an idea of the processes in the surveyed companies – also to enable the proper assessment of the companies' answers concerning process modeling. We were mainly interested in the concepts needed to describe processes (Questions (Q8) to (Q16)). These concepts may not only depend on the processes' characteristics, but also on the modeling style. Therefore we first asked how the companies proceed in modeling processes **(Q7)**, in particular, what constitutes the starting point for creating process models. According to their responses (N=120; na=10), the companies start to model processes from:

1. The interactions (including data flow) between departments or responsible persons (72/60%)
2. Some goal and the activities to achieve it (57/48%)
3. The products of the company and their transformation from an initial to a final form (44/37%)

Several companies (Σc_o =17) named other procedures: Most often, 'top-down approaches' (4/3%) and 'customer touch points' (4/3%) were mentioned, followed by the reuse of results from existing ERP projects (3/2.5%). Top-down approaches are similar to our alternative 'starting from goals', whereas customer touch points are close to 'interactions'.

Our research goal was to find out which concepts companies need to describe their processes, and we posed this as a question **(Q8)**. Table 4 shows the results (rating scale of Type 1). Tasks are the by far most important concept, followed by relationships, persons or roles, documents, events and resources. Time information and timely distances between tasks are rated as least important. When asked openly to list other concepts needed (Σc_o =19), constructs to describe exceptions and exception handling (3/2%) were mentioned most often, immediately followed by business rules (2/2%) and the data flow between tasks (2/2%). Here, the companies did not recognize that exceptions are a special type of events.

Table 4. Concepts needed to describe processes in the company/department (N= 130, na=0)

Answer	μ	σ	Rank
Tasks to be done in a process	1.34	.63	1
Relationships between tasks	1.60	.84	2
Persons or roles who execute the tasks	1.68	.86	3
Documents or objects relevant for a task	1.91	.84	4
Events that happen and influence a task	2.22	.86	5
Resources needed to fulfill the tasks	2.24	.89	6
Time information related to a task	2.42	.89	7
Timely distances between tasks	2.83	.94	8

The previous question has shown that tasks and their relationships are the most important concepts. Question **(Q9)** investigated the typical relationships between tasks in real-life processes; Table 5 gives the results (rating scale of Type 2) for each alternative. The most frequent relationships between tasks are sequences, followed by start conditions (applicable to any form of relationship) and process splits. Loops and alternative tasks are rare.

Table 5. Frequency of observations that apply to tasks in processes (N= 130, na=0)

Answer	μ	σ	Rank
Tasks follow in strict sequence	2.32	.72	1
Condition must be satisfied in order to start a task	2.40	.64	2
Task has more than one immediate successor (split)	2.47	.61	3
Task has more than one immediate predecessor (merge)	2.58	.64	4
Task is repeated till some criterion is satisfied (looping)	2.82	.78	5
Tasks are alternative to each other	3.16	.82	6

According to Question (Q9) and Table 5, start conditions are common. We also inquired into the nature of these conditions **(Q10)**. As reported by the companies (N=129; na=1), start conditions relate to:

- The results of adjacent tasks (92/71%),
- The overall state of the process (82/64%),
- The task only (e.g. the availability of resources) (70/54%)
- Information external to the process (60/47%)
- Time (50/39%)

Splits, which are the third most important relationship between tasks (see Question (Q9) and Table 5), raise the situation that a task has several successors. Table 6 summarizes the companies' answers (rating scale of Type 2) about what happens with the successors **(Q11)**. Mostly, all of the successors are executed, either concurrently (with synchronization) or independent of each other (without synchronization). If not all succeeding tasks are executed, the selection mostly depends on a condition (see Question (Q10) for the nature of conditions). In line with the results of Question (Q9), exclusive choices are not very frequent.

Table 6. Execution of succeeding tasks

Answer	N/na	μ	σ	Rank
All (AND) … and concurrently	108/22	2.64	.76	1
….independent of each other	109/21	2.97	.70	2
Not all tasks are executed (OR)	106/24	3.25	.85	3
Which tasks are executed depends on…				
… a random selection	99/31	4.09	.87	3-IV
… some selection condition is satisfied	107/23	2.55	.73	3-I
… subjective experience	106/24	3.22	.82	3-II
As soon as one task has been executed, no other task is performed (XOR)	98/32	3.29	.92	3-III

In addition to tasks and their relationships, persons, events and resources are important to describe processes (see Question (Q8), Table 4). In the questionnaire, these concepts were investigated in more detail by the Questions (Q12) to (Q14):

Table 7 lists the resources companies wish to include in process models (Question **(Q12)**, rating scale of Type 1). Information is most frequently needed, followed by personnel and documents. Note that 'information' (e.g., a database listing the availability of items in store) and 'document' do not necessarily coincide.

Table 7. Resources needed in process models (N= 130, na=0)

Answer	μ	σ	Rank
Information	1.53	.61	1
Personnel	1.72	.86	2
Documents	1.77	.70	3
Software systems	1.98	.95	4
Material	2.75	1.11	5
Machines, appliances	3.04	1.13	6

If persons are used in process models, the companies would like to represent the following details (Question **(Q13)**, N=130; na=0):

1. Required capabilities, roles, positions (110/85%)
2. Organizational units the persons belong to (94/72%)
3. Names (46/35%)
4. None information at all (5/4%)

When asked to mention other details wished to be included, (2/14%) of the respondents (Σc_o =10) mentioned the availability of the person executing the task.

Finally, information is not only an important resource (see Question (Q12), Table 7), but also the most important trigger of processes or tasks – Table 8 shows the results of Question **(Q14)** on a rating scale of Type 3. Other common event types are timing and human intervention.

Table 8. Triggers of processes or tasks (N= 130, na=0)

Answer	μ	σ	Rank
Company internal *information*	2.58	.87	1
Information from business partners	2.62	.81	2
Timing (date, cycles)	2.82	.89	3
Human judgment and intervention	2.84	.79	4
State of process or task	3.10	.87	5
Deviations from targets or failures	3.45	.92	6
Signals from machines and sensors	3.91	.81	7

The Questions (Q8) to (Q14) dealt with constructs provided by most process modeling languages. We also wanted to know **(Q15)** whether the companies need additional concepts to describe their processes or tasks. As it can be seen in Table 9, the execution status (started, cancelled, etc.) is the concept most needed for the task, followed by priorities, goals, cost and planning status, whereas goals are the most important information for processes. The open Question **(Q16)** aimed at getting examples for process goals. Among the Σc_o =127 answers obtained, timing goals (19/15%) were the most frequently mentioned category. Examples of timing goals are the reduction of cycle time, throughput time etc. Timing goals were followed by customer satisfaction (8/6%) and quality-related goals (7/5%). Other named goals included the resolution of problems (6/4%) as well as the reduction of error rates (5/4%).

Finally, Table 10 summarizes how the sampled companies currently document their processes **(Q17)**: Most companies combine text (55.9%) and some (modeling) language (altogether 55.9%), but also tables are widespread (31.5%). Among the process modeling languages, the BPMN dominates, followed by the Unified Modeling Language (UML) and Event-driven Process Chains (EPC). Other ways (Σc_o =124) of describing processes are value chains (3/2%), flow charts (3/2%) and company-specific notations (4/2%).

Table 9. Additional concepts needed in process models (N=111, na=19)

Answer	Execution status		Priorities		Goals		Cost		Planning status	
	Task	Process	Task	Process	Task	Process	Task	Process	Task	Process
Needed (c_i)	71	68	61	54	54	80	48	58	44	56
Percentage responses (c_i / N)	64%	61%	55%	49%	49%	72%	43%	52%	40%	50%
Rank	1	2	2	5	3	1	4	3	5	4

Table 10. Documentation of processes (N = 127; na = 3)

Answer		Count c_i	Percentage responses (c_i / N)
As text		71	55.9%
As tables		40	31.5%
With languages	BPMN	27	21.3%
	UML	19	15.0%
	EPC	16	12.6%
	BPEL	5	3.9%
	IDEF	4	3.1%
Other		14	11.02%

4. DISCUSSION

4.1 Comparison with other Empirical Results

In this section we compare our findings concerning the concepts needed to describe processes with other empirical results that have been obtained for the BPMN: Based on an analysis of 120 BPMN models, it was revealed that the most often used BPMN constructs are the following (ordered by decreasing occurrence frequency): sequence flow, task, end/start/general event, pool (to express organizational information or roles) and gateway [32]. Our Table 4 agrees with these results in the (naturally) high importance of the construct 'task' (Rank 1 in Table 4 as opposed to Rank 2 in [32]), but ranks organizational information (persons, roles) higher than events. Moreover, there is also some agreement concerning the relationships between tasks: Though the occurrence frequency of the gateway types (AND occurs more frequently than XOR [32]) depends on the particular processes modeled, our results are analogous (see Table 6). Loops of activities play a subordinate role in process models, both according to our results (Table 5) and according to [32].

The contradictions stated above probably result from the fact that the ranking obtained in [32] also reflects the BPMN syntax and not only process modeling requirements. For example, the sequence flow is the most frequent construct in BPMN models [32] because it is needed to connect tasks and events, which, however, have to be defined before. Moreover, the comparatively low importance of events in our results can be explained by the fact that we inquired about the constructs needed to describe processes irrespective of any modeling language (also textual descriptions were allowed); thus, our results are free from any syntactic restrictions that require the use of events (e.g., start and end events).

Interestingly, the dominance of the task in our results cannot be derived from the modeling style since most companies use the interactions between departments as the starting point to model processes (Q7).

By testing hypotheses derived from the BWW representation model, another research [19] also aimed at identifying needed or excess (and, thus, unused) BPMN constructs. The following was discovered [19]: First, there is only moderate need for business rules, which supports our results of Question (Q8) where business rules were mentioned only two times as 'other' concepts. Secondly, data objects turned out to be the fourth most useful of the hypothetically 'excess' BPMN constructs, which agrees with the fourth rank of documents in our Table 4. Thirdly, activity looping rated comparatively high in perceived usefulness, which contradicts both our results in Table 5 and [32]. Altogether, the results of [19] are restricted by the BWW representation model and the BPMN, and our research is not subject to these restrictions; thus, a detailed comparison is difficult.

Our research augments the existing results for the BPMN by a language-neutral view, and language neutrality explains most of the deviations from the construct rankings observed by other researchers. Whether or not construct rankings are important for selecting among process modeling languages is discussed in the next section.

4.2 Comparison with Process Modeling Languages and Evaluation Frameworks

Our empirical investigation (see Section 3.1) aimed at discovering the concepts that must be represented by process modeling languages to describe real-life processes. The first column of Table 11 summarizes our results. The order of the concepts in bold print corresponds to the ranks from Table 4. Most of these concepts were investigated in more detail by further questions (e.g., Question (Q9) for relationships; see Section 3.3), whose numbers are given and whose rankings determine the order of the more detailed concepts. Concepts with a rounded mean rank below three are omitted in Table 11.

The final goal of our research (see Section 1) was to support companies in selecting 'the best' process modeling language. Our empirical investigation enables an assessment of process modeling languages by the required expressiveness. Valid references for expressiveness must be consistent sets of independent, atomic statements [15]. For that reason the first column of Table 11 differs as follows from the tables in Section 3.3:

- Personnel and documents are no subtypes of the concept 'resource' (see Table 7), but distinct concepts (see Table 4).
- Information (see Table 7) is not explicitly listed as a resource because it is provided by software systems, documents or events, which are separate concepts.
- Time information for tasks (see Table 9) does not appear in Table 11 as it can be expressed by time events for durations and deadlines (see also [14]).
- The concept ‚material' (Table 7) can be seen as a special type of ‚object' and is, thus, omitted in Table 11.

Altogether, the first column of Table 11 is a valid reference for measuring expressiveness. The expressiveness of the process modeling languages that were most frequently used in our sample (see Table 10 in Section 3.3) is apparent from Table 11: We count a line of the reference (*reference item*) as expressible (count: 1) by a process modeling language if it is directly or indirectly supported. *Direct support* means that the process modeling language provides a construct that readily represents the semantics of the reference item, whereas *indirect support* requires the combination or appropriate definition of existing constructs. Indirect support by adapting more general constructs is typical for the reference items representing detailed concepts in Table 11. For example, the general constructs 'pool' and 'lane' of the BPMN or 'partition' of UML Activity Diagrams can be used to express organizational units, capabilities or even names. Moreover, the generic construct 'event' of the EPC can be used to represent all specific event types such as messages, timing etc.

Table 11. Expressiveness of prominent process modeling languages and evaluation frameworks

Our empirical results Concepts needed to describe processes (Q8)	**Existing process modelling languages**			**Standards**		**Evaluation frameworks**				
						Newly proposed			**Widespread**	
	BPMN [14]	UML Activity [13]	EPC for ARIS [23]	WfMC [28]	XPDL [30]	[11]	[24]	[9]	BWW [5]	WF-Pattern [22] [26]
Tasks	✓	✓	✓	Activity	Activity	Activity	Activity	Activity	Trans-formation	♦
Relationships between tasks (Q9)	✓	✓	✓	Transition	Transition	Relation	Depen-dency	Control flow	State law	♦
Sequence	✓	✓	✓	Transition	Transition	Relation	✓	Control flow	♦ State law	WFCP-1
Start Condition (of task)	♦	✓	♦	Pre-condition	TC	(Behavior)	♦	—	Lawful trans-formation	WFDP-33, 34, 35, 38, 39
Split (Q11)	✓	✓	✓	TC	Route-A	Behavior	♦	✓	— (s. GrRo)	WFCP-2, 6
AND	✓	✓	✓	TC	Route-A	Behavior	Fork	✓	—	WFCP-2
OR	✓	✓	✓	TC	Route-A	Behavior	Selection	✓	—	WFCP-6
XOR	✓	✓	✓	TC	Route-A	Behavior	Selection	✓	—	WFCP-4
Merge	✓	✓	✓	TC	Route-A	Behavior	✓	Join	—	WFCP-3, 5, 7, 8, 9, 30 to 38
Loop	✓	✓	♦	TC	Activity	Behavior	—	♦	—	WFCP-10, 21
Persons or Roles (Q13)	✓	✓	♦	Role	Pool, Participant	Agent	Role	Role	Thing	R-RBA
Capability	♦	♦	♦	—	Participant	—	♦	♦	Property	R-CBA, R-DE
Organizational Unit	♦	♦	✓	—	Pool, lane	—	♦	✓	Thing	R-OA
Name	P	♦	♦	—	Participant	—	(Actor)	—	Property	R-DBAS
Documents or objects	♦*	✓	✓	WF-relevant data	(Data Object)	Entity, Information	Information	Data Object	Thing	WFDP-2, 3, 5, 6
Events (Q14)	✓	♦	✓	— See Text	✓	✓	✓	✓	✓	WFDP-37
Information	✓	♦	✓	—	Message	✓	✓	♦	♦	WFDP-1, 5, 6, 8 to 14, 16 to 18, 20 to 22, 24, 25, 38
Timing	✓	—	♦	P	Timer	♦	✓	♦	♦	(WFEH Deadline)
Human intervention	♦	♦	♦	—	—	♦	—	♦	♦	WFDP-7, 14 to 25 (WFCP-16; WFEH External trigger)
State of process/taks	♦	♦	♦	—	♦	♦	✓	♦	State	(WFDP-38;WFEH)
Deviations	✓	✓	♦	—	Error	♦	—	♦	♦	WFEH (except for External Trigger)
(Additional) Resources** (Q12)	✓	✓	✓	—	♦	✓	✓	✓	Thing	R-DA, -FBA, -RA, -CH, -RF, -HBA, -DBOS, -DBOM, -RMA, -RRA, -SHQ, -ED, -LD, -D, -E, -SD, -AR
Software Systems	♦	♦	✓	Invoked application	Application	Information	♦	Application	Thing	♦
Machines, Appliances	♦	♦	✓	—	Participant	♦	♦	♦	Thing	♦
Additional information for a task (Q15)										
Execution status	✓	♦ S	—	✓	♦ (Extension Attribute)	—	State	—	State	WFDP-1, (WFDP-3)
Priority	♦ P	♦ S	—	P		—	—	—	Property	
Goal	♦ P	♦ S	✓	—		—	✓	✓	Property	
Cost	♦ P	♦ S	—	—		—	—	—	Property	
Excess constructs or elements provided by the process modeling languages or frameworks										
≈ 25	8	—	—	9	—	3	≈ 15	≈ 22	≈ 57	
Expressiveness Core/Total Reference										
1 / 1	0.96 / 0.96	1 / 0.89								

Legend: ✓ Directly supported (count: 1) ♦ Indirectly supported (count: 1) —: Not supported (count: 0)
* No influence on control flow.
** Only if not already listed before; rearrangement due to the calculation of expressiveness (see text).

Abbreviations: P/S: Can be defined by a construct's properties/stereotypes, Route-A: Route Activity, R: Workflow resource pattern.
TC: Transition condition, WFCP/DP/EH:Workflow control pattern/data pattern/exception handling pattern

Expressiveness is calculated by relating the sum of all expressible reference items (irrespective of their order/rank) to the sum of all items of the reference [15]. The *total reference* of Table 11 consists of 28 items[2]; if the additional information for a task is excluded, 24 items remain in the *core reference*. For the core reference, the BPMN and EPCs are equally expressive (see Table 11); UML Activity Diagrams slightly lag behind because of their inability to express timing. For the total reference, the BPMN is the most expressive language; both the BPMN and UML Activity Diagrams benefit from the languages' extension mechanisms such as freely definable attributes or stereotypes.

Altogether, the differences in expressiveness are marginal, especially within the core reference. For the companies this result implies that the BPMN, EPCs and UML Activity Diagrams are interchangeable. Technically, this interchangeability is (except for events that represent human intervention) supported by the XML Process Definition Language (XPDL) [30], a standard for the exchange of process models (diagrams) between tools; see Table 11. The XPDL support even comprises the total reference as additional information for tasks can be defined by 'extension attributes'.

If we measure complexity by the count of constructs of a process modeling language that do *not* represent reference items ('excess constructs' in Table 11), our findings support the larger complexity of the BPMN compared to UML Activity Diagrams as reported in [21].

To sum up our results, the most widespread process modeling languages equally satisfy the companies' requirements concerning process modeling. Thus, our evaluation of process modeling languages is far more positive than the existing ones (see Section 2). This can be explained as follows: First of all, expressiveness always depends on the purpose ('What must be described?'), which also dictates the evaluation basis (reference). Distinct references will naturally provoke different assessments of process modeling languages and their expressiveness. The references used in the field of process modeling languages (see Section 2) are given in Table 11. On the one hand, they are very generic (e.g., [11], [5]) and, thus, do not adequately cover the requirements of process modeling. On the other hand, the number of excess elements hints at 'over-engineering' of the references (especially [9]) or at the inclusion of purposes beyond process modeling, namely workflow execution and the corresponding tool support [22]. The findings of our empirical investigation suggest that modeling business processes requires far less than the workflow patterns, but more than the WfMC reference model [28], which in fact corresponds to the 'least common denominator'. For example, though 'events' are mentioned several times in the document of the standard, they are no part of the process definition [28].

4.3 Limitations of our Research

Our results can be assumed to be representative for large companies that operate worldwide in any sector. However, the list 'Forbes Global 2000' we used suffers from three limitations: First, it disregards large non-American companies that don't have commercial relations with the USA. Secondly, ranking companies based on sales, profit and market value favors sectors where borrowed capital is important (e.g., banking and insurance companies). Thirdly, non-profit organizations (i.e., public administration, universities) are completely excluded. These limitations might well affect the validity of our results.

The ratings we have obtained from the companies reflect the subjective experiences of our participants, which is a common limitation of such surveys (e.g., [19]). Our rating scales (see Section 3.2) cover perceived need, importance or frequency – especially in the latter case we cannot expect that the number of tasks in all process models of a company was counted. This could easily be mistaken for a disadvantage, but also the representativity of process models in a sample cannot be guaranteed [32].

From a methodical point of view, the calculation of mean ratings is only valid if the ratings of the underlying scale are equidistant (interval scale). This assumption is generally made [8]. Additionally, our main conclusions rely on these means only to exclude two items (signals, planning status) from the reference in Table 11; thus, the effect of this methodical assumption is negligible.

Process modeling can serve different purposes; the most important ones for companies are Business Process Reengineering, documentation, understanding and communication [7], [16]. These purposes need expressiveness concerning description, which we have measured here. Nevertheless we acknowledge that distinct purposes (e.g. workflow execution by BPM tools) may require other expressiveness, i.e., another reference, which will lead to deviating evaluation results.

Finally, gathering requirements is usually not done by questionnaires. For that reason we currently conduct qualitative interviews with selected companies from the sample.

5. CONCLUSIONS AND FUTURE RESEARCH

From an inquiry of 130 public companies from all over the world we gathered a common set of requirements that reflect the concepts needed to describe real-life business processes. We used these requirements as a reference to assess the expressiveness of the currently dominating process modeling languages, i.e. the BPMN, UML Activity Diagrams and EPCs. On the whole, the expressiveness of these process modeling languages is equal and, thus, they can be used interchangeably – if the process models are created for documentation, understanding, and communication or to support Business Process Reengineering.

The required expressiveness also depends on the purpose of modeling. Since our survey gathered the reasons to describe processes, we will conduct a more detailed analysis on the correlation between needed language constructs and process model usage.

If we accept the core reference we have found as the common notion of 'business process', the results of [18] indicate that the evaluated process modeling languages should be equally understandable. This is another topic for future research.

6. APPENDIX

In the following, we give the wording of the questions that are discussed in this paper as they appeared in the questionnaire:

(Q1*) To what are your processes related?* Product of our company, Customer contact, Administration, Systems integration; data transformation, Systems development, Emergency procedure, Other.

[2] The line ‚Additional information for a task' is a heading and, thus, not counted.

(Q2) *List at least one "title" or short description of typical processes in your company/department:*

(Q3) *Please assess the following statements about the distribution of the processes in your company!* (Scale Type 3) Processes are executed within one department by one person, The processes are executed within one department but involve more than on responsible person, The processes are executed within the company by more than one department, The execution of the processes involves also other companies.

(Q4) *How do you measure the average run time of the processes in your company?* (Scale Type 3) In days, In weeks, In months, In years.

(Q5) *How often do you execute the processes in your company?* (Scale Type 3) Several times a day, Several times a week, Several times a month, Several times a year.

(Q6) *Please estimate the following numbers for an average process in you company (wherever it applies)!* The number of persons from the same department involved in the execution of an average process, The number of departments involved in the execution of an average process, The number of application systems involved in the execution of an average process, The number of tasks or activities of an average process.

(Q7) *How did you proceed in describing your processes?* We started from the products of our company and concentrated on their transformation from an initial to a final form, We started from some goal and grouped the activities to achieve it, We started from the interactions (including data flow) between departments or responsible persons.

(Q8) *Which entities do you need to describe the processes in your company/department?* (Scale Type 1) The tasks that have to be done in the process, The relationship between the tasks, The persons or roles who execute the tasks, Resources needed to fulfill the tasks (e.g. machines, material, documents, software systems, etc.), Time information related to the task (e.g. durations, start/end time points), Timely distances between the end of a task and the start of the following task.

(Q9) *How often do the following observations apply to the tasks in your processes?* (Scale Type 2) The tasks follow each other in a strict sequence. In other words, each task has at the most one predecessor and one successor, A task has more than one immediate successor (process splits), A task has more than one immediate predecessor (process merge), Some condition (other than the end of the preceding task) must be satisfied in order to start a task, A task (or a group of tasks) is repeated, till some criterion is satisfied, Tasks are alternatives to each other.

(Q10) *If the start or the selection of a task depends on a condition, the condition refers to:* The tasks only, e.g., the availability of resources, The results of adjacent tasks, The overall state of the process, Information external to the process, Time, Other.

(Q11) *If more than one task can be started, they are performed:* (Scale Type 2) All and concurrently, All independent of each other and in any order, Not all tasks are executed, Which task(s) is/are chosen for execution, depends on: -a random selection, -whether or not some condition is satisfied, -the subjective experience of some person in charge, As soon as one task has been executed no other tasks from the ones that can be started is performed.

(Q12) *Which resources do you (wish to) include in the models or descriptions of your processes?* (Scale Type 1) Machines, appliances, Personnel, Material, Documents, Information, Software Systems.

(Q13) *Which information about the persons executing the task do you (wish to) include in the models or descriptions of your processes?* None, The required capabilities (roles, positions), Names of the responsible persons, Organizational units they belong to, Other.

(Q14) *What triggers the execution of your processes or their tasks?* (Scale Type 3) Information sent from business partners (customers, suppliers, etc.), Company-internal information, Timing (date, cycles), Signals from machines or sensors, Deviations from targets or failures, The state of some process or task, Human judgment and intervention.

(Q15) *What kind of additional information do you want to gather in your models?* Needed for the task & Needed for the process: Priorities, Cost, Goals, Execution status (e.g., cancelled…), Planning status (e.g., plan/actual; strategic/tactical).

(Q16) *Could you give us at least one example of goals you wish to express?*

(Q17) *How do you describe the processes in your company/ department?* As text in normal language, As tables, By using a (process) modeling language, e.g., BPEL, BPMN, EPC, IDEF, UML, Other (process) modeling language, please specify.

7. REFERENCES

[1] Becker, J., Kugeler, M. and Rosemann, M. 2003. *Process Management – A Guide for the Design of Business Processes*. Springer, Berlin et al.

[2] Economist Intelligence Unit (in cooperation with the IBM Institute for Business Value): *E-readiness rankings 2008: Maintaining momentum.* A white paper from the Economist Intelligence Unit. London et al., 2008. http://www.eiu.com/site_info.asp?info_name=ibm_ereadiness&page=noads&rf=0.

[3] Forbes: *The Forbes Global 2000*. http://www.forbes.com/lists/2008/18/biz_2000global08_The-Global-2000_Rank.html.

[4] Gravetter, F. and Wallnau, L. 2009. *Statistics for the Behavioral Sciences*. 8th ed., Wadsworth, Belmont.

[5] Green, P. and Rosemann, M. 2000. Integrated process modeling: An ontological evaluation Information Systems. In *Proceedings of the 11th International Conference of Advanced Information Systems Engineering*. (Stockholm, Sweden, June 05 - 09, 2000). CAiSE 2000. Elsevier. Vol. 25, No. 2, 73-87. DOI= http://dx.doi.org/10.1016/S0306-4379(00)00010-7.

[6] Holt, J. 2009. *A Pragmatic Guide to Business Process Modeling*. 2nd ed., BCS, Swindon.

[7] Indulska, M., Recker, J., Rosemann, M. and Green, P. 2009. Business Process Modeling: Perceived Benefits. In *Proceedings of the 28th International Conference on Conceptual Modeling* (Gramado, Brazil, November 09 - 12, 2009). ER 2009. LNCS, Vol. 5829, 458-471. Springer, Berlin.

[8] Jackson, S. 2009. *Research Methods and Statistics: A Critical Thinking Approach*. 3rd ed., Wadsworth, Belmont.

[9] List, B. and Korherr. B. 2006. An evaluation of conceptual business process modelling languages. In *Proceedings of the ACM Symposium on Applied Computing* (Dijon, France,

April 23 – 27, 2006). ACM, New York, NY, 1532-1539. DOI= http://doi.acm.org/10.1145/1141277.1141633.

[10] LimeSurvey. Version 1.85. http://www.limesurvey.org/.

[11] Lin, F., Yang, M. and Pai, Y. 2002. A Generic Structure for Business Process Modeling. *Business Process Management Journal*. 8, 1 (2002), 19-41.

[12] Mendling, J., Reijers, H.A. and Cardoso, J. 2007. What Makes Process Models Understandable? In *Proceedings of the 5th International Conference on Business Process Management* (Brisbane, Australia, September 24 – 29, 2007). BPM 2007. LNCS Vol. 4717, 48-63. Springer, Berlin.

[13] Object Management Group (OMG): *OMG Unified Modeling Language (OMG UML), Superstructure, Version 2.2.* OMG Document Number: dtc/2010-06-05, http://www.omg.org/technology/documents/formal/uml.htm.

[14] Object Management Group (OMG): *Business Process Model and Notation (BPMN) Version 2.0*. OMG Document Number: dtc/2010-06-05, http://www.omg.org/cgi-bin/doc?dtc/10-06-04.

[15] Patig, S. 2004. Measuring Expressiveness in Conceptual Modeling. *In Proceedings of the 16th International Conference of Advanced Information Systems Engineering*. (Riga, Latvia, June 07 – 11, 2004). CAiSE 2004. Springer, Berlin et al. LNCS Vol. 3084, 127-141.

[16] Patig, S., Casanova-Brito, V. and Vögeli, B. 2010. IT Requirements of Business Process Management in Practice – An Empirical Study. In *Proceedings of the 8th International Conference on Business Process Management* (Hoboken, USA, September 13-16, 2010). BPMN 2010.

[17] Recker, J. 2008. *BPMN Modeling – Who, where, how and why. BP Trends, March 2008*. http://www.bptrends.com/publicationfiles/05-08-ART-BPMN%20Survey-Recker-JR%20final.pdf.

[18] Recker, J. and Dreiling, A. 2007. Does It Matter Which Process Modeling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages without Formal Education. In *Proceedings of the 18th Australasian Conference on Information Systems* (Toowoomba, Australia, December 05 - 07, 2007). ACIS 2007.

[19] Recker, J., Indulska, M., Rosemann, M. and Green, P. 2006. How Good is BPMN Really? Insights from Theory and Practice. In *Proceedings of the 14th European Conference on Information Systems* (Göteborg, Sweden, June 12 – 14, 2006). ECIS 2006.

[20] Recker, J.; Rosemann, M. and Krogstie, J. 2007. Ontology-Versus Pattern-Based Evaluation of Process Modeling Languages: A Comparison. In *Communications of the Association for Information Systems*. 20, Article 48 (2007). http://aisel.aisnet.org/cais/vol20/iss1/48.

[21] Recker, J., zur Muehlen, M., Siau, K., Erickscon, J. and Indulska, M. 2009. Measuring Method Complexity: UML versus BPMN. In *Proceedings of the 15th Americas Conference on Information Systems* (San Francisco, California, August 06 – 09 2009). AMCIS 2009. AISeL, Paper 541. http://aisel.aisnet.org/amcis2009/541.

[22] Russell, N., ter Hofstede, A.H.M., van der Aalst, W. and Mulyar, N. 2006. *Workflow control-flow patterns – A revised view.* BPM Center Report BPM-06-22, BPMcenter.org. 2006. http://www.workflowpatterns.com/documentation/documents/BPM-06-22.pdf

[23] Scheer, A.-W., Thomas, O. and Adam, T. 2005. Process Modeling Using Event-driven Process Chains. In *Process-Aware Information Systems*. Dumas, M., van der Aalst, W. and ter Hofstede, A.H.M., Ed. Wiley, Hoboken, New Jersey, 119-145.

[24] Söderström, E., Andersson, B., Johannesson, P., Perjons, E., and Wangler, B. 2002. Towards a Framework for Comparing Process Modeling Languages. In *Proceedings of the 14th International Conference of Advanced Information Systems Engineering*. (Toronto, Canada, May 27 - 31, 2002). CAiSE 2002. LNCS Vol. 2348, 600-611. Springer, Berlin.

[25] van der Aalst, W. and van Hee, K. 2004. *Workflow Management: Models, Methods and Systems*. MIT Press, Cambridge.

[26] van der Aalst, W. and ter Hofstede, A. 2010. *Workflow Patterns Home Page* http://www.workflowpatterns.com/patterns/index.php.

[27] Weske, M. 2007. *Business Process Management: Concepts, Languages, Architectures*. Springer, Berlin.

[28] The Workflow Management Coalition (WfMC). 1995. *The Workflow Reference Model*, Document Number TC00-1003, Issue 1.1., 19 November 95. http://www.wfmc.org/standards/docs/tc003v11.pdf.

[29] The Workflow Management Coalition (WfMC). 1999. *Terminology & Glossary*, Document Number WFMC-TC-1011, Issue 3.0, February 99. http://www.wfmc.org/standards/docs/TC-1011_term_glossary_v3.pdf.

[30] The Workflow Management Coalition (WfMC). 2010. *Process Definition Interface - XML Process Definition Language (XPDL)*, Document Number WFMC-TC-1025, 10 October, 2008, Version 2.1a. http://www.wfmc.org/xpdl.html.

[31] Wolf, C. and Harmon, P. 2010. *The State of Business Process Management 2010*. BPTrends Reports, February 2010. http://www.bptrends.com/surveys_landing.cfm.

[32] zur Muehlen, M., Recker, J. and Indulska, M. 2008. How Much Language is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In *Proceedings of the 20th International Conference of Advanced Information Systems Engineering*. (Montpellier, France, June 16 - 17, 2008). CAiSE 2008. LNCS Vol. 5074, 465-479. Springer, Berlin. DOI= http://dx.doi.org/10.1007/978-3-540-69534-9_35

Reinventing the Wheel?! Why Harmonization and Reuse Fail in Complex Data Warehouse Environments and a Proposed Solution to the Problem

Torsten Priebe
Teradata GmbH
Storchengasse 1
A-1150 Vienna, Austria
torsten.priebe@teradata.com

Andreas Reisser
University of Regensburg
Universitätsstrasse 31
D-93053 Regensburg, Germany
andreas.reisser@wiwi.uni-regensburg.de

Duong Thi Anh Hoang
Vienna University of Technology
Favoritenstrasse 9-11/188
A-1040 Vienna, Austria
htaduong@ifs.tuwien.ac.at

ABSTRACT

Enterprise or group data warehouses are often introduced in complex multi-national organizations in order to foster harmonization, integrate heterogeneous source systems and hide the heterogeneity from analytical systems. Industry reference data warehouse logical data models such as Teradata's FS-LDM or IBM's BDW are promoted as accelerators for the development of such large data warehouses. However, this paper shows that logical data models alone are not sufficient to ensure reusability in those environments. In order to provide a solid basis for standardization, the logical data model needs to be accompanied by a semantic business information model used as an anchor point for the mappings and for communication with business users. Such a model allows a model-driven approach for specification of the data transformations, which usually accounts for at least half the total effort of large data warehouse projects. The paper presents an approach building upon the Teradata Business Data Element (BDE) concept giving practical examples from project experience in the financial services industry. A research prototype is presented, utilizing Semantic Web technologies such as the Web Ontology Language (OWL) to facilitate the traceability of data requirements, business terms and physical data elements in the different layers of a complex data warehouse architecture.

Keywords

Data Warehouse, Harmonization, Reuse, Traceability, Semantics, Data Modeling, Model-driven Engineering

1. INTRODUCTION

Large and complex multi-national organizations are increasingly aiming for enterprise or even group-wide data warehouse approaches to support their analytical needs efficiently and to facilitate harmonization and integration. Besides the advantages that true enterprise data warehouses (EDWHs) can deliver such as lower total-cost-of-ownership or a holistic view of the business, there is a high effort associated with the development and maintenance of a common data model and the mappings to source systems, analytical applications, semantic BI tool layers, etc., especially if the operational IT landscape is diverse and heterogeneous.

Real-world EDWHs usually contain thousands of data elements, which leads to several thousands of data mapping rules. In many cases these specifications are held in Excel spreadsheets or small databases, which are typically very hard to maintain. Often, specification and mapping account for about half the total effort and budget of large DWH projects. There are industry reference logical DWH data models, which accelerate and assure quality of the DWH evolution over time. However, as we will show in this paper, such logical data models are not sufficient to ensure traceability and enable harmonization and reuse in such complex environments. Experience shows, that standardized group or enterprise data warehouse projects in heterogeneous multi-national organizations often do not show the expected business value or leverage the expected synergies in the foreseen timeframe.

Furthermore, agile and self-service business intelligence approaches aim at accellerating analyses and thus reducing time-to-market by enabling business users without being dependent on big and sluggish IT projects. However, research by The Data Warehouse Institute (TDWI)[1] shows that 80% of business users are incapable of creating their own reports. One major barrier that keeps them from independently performing analyses is a lack of understanding of the logical and physical data models in place. Highly normalized EDWH data models are often hard to comprehend and query for business users and do not cover all semantics (in business terminology) needed to describe which data is available where and how it can be accessed.

It should be noted that this paper presents work in progress. Data warehouses are integrating information about business concepts from various sources and sharing it between dif-

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

[1] http://www.tdwi.org

ferent target user groups and systems. In order to enable harmonization, it is necessary to map the different representations, of which one is the DWH itself (with its logical and physical data model), to a central point of reference which is the pure business concept. The main contribution of this paper is to examine the usefulness of a such a semantic business information model on top of logical DWH data models and to define its role within the data warehouse engineering process. It also shows how the Teradata Business Data Element (BDE) concept and Semantic Web Technologies can provide a basis for supporting this approach. However, Semantic Web technologies are currently really just used as a framework to implement a traceability prototype, based on the work presented in [13]. Our current follow-up research focuses on defining in detail, how this model should be built (e.g. using OWL modeling constructs instead of representing BDEs as OWL individuals) in the context of a model-driven data warehouse engineering approach.

The rest of this paper is organized as follows: Section 2 gives a thorough problem definition backed by some real-world examples from the authors' project experience. Section 3 discusses related work, both scientific and by DWH software vendors. In section 4 we outline our proposed solution, which we base on the Business Data Element (BDE) concept recently defined by Teradata and our preliminary research on data tracing with Semantic Web technologies [13]. We will give a thorough coverage of our future work and a conclusion of the paper in section 5.

2. PROBLEM DEFINITION

2.1 Harmonization and Reuse in Complex DWH Environments

Enterprise data warehouses are often introduced as a means of integrating heterogeneous source system landscapes and to hide this heterogeneity from analytical systems. The DWH data model is used to decouple the analytical world from the various different source systems with their different data models. This approach can be found particularly in multi-national organizations, which went through mergers and acquisitions and whose operational IT landscapes cannot be harmonized that easily. For example, the authors have worked with different groups in the financial services industry which have grown rapidly in emerging markets such as Central and Eastern Europe.

Figure 1 shows different architecture schemes for such enterprise or group data warehouses. Scheme (a) represents the approach of defining a standardized local DWH in order to reuse analytical applications and transformations from the DWH to application data marts. Scheme (b) adds to this a group layer in form of a central enterprise data warehouse (EDWH) for group-level reporting and applications. Finally, scheme (c) depicts a purely central approach without local DWHs (or with an unstandardized local DWH landscape, the point is that the reuse is accomplished through the central EDWH). In all three cases there is a standardized DWH data model, which is usually built with the following requirements in mind[2]:

- It should be independent of particular source systems, which may change over time or even be replaced due to harmonization and standardization initiatives.
- The data model should enable the standardization of data across the different entities (with different source systems, and to some degree even different business processes).
- It should allow business users to easily understand, which data is included in the DWH and where it can be found. Data requirements should be easily traceable to their physical location in the DWH and back to the source systems they are originating from.
- As changes in the physical data model impact a large number of data load jobs (consider, say, 10 different source systems), it should be able to cope with changing data requirements without having to change the the physical database schema as frequently.

These requirements are often addressed by basing the data model on an industry reference model, such as the Teradata Financial Services Logical Data Model (FS-LDM)[3] or the IBM Banking Data Warehouse (BDW) model[4]. These data models are built in a way that they combine regular modeling techniques with a meta-modeling approach for areas, where business requirements change frequently.

For example, business KPIs (e.g., profitability figures or other calculated data) and product features (e.g., interest rates and interest calculation methods) are not modeled as individual attributes but stored as code/value pairs, i.e. using a single value attribute combined with a code attribute that identifies the actual business attribute stored. The advantage of this approach is that new business attributes can be added (e.g., when a new product is introduced) without changing the logical or physical data model – existing load jobs will run unchanged. More thorough examples are given in the next subsection.

2.2 Practical Examples from the Financial Services Industry

As mentioned before, the authors have worked on various projects in different banking groups in Central and Eastern Europe, some of them based on the Teradata FS-LDM, of which a small simplified excerpt is shown in figure 2. Basically, the main entities shown are *PARTY* (representing customers, employees, but also organizational units) and *AGREEMENT* (representing all sorts of accounts and other contracts a customer may have with the bank).

As mentioned before, those areas of the data model dealing with less dynamic business requirements are modeled using a "traditional" approach, representing each business attribute as an attribute in the logical data model. An example is the customer master data (for individuals in the *INDIVIDUAL* and *INDIVIDUAL NAME HIST* entities, for orga-

[2]It turns out that these requirements are to some degree conflicting. This will be discussed in detail in the next subsection.

[3]http://www.teradata.com/t/logical-data-models/financial-services/

[4]http://www.ibm.com/software/data/industry-models/banking-data/

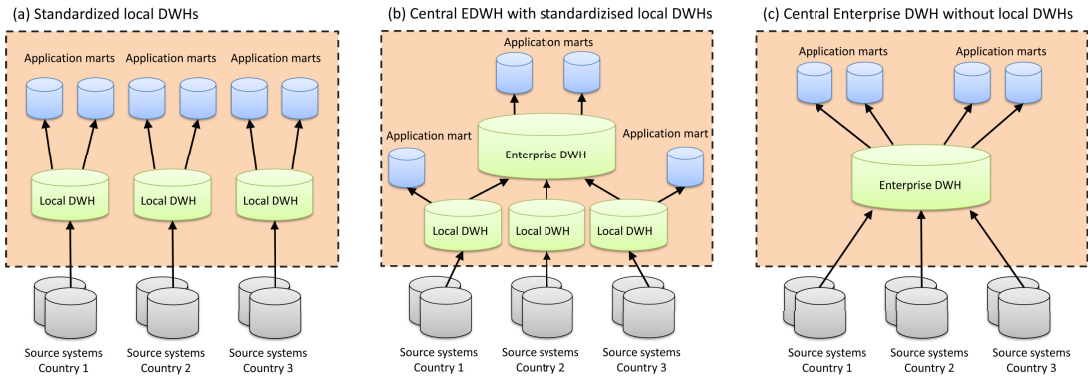


Figure 1: Data warehouse architectures for distributed organizations

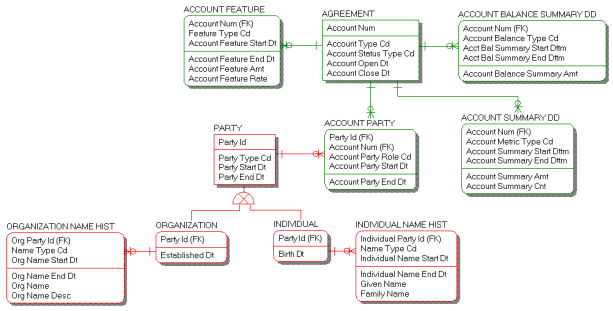


Figure 2: Excerpt of the Teradata FS-LDM

nizations in the *ORGANIZATION* and *ORGANIZATION NAME HIST* entities).

More dynamic business requirements are represented using a meta-modeling approach. For example features of an account (e.g. interest rates, nominal amounts) as well as derived KPIs (e.g. balances and other summary data) can be found as code/value pairs in the *ACCOUNT FEATURE*, *ACCOUNT BALANCE SUMMARY DD* and *ACCOUNT SUMMARY DD* entities. To cater for the different data types that may be used for different features, etc., a value attribute for each data type (e.g. *Account Feature Amt* for amounts, *Account Feature Rate* for rates) is used.

Now, recall the requirement to use the DWH data model to enable the standardization among different entities with different source systems. Basing the standardization solely on the LDM is not sufficient. As the business semantics are not fully covered by the model, different teams would map source data differently to the LDM.

The following are some real-world examples of problems, that can occur:

- **Same information (from different source systems) mapped differently to the DWH LDM/PDM.** For example, the current outstanding amount of a loan may be mapped to *ACCT SUMMARY DD.Account Summary Amt* by one team and to *ACCOUNT BALANCE SUMMARY DD.Account Balance Summary Amt* by another team. Actually, it can be discussed, whether the outstanding amout is a balance or not. As a result, an integrated use of the data (resp. reuse of mappings and transformations) becomes impossible.

- **Same information mapped to the same tables, but different codes used.** For example, the nominal interest rate of a loan may be correctly mapped to *ACCOUNT FEATURE.Account Feature Rate*, but one team may use 'ITR_NOM' as *Feature Type Cd*, while the other mapping team may use 'ITR'. Again, as a result the possibility of reuse becomes limited.

- **Different information mapped to the same place in the DWH.** While the first two points appear if semantically identical information is not recognized as such and thus not mapped consistently, the opposite can also occur. For example, there may be different days-past-due (DPD) counters, counting how many days a loan debt has not been paid according to the installment plan. There are different (all valid) business definitions, e.g. if holidays should be counted. If information from different sources with different business definition is mapped to the same place in the DWH (e.g. *ACCOUNT SUMMARY DD.Account Summary Cnt*using 'DPD' as *Account Metric Type Cd*, using this information in different application data marts will lead to unexpected results.

- **Different domain values used.** For example, the status of an account may be unambiguously mapped to *AGREEMENT.Account Status Type Cd*, but one source may use the value 'O' for open and 'C' for closed, while another may use 'A' for active and 'I' for inactive accounts. If this difference is not identified (and dealt with in the source to DWH transformation), the reusability of the data is again limited.

- **Lineage is only captured on LDM/PDM attribute level (not on code level).** Say, the nominal interest rate is stored in a data mart field *NOM_INTEREST_RATE*. It is mapped to *ACCOUNT FEATURE. Account Feature Rate*, which in turn is mapped to various source fields (i.e. various features with different

codes). In this case, the traceability of data requirements resp. application layer attributes to source attributes is ambiguous and thus limited.

Basically, in order to provide a solid basis for standardization, the LDM needs to be combined with a catalog of the codes used in those areas of the data model that are following a meta-modeling approach. Furthermore, it turns out that DWH LDMs, which also include a number of technical attributes to deal with data history, etc., are hard to comprehend by business users. As a consequence, they are of limited suitability to address the traceability requirement mentioned above. Business users will not be able to easily locate their data requirements in a catalog of LDM entities and attributes.

This all calls for a semantic business information model on top of the LDM to be used as an anchor point for the mappings and for communication with business users. Section 4 will discuss our approach how to build such a model as well as the use of semantic technologies to provide the necessary traceability.

3. RELATED WORK

As a matter of fact, we cannot state that traceability and data lineage are new issues; in the literature we can find various research approaches and published papers dating back to the early 1990s with proposed methodologies for software traceability [12]. The problem of data lineage tracing in data warehousing environments has been formally founded by Cui and Widom [3]. Moreover, based on the AutoMed notion of data lineage tracing, Fan and Poulovassilis [5, 10] developed algorithms for deriving affected data items along the transformation pathway.

These approaches formalize how to trace tuples (resp. attribute values) through rather complex transformations, given that these transformations are known on schema level. In practice, this assumption often does not hold. Transformations may be documented in source-to-target matrices (before they are implemented, the so-called specification lineage) and implemented in ETL tools (the so-called implementation lineage). Our work concentrates on how to properly define this specification lineage, which is a huge problem in large-scale DWH projects, especially if different sources have to be consistently mapped to the same target. The contribution of this paper is the introduction of a business information model as the central mapping anchor point.

From a commercial product perspective, data lineage and impact analyses are usually provided by metadata repositories. There are two categories of such metadata management tools. Firstly, most ETL software products (e.g. Informatica PowerCenter[5], IBM DataStage[6], Microsoft SQL Server Integration Services[7]) include a repository storing the metadata created and managed by the tool. Most of those vendors also provide interfaces to import metadata from database and BI tools. A prominent example of this category is the IBM Metadata Workbench[8], part of the InfoSphere Information Server product family, which allows tracing the transformations implemented in DataStage. The second group of commercially available tools are specialized products which are not related to a particular ETL tool and thus need to import the metadata using standard Common Warehouse Metamodel (CWM)[9] or tool-specific interfaces. They are of course less tightly integrated and may therefore have limitations in the metadata they can import from a particular tool. On the other hand, they are usually more flexible in specifying a customized metadata model tailored to the specific requirements of an organization. Examples in this group of tools are ASG Rochade[10], Adaptive Metadata Manager[11] or Teradata Metadata Services (MDS)[12].

All these commercial tools meanwhile provide quite sophisticated mechanisms for data lineage and impact analyses based on the transformations extracted from the ETL tool (implementation lineage). However, they are rather limited in combining and comparing this information with business information models, mapping specifications (specification lineage) and data requirements from a requirements management process. A first promising (but in practice also limited) approach is the one provided by the IBM Metadata Workbench (see above) in combination with the other Information Server components Business Glossary[13] (providing access to a catalog of business terms and categories) and FastTrack[14] (a mapping tool which allows generating DataStage jobs from a mapping specification). Similarly, ASG metaGlossary[15] provides mechanisms for defining and categorizing business terms and mapping them to physical data elements, based on the Rochade repository (see above).

The use of (business) models as a basis for development is a well-known concept in the software engineering world, referred to as Model-Driven Engineering (MDE) [15]. The most prominent MDE approach is OMG's Model-Driven Architecture (MDA) initiative[16], which defines system functionality using a platform-independent model (PIM). In fact, our business information model as presented in section 4 is such a PIM. Similar to MDE approaches in software engineering, we aim at deriving the development specification (source-to-target matrices) from this model.

Various research projects have been presented based on a mediated DWH architecture with an ontology infrastructure, providing ontology-based specification of relationships

[5] http://www.informatica.com/products_services/powercenter/
[6] http://www.ibm.com/software/data/integration/datastage/
[7] http://www.microsoft.com/sqlserver/
[8] http://www.ibm.com/software/data/infosphere/metadata-workbench/
[9] http://www.omg.org/cwm/
[10] http://www.asg.com/products/product_details.asp?code=ROC
[11] http://www,adaptive.com/products/mm.html
[12] http://www.teradata.com/t/tools-and-utilities/meta-data-services/
[13] http://www.ibm.com/software/data/infosphere/business-glossary/
[14] http://www.ibm.com/software/data/infosphere/fasttrack/
[15] http://www.asg.com/products/product_details.asp?code=AMG
[16] http://www.omg.org/mda/

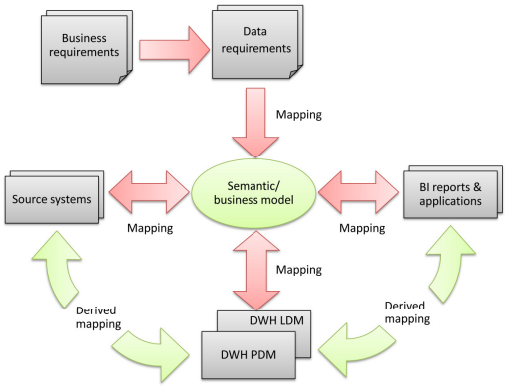


Figure 3: Business information model as the central mapping anchor point

between sources and the DWH on a conceptual level. For example, Romero and Abelló [14] address the design of a DWH multidimensional analysis schema starting from an OWL ontology that describes the data sources. They identify the dimensions that characterize a central concept under analysis (i.e. fact) by looking for concepts connected to it through one-to-many relationships. The relation between requirements, multidimensional design and underlying data sources were addressed by Mazón et al. [9] which applied a goal-oriented approach to requirements analysis for DWHs based on the TROPOS methodology. In [6] Giorgini et al. introduce an approach where conceptual multidimensional models capturing the various user requirements can be obtained. All three approaches are however not suitable for non-dimensional (ER-modeled, normalized) EDWH.

To the best of our knowledge, in current approaches, the use of semantic inference mechanisms for optimizing data lineage and impact analyses is not considered. Furthermore, there is still no comprehensive approach to support the representation and analysis of links over multiple layers of complex DWH architectures that would include conceptual business terms and data requirements. End-to-end tracing is however critical to enable successful reuse of mappings and transformations. Moreover, there are little resources on semantic mismatches and disparate use of terminology across DWH environments. In this context, ontologies can be used to represent the precise semantics that are required to support harmonization and ensure reusability.

4. SOLUTION OUTLINE

In section 2 we have shown that a DWH LDM or PDM alone is not suitable to cope with the traceability and standardization requirements to enable reuse in complex DWH environments. In order to address those requirements properly, the DWH data model has to be accompanied by a semantic or business information model. This model needs to capture the definitions and business rules plus the mappings to the different representations of the corresponding data artifacts.

The semantic model should provide a comprehensive and unique list of business concepts (with their attributes) needed to satisfy the information needs of all relevant user communities. As shown in figure 3, it is intended as an anchor point for all mappings. The semantic model should act as:

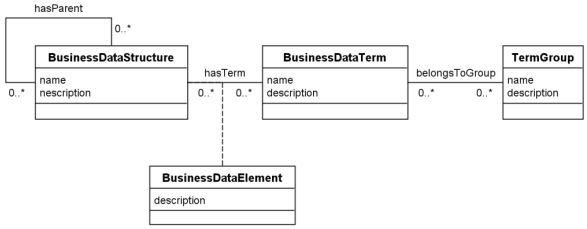


Figure 4: Business Data Elements (BDE) model

- **The main source for business-related metadata.** It should enhance the business users' understanding of the data by communicating the meaning and context, achieving traceability between the business concepts and their physical representations.
- **The main source to generate the DWH development specification.** As you can see in figure 3, by mapping the semantic model elements to the data sources, the DWH and targets such as application marts or semantic layers of a BI tool, the actual development specification for the transformations, the source-to-target matrices (STMs) can to a large degree be derived automatically.
- **The foundation for requirements management.** The model should show, which business concepts are already included in the DWH, which are requested but not yet implemented, who requested them, etc.
- **The foundation for enterprise information integration, enterprise data quality initiatives and increased usability.** It should facilitate consistent data modeling, consistent data definitions, allow to determine data redundancies, etc.
- **The foundation for improved time-to-market.** It should help to leverage synergies and enable higher development efficiency by reusing mappings, definitions, transformation jobs and analytical applications. It should reduce complexity by decomposition, etc.

4.1 Business Data Element Concept by Teradata

As an answer to the need for a semantic or business model, Teradata has recently defined a concept called Business Data Elements (BDE). The BDE concept is based on the metamodel shown in figure 4. Objects of interest to the organization are captured as a *BusinessDataStructure*, organized in a parent-child hierarchy – similar to the entity construct in the entity relationship model. Organization-wide unique business terms are captured as a *BusinessDataTerm*, in fact these are similar to the attribute construct in the entity relationship model. Finally, a *BusinessDataElement* is the intersection of a term in a given structure. In this regard, the main difference to the entity relationship model is the fact

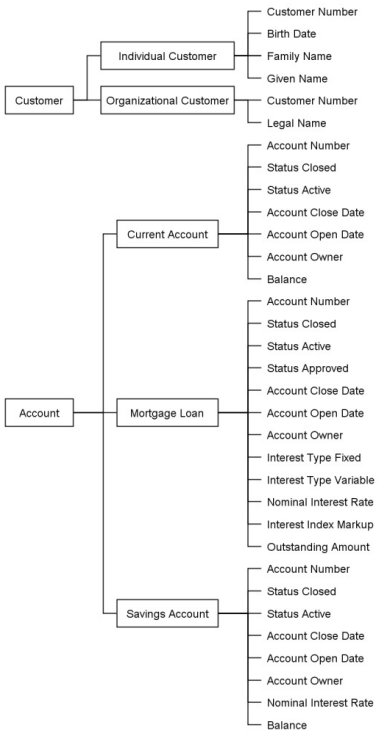


Figure 5: Simplified example BDE model for a bank

that the same term can be linked to more than one structure, i.e. uniquely defined terms can be reused in different structures independently from their hierarchy. Additional a *TermGroup* can be used to create user-specific views to the BDE model (e.g. a group of Basel II relevant BDEs).

Figure 5 shows a (simplified) excerpt of an example BDE model from the banking industry using the main structures *Customer* (with substructures *Individual Customer* and *Organizational Customer*) and *Account* (with substructures based on the products the bank offers, in this case *Current Account*, *Savings Account* and *Mortgage Loan*). Recalling the problems stated in section 2.2, note that *Outstanding Amount* and *Nominal Interest Rate* are explicitly listed as business terms. Also note, that *Status Active* and *Status Closed* are explicitly included (rather than a single *Status* term) in order to capture the different status values.

Now, how can such a BDE model be developed for a specific organization? A business model is always organization-specific, as the terminology and KPIs used will differ from organization to organization, especially when spanning country borders. Experience shows that even within the same language area (e.g. two German-speaking subsidiaries of the same banking group) terminology may differ significantly. Nevertheless, there are a number of sources than can help define the organization-specific BDE model.

In the case of Teradata, the main source of course is the reference LDM of the respective industry (providing explicit constructs for traditionally modeled business areas and example codes for areas that are represented using a meta modeling approach, see section 2), or rather the version of it that has been customized for the organization (mainly by omitting irrelevant parts). As shown in figure 6, further inputs that add content specific to the organization are corporate standards like a product data catalog and existing source data models – or legacy DWH or data mart data models in case of migration or data mart consolidation projects.

4.2 Ontologies and Semantic Technologies

As the connections between data elements in the different layers of a DWH system form graph structures, it suggests itself to utilize Semantic Web technologies which provide proper and mature methods and technologies for handling and operating on graphs. As we presented in [13], these technologies are highly suitable for performing lineage tracing operations in DWH environments based on (in [13] still solely implementation linage) metadata.

One of the basic concepts used in our approach and prototype (see next subsection), is the Resource Description Framework (RDF) [8], which uses a triple-based model (subject-predicate-object) meaning that resources have a number of properties with certain values. Values can be be again resources or literals. Resources and properties are uniquely identified by a Uniform Resource Identifier (URI). There is also a graphical notation, where an RDF document describes a graph having nodes (resources or literals) connected to each other via directed arcs (properties) and an XML-based syntax called RDF/XML which enables machines to process RDF metadata.

To enrich the representation formalism of RDF, the World Wide Web Concortium (W3C) introduced simple ontological modeling primitives (i.e. classes and subclasses) with RDF Schema (RDFS) [1]. In order to enable machines to fulfill reasoning tasks, stronger ontological conepts are needed. For example, one of our key concepts for tracing data elements is the transitivity of transformations through the different DWH layers. The Web Onlology Language (OWL) [4] extends RDFS by adding concepts for describing relations between classes (e.g. synonyms), cardinality and characteristics of properties (e.g. transitivity).

To be able to infer information from the ontology automatically, e.g. implicit transformation arcs arising from the transitivity property, inference rules have to be processed by an inference engine or reasoner. In addition to the predefined

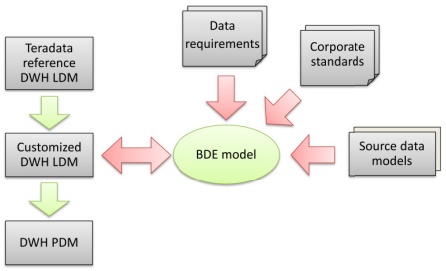


Figure 6: Creation of a specific BDE model

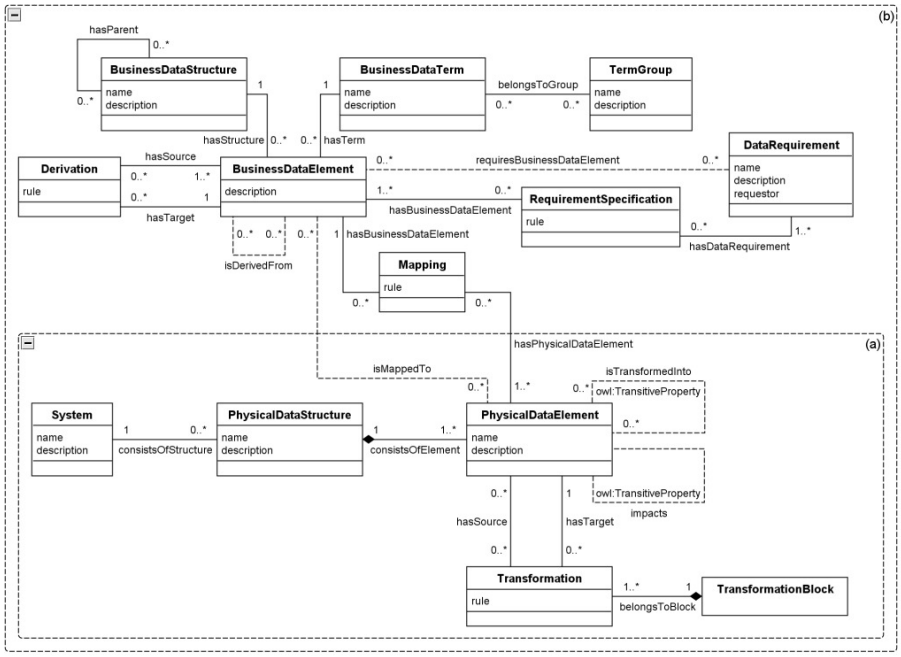


Figure 7: Prototype metamodel

inference rules in OWL, we use custom inference rules based on the Semantic Web Rule Language (SWRL) [7]. In order to efficiently extract information from the ontology we use the SPARQL Protocol and Query Language [11].

The suitability of Semantic Web Technologies in this environment may actually be questioned due to the open world assumption that they are based on. This means that formally, if a data element x does not have an *isTransformedInto* property to element y (see next subsection), we could not conclude that x is not transformed into y. But as we use the Semantic Web technologies in a closed environment with a closed repository, this is actually not an issue. We would even argue that an open world assumption is OK for most tracing problems (e.g. impact analysis, it is better to find an incomplete subset of impacted elements than none).

In addition to the suitability of semantic technologies to enable traceability among different data elements, ontology languages are a promising basis for extending the expressiveness of the business model. OWL offers a range of very useful constructs, like for describing synonyms or the disjunction of classes. Furthermore, there are domain ontologies freely available for different industries.[17] Representing the business model directly as an OWL ontology (i.e. business structures and terms as OWL classes and properties) is however subject to our future work (see section 5).

4.3 Research Prototype

In [13] we presented an approach and prototype utilizing Semantic Web Technologies to perform data lineage and impact analyses efficiently by infering additional traceability information from transitive properties and custom rules. Like most commercially available metadata management tools, this prototype was focussing on tracing physical data elements through transformations that may be implemented in an ETL tool (implementation lineage).

The part marked as (a) in the metamodel in figure 7 shows the coverage of the prototype that was already presented in [13]. As described in more detail below, it is implemented as an OWL ontology using the HP Jena framework[18] for persistent storage, inference and querying. In a nutshell, it shows physical data models and the transformations between elements within those models. A *System* consists of tables or flat files represented by the *PhysicalDataStructure* class which contains at least one *PhysicalDataElement* (columns in a table resp. fields in a flat file). The *System* construct can also be used to represent different layers (e.g. stage, core, etc.) in a DWH. The inferred transitive relationship *isTransformedInto* states that one column or field is transformed into another one. This relationship is automatically derived by a custom SWRL rule based on the fact that the physical data elements are specified as source and target of a *Transformation*. Moreover, *Transformations* can have transformation rules (e.g. a WHERE clause to limit the source tuples selected, or the assignement of a constant value to a target field) and can be grouped in *TransformationBlocks*. The also automatically derived relationship *impacts* reflects the fact, that a modification being made to a certain element may have an impact on another element. The *isTransformedInto* and *impacts* properties can be used to easily perform data lineage and impact analyses as simple SPARQL queries [13].

[17] e.g. `http://www.fadyart.com` for financial services

[18] `http://jena.sourceforge.net/`

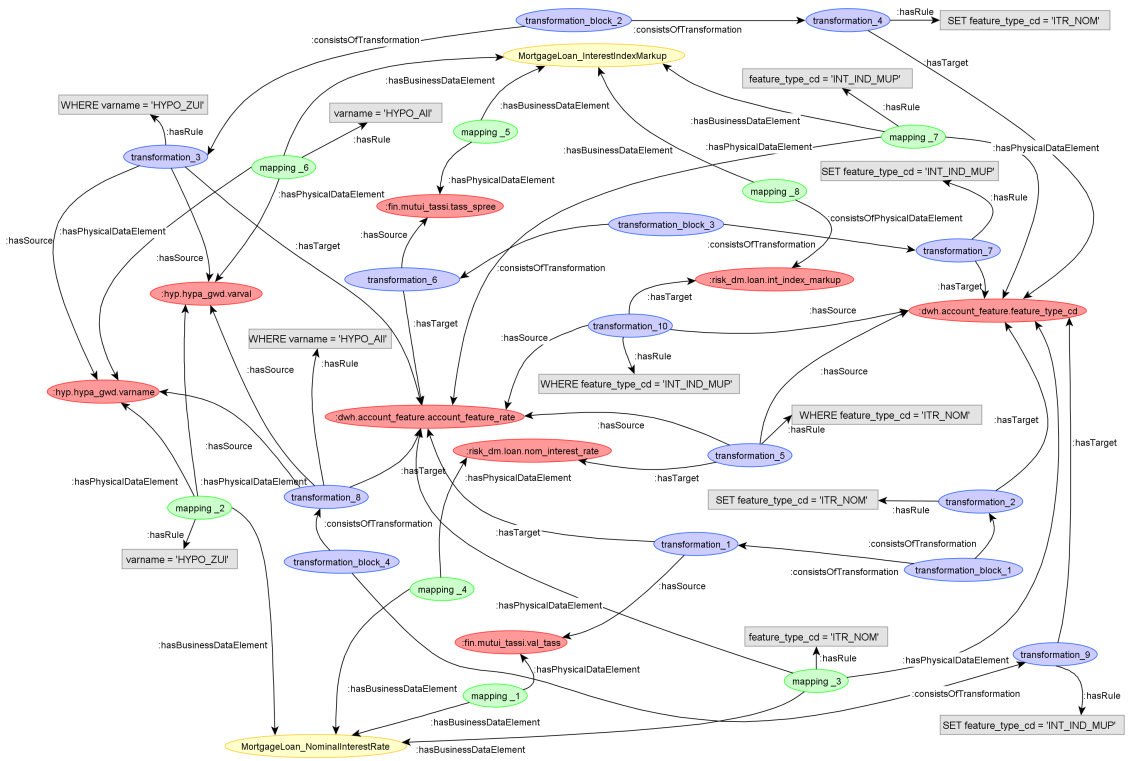


Figure 8: Example in RDF notation

Meanwhile, we have extended our prototype to also cover a business information model and data requirements as shown in figure 7 (b). This also distinguishes our approach from what most currently available commercial systems offer, which usually focus on the implementation rather than specification lineage (see section 3). Currently we have explicitly represented the business data elements metamodel (see section 4.1), i.e. business data structures become individuals of a class *BusinessDataStructure*, which can form hierarchical structures, and business data terms become individuals of a class *BusinessDataTerm*, which can be grouped in *TermGroups*. As future work, we will investigate using OWL classes and properties themselves to represent the business data structures and terms, i.e. represent the BDE model as an OWL ontology (rather than individuals of a metamodel represented in OWL).[19] The intersection of a business data term and a business data structure is represented as a *BusinessDataElement*. Note that the modeling is slightly different than presented in section 4.1 (*BusinessDataElement* is modeled as a regular class rather than an association class) in order to be able to represent this part in OWL.

Part (b) is linked to part (a) via the mapping of business data elements to a physical data elements. The relationship *isMappedTo* will be inferred if a BDE is connected to a physical data element via a *Mapping*, which can have mapping rules (similar to the transformation rules described above), along with the relationships *hasBusinessDataElement* and *hasPhysicaldataElement*. The fact that a BDE can be derived from other BDEs (according to a defined derivation rule) is reflected by the class *Derivation*.

Data requirements are represented using the class *DataRequirement* with a name, a description and the assignment of a requestor. Data requirements are formally specified by selecting one or more BDEs (using the *RequirementSpecification* class). Again, the relationship *requiresBusinessDataElement* will be derived for easier querying.

Figure 8 shows an excerpt of the resulting ontology of our banking example. It covers the *MortgageLoan_NominalInterestRate* and *MortgageLoan_InterestIndexMarkup* (i.e. the markup on a reference interest rate like Euribor) BDEs (shown yellow) and its physical representations (shown red) in two source systems (called *HYP* and *FIN*) and a data mart used for risk management (called *RISK_DM*). As discussed in section 2.2, *MortgageLoan_NominalInterestRate* is represented in the DWH as a *ACCOUNT_FEATURE.ACCOUNT_FEATURE_RATE* using 'ITR_NOM' as the *FEATURE_TYPE_CD*. Note that the inferred properties *isTransformedInto* and *impacts* are not shown in figure 8 for readability reasons.

Also note, that the view in figure 8 is not meant to be exposed to the users. Instead, figure 9 (a) shows the result set and the inferred *isTransformedInto* properties of tracing the physical column *RISK_DM.LOAN.NOM_INTEREST_-*

[19]Note that doing this properly is a non-trivial task as coupling the BDE ontology with the rest of our metamodel would require RDF reification mechanisms as different levels of modeling abstractions would be mixed.

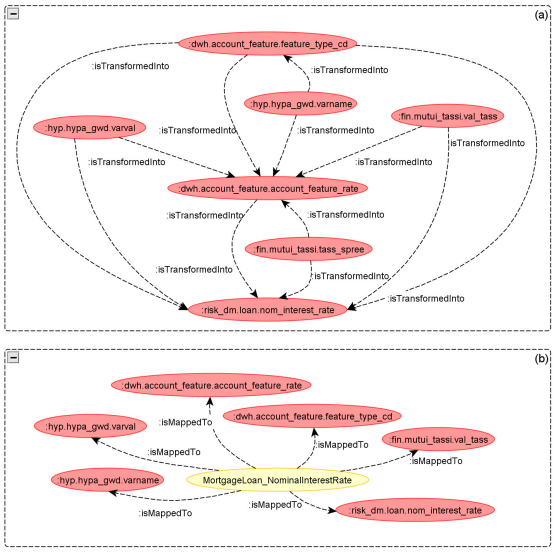


Figure 9: Tracability queries with derived properties

RATE back to its origin (based solely on the physical transformations, i.e. not considereing the business model). Note that the source column *FIN.MUTUI_TASSI.TASS_SPRE* is also shown, although it actually contains a different rate value (the *MortgageLoan_InterestIndexMarkup* BDE). It is detected by the lineage query as it is also stored in the same DWH column *ACCOUNT_FEATURE.ACCOUNT_FEATURE_RATE*, but using a different *FEATURE_TYPE_CD*. This is exactly the lineage ambiguity problem we have mentioned in section 2.2. In turn, figure 9 (b) shows the tracing information when a business model is used. In this case, the BDE *MortgageLoan_NominalInterestRate* and their physical representations are shown, using the mapping property *isMappedTo* rather than the physical transformations.

Finally, figure 10 presents the architecture of our research prototype. The representation of the presented metamodel in OWL is realized by using Protégé[20] and then imported into HP Jena, containing a rule-based inference engine capable of processing custom inference rules and mechanisms to store, manage and query ontologies [2]. The ontology is persistently stored in a MySQL DB. In order to define the mappings and transformations between data elements (both business and physical), we developed a simple mapping editor which creates RDF/XML code ready to be processed by Jena. Finally, we revert to the JUNG framework [21] for visualizing the query result sets as graph structures.

5. CONCLUSION & FUTURE WORK

We have shown that semantic business models are inevitable for enabling harmonization and reuse in complex DWH environments. The Teradata Business Data Element (BDE) concept can serve as a basis for the definition of such a model. Furthermore, we have presented an approach which utilizes Semantic Web standards and technologies to manage business as well as physical data elements and their connections through mappings and transformations. We have extended our prototype presented in [13] accordingly and introduced the notion of data requirements.

One further area of research is the definition of a proper foundation, rules and best practices for the semantic resp. BDE model. Besides investigating the use of OWL classes and properties to represent this model, main questions that remain answered are:

- **How to deal with the parent-child hierarchy of the business data structures?** Should business terms be assigned to all levels? What impact does this have on the transformations that can be derived from the mapping. For example, consider a the *Account Number* term from the example in section 4.1. Should there be only BDEs for the individual account types (i.e. *Current Account.Account Number*, *Savings Account.Account Number*, *Mortgage Loan.Account Number*) or only for the generalized account (i.e. *Account. Account Number*), or both?[22] There may be a source system that delivers data for all types of accounts, but also individual source systems with different mappings. Hence, this question is not trivial.

- **How should relationships between business structures be modeled?** Consider the business term *Account Owner* from the example. There could also be a corresponding *Owned Accounts* term linked to the *Customer* structure. In this case the two BDEs should be linked with a derivation rule. Another approach would be to explicitly model relationships as individual structures (i.e. introduce a *Account Customer Relationship* structure). Again, the impact on the usability for BDE-to-physical mappings and derivation of transformations has to be considered.

- **How should the mapping, derivation and transformation rules be specified?** In the example in section 4.3 we used SQL-like rules specifying the codes like 'ITR_NOM', etc. Further expressiveness will be needed for more complex rules. In order to allow the automatic derivation of physical transformations from the BDE mappings, this needs to be (at least semi-) formally defined.

- **How detailed should a business structure model be?** Or in other words, how many business structures are manageable and can be properly browsed by business users? So far we have worked in projects with business models containing 80-100 structures. It seems that this is a somehow practicable number, but this needs to be verified more thoroughly.

- **Which categorization schemes are practicable as additional term groupings?** The authors have experimented with both business categories as well as "technical" categories such as distinguishing master data (attributes) from measures, but further research is needed. Furthermore, it needs to be checked

[20] http://protege.stanford.edu
[21] `http://jung.sourceforge.net/`
[22] The use of OWL would actually automatically duplicate the properties down the hierarchy through inheritance.

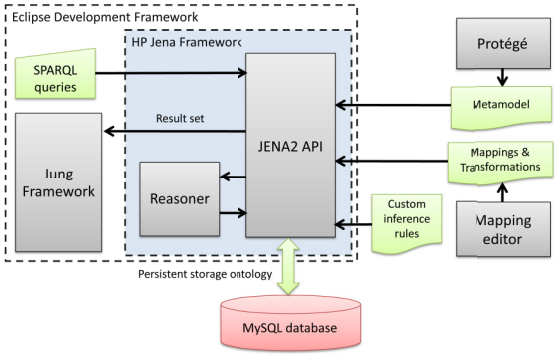


Figure 10: Prototype architecture

if the model needs to be enhanced to also include BDE groups (in addition to term groups).

- **How to setup governance for business data structures and terms?** Business structures should be aligned across the enterprise (similar to the entities in an an enterprise-wide ER model). Business terms (cf. the attributes in an ER model) can be defined iteratively as they get requested by a certain business department. Alignment then needs to take place as part of the iterative process to avoid redundancies. Corresponding data governance roles and responsibilities need to be defined in detail.

Another field of work is the definition of a proper tool support, both for BDE management and for defining source-to-BDE, BDE-to-DWH and BDE-to-target (e.g. data mart) mappings and to derive the development specification for the transformations. As mentioned in the related work in section 3, there are bits and pieces available on the market (e.g. IBM FastTrack with Business Glossary, Teradata Mapping Manager, etc.), but none of them fulfills all needed requirements. We will continue to extend our research prototype as a proof of concept.

Furthermore, based on the presented work, we plan to develop techniques to perform impact analyses on requirement changes in a timely manner to support DWH evolution and change management. Especially, the semantic representation and tracing of data requirements and their links to DWH architecture components (business and physical data elements, transformations) will be studied further. Moreover, we will examine if and how security requirements can be integrated into the semantic model and how traceability also for security requirements can be realized.

6. REFERENCES

[1] D. Brickley and R. Guha. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation 10 February 2004. http://www.w3.org/TR/rdf-schema/, 2004.

[2] J. J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. Jena: implementing the semantic web recommendations. In *Proc. of 13th Intl. World Wide Web Conference on Alternate track papers & posters (WWW '04)*, pages 74–83, New York, NY, USA, 2004. ACM.

[3] Y. Cui and J. Widom. Lineage tracing for general data warehouse transformations. *The VLDB Journal The International Journal on Very Large Data Bases*, 12(1):41–58, 2003.

[4] M. Dean and G. Schreiber. OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-owl-ref-20040210/,2004.

[5] H. Fan and A. Poulovassilis. Using AutoMed metadata in data warehousing environments. In *Proc. of the 6th ACM Intl. workshop on Data warehousing and OLAP (DOLAP'03)*, 2003.

[6] P. Giorgini, S. Rizzi, and M. Garzetti. Goal-oriented requirement analysis for data warehouse design. In *Proc of the 8th ACM Intl. Workshop on Data Warehousing and OLAP (DOLAP'05)*, pages 47–56. ACM, 2005.

[7] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission 21 May 2004. http://www.w3.org/Submission/SWRL/, 2004.

[8] F. Manola and E. Miller. RDF Primer, W3C Recommendation 10 February 2004. http://www.w3.org/TR/rdf-primer/, 2004.

[9] J.-N. Mazón, J. Trujillo, and J. Lechtenbörger. Reconciling requirement-driven data warehouses with data sources via multidimensional normal forms. *Data & Knowledge Engineering*, 63(3):725–751, 2007.

[10] A. Poulovassilis. Tracing Data Lineage Using Schema Transformation Pathways. In *Knowledge Transformation for the Semantic Web*, volume 95 of *Frontiers in Artificial Intelligence and Applications*, pages 64–79. IOS Press, 2003.

[11] E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF, W3C Recommendation 15 January 2008. http://www.w3.org/TR/rdf-sparql-query/, 2008.

[12] B. Ramesh and M. Jarke. Toward reference models of requirements traceability. *IEEE Trans. Software Eng* , 27(1):58–93, 2001.

[13] A. Reisser and T. Priebe. Utilizing Semantic Web Technologies for Efficient Data Lineage and Impact Analyses in Data Warehouse Environments. In *Proc. of the 8th Intl. Workshop on Web Semantics (WebS '09), in conjunction with DEXA '09*, pages 59–63, Washington, DC, USA, 2009. IEEE Computer Society.

[14] B. Romero and A. Abelló. Automating multidimensional design from ontologies. In *Proc. of the 10th ACM Intl. workshop on Data warehousing and OLAP (DOLAP '07)*, 2007.

[15] D. C. Schmidt. Guest Editor's Introduction: Model-Driven Engineering. *IEEE Computer*, 39(2):25–31, 2006.

Integrating Innovation into Enterprise Architecture Management

Michael Rohloff
University of Potsdam
August-Bebel-Str. 89
14482 Potsdam, Germany
michael.rohloff@wi.uni-potsdam.de

ABSTRACT

The ability of organizations to innovate is acknowledged as an essential capability to compete in a competitive market. This paper proposes to use enterprise architecture management as a systematic approach to innovate the enterprise. The enterprise architecture approach is based on a comprehensive architecture framework which aligns the domains of business -, application - and infrastructure architecture. The framework addresses all dimensions relevant for enterprise innovation like business model, organization, processes, and technology and provides appropriate design techniques.

A comprehensive architecture development process is introduced which integrates innovation as a central element for the enterprise architecture design. The process encompasses all activities from business vision & strategy, architecture design to implementation. In addition, the organization of architecture management and the main stakeholders from the business as well as IT side and their respective role are explained.

Keywords

Enterprise transformation and innovation, enterprise architecture framework, architecture development process, enterprise architecture management, stakeholders

1. INTRODUCTION

The ability of organizations to innovate is acknowledged as an essential capability to compete in a competitive market. There are several dimensions of innovation discussed in literature and practice. Besides technological innovation there can be also an organizational innovation in terms of new processes or channels to communicate with customers and suppliers for instance. Davila for example identify six levers of innovation [15]: value proposition, supply chain, target customer as part of business model innovation and product & services, process technologies and enabling technologies as part of technology innovation.

Innovation has as many definitions as it has contexts; we follow Sørensen in his definition of innovation as a process of creativity, ability to implement and to succeed [43]. Thus, innovation is not only creating ideas but the ability to implement new ideas rapidly, and the ability to succeed in the market place (products & services) or within the organization (services or processes).

In this understanding fostering innovation is not only providing a creative environment but also to provide the ground for implementation and success in the market or the organization. The importance of an innovative culture of the organization and the innovative attitude of individuals is argued extensively and has to be paid high attention [15, 32]. Nevertheless providing creativity and freedom is only one important side of it which should be supplemented by a systematic approach fostering the implementation and success of these innovative ideas.

Here, Enterprise Architecture Management comes into place; it not only addresses all dimensions of enterprise innovation with its architecture design it also provides a systematic approach. This paper argues to leverage this potential for innovation.

Enterprise Architectures are in the scope of interest in the recent years in academia as well as practice [1, 10, 16, 35, 36, 39, 45]. It is recognized as an instrument for business/ IT alignment [5, 7, 24, 30, 31, 47]. Architecture is a commonly used term in the design of information systems. However understanding and structuring of enterprise architecture and their basic elements differs (see the discussion in [151, 11, 17, 18, 48]).

IEEE Standard 1471-2000 defines architecture as "„.. the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution" [25].

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Enterprise Architecture comprises the entire business with all its constituents. The alignment of the business and organizational design of the enterprise (business architecture) with the IT architecture is fundamental. The wide range of different domains and scope of enterprise architecture and its high complexity are characteristics of enterprise architectures.

In contrast to information systems architecture, which is widely discussed under the aspect of a single information system being integrated in an organization and aligned with business processes, enterprise architecture takes the entire IT landscape into focus. In comparison architecture is understood as city planning and not only as planning the architecture of a house [12, 21]. It requires the definition of development plans for an entire area and not only the construction plan for a building. The development of the IT landscape in contrast to the information system architecture of a single system is architecture design on a large scale.

To summarize, the extension of architecture design to the business and not only IT architecture focus and the large scale development encompassing the entire enterprise give a new perspective to Enterprise Architecture Management. The intended scope (opposed to the Enterprise Architecture Management practice which is still IT driven, see [10, p. 11f.]) is truly the architecting of the entire enterprise or organization, institution or value network etc. in focus.

The following sketches some objectives to be pursued with Enterprise Architecture Management [4, p. 9f., 34, p. 9f., 35, p. 6f., 49f., 46, p. 6f.]:

- Business/ IT alignment and leverage of IT
- Governance, planning and controlling of I&C program and compliance to regulations & standards
- Implement the I&C landscape in a systematic and efficient manner
- Provide for Scalability, Growth and Adaptability
- Transparency of architecture building blocks and better communication of stakeholders

Enterprise Architecture is a means to support business and IT alignment. Architecture planning is the ground for IS strategy implementation and the development of the IT landscape. At the same time an enterprise architecture approach provides the agility to react fast to market requirements.

However, while strategy and business orientation and the leverage of IT is widely addressed little attention is given to use architecture for a systematic approach towards innovation of the enterprise and its business. Enterprise Architecture constitutes all important building blocks of the enterprise like organization, processes and technology. Thus, it addresses all dimensions of enterprise innovation. Enterprise Architecture Management provides a structured and systematic approach towards the design of the enterprise. Enterprise Architecture Frameworks and supporting methods and tools provide transparency on the building blocks and their interrelationships and ease of communication among the stakeholders.

At present in many companies this potential of Enterprise Architecture Management to leverage innovation is only partly used. A major drawback is that Enterprise Architecture design is often seen as an IT project with focus on technology implementation and not as a business issue to design and innovate the entire enterprise [10, p. 11f.].

This paper proposes to use enterprise architecture management as a systematic approach to innovate the enterprise. It is based on an enterprise architecture framework and corresponding methods which were developed in a number of architecture projects (for details of the framework and its application [38]). The next section gives an outline of the framework and methods in order to show how it can be used for the proposed innovative enterprise architecture development. It is followed by a section which introduces a comprehensive architecture development process designed to integrate innovation into Enterprise Architecture Management. The process encompasses all activities from business vision & strategy, architecture design to implementation. In addition, the organization of architecture management and the main stakeholders from the business as well as IT side and their respective role and usage of the design techniques are explained.

2. FRAMEWORK AND REFERENCE FOR ENTERPRISE ARCHITECTURE DESIGN

2.1 TOGAF and other frameworks

An architecture framework is a foundational structure, or set of structures, which can be used for developing a broad range of different architectures. It should describe a method for designing a target state of the enterprise in terms of a set of building blocks and how these building blocks fit together. In addition, it should contain a set of tools and provide a common vocabulary [46, p. 7].

For an overview on enterprise architecture frameworks see [18, 26, 28, 29, 41, p. 85f., 42]. Examples for frameworks are: The Open Group Architecture framework (TOGAF [46]), Federal Enterprise Architecture Framework (FEAF [19]), Generic Reference Enterprise Architecture Methodology (GERAM [22]), Gartner and META Group Enterprise Architecture Frameworks [20, 35], and Zachman Framework [49, 50]. Regardless which framework will be selected, it is important to base architecture development on a framework. It is an essential means for transparency, communication, and a systematic approach.

The Generic Reference Enterprise Architecture Methodology, as the name implies, aims for a generic approach with a common framework embracing the other frameworks. As a result it remains rather abstract.

The Open Group Architecture framework (TOGAF Version 9 [46]) is widely recognized and plays a prominent role. It provides the methods and tools for assisting in the acceptance, production, use, and maintenance of enterprise architecture. It is based on an iterative process model supported by best practices and a re-usable set of existing architecture assets.

TOGAF has been developed through the collaborative efforts of 300 Architecture Forum member companies and represents best practice in architecture development. Using TOGAF as the architecture framework will allow architectures to be developed that are consistent, reflect the needs of stakeholders, employ best practice, and give due consideration both to current requirements and to the likely future needs [46, p. 7].

It is designed as an open framework which may be used in parts, enhanced with on deployments, and adapted to company specific needs. It serves as a guideline and best practice collection thus leaving freedom and giving a frame in terms of an outline and general descriptions.

The core concepts of TOGAF are [46, literally extracted from p. 9-16.]: The Architecture Development Method (ADM) provides a tested and repeatable process for developing architectures. The ADM includes establishing an architecture framework, developing architecture content, transitioning, and governing the realization of architectures. Architects executing the ADM will produce a number of outputs as a result of their efforts, such as process flows, architectural requirements, project plans, project compliance assessments, etc.

TOGAF includes the concept of the Enterprise Continuum, which sets the broader context for an architect and explains how generic solutions can be leveraged and specialized in order to support the requirements of an individual organization.

Supporting the Enterprise Continuum is the concept of an Architecture Repository which can be used to store different classes of architectural output at different levels of abstraction, created by the ADM. In this way, TOGAF facilitates understanding and co-operation between stakeholders and practitioners at different levels.

By means of the Enterprise Continuum and Architecture Repository, architects are encouraged to leverage all other relevant architectural resources and assets in developing an Organization-Specific Architecture.

In order to carry out architectural activity effectively within an enterprise, it is necessary to put in place an appropriate business capability for architecture, through organization structures, roles, responsibilities, skills, and processes."

The TOGAF framework consists of four architecture domains: The business (or business process) architecture defines the business strategy, governance, organization, and key business processes. The data architecture describes the structure of an organization's logical and physical data assets and data management resources. The applications architecture provides a blueprint for the individual application systems to be deployed. Finally the technology architecture describes the logical software and hardware capabilities that are required to support the deployment of business, data, and application services.

However, since the data architecture is defined as the 'logical and physical data assets of an organization it mixes business requirements with technical implementation.

2.2 The Enterprise Architecture Triangle

In this section a framework is introduced which structures enterprise architecture in key domains and building blocks in order to give a comprehensive view on all relevant aspects of enterprise architecture. The framework and the architecture modeling approach briefly introduced in the next section have been successfully implemented in a number of projects (for details of the framework and its use in architecture projects, see [38]).

In difference to TOGAF the architecture framework introduced in this paper clearly separates the domains of business and IT architecture. This provides for a clear distinction of the business oriented description of the enterprise architecture and the derived technological implementation. Hence, in the proposed framework, in difference to TOGAF, the information architecture is not described as a separate architecture domain but split in a building block of the business architecture in terms of logical information structures and a building block of applications architecture in terms of implementation of data repositories. Furthermore, the framework details the domains in architecture building blocks to give a comprehensive overview of all constituents of enterprise architecture.

The framework is composed of three basic domains each with distinct architecture building blocks.

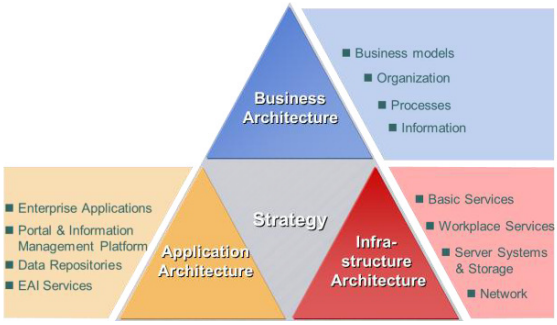


Figure 1: Enterprise Architecture Framework

The business architecture describes the fundamental organization and requirements of the business based on business strategy and objectives. It is composed of the four building blocks business model, organizational architecture, process architecture, and information architecture.

The business model gives a high level view on the nature of the business in terms of products & services offered in the market, the value chain, business partners, market channels utilized, and the combination of resources and information needed for value creation.

The organizational architecture describes the organizational design of the enterprise and the principal cooperation with customers and suppliers.

The process architecture classifies and describes all processes of the business and their respective value creation. It is the core building block of the business architecture. The process architecture can be classified in the core business processes customer relationship management, supply chain management,

product life cycle management and the management and support processes.

The information architecture shows the logical structure of all information entities like products, business partners, logistic information etc.

The application architecture gives an overview on all applications supporting the processes of the business with the building blocks enterprise applications, portal & information management platform, data repositories, and EAI Services.

Enterprise applications are supporting the automation of business processes and can be assigned to the respective process in terms of their functional support.

The portal and information management platform is the universal access to all company information and knowledge. Portals are designed to specific user groups like customer, business partner, and employees.

Data repositories are the physical storage of all relevant company data and provide an integrated view on product -, customer and business partner - , logistic -, or financial data.

Enterprise application integration services provide the integration of applications and data across the company. Integration technology spans from message exchange, data exchange to process integration.

The infrastructure architecture, also referred to as technology architecture, comprises the software, hardware and network infrastructure required for operations of all applications. Infrastructure building blocks are basic services, workplace services, server systems & storage, and the network.

Basic services are essential applications providing a specific functional support which are independent from business processes. Examples are communication services like Email or Telco, directory or search services, and administration services like single sign on or PKI. Web Services are emerging services which transfer the concept of basic service to the internet. Basic services are modules to be used in different applications.

Workplace services provide for presentation and work with information and for productivity support at the work place (office desk, shop floor etc.). They comprise fixed and mobile devices at the client site and include basic office applications and browser.

Server systems support all back end resources like applications, data repositories, integration services etc. Storage provides all memory capacity for running the applications on the server.

Networks provide the communication links in the LAN and WAN, within the company, between companies and in the Internet.

Security is integral part of all architecture building blocks and described in an overlaying structure.

This paper can only give an outline on the domains of the enterprise architecture framework and sketch the main building blocks at a high level. All building blocks of the framework are detailed down to the level of modules, systems and components. The framework gives a comprehensive description of all relevant elements of enterprise architecture providing a principal structure and classification schema used as a reference for architecture development.

2.3 Enterprise Architecture method and artifacts

The projected method for architecture design consolidates the modeling of enterprise architectures to a few essential and ample techniques (for details of the architecture models see [38]). For the design of architecture a "Service Oriented Architecture" (SOA) approach is followed (for SOA principle e.g. [2, 9, 33, 37]. Basically, IT-architecture can be always seen as providing services to the business e.g. an application supporting a business process, office and communication services at the workplace supporting each individual employee. Taking these perspective puts the value adds of IT in the focus. Consequently the building blocks of the architecture framework are structured in service groups, core services, and service modules.

Enterprise Architecture is more than the collection of the constituent architectures. The inter-relationships among these architectures, and their joint properties, are essential to the enterprise architecture. Thus, the architecture domains should not be approached in isolation. Key element of architecture design is to account for interdependencies among the building blocks of architecture. Blueprints are introduced as a means in planning the deployment of architecture on a large scale. Blueprints give a comprehensive view on the building blocks and how the interact. They show the effects of architecture design between business, application, and infrastructure architecture.

In the focus of enterprise architectures is the alignment of business and IT. In other words, the design of the business architecture determines the IT architecture which has to support and enable business. The building blocks of business architecture with the process architecture as the core define the frame for the design of the IT landscape. The dependencies between the different architectures can be described in blueprints.

A blueprint is a plan which describes the deployment of an architecture building block across the enterprise. It pictures the landscape of this building block in a matrix of two business dimensions.

- The application landscape describes for each business process how it is supported by applications. The second dimension shows the deployment in organizational units, like divisions, business segment etc.
- The data repository landscape describes the deployment with databases and how the support defined information clusters of the information architecture. The second dimension shows the deployment of the databases in organizational units.
- The service landscape shows the deployment of infrastructure services and the support of applications. The second dimension shows the deployment in organizational units.

In general, different types of blueprints can be generated depending which dependency of business -, applications -, and IT infrastructure architecture or building block is in focus. Also, the matrix dimensions can be chosen in different level of detail, although the experience in architecture projects shows that a high level is sufficient in order to derive decisions for architecture development. Generating to detailed blueprints involves a lot of resources and time without enriching the decision base in the same degree. The three blueprints introduced above provide a good information base for management decisions. Blue prints are used for presentation of "as is" as well as target architecture.

In addition to blueprints, each enterprise architecture domain can be described by views [13], which look at the architecture, its structure and elements from a specific perspective. Three basic views were identified which are sufficient to describe all relevant aspects of enterprise architectures:

- Component view: The view describes the logical and functional structure of the architecture in scope. All building blocks and their systems and components are described in terms of composition, structure and relationships among one another. The component view allows for different level of detail of the architecture. Components, systems, subsystems, building blocks can be grouped or decomposed. The segmentation of the diagram is in building blocks based on the respective architecture in scope.
- Communication view: The view describes the communication (interaction) between systems and components. The relationship among the systems is decomposed in the interaction of components within a system and to other systems. Different types of communication can be described with distinct notation for communication lines. The segmentation of the diagram is in communication areas based on the respective architecture in scope.
- Distribution view: The view describes the allocation of systems or components in terms of geographical or organizational distribution. The diagram is segmented in organization or location based on the respective architecture in scope.

These three views can be applied to all domains of enterprise architecture. However, the segmentation of the diagram is selected according to the appropriate context of the domain and chosen level of detail. The high complexity of enterprise architectures can be reduced by taking particular views which focus on specific aspects of architecture. The three views facilitate the reduction to core entities and construction principles and the understanding of their behavior.

In a number of projects these core design techniques, the blueprints and three views used for all domains and architecture building blocks, proved to be capable to advance the communication and architecture design of the different parties involved in architecture development [38].

In the following section we show the interrelationship of business and IT architecture and focus on the process for architecture development. Also, we explain how stakeholders can use the artifacts for innovation of the enterprise.

3. ARCHITECTURE DEVELOPMENT AND ENTERPRISE INNOVATION

3.1 The Innovation Potential

The objectives and success factors for enterprise architecture discussed in the first section show the relevance of architecture management and a defined architecture development process. In particular the transparency of architecture and the communication of objectives and results of all activities and artifacts along the process of architecture development are in the focus. Enterprise architecture is even seen as a change agent [35, p. 8].

The following summarizes the key features of Enterprise Architecture Management explained in the prior sections which bear a high potential to be used for innovation of the business:

- Enterprise Architecture is a comprehensive approach which constitutes all important building blocks of the enterprise like business models, organization, processes and technology.
- It addresses all dimensions of enterprise innovation.
- Enterprise Architecture Frameworks and supporting methods and tools provide
  - transparency on the building blocks and their interrelationships
  - easier communication among the stakeholders.
- Enterprise Architecture Management provides a structured and systematic approach towards the design of the enterprise.

At present in many companies this potential of Enterprise Architecture Management to leverage innovation is only partly used. A major drawback is that Enterprise Architecture design is often seen as an IT project with focus on technology implementation and not as a business issue to design and innovate the entire enterprise [10, p. 11f.]. Hence, based on reported deficiencies, own experience in architecture projects and guidelines for Enterprise Architecture Management, we design and propose an improved architecture development process which integrates innovation in a systematic manner.

3.2 The TOGAF Architecture Development Method (ADM)

At first we take a look what TOGAF recommends for Enterprise Architecture Management, in order to contrast the traditional approach. The TOGAF Architecture Development Method (ADM) provides a tested and repeatable process for developing architectures [46, p. 49f.]. The ADM includes establishing an architecture framework, developing architecture content, transitioning, and governing the realization of architectures. All of these activities are carried out within an iterative cycle of continuous architecture definition and realization that allows organizations to transform their enterprises in a controlled manner in

response to business goals and opportunities. Phases within the ADM are as follows:

The Preliminary Phase describes the preparation and initiation activities required to prepare to meet the business directive for a new enterprise architecture, including the definition of an organization-specific architecture framework and the definition of principles.

Phase A: Architecture Vision describes the initial phase of an architecture development cycle. It includes information about defining the scope, identifying the stakeholders, creating the Architecture Vision, and obtaining approvals.

Phase B: Business Architecture describes the development of a Business Architecture to support an agreed Architecture Vision.

Phase C: Information Systems Architectures describes the development of Information Systems Architectures for an architecture project, including the development of Data and Application Architectures.

Phase D: Technology Architecture describes the development of the Technology Architecture for an architecture project.

Phase E: Opportunities & Solutions conducts initial implementation planning and the identification of deliver y vehicles for the architecture defined in the previous phases.

Phase F: Migration Planning addresses the formulation of a set of detailed sequence of transition architectures with a supporting Implementation and Migration Plan.

Phase G: Implementation Governance provides an architectural oversight of the implementation.

Phase H: Architecture Change Management establishes procedures for managing change to the new architecture.

Requirements Management examines the process of managing architecture requirements throughout the ADM.

Taking a closer look to the activities listed for each of these phases reveals that the scope and focus is on IT architecture design which has to match the business architecture. The design of the business architecture is not given the same attention. Here the potential of Enterprise Architecture Management especially for the innovative architecture design of the business is not used.

Preliminary phase and architecture vision provide the ground for architecture projects in defining the scope and baseline of projects, management buy in, identifying stakeholders etc.

Phase B – D have the focus on analyzing the architecture domains, documenting it and designing a future state.

Phase E – H deal with the implementation of new architecture by governance, portfolio management and project execution.

Thus, the Architecture Development Method (ADM) provides a good ground for EAM activities with IT architecture in its core but needs to be enhanced for innovation and business design issues.

3.3 Enterprise Architecture Development

Based on the potential of enterprise architecture management for innovation we propose and introduce in the following an architecture development process which aims to leverage that potential. Key features of the extended architecture development process are:

- to use enterprise architecture management and the architectural artifacts for a systematic approach towards innovation
- to integrate creative brainstorming and innovative design at well-defined stages into the process
- to use open innovation for the architecture design [14]

Figure 2 gives an overview of the overall structure and steps of the architecture development process. It advances the understanding of architecture management by extending the scope to business design and innovation (compare the architecture process in [35, p. 5, 57f., 46, p. 51f.].

In addition, the following principles guided the design of the process:

- Scope is business engineering not only IT architecture
- Comprehensive process from visioning to implementation covering the entire business
- Balanced involvement of business people and IT

It is proposed to start with an initial “Visionate” phase. It extends the traditional architecture management by an open, informal, and unstructured process to generate new ideas how to set up and run the business. Visionate stands for a combination of vision and innovate. This phase brings together business people who know the market, the own business capabilities in terms of skills, organization, and processes and IT people who know the information systems and I&C technology. Objective is to develop in cooperation business vision and innovation ideas in all dimensions based on a deep understanding of the capabilities and the market.

Ideas should be generated free of restrictions of the current business design. Sources are business capabilities, market & technology trends, and blueprints of architecture. The “visionate team” elaborates on the business model building block, sketches business scenarios and envisions blueprint options for both business architecture and IT architecture. Here, an open innovation is applied, e.g. integrating customers to develop new business model ideas. The results are worked out independent from the current architecture in order to be open minded for innovation and free of any restrictions. However, architecture blueprints can be used as a powerful means to understand the current business.

It is followed by the “Definition of the Business Strategy”. Here, the results of the Visionate are matched against the current business architecture and evaluated, resulting in the formulation of a business strategy.

The development of the business strategy is followed by the two phases of architecture design, the “definition of business architecture” and “definition of IT-architecture”. Both phases are

closely linked and perform the same activities for their respective domains.

Architecture development is linked to business strategy which is the starting point for the definition of the IT strategy taking environmental and technological trends into account. This strategy alignment is basically the first cycle of business/ IT alignment. The strategy is detailed to features of the enterprise architecture. Here the architecture framework comes into place, which links business -, applications -, and infrastructure architecture and the respective building blocks. This is the kernel of architecture development where techniques like blueprints are applied.

Based on "as is blueprints" of the business and IT landscape "target blueprints" for all architecture building blocks are defined. They are derived from the business and IT strategy. In addition, an adaptability analysis is performed in order to assess how the information systems in place can adapt to a changing environment [23]. The results are used for the definition of target blueprints. In this phase architects collaborate in teams which range from enterprise to the specific domains of business or IT architecture. These activities correspond to phase B - D of TOGAF.

An Enterprise Architecture however, includes not only the three domains for the "as is" architecture (baseline architecture) and the target architecture. It is built on a strategic information base with a clear definition of business objectives and strategy. The strategy is needed for the transitional processes in order to implement new business design and technologies in response to the changing business needs. That means the enterprise architecture management includes also the process to create, update and manage the evolution of the architecture domains in line with business strategy. These activities correspond to Architecture Change Management and Requirements Management of TOGAF.

In addition to the initial "Visionate" an innovative design of the architecture domains is encouraged based on a clear understanding of business strategy and a transparency on the interrelation of the architecture domains. Here, open innovation is applied again, e.g. integrating I&C technology experts to develop technology options to best support business.

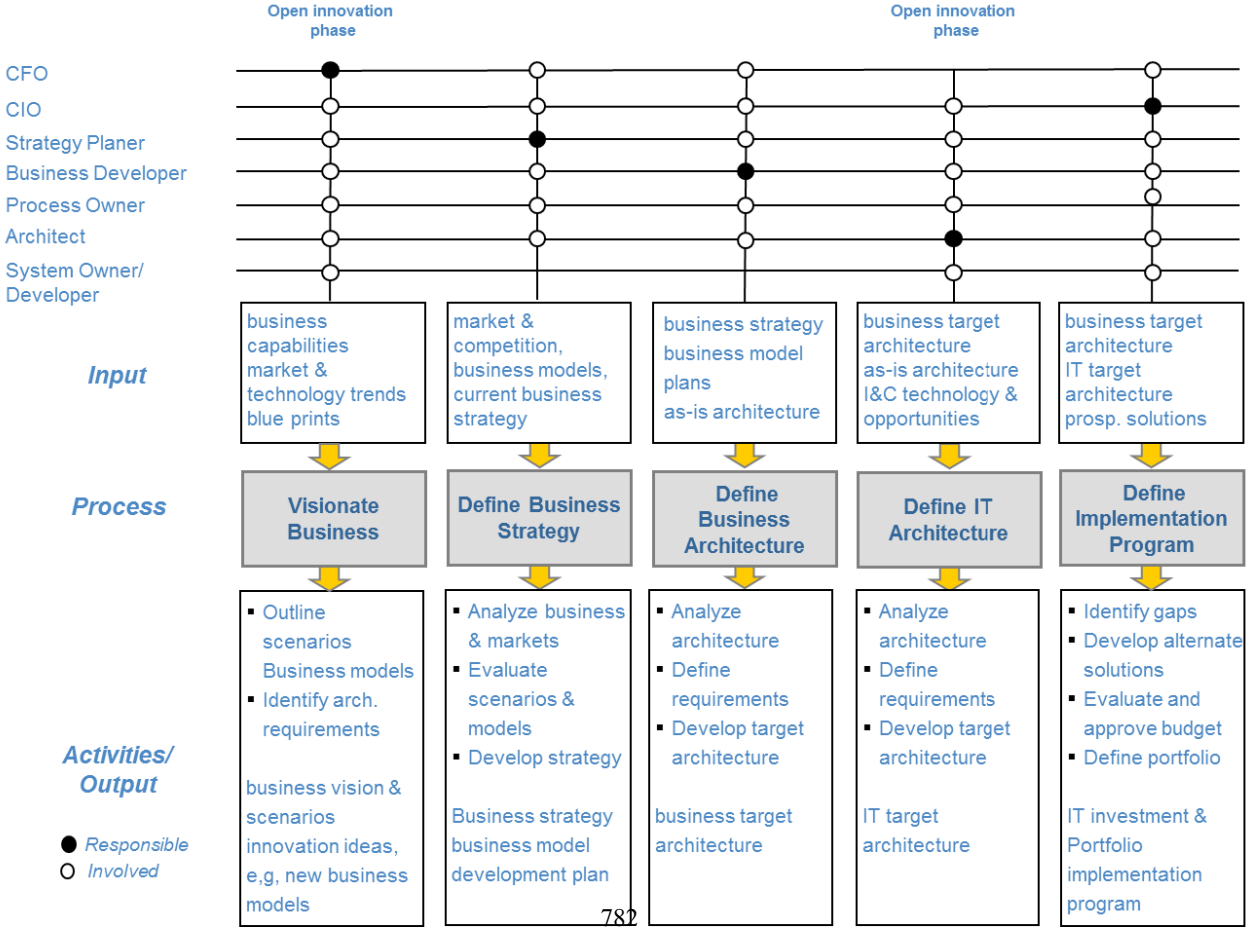
The architecture design phases are followed by the "Definition of an Implementation Program" before hand over to project execution. Architecture blueprints describe the deployment plan to implement IT strategy. From the gap analysis of "as is" and target architecture potential projects are derived.

Figure 2: extended enterprise architecture development process

In an investment & implementation planning the projects are prioritized, migration strategies are worked out and the overall program and portfolio is defined in order to implement the architecture. A business oriented program and portfolio management is of central importance [27, 35, p. 113f.].

The execution of the respective program and the projects finally result in changes to the current architecture and IT service operations. The implementation of the target architecture forms the second cycle of business/ IT alignment. These activities correspond to phase E - H of TOGAF.

In addition to the main focus to drive the architecture development process, architecture management deals with activities



like documentation and maintenance of architecture, communication of architecture benefits, management buy in, and securing compliance to legislation and company standards. All these activities require the involvement of stakeholders from business as well as IT. The following section shows how the architecture development can be supported by an appropriate architecture management organization

3.4 Organization and Stakeholders in Architecture Management

The development of architectures is complex and extensive. It involves a number of participants with different professional background. Many of them have other permanent tasks in addition to architecture design. Thus, it is recommended to set up a project organization for architecture development at high company level, besides a permanent staff within CIO which is responsible for architecture management [16, p. 108f., 35, p. 69f., 46, p. 51f.].

The following proposed project organization is based on Meta Group clients architecture projects experiences and company internal architecture projects experience, [35, p. 69 f., 38].

In practice, for most projects IT is the driver of the architecture development and the business is often not adequately involved [10, p. 11f.]. In consequence the following objectives are important for architecture management success:

- Set up as a business project with clear business goals no technical focus, high level management back up
- High involvement of business people (e.g. resources and management functions)
- Cooperative teams with a good mix of expertise and skills in business and IT
- Awareness building and communication of project scope across the enterprise
- and providing an innovative culture

Figure 3 outlines the overall project organization (compare [35, p. 76]. The core is the architecture team led by an architect at enterprise level. The team is composed of architects of all architecture domains ([35, p. 69f.] recommends not to exceed 8 persons [46] recommends about 5-10 max.). Additional teams are set up for diverse architecture domains. These teams drive the overall process, create and maintain deliverables and oversee all infrastructure projects. They develop an architecture requirements definition and other key artifacts and present it to the steering committee and architecture review board for approval.

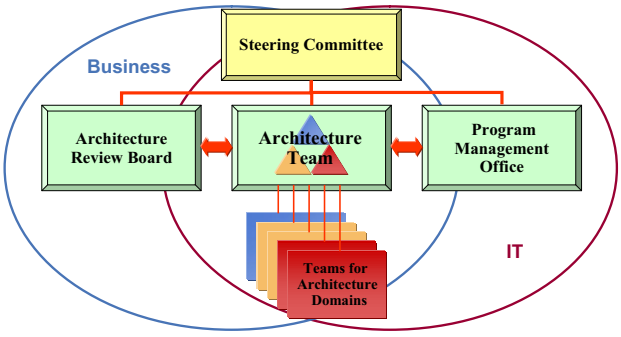


Figure 3: Overall architecture management project organization

A steering committee representing seniority and governance sets the general regulations and gives guidance for the project and the overall architecture development process. The steering committee is set up with business and IT managers. The architecture review board provides input for, reviews, and approves the overall conceptual architecture and standards set. The program management office is responsible for the planning and execution of the entire project. It governs the architecture development from analysis, innovation, and target design to implementation.

The entire project organization should be a good mix of business and IT people. The scope of enterprise architecture and the architecture development process in mind, ranging from visioning to implementation, it is obvious that a number of people coming from different organizational backgrounds and disciplines are involved. The enterprise architecture framework and the design techniques outlined in the sections 2.2 and 2.3 are in the core for architecture design.

The stakeholders like CEO/CFO; CIO, IT strategy planer, and program manager are involved in the business/ IT strategy and decisions for direction, objectives of architecture, and IT program. They merely use methods of business IT alignment, IT impact, portfolio techniques and blueprints. Architecture principle and pattern are partly used.

The central role plays the enterprise architect who leads the architecture development and coordinates all respective activities. He is responsible for as is/ target-architectures and dependencies among architectures. In addition, there are other domain related architects involved. Responsibility of architects can be on diverse domains ranging from system to enterprise architecture. Architects use the entire range of techniques with different levels of detail depending on their respective domain.

It is essential that business and IT architects are involved in the overall enterprise architecture team and in the architecture domain teams. For innovation of the business architecture it is important to involve business development manager and also employees who execute the business processes in order to integrate experience and skills across the enterprise. Furthermore in line with an open innovation consideration should be given to use the potential of external innovation knowledge [14].

We recommend open innovation at two well-defined stages of the process: integration of customers to the Visionate phase and I&C experts for the IT architecture design (see section 3.3 and figure 2). Practice shows that at present it is often limited to analysts for the evaluation of I&C trends.

Process owners focus on blueprints which show how processes are supported by applications and services. Principles and patterns for business architecture are also used.

System owners and system developers use component -, communication - , and distribution diagrams with focus at system level. Defined principles and pattern are basis for their work.

The respective techniques are used differently by the stakeholders depending on the respective scope of work. The alternate ways of architecture description are an important means of com-

munication among the stakeholders involved in the architecture development process.

4. SUMMARY AND OUTLOOK

This paper addressed the management of enterprise architectures. The first section introduced a comprehensive architecture framework which classifies the basic domains and building blocks of enterprise architecture. It is used as reference for the constituents of enterprise architecture development. The framework addresses all dimensions relevant for enterprise innovation like business model, organization, processes, and technology and provides appropriate design techniques. This paper recommends to use enterprise architecture management as a systematic approach to innovate the enterprise.

Based on the projected method for architecture design and experiences in a number of architecture projects the second part outlined the organization of architecture management and the principal architecture development process. An extended architecture development process was introduced which integrates innovation as a central element for the enterprise architecture design. The process encompasses all activities from business vision & strategy, architecture design to implementation. Enterprise architecture needs to integrate business and IT. The main stakeholders from the business as well as IT side and their respective role in the process were explained.

In a next step the proposed extended architecture development process will be applied in a reference project. Based on the project experience, future work will enhance the process with a set of guidelines for architecture management and innovation. This will be added with methods to measure enterprise architecture success and an architecture management maturity model.

5. REFERENCES

[1] Aier, S./ Riege, C./ Winter, R.(2008): Unternehmensarchitektur – Literaturüberblick und Stand der Praxis, in: Wirtschaftsinformatik 50(2008)4, p. 292-304

[2] Allen, P. (2006). Service Orientation, winning strategies and best practices, Cambridge, UK 2006

[3] Arbab, F./ de Boer, F. / Bonsangue, M./ Lankhorst, M./. Proper, H./ van der Torre; L. (2007). Integrating Architectural Models. Enterprise Modelling and Information Systems Architectures, 2(2007)1, p. 40-57

[4] Aranow, E. (2002). Enterprise Integration Strategies, Cutter 2002

[5] Avison, D./ Jones, J./ Powell, P./ Wilson, D. (2004). Using and validating the strategic alignment model, in: The Journal of Strategic Information Systems 13(2004)3, p. 223–246.

[6] Bachmann, F. (2000). Software Architecture Documentation in Practice: Documenting Architectural Layers, Special Report CMU/SEI-2000-SR-004, March 2000

[7] Baumöl, U. (2006). Methodenkonstruktion für das Business/ IT Alignment, in: Wirtschaftsinformatik 48(2006)5, p. 314-322

[8] Bernus, P./ Mertins, K./ Schmidt, G. (Eds.): Handbook of Information Systems, Berlin 1998

[9] Bieberstein, N/ Bose, S./ Fiammante, M./ Jones,K./ Shah, R. (2006). Service-Oriented Architecture Compass - Business Value, Planning and Enterprise Roadmap, Upper Saddle River (Pearson) 2006

[10] Buckl, S./ Ernst, A. M./ Lankes, J./ Matthes, F./ Schweda, C. M. (2009). State of the Art in Enterprise Architecture Management - 2009. sebis, Technical University Munich 2009

[11] Buhl, U.; Heinrich, B. (Ed.) (2004). Meinung/Dialog: Unternehmensarchitekturen in der Praxis – Architekturdesign vs. situationsbedingte Realisierung von Informationssystemen, Wirtschaftsinformatik 46(2004)4, p. 311-321

[12] Burke, B. (2003). Enterprise Architecture or City Planning?, META Group, Report 2638, 2003

[13] Clements, P.; Bachmann, F.; Bass, L. (2003). Documenting Software Architectures: Views and Beyond, Addison-Wesley, 2003.

[14] Chesbrough, H. (2003). The era of open innovation, MIT Sloan Management Review, 44(2003)3, p. 31-41

[15] Davila, T./ Epstein, M./ Shelton, R. (2006). Making Innovation Work. How to manage it, measure it, and profit from it, Wharton School Publishing 2006

[16] Dern, G. (2006). Management von IT-Architekturen, Wiesbaden 2006

[17] EMISA Portal. Enterprise Modelling and Information Systems Architecture, http://www.wi-inf.uni-duisburg-essen.de/MobisPortal, called 2010-07-20

[18] Esswein, W./ Weller, J. (2008). Unternehmensarchitekturen – Grundlagen, Verwendung und Frameworks. in: HMD Praxis der Wirtschaftsinformatik, 45(2008)262, p. 6-18

[19] FEAF: Federal Enterprise Architecture Framework (FEAF), http://www.whitehouse.gov/omb/egov/, called 2010-07-20

[20] Gartner: The Gartner Enterprise Architecture Framework, see http://www3.gartner.com/Init, called 2010-07-20

[21] Gartner Group (2002). Enterprise Architecture and IT “City Planning”, July 2002

[22] GERAM. Framework for a Generic Reference Enterprise Architecture Methodology, http://www.cit.gu.edu.au/bernus/taskforce/geram/report.v1/report/report.html, called 2010-07-20

[23] Gronau, N./ Rohloff, M. (2007). Managing Change: Business/ IT Alignment and Adaptability of Information Systems, in: Österle, H.; D'Atri, A.; Bartmann, D.; Pries-Heje, D.; Golden, W., Whitley, E. (Eds.): Proceedings ECIS 2007, 15th European Conference on Information Systems, "Relevant rigour - Rigorous relevance", St. Gallen 2007

[24] Henderson, J. C./ Venkatraman, N. (1999). Strategic alignment: Leveraging information technology for transforming

organizations, IBM Systems Journal 38(1999)2/3, S. 472–485

[25] IEEE (2000). IEEE Standard 1471-2000, Recommended Practice for Architectural Description of Software-Intensive Systems. IEEE Computer Society, New York, October 2000

[26] James, G. (2004). Architecture Frameworks: Tool Support, Gartner Research, November 2004

[27] Jeffery, M./ Leliveld, I. (2004). Best Practices in IT Portfolio Management, in: MIT Sloan Management Review, Spring 2004, p. 41-49

[28] Lapkin, A. (2004a). Architecture Frameworks: How to Choose, Gartner Research, November 2004

[29] Lapkin, A. (2004b). Architecture Frameworks: Some Options, Gartner Research, November 2004

[30] Luftman, J. (2005). Key issues for IT executives 2004, in: MIS Quarterly Executive 4(2005)2, p. 269–285

[31] Luftman, J. (2003). Assessing IT/business alignment, in: Information Strategy 20(2003)1, p. 7

[32] Lundin, S. (2009). Cats – The nine lives of Innovation, McGrawHill 2009

[33] Marks, E./ Bell, M. (2006). -Service Oriented Architecture: A Planning and Implementation Guide for Business and Technology. Hoboken 2006

[34] Masak, D. (2005). Moderne Enterprise Architekturen, Berlin et al. 2005

[35] Meta Group (2002). Executive Insights: Enterprise Architecture Desk Reference, 2002

[36] Nieman, K. (2006). From Enterprise Architecture to IT-Governance, Wiesbaden 2006

[37] Pulier, E./ Taylor, H. (2005). Understanding Enterprise SOA, Greenwich 2005

[38] Rohloff, M.: Ein Beitrag zur Konsolidierung der Modellierung von Unternehmensarchitekturen, in Kühne, T./ Reisig, W./ Steimann, F. (Hrsg.): Proceedings „Modellierung 2008“, Lecture Notes in Informatics, Vol. P-127, Bonn 2008, S. 105-120

[39] Ross, M./ Weil, P./ Robertson, D. (2006). Enterprise architecture as strategy: Creating a foundation for business execution, Harvard Business School Press, Boston 2006

[40] Scheer, A.-W (1999). Architecture of Integrated Information Systems: Business Process Framework, Berlin et al. 1999

[41] Schekkermann, J. (2006). How to survive in the Jungle of Enterprise Architecture Frameworks: Creating or Choosing an Enterprise Architecture Framework, Trafford 2006

[42] Schulman, J. (2004). Architecture Frameworks: Provide System Road Maps, Gartner Research, November 2004

[43] Sørensen, J. (2009) Jacob-S.NET Innovation, see http://www.jacob-s.net/, called 2010-07-20

[44] Sowa, J.F./ Zachman, J. (1992). Extending and Formalizing the Framework for Information Systems Architecture, in: IBM Systems Journal 31(1992)3

[45] Spewak, C./ Hill, S (2006). Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology, New York 2006

[46] TOGAF (2009). The Open Group Architecture Framework: Version 9 "Enterprise Edition", 2009, see http://www.opengroup.org/togaf/, called 2010-07-20

[47] Weiss, J. W./ Anderson, D. (2004). Aligning Technology and business strategy: Issues & Frameworks, A field study of 15 Companies, in: Proceeding of the 37th Hawaii International Conference on System Sciences 8 (2004) 8, S. 1–10.

[48] Winter, R./ Fischer, R.(2007). Essential Layers, Artifacts, and Dependencies of Enterprise Architecture, in: Journal of Enterprise Architecture 3(2007)2, p. 7-18,

[49] Zachman, J. (1987). A Framework for Information Systems Architecture, in: IBM Systems Journal 26(1987)3

[50] Zachmann Framework, see http://www.zifa.com/, called 2010-07-20

Usability of Modelling Languages for Model Interpretation: An Empirical Research Report

Christian Schalles
Department of Computing
Cork Institute of Technology
Rossa Ave, Cork, Ireland
christian.schalles@mycit.ie

John Creagh
Department of Computing
Cork Institute of Technology
Rossa Ave, Cork, Ireland
john.creagh@cit.ie

Michael Rebstock
Faculty of Economics and Business Administration
Hochschule Darmstadt University of Applied Sciences
64295 Darmstadt, Germany
michael.rebstock@h-da.de

ABSTRACT
Models offer visual support for analyzing complex domains such as business processes and information systems. In both cases, models are developed using graphical modelling languages. In our study we focus on usability evaluation of modelling languages for the model interpretation scenario. The study is based on a causal model of hypotheses, which was developed under consideration of psychological cognitive theories and usability theory. Survey data is collected and the causal relations hypotheses are assessed using a structure equation modelling approach. Our study shows important findings for practical and theoretical issues of how differing modelling languages are influencing usability attributes on causal stages in model interpretation.

Keywords
Modelling Languages, Usability, Perceptibility

1. INTRODUCTION
In organizations models are important for documenting business processes and specifying information system requirements under development. Models are represented by using graphical modelling languages such as BPMN, EPC and UML providing a set of elements, relations and rules for combining them. In general, graphical modelling languages aim to support the expression of relevant aspects of real world domains such as business processes or application system structures [1]. For accurate human interpretation it is important that a model reproduces the knowledge contained in a clearly arranged and well-structured manner. When evaluating the usability of modelling languages it is necessary to distinguish between model interpretation and model development scenarios [2]. A model developer needs (1) to learn a modelling language, (2) to remember the language's elements and syntax to ensure correct models, (3) to reach a fast and correct task accomplishment and (4) to be satisfied with the modelling language.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

A model interpreter needs to recognize the meaning of a model. Due to this fact a model interpreter requires an intuitive and well-defined knowledge regarding shapes, model structure and syntax [3].
This summary research report focuses on empirical usability evaluation of graphical modelling languages in model interpretation. We define underlying background theories connected with our research. Based on this we are theoretically deriving a causal model of hypotheses, which is validated with empirical data collected in a follow-up experimental data collection. Finally, we conclude and interpret the survey results and consequently derive theoretical and practical implications.

2. BACKGROUND THEORIES
In general, usability theory has its roots in cognitive psychology and is a relatively young branch of computer science. While some of the principles of usability theory are gradually making their way to the mainstream software applications the underlying research is less known [4]. However, our research model integrating usability determinants in the field of business modelling is based on two centre theories adopted by usability research. First we underlie cognitive theory, which generally defines the external impact of human learning and acting. The theoretical constructs of cognitive psychology have direct analogies in model interpretation scenarios. From the traditional cognitive point of view, the usability system in our study is composed of three basic information generating and processing units, (1) the human being such as model interpreter, (2) the model, which contains the information interpreted and (3) particular language the graphical model is based on [5].
Secondly we underlie a development of cognitive theory called cognitive load theory [6]. This theory is focusing on the impact of memory load to human learning and knowledge acquisition. Figl et al. (2010) mapped cognitive theory to the context of modelling languages [7]. Cognitive theory differs between three types of cognitive load. The extraneous cognitive load is influenced by the way the information is represented. The intrinsic cognitive load is determined by information complexity. Finally germane cognitive load is strongly connected with learning processes and especially the load expended for learning [8]. As a result, the cognitive load referring to learning and understanding should be expanded. Extraneous cognitive load should be held low by minimizing irrelevant information. Transferring this to our approach, we conclude that language specific properties categorized in three loads are influencing the usability in model interpretation.
The variety of definitions and measurement models of usability

complicates the extraction of capable attributes for assessing the usability of modelling languages. A usability study would be of limited value if it would not be based on a standard definition and operationalization of usability [9]. The International Organization for Standardization (ISO) defines usability as the capacity of the software product to be understood, learned and attractive to the user, when it is used under specified conditions [10]. Additionally, the ISO defined another standard which describes usability as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use [11]. The Institute of Electrical and Electronics Engineers (IEEE) established a standard, which describes usability as the ease a user can learn how to operate, prepare inputs for, understand and interpret the outputs of a system or component [12]. Dumas and Redish (1999) define that usability means quickness and simplicity regarding a user's task accomplishment. This definition is based on four assumptions [13]: 1. Usability means focusing on users, 2. Usability includes productivity, 3. Usability means ease of use, 4. Usability means efficient task accomplishment. Shackel (1991) associates five attributes for defining usability: speed, time to learn, retention, errors and the user specific attitude [14]. Preece et al. (1994) combined effectiveness and efficiency to throughput [15]. Constantine and Lockwood (1999) and Nielsen (2006) collected the attributes defining usability and developed an overall definition of usability attributes consisting of learnability, memorability, effectiveness, efficiency and user satisfaction [16], [17]. The variety of definitions concerning usability attributes led to the use of different terms and labels for the same usability characteristics, or different terms for similar characteristics, without full consistency across these standards; in general, the situation in the literature is similar. For example, learnability is defined in ISO 9241-11 as a simple attribute, "time of learning", whereas ISO 9126 defines it as including several attributes such as "comprehensible input and output, instructions readiness, messages readiness […]" [18], [19], [11]. As a basis for our following up research we are underlying usability definition for modelling languages in model interpretation scenario including attributes as follows: The usability of modelling languages is specified by learnability, memorability, effectiveness, efficiency, user satisfaction and perceptibility. The learnability of modelling languages describes the capability of a modelling language to enable the user to learn interpreting models based on particular language. The modelling language and its semantics, syntax and elements should be easy to remember, so that a user is able to return to the language after some period of non-use without having to learn the language and especially the interpretation of models developed with specific language again. Effective model interpretation should be supported by particular language for reaching a successful task accomplishment. Modelling languages should be efficient to use, so that a high level of working productivity is possible. Users have to be satisfied when using the language. The language should offer a convenient perceptibility regarding structure, overview, elements and shapes so that an interpreter is able to search, extract and process available model information in an easy way [2, 20].

3. THEORY DEVELOPMENT

The usability concept in our research is specified by learnability, memorability, effectiveness, efficiency, user satisfaction and perceptibility. We state that these attributes and especially their causal interaction influence usability of model interpetation based on different modelling languages.

3.1 Structural Model

Usability literature and transferred theories only set the different attributes on one causal level. For example, Nielsen (2006) and Abran et al. (2003) state that usability is affected by attributes with same weightings [17, 18]. We argue that the usability of modelling languages is defined by chosen attributes on different stages. Furthermore we state a causal connection between usability attributes, which is examined in our empirical research. Adopting the background theories we propose our research model in figure 1. The research model includes two basic parts, the metamodel properties and the attributes defining usability. Metamodel properties are set in language's metamodel. They are language specific attributes, which affect the usability attributes on different stages.

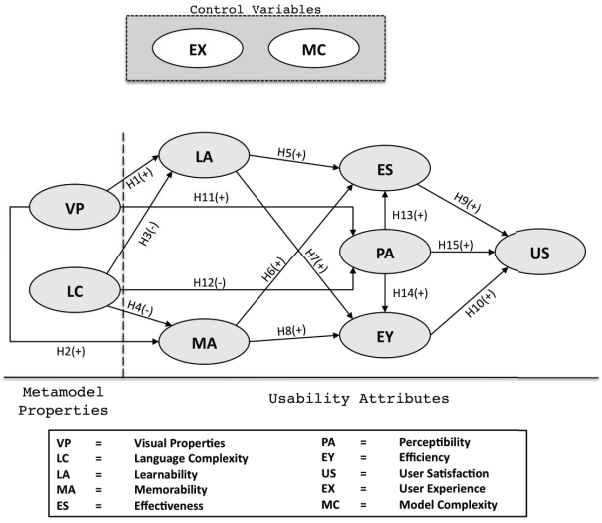


Figure 1 Research Model

HYPOTHESIS 1. *The range of different element colours and geometrics set in the language's metamodel (VP) are positively influencing user's ability to learn the application of the modelling language (LA)*

With considering perceptive factors affecting modelling languages' usability visual based metrics such as the number of different element shapes and the number of different element colours were defined [21]. Hall and Hanna (2004) analyzed the impact of colour on web usability attributes in an empirical survey. They concluded that the application of different colours results in a higher grade of website structuredness, which leads to more efficient information processing in the user's brain [22]. Transferring that, we can assume that more element colours set in the language's metamodel are leading to more information structuredness, which is influencing the learnability of modelling languages connected with model interpretation in a positive way. Furthermore we assume that the variance of different geometric shapes depicting different element types is positively influencing language learnability. The theoretic basis for this assumption is initially given by Comber et al. (1997). They concluded that screen complexity including the application of various geometric

shapes is a positive influencing variable of usability and especially learnability. However, they additionally underlay a positive trade-off between screen complexity and learnability [23].

HYPOTHESIS 2. *The range of different element colours and geometrics set in the language's metamodel (VP) are positively influencing user's ability to remember the elements and syntax of the modelling language (MA)*

Hall and Hanna (2004) analyzed a strong impact of visual properties on website structuredness [22]. Furthermore Nembhard and Napassavong (2002) found a positive correlation between structured information and information storage in human's brain [24]. Deducing this to our model we state that visual variability of modelling languages is positively influencing the user's ability to remember elements and syntax of modelling languages.

HYPOTHESIS 3. *The complexity of a modelling language (LC) affects negatively the proband's ability to learn this language (LA)*

Referring to Rossi and Brinkkemper (1996) elements, relations and properties can be abstracted and defined as modelling language complexity. The language complexity influences the usability attributes [25]. For analyzing the language's complexity Welke (1992) and additionally Rossi and Brinkkemper (1996) developed metrics based on the OPRR data model [26], [27]. Transferring this to our approach metrics such as the number of object types (i.e. class), number of relationship types (i.e. association) and the number of property types (i.e. class name) are relevant for analyzing the complexity of a modelling language. The more elements, relations and properties a modelling language consists of, the more difficult a user can learn the application due to high semantically and syntactical power. We suppose, for example, that a high range of BPMN-elements is negatively influencing the user's ability of learning the interpretation of BPMN-models.

HYPOTHESIS 4. *Language complexity (LC) affects negatively the user's ability to remember elements, relations and syntax within a period of non use/training (MA)*

According to Kintsch 1998 cognitive processes underlie comprehension of a specific domain [28]. Nembhard and Napassavong (2002) found out that the complexity of a special domain influences memorability negatively [24]. According to our approach we assume that metamodel complexity is negatively related to memorability of modelling languages. A high semantically and syntactical complexity of language's metamodel is complicating model interpretation due to hindered ability of remembering elements, relations and their specific way of interpreting them.

HYPOTHESIS 5. *The gradient of a language's learning curve (LA) is positively related to the ability of completing a task with minimal errors and maximal completeness (ES)*

The ability of learning a modelling language in an easy or difficult way influences language's effectiveness in model interpretation when the language is applied. On the one hand we imply that low learnability values of a modelling language result in rising error rates and decreasing task completion rates. On the other hand we assume that an easy to learn modelling language support task completion rates and lowers error rates. In cognitive psychology low gradients of learning curves causes ineffective application of a construct in a specific domain [29]. Therefore our underlying assumption is that modelling languages, which are difficult to learn, are offering a limited user individual application. This fact influences task completion rates and task error rates, which are manifest variables for measuring the latent construct effectiveness.

HYPOTHESIS 6. *The user's ability to remember the range of elements, relations and syntactic regulations (MA) is positively related to the user's ability of performing tasks with minimal errors and maximal completeness (ES)*

Memorability describes the "remembering rate" of a modelling language. Overall it describes the fact that a modelling language should be easy to remember regarding its elements, syntax and semantics [30], [31]. Memorability is a very important attribute for measuring the usability of modelling languages considering that users may not be using a modelling language all the time [17]. Hence, we hypothesize that an easy to remember modelling language results in less errors and higher completion rates regarding model interpretation tasks.

HYPOTHESIS 7. *The gradient of a language's learning curve (LA) is positively related to the efficiency (EY) that is offered by modelling languages during applying them.*

Learnability is probably the most important attribute of usability, preferably a modelling language is easy to learn. Learning to use a modelling language in interpretation scenario seems to be the first experience most users are confronted with when using a new modelling language [25], [30]. Easy to learn languages offer a higher user-individual learning growth and consequently higher curve gradients based on task completion time values than difficult to learn modelling languages [32]. We state that this effect is supporting efficiency in interpreting models.

HYPOTHESIS 8. *The user's ability to remember the range of elements, relations and syntactic regulations (MA) is positively related to efficient task accomplishment (EY) offered by the modelling language*

Usability research shows that memorability is an initial basis for applying a system or a website [17]. Transferring this we state that some modelling languages are easier to remember than other. For example, it seems that BPMN elements are not easy to remember because of its high range of different element types. From this fact can be deduced that an efficient use and consequently a fast task completion is influenced by the memorability of the different metamodel properties a language consists of.

HYPOTHESIS 9. *The ability to perform a task with minimal errors and maximal completeness (ES) is positively related to user's individual satisfaction (US) with a modelling language*

Effectiveness characterises the fact, that it should be possible to reach a successful task accomplishment. In this regard, a user should be able to develop and comprehend models with low error rates and high task completion rates [33], [34]. Regarding the usability of modelling languages we imply that languages offering high effectiveness result in higher satisfaction values. In contrast we state that languages offering low effectiveness values are affecting user's individual satisfaction negatively.

HYPOTHESIS 10. *The Efficiency of task completion (EY) is positively related to user's individual satisfaction (US) of modelling languages*

A modelling language is efficient to use when the users are able to develop or comprehend a model relatively quickly and correctly regarding the regulations of the modelling language. Once a user has learned a modelling language it should be possible to reach a high level productivity regarding task completion time [35], [34]. Hence we hypothesize, that languages which afford an efficient interpretation completion result in higher values concerning user satisfaction.

HYPOTHESIS 11. *The variance of visual language properties (VP) set in the metamodel of the modelling language is positively influencing language's perceptibility (PA)*

Many researchers analyzed the influence of visual differentiation caused by varying geometric shapes and colours in usability and primarily neurophysical research. For example, Westphal and Würtz (2009) investigated that visual differentiation is supporting object recognition and consequently information search and information extraction [36]. However, in our research model language's perceptibility is measured by values indicating cognitive processes e.g. information search and information extraction [37]. Furthermore, Underwood (2009) corroborates the hypothesis that visual characteristics of an image are influencing eye movements [38]. From this we can deduce, that visual language properties, i.e. colours, geometric shapes, are positively influencing language's perceptibility due to stronger visual differentiation in model diagrams.

HYPOTHESIS 12. *The complexity of modelling languages (LC) is negatively influencing visual perceptibility (PA)*

The complexity of modelling languages, which is set in the language's metamodel, is strongly connected with syntactical and semantical complexity. For example, UML-class-diagrams contain a high range of syntactically different relations (e.g. association, aggregation etc.), which can be expanded by cardinalities. Furthermore, a class diagram generally includes two different class types: standard and abstract classes. Pan et al. (2004) analyze the viewing behaviour of web pages by using an eye-tracker [39]. They come to the conclusion that visual complexity negatively contributes to eye-movement behaviour due to difficulty of information search and information extraction. In our research model we state, that syntactic and semantic language properties are negatively influencing the perceptibility of a diagram developed by the application of specific modelling languages.

HYPOTHESIS 13. *The visual perceptibility (PA) of modelling languages is positively contributing to effective model interpretation (ES)*

With analyzing visual perceptibility we aim to measure processes of information search, information extraction and information processing in user's brain during model interpretation. For example, a low visual perceptibility of a model results in difficult information search and information extraction. Consequently we deduce that this fact is especially influencing task completion rate and subsequently effectiveness of model interpretation. Finally we hypothesize that visual perceptibility is influencing user's ability of ending an interpretation task with minimal errors and maximal completeness.

HYPOTHESIS 14. *The visual perceptibility of modelling languages (PA) is positively contributing to efficient model interpretation (EY)*

Goldberg and Kotval (1999) concluded that the number of overall fixations is negatively correlating with search efficiency. We state that this effect is influencing interpretation time and consequently interpretation efficiency [40]. Furthermore, high fixation durations implicate participant's difficulty of extracting information from a model [41]. Accordingly, this effect leads to increasing interpretation times and consequently lower efficiency.

HYPOTHESIS 15. *The visual perceptibility (PA) of models developed by the application of modelling languages affects positively the user's satisfaction (US) of specific modelling languages*

Many researchers concluded a strong impact of design (screen, website etc.) and especially layout and order of elements on target individual's satisfaction [42], [43]. Lindgaard 2007 states a positive link between user satisfaction and visual screen design [44]. Subsequently, in our research model we assume that a high language's visual perceptibility results in higher user satisfaction.

Furthermore we include additional variables as controls recognizing their effects on key constructs in our research model. The users of modelling languages differ regarding modelling experience. This fact influences the task accomplishment and consequently the usability and has to be considered in our research model [17]. Hence, the user and his/her individual modelling experience must be treated as control variable. The level of difficulty and complexity of a particular model affects understandability and consequently the usability of the applied modelling language [45]. When conducting a survey on usability evaluation of modelling languages, the complexity of a particular model applied in an experiment i.e. task complexity must be controlled for minimizing its influence on the outcome.

3.2 Measurement Model

In this section we theoretically underlie chosen manifest variables working as indicators for latent constructs in our research model. Evaluating **effectiveness** requires analysis of task output with measuring quantity and quality of goal achievement [46]. Quantity is defined as the proportion of task goals represented in the output of a task. Quality is the degree to which the task goals represented in the output have been achieved [47]. Bevan (1995) defined effectiveness as a product of quantity and quality [48]. Transferring this to our model, indicating manifest variables for measuring effectiveness are the grade of completeness and the grade of correctness of a model interpretation task.

The **efficiency** is the amount of human, economical and temporal resources. Measures of efficiency relate to the level of effectiveness achieved to the expenditure of resources [47]. Measure values of efficiency include time taken to complete tasks, i.e. duration time for performing a model interpretation task [49].

Learnability describes the ease of learning the application (i.e. interpretation) of modelling languages. For this characteristic, the standard measure values are based on task completion rates and the task accuracy [50]. In general, learnability is a development and can be graphically described by learning curves [32]. Hence, learnability can be measured by the rate of difference when the user repeats evaluation sessions [48]. Nielsen 2006 insists that highly learnable systems could be categorized as "allowing users to reach a reasonable level of usage proficiency (…)"[17].

Furthermore, Nielsen (2006) proposes measuring proficiency by quantity and quality and of task fulfillment [17]. Thus, we chose grade of completeness and grade of correctness as basic variables for measuring learnability. With conducting two measuring points mp and mp+1, it is possible to analyze the relative difference between mp and mp+1 for indicating Δ learnability, i.e. individual learning progress in percent [24], [51].

The **visual perceptibility** is measured by using the method of eye-tracking with analyzing the user's visual attention [52]. In our research we aim to include eye-tracking for measuring user's cognitive processes i.e. information search and information extraction during model interpretation process. The pioneering work regarding the use of eye-tracking was first carried out by Fitts et al. (1950) [41]. They proposed that fixation length is a measure of difficulty of information extraction and interpretation. Fixations are eye movements that stabilize the gaze over an object of interest. During this, the brain starts to process the visual information received from the eyes [53]. The number of fixations overall is thought to be negatively correlated with search efficiency [40]. Consequently, a larger number of fixations indicates less efficient search in a model. Concerning an eye-tracking experiment for evaluating the visual perceptibility of modelling languages a large number of fixations implies an intensive search to explore the model's diagram structure. This fact complicates the interpretation of a model. Furthermore, we aim to analyze the difficulty of information extraction in a model. Byrne et al. 1999 [54] propose tracking fixation duration time as a measure for information extraction. From this follows that longer fixations times during an interpretation process are indicating a participant's difficulty extracting information from a model. Compared to the other latent variables in our research model, the **individual satisfaction of a user** while interpreting a model is a user subjective criterion that can be measured best by using standardized questionnaires [49]. Currently no standardized method for measuring user satisfaction in the modelling domain exist. Therefore, we mapped questionnaires focusing on system and website usability [55], [56]. For evaluating user satisfaction we developed a questionnaire, which consists of thirty items structured in 1) General impression, 2) Recommendation rate and 3) Language application. We measured the constructs with 5-point Likert-scales. The development of this questionnaire is generally contributing to the Questionnaire for User Interaction Satisfaction (QUIS) and additionally the Software Usability Measurement Inventory (SUMI) [57] [55].

Memorability is best measured as proficiency after a period of non-use provided a user has already learned a language [58]. The non-use period can be minutes for simple element meanings, hours for simple syntactic regulations and days or weeks for measuring a complete modelling language [50]. Accordingly, the measure values for memorability are neglect curves and time-delayed knowledge tests [59]. Concerning the usability of modelling languages, the user must remember the different elements and its intended meaning (semantics), the syntax and the application. In due consideration of Nielsen 2006, the measuring points interval should be several weeks regarding memorability [17]. Thus, for measuring memorability we decided to use a knowledge test consisting of items focusing on 1) elements and relations, 2) syntax and 3) application of particular language. For measuring exogenous variable language complexity we track number of elements, number of relations and number of properties (LC) under consideration of Rossi and Brinkkemper's (1996) OPRR-model and particular expansions by Recker et al. (2009) and Indulska et al. (2009) [27, 60, 61]. Furthemore, for indicating visual properties we are analyzing different colours and different geometric shapes set in language's metamodel.

For measuring model experience we track participant's **individual experience** in 1) general modelling experience and 2) language experience on a 5-point Likert-scale. Finally, we operationalized **model complexity** by three indicator-variables: number of elements and relations (size), connectivity degree and semantic spread. With running causal analysis we include controls as moderator effects.

4. RESEARCH METHODOLOGY AND DATA COLLECTION

This study uses a various data collection methods for measuring manifest variables of latent usability attributes. Furthermore, we introduced an experimental design consisting of two data collection sessions per modelling language. The experiment focused on model interpretation tasks. Within these experiments we collected error rates, grade of completeness and task finishing time values for measuring efficiency, effectiveness and learnability, which is the relative learning growth between two data collection sessions. Additionally, we introduced the method of eye-tracking for analyzing visual perceptibility of modelling languages. The instruments were either adapted from traditional usability research or we developed new measuring instruments on modelling languages. A pretest was conducted prior collecting data for the field test. The research instruments were tested for reliability, content validity and construct validity. Necessary changes were made to improve measuring instruments. All pilot test participants were excluded from the analysis sample.

4.1 Measurement Scales

Multiple indicators measured all but one construct. The exception was EY, which represents a discrete value and therefore can be appropriately measured with a single item focusing on task completion time. We conceptualized and measured Language Complexity, Memorability, Learnability and Effectiveness as aggregations of different manifestations; thus the direction of causality is from indicator to construct (i.e. formative). The other constructs were operationalized as reflective indicators.

4.2 Data Collection

The sample includes third year students of business informatics. The experimental data collection, the questionnaire and the knowledge test were conducted with these students. The overall sample size amounts 57 students, 47% female and 53% male. The data collection was based on two different modelling concepts and connected languages. On the one hand process based languages, Event driven Process Chains (EPC), UML Activity Diagrams and on the other hand structure based modelling languages, UML Use Case and UML Class Diagrams were included in our survey. For developing variables measuring the latent construct learnability we introduce a second measuring point. In the session the students are confronted with one experimental task: the interpretation of given models. The interpretation scenario is structured in two parts. The first part is focusing on general observation while the second part includes verbal interpretation of given model. However, the interpretation task generates time, error, completeness and additionally eye-tracking values for measuring ES, EY, LA and PA. At the beginning of second collection phase we distributed the knowledge tests for measuring the ability of

remembering specific metalevel properties *(MA).* Subsequently, the user satisfaction *(US)* questionnaire was administered to the participants.

5. DATA ANALYSIS AND RESULTS

To test the proposed research model, data analyses for both the measurement model and the structural model were performed using partial leased squares (PLS), bootstrapping and the blindfolding method [62]. For calculating we took SmartPLS version 2.0 M3. Chin et al. (2003) defined various strengths of the PLS-approach. Partial Least Squares (PLS) gives reliable results and should being preferred to competing LISREL approach if 1) phenomena explored are new without existing construct and measuring theories, 2) structural model includes a large number of indicating variables, 3) relative small sample size and 4) detection of causal paths and predictions is focused on [63]. PLS is a powerful method of analysis because of the minimal demands on measurement scales, sample size, and residual distributions [64]. Although PLS can be used for theory confirmation, it can also be used to suggest where relationships might or might not exist and to suggest propositions for later testing [65].

5.1 Validity and Reliability

We conducted an exploratory factor analysis in SPSS for each construct of our models including all defined items using a Promax rotation. In all cases the Bartlett-test of sphericity indicating independency of construct items among was accepted. Consequently we analyzed different factors and assigned variables to specific factors considering Kaiser's criterion [66]. Indicating acceptable validity items with loadings smaller than 0.5 were excluded from our model. By doing so we assure that our models include construct items, which are loading sufficiently on specific factors.

5.2 Testing the Measuring Model

Internal consistency reliability was evaluated using Cronbach's Alpha, corrected item total correlation and average variance extracted (AVE) [67]. Cronbach's Alpha coefficients were all but one higher than the proposed minimum cutoff score of 0.70 [68]. The alpha value for experience is 0.68. Barker et al. (1994) conclude that values between 0.60 and 0.70 are marginal and can be accepted as well [69]. Values for composite reliability are all higher than desired threshold of 0.60 [70]. Furthermore all reflective constructs had an minimum AVE (Average Extracted Variance) of 0.5, indicating adequate internal consistency of our model [67]. For testing reliability of formative constructs we analyzed R2-value proposed by Chin (1998) with a minimum cutoff of 0.19 [65]. Furthermore, Diamantopoulos and Winklhofer (2001) concluded that sufficient significant regression weights between formative constructs and other constructs in the path model are indicating formal construct validity [71]. As shown in the following section all relevant path regression weights are at least significant at 0.05-level. According to Fornell and Larcker (1981), constructs have adequate discriminant validity if the square root of AVE is higher than variance shared between construct and other constructs in the model [67]. In all cases the correlations between each pair of constructs were lower than the square root of the AVE for specific construct. In conclusion, these results as well as the factor analysis confirm that all constructs in our model are empirically distinct. Table 1 shows detailed values for each construct of our research model.

Table 1 Reliability and Validity of our Research Model

	Measuring Model quality metrics			**Structural Model quality metrics**		
	Type	Alpha	Composite Reliability	AVE	R^2	Q^2
Threshold		≥ 0.6	≥ 0.6	≥ 0.5	≥ 0.19	≥ 0.0
Visual Properties (VP)	R	0.96	0.98	0.97	NA*	0.78
Language Complexity (LC)	F	NA	NA	NA	NA*	0.58
Memorability (MA)	F	NA	NA	NA	0.47	0.24
Learnability (LA)	F	NA	NA	NA	0.20	0.10
Efficiency (EY)	R	0.72	0.75	0.60	0.19	0.08
Effectiveness (ES)	F	NA	NA	NA	0.42	0.16
User Satisfaction (US)	R	0.89	0.90	0.68	0.19	0.07
Perceptibility (PA)	R	0.78	0.88	0.88	0.20	0.09
Task Complexity (TC)	R	0.70	0.83	0.63	NA*	0.31
Experience (EX)	R	0.68	0.62	0.52	NA*	0.66

Notes. R: reflective, F: formative; n=114 for all constructs; NA: not applicable: because formative measures need not covary, the internal consistency of formative items is not applicable [65]. NA*: not applicable: because R2 value is only relevant for assessing endogenous latent variables in the inner structural model [65].

5.3 Testing the structural model

Figure 2 presents the results of structural model testing including regression weights and significance of the paths. According to Lohmöller (1989) path regression weights should be at least 0.10 in order to be considered meaningful for discussion [72]. Our results confirmed the general assumption that language's metamodel properties are influencing usability attributes on different stages. According to Chin (1998) and for ensuring complete model assessment we additionally show effect size f^2, which is indicating whether a path's latent exogenous variable has a significant influence (effect) on latent endogenous variable or not. Thresholds for f^2 are 0.02 (weak), 0.15 (medium) and 0.35 (strong) [65].

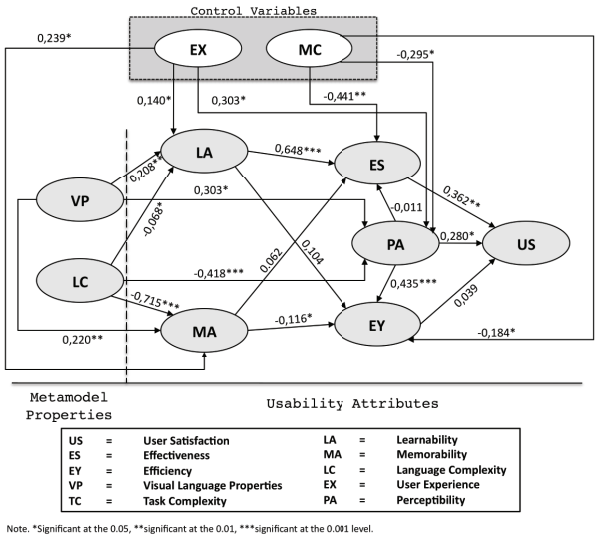


Figure 2 Structural Model Results

LC has a strong negative and highly significant influence on MA (beta=-0.715, f^2=0,80, p<0,001). This empirical result supports our hypothesis H4. LC has also a strong significant negative impact on PA underlining H12 (beta=-0.418, f^2=0.16, p<0.001). Furthermore, LC has a negative significant relation to LA contributing to H3 (beta=-0.068, f^2=0.02, p<0.05). However, this path disposes not to Lohmöller's (1989) proposed threshold for path weighting of 0.1. VP are positively influencing LA of applying modelling languages concerning to model interpretation (beta=0.208, f^2=0.02, p<0,01). In addition to that VP is positively influencing PA (beta=0.303, f^2=0.03, p<0.05). Considering this, all hypotheses in our research model connected with VP are accepted. Additionally, LA is strongly positively related to ES on a high significance level (beta=0.648, f^2=0.72, p<0.001), which is contributing to H5. Furthermore, LA is positively affecting time based latent construct EY, also, MA is positively correlating with ES. These path regression weights are not significant (p>0.05). Deducing from that, we cannot reject null hypothesis with probability level of 0.05. Consequently, we assume that these paths are not empirically explaining our research model. H6 and H7 are not empirically supported. MA has a weak negative impact on EY (beta=-0.116, f^2=0.01, p<0.05). This relation is not contributing to H8. As a consequence we state, that in modelling domain MA is negatively influencing the time used for model interpretation. PA is positively influencing EY (beta=0.435, f^2=0.24, p<0.001) and US (beta=0.280, f^2=0.075, p<0.05). Users ability of complete and correct model interpretation is positively influencing US (beta=0.362, f^2=0.11) p<0.01). From this, we can deduce that H9 is accepted. Turning to model fit, the R-square values for MA, LA, EY, ES, US and PA were 0.473, 0.202, 0.194, 0.420, 0.192 and 0.196 respectively, indicating that the model explains substantial variation in these variables. For example, the R-square value for MA implies that the causes specified in this model, VP and LC, jointly explain 47% of the total variance in MA.

In summary, the results show that most hypotheses in our research model are fully supported. However, H8 is not supported by our results. Furthermore, H6, H7 and H13 could not be confirmed by significant results. As a consequence, particular hypotheses are not confirmed for further comparable samples. The resulting regression weights of H6, H7 and H13 are valid for our specific sample and should be proved in further surveys based on our research model.

6. DISCUSSION

This study provides several important findings supporting the understanding of usability attributes. We focus on the model interpretation scenario. The two major influencing areas are (1) complexity of a language and (2) causal impact of visual properties.

Firstly, our results show that the complexity of language's metamodel, i.e. variability in elements, relations and properties, is strongly influencing user's ability to remember them. Usability research shows that memorability is an initial basis for applying a system or a website effectively [17]. However, with our results we cannot confirm those theses for model interpretation scenario. Additionally, memorability is weakly influencing effectiveness of model interpretation. Furthermore, we find that memorability is weakly influencing effectiveness and that memorability weakly impacts interpretation time negatively. Our research findings for the causal path between memorability and efficiency of model interpretation are inconclusive. Concerning this, further research into this area will be required and may lead to more conclusive findings. However, it seems that memorability plays a secondary role in model interpretation scenario.

Metamodel complexity is strongly influencing language's visual perceptibility. This result provides evidence that languages based on complex metamodels are not supporting user's ability of easy information search and extraction when interpreting a model. Additionally the visual perceptibility of modelling languages is strongly connected with duration time of information search and extraction. Concerning this, we deduce that languages offering a good perceptibility afford fast information search and information extraction times leading to an efficient model interpretation process. Considering model complexity as control variable, a process model developed with BPMN including a high range of different elements offers lower visual perceptibility and accordingly results in higher time values for information search and extraction compared to an EPC-model. Moreover, the visual perceptibility of a modelling language is positively supporting user's individual language satisfaction. From this result we infer, that visual perceptibility is one important base of user satisfaction. User acceptance is strongly connected with user satisfaction [17]. This relationship underlines the fact that visual perceptibility concerning particular languages is obviously a basic result of user satisfaction and consequently user acceptance. In other words, visual perceptibility may decide whether a modelling language is accepted or not by users concerning model interpretation.

Obviously, the positive impact of interpretation time on user satisfaction is not as much as expected. This might be underlining former findings of Walker (1998). In their studies they found out that users have demonstrated preferences for systems with which they performed less efficiently [73]. It shows that the ability for finishing interpretation tasks completely and correctly and the ability for convenient information search and information extraction out of a model are more important to satisfy users than the commonly assumed performance factors of efficiency.

Secondly, an important result of our survey is the causal impact of visual language properties, i.e. variability in shape geometrics and shape colours, in the field of model interpretation. The output of our study shows that visual language properties are positively

influencing the visual perceptibility of modelling languages. This result underlines the finding that visual differentiation supports object information search and information extraction [36]. As a consequence, the application of different colours and geometrics in a model supports interpreting users in searching and extracting information. Furthermore, the variability in shape colour and geometrics is positively influencing learnability of model interpretation and memorability of language's elements and relations. Consequently, languages offering higher variability in geometrics and colours are easier to learn concerning model interpretation.

The learnability of interpreting a model based on a certain language is strongly impacting the ability of performing an interpretation task completely and correctly. For example, in industry and education it is important that users can interpret developed models with a high level of completeness and correctness [74]. With our study we found out that learnability, which is positively influenced by visual language properties acts as a basic independent variable strongly impacting on user's ability of complete and correct model interpretation. Furthermore, learnability is positively influencing efficiency of model interpretation. We conclude that learnability is a basic construct in model interpretation scenario. A theoretical basis might be cognitive load theory and especially intrinsic cognitive load [8]. The intrinsic cognititve load is determined by information complexity. The interdependency of information to be learned is positively impacting cognititve load and consequently the more important learnability appears in a causal system. Concerning modelling languages and model interpretation, the cognitive load is high because of strong information interdependency occurring in models. Considering our results and cognitive load theory the importance of learnability in model interpretation is emphasized. In due consideration of our results it consequently becomes clear that learnability is positively impacted by visual language properties. From this follows that languages offering high visual variability are easier to learn than other. As a consequence languages containing high visual variability allow higher task completion and accuracy rates in model interpretation. In conclusion, if a language should support effectiveness of model interpretation, the metamodel should offer high visual variability in elements and relations.

7. CONCLUSION

In this paper we propose a study of usability assessment of modelling languages using a structural equation modelling approach. The study focuses on model interpretation scenario. Our causal path shows that in model interpretation memorability of language's elements and relations plays a secondary role. It becomes clear that visual perceptibility and effectiveness are fundamental attributes for reaching high values in user satisfaction. Furthermore, the model supports our idea that language's metamodel properties are influencing usability attributes on different causal stages. In the following, we derive concluding implications for both theoretical and practical needs.

7.1 Implications for Theory

First, our results confirm most of our hypotheses deduced from theory. In usability research, a theoretical embedment of usability attributes in a causal model is missing up to now. In our study we show interesting causal relations between usability attributes. Thus, there might be important results for usability research concerning the causal impact of different usability attributes. Further studies are required for testing our structural model in other usability domains (e.g. website usability etc.).

7.2 Implications for Practice

Our structural model delivers important results showing how modelling languages affect usability attributes on different causal stages in the model interpretation scenario. We structure practical implications in two parts focusing on 1) industry and 2) language specification/development organizations.

In companies the importance of business process and application system modelling has steadily risen. Consequently, the interpretation of models becomes an issue of organizational concerns. How efficiently can an employee extract information out of a model? Does he/she understand the information, i.e. does he/she interpret the model accurately? These might be basic questions connected with decision-making for or against the use of particular modelling languages in organizations. With considering our first results, the structural model can support the process of decision-making focusing on language usability in model interpretation. Thus, companies aiming for fast, complete and correct model interpretation, e.g. business process consulting companies, should apply modelling languages with high variability in visual properties.

Our second practical implication deduced from our results is focusing on modelling language specification and development organizations. For example, an important finding in our study is that visual variability of elements and relations is supporting accuracy, completeness and speed in model interpretation processes. In this regard, we conclude that UML activity diagrams (i.e. low visual variability) are not as usable as EPCs (i.e. high visual variability) in model interpretation. For optimizing UML activity diagram's usability in the model interpretation scenario it might be worth increasing visual variability in the meta-model by adding colours and various geometric shapes. Furthermore, to improve user satisfaction values in the model interpretation scenario it is necessary to decrease language complexity (e.g. by reducing number of different elements and relations) and increase visual variability. We are aware that complexity reduction possibly may impact the expense of explanatory power offered by particular language. The results in this paper provide a starting point for further empirical based discussions on usability of graphical modelling languages.

7.3 Limitations and Future Directions

The study is limited to the model interpretation scenario. Henceforth, we are expanding our study to model development cases. A comparison of results for interpretation and development scenarios may lead to a greater understanding for usability and particular attributes in the domain of modelling languages.

REFERENCES

[1] Ludewig, J. 2003. Models in software engineering - an introduction. *Software and Systems Modeling*, 2, (1) 5-14.

[2] Schalles, C., Rebstock, M. and Creagh, J. 2010. Ein generischer Ansatz zur Messung der Benutzerfreundlichkeit von Modellierungssprachen. In Modellierung 2010 (Klagenfurt, 2010), Gesellschaft für Informatik (GI), Klagenfurt, 15-30.

[3] Siau, K. and Wang, Y. 2007. Cognitive evaluation of information modeling methods. *Information and Software Technology*, 49, (5) 455-474.

[4] Ilomäki, T. 2008. The Usability of Music Theory Software: The Analysis of Twelve-Tone Music as a Case Study. Computer Music Modeling and Retrieval. Sense of Sounds: 4th International Symposium, CMMR 2007, Lecture Notes in Computer Science (LNCS). Springer-Verlag. 98-109.
[5] Zhang, P. and Li, N. 2004. An assessment of human–computer interaction research in management information systems: topics and methods. *Computers in Human Behavior*, 20, (2) 125-147.
[6] Plass, J., Moreno, R. and Brünken, R. 2010. *Cognititve Load Theory*. Cambridge University Press, Cambridge.
[7] Figl, K., Mendling, J., Strembeck, M. and Recker, J. C. 2010. On the cognititve effectiveness of routing symbols on process modeling languages. In Business Information Systems (BIS) (Berlin, 2010), Springer, Berlin, 18-28.
[8] Sweller, J. 1988. Cognititve load during problem solving: Effects on learning. *Cognititve Science: A Multidisciplinary Journal*, 12, (2) 257-285.
[9] Coursaris, C. and Kim, D. 2006. A Qualitative Review of Empirical Mobile Usability Studies. In *Proceedings of the Twelfth Americas Conference on Information Systems* (2006), City,
[10] International Organization for Standardization (ISO). 2006. Ergonomics of Human-System-Interaction; Part 110: Dialogue Principles.
[11] International Organization for Standardization (ISO). 1998. Ergonomic Requirements for Office Work with visual Display Terminals (VDTs); Part 11: Guidance on Usability.
[12] Institute of Electrical and Electronics Engineers (IEEE). 1990. Standard Glossary of Software Engineering Terminology. http://www.idi.ntnu.no/grupper/su/publ/ese/ieee-se-glossary-610.12-1990.pdf.
[13] Dumas, J. and Redish, J. 1999. *A practical guide to usability testing*. Greenwood Publishing Group, Westport.
[14] Shackel, B. 1991. Usability - Context, framework, definition, design and evaluation. B. Shackel and S. Richardson. Human Factors for Informatics Usability. University Press. Cambridge, 21-38.
[15] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. 1994. *Human Computer Interaction*. Addison-Wesley, Wokingham.
[16] Constantine, L. L. and Lockwood, L. A. D. 1999. *Software for Use: A practical Guide to the Models and Methods of Usage-Centered Design* Addison-Wesley, New York.
[17] Nielsen, J. 2006. *Usability engineering*. Kaufmann, Amsterdam.
[18] Abran, A., Khelifi, A., Suryn, W. and Seffah, A. 2003. Consolidating the ISO Usability Models. In *Proceedings of 11th International Software Quality Management Conference* (Glasgow, 2003). Springer, New York.
[19] International Organization for Standardization (ISO). 2004. Software Engineering - Product Quality; Parts 1-4.
[20] Schalles, C., Rebstock, M. and Creagh, J. 2010. Developing a Usability Evaluation Framework (FUEML) for Modeling Languages. In Proceedings of the IASTED International Conference on Software Engineering (SE) (Innsbruck, 2010), Acta Press, Innsbruck, 126-135.
[21] Elsuwe, H. and Schmedding, D. 2001. Metriken für UML-Modelle. *Informatik Forschung und Entwicklung*, 18, (1) 22-31.
[22] Hall, R. and Hanna, P. 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention *Behaviour and Information Technology*, 23, (3) 183-195.
[23] Comber, T. and Maltby, J. 1997. Layout complexity: does it measure usability? In *Human-Computer Interaction: Interact '97, International Conference on Human-computer Interaction* (Sydney, 1997). Springer, New York.
[24] Nembhard, D. and Napassavong, O. 2002. Task complexity effects on between-individual learning/forgetting variability. *International Journal of Industrial Ergonomics*, 29, (2) 297-306.
[25] Siau, K. and Rossi, M. 2008. Evaluation techniques for systems analysis and design modelling methods ; a review and comparative analysis. *Information Systems Journal*.
[26] Welke, R. 1992. The case repository: more than another database application. W. Cottermann and J. Senn. Challenges and strategies for research in systems development Wiley Inc. 181-218.
[27] Rossi, M. and Brinkkemper, S. 1996. Complexity Metrics for Systems Development Methods and Techniques. *Information Systems*, 21, (2) 209-227.
[28] Kintsch, W. 1998. *Comprehension: A Paradigm for Cognition* Cambridge University Press, Cambridge, Melbourne.
[29] Anderson, J. R. 1985. *Cognitive psychology and its implications*. Freeman, New York.
[30] Mayer, R. E. 1989. Models for Understanding. *Review of Educational Research*, 59, (1) 43-64.
[31] Recker, J. C. and Dreiling, A. 2007. Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education In *Australasian Conference on Information Systems* (Toowoomba, 2007). University of Southern Queensland.
[32] Tamir, D., Komogortsev, O. V. and Mueller, C. J. 2008. An effort and time based measure of usability. In *Proceedings of the 6th international workshop on Software quality* (Leipzig, Germany, 2008). ACM.
[33] Bobkowska, A. 2005. A framework for methodologies of visual modeling language evaluation. *ACM International Conference Proceeding Series*, 214, (2).
[34] Wand, Y. and Weber, R. 1993. On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, 3, (4) 217-237.
[35] Bobkowska, A. 2005. Modeling Pragmatics for Visual Modeling Language Evaluation. In *Proceedings of the 4th international workshop on Task models and diagrams* (Gdansk, 2005).
[36] Westphal, G. and R.G, W. 2009. Combining feature-and correspondence-based methods for visual object recognition. *Neural Computation*, 21, (7) 1952-1989.
[37] Underwood, G. D. M. 2005. *Cognititve Processes in Eye Guidance*. Oxford University Press, New York.
[38] Underwood, G. D. M. 2009. Cognitive Processes in Eye Guidance: Algorithms for Attention in Image Processing. *Cognititve Computation*, 1, (1) 64-76.
[39] Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K. and Newman, J. K. 2004. The determinants of web page viewing behavior: an eye-tracking study. *Proceedings of the 2004 symposium on Eye tracking research and application*, 147-154.

[40] Goldberg, J. and Kotval, X. 1999. Computer interface evaluation using eye movements: methods and constructs *International Journal of Industrial Ergonomics*, 24, (6) 631-645.
[41] Fitts, P. M., Jones, R. E. and Milton, J. L. 1950. Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9, (2) 24-29.
[42] Sonderegger, A. and Sauer, J. 2009. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, 41, (3) 403-410.
[43] De Angeli, A., Sutcliffe, A. and Hartmann, J. 2006. Interaction, usability and aesthetics: what influences users' preferences? *Proceedings of the 6th conference on Designing Interactive systems*, 271-280.
[44] Lindgaard, G. 2007. Aesthetics, Visual Appeal, Usability and User Satisfaction: What Do the User's Eyes Tell the User's Brain? . *Australian Journal of Emerging Technologies and Society*, 5, (1) 1-14.
[45] Melcher, J., Mendling, J., Reijers, H. A. and Seese, D. 2009. On Measuring the Understandability of Process Models. http://digbib.ubka.uni-karlsruhe.de/volltexte/1000011993.
[46] Rengger, R., Macleod, M., Bowden, R., Blaney, M. and Bevan, N. 1993. *MUSiC Performance Measurement Handbook*. National Physical Laboratory, Teddington, UK.
[47] Bevan, N. and Macleod, M. 1994. Usability Measurment in Context. *Behaviour and Information Technology*, 13, (1) 132-145.
[48] Bevan, N. 1995. Measuring usability as quality of use. *Software Quality Journal*, 4115-150.
[49] Vuolle, M., Aula, A., Kulju, M. and Vainio, T. 2008. Identifying Usability and Productivity Dimensions for Measuring the Success of Mobile Business Services. *Advances in Human-Computer Interaction*.
[50] Seffah, A., Donyaee, M., Kline, R. and Padda, H. 2006. Usability measurement and metrics: A consolidated model. *Software Quality Control* 14, (2) 159-178.
[51] Grossman, T., Fitzmaurice, G. and Attar, R. 2009. A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the 27th international conference on Human factors in computing systems* (Boston, MA, USA, 2009). ACM, City, 649-658.
[52] Gordon, I. E. 2004. *Theories of visual perception*. Psychology Press, Hove.
[53] Duchowski, A. T. 2007. *Eye Tracking Methodology - Theory and Practice*. Springer, New York.
[54] Byrne, M. D., Anderson, J. R., Douglass, S. and Matessa, M. 1999. Eye Tracking the Visual Search of Click-Down Menus. In Proceedings of CHI'99 (1999), 402-409.
[55] Kirakowski, J. and Corbett, M. 1993. SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24, (1) 210-212.
[56] Armstrong, B., Fogarty, G., Dingsdag, D. and Dimbley, J. 2005. Validation of a user satisfaction questionnaire to measure IS success in Small Business. *Journal of Research and Practice in Information Technology*, 37, (1) 27-48.
[57] Chin, J., Diehl, V. and Norman, K. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 213-218.
[58] Olle, T. W., Sol, H. G. and Verijin-Stuart, A. A. 1986. A comparative evaluation of system development methods. In *Proc. of the IFIP WG 8.1 working conference on Information systems design methodologies: improving the practice* (Noordwijkerhout, Netherlands, 1986). North-Holland Publishing Co.,Amsterdam, 19-54.
[59] Nembhard, D. and Uzumeri, M. 2000. Experimental learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics*, 25, (2) 315-326.
[60] Recker, J. C., Zur Muehlen, M., Keng, S., Erickson, J. and Indulska, M. 2009. Measuring Method Complexity: UML versus BPMN. *Proceedings of the Fifteenth Americas Conference on Information Systems, San Francisco, California*
[61] Indulska, M., Zur Muehlen, M. and Recker, J. C. 2009. Measuring Method Complexity: The Case of the Business Process Modeling Notation http://is.tm.tue.nl/staff/wvdaalst/BPMcenter/reports/2009/BPM-09-03.pdf.
[62] Tenenhaus, M., Vinzi, V. E., Chatelin, Y. and Lauro, C. 2005. PLS path modeling. *Computational Statistics and Data Analysis*, 48, (1) 159-205.
[63] Chin, W. W. and Newsted, P. R. 1999. Structural Equation Modeling Analysis with small Samples using PLS. R. Hoyle. Statistical Methods for small sample research. Sage Publications. Thousand Oaks,
[64] Chin, W. W., Marcolin, B. L. and Newsted, P. R. 2003. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic–Mail Emotion/Adoption Study *Information Systems Research*, 14, (2) 189-217.
[65] Chin, W. W. 1998. Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, 22, (1) 7-16.
[66] Kaiser, H. F. 1974. An index of factorial simplicity. *Psychometrika*, 39, (1) 31-36.
[67] Fornell, C. and Larcker, D. F. 1981. Evaluating Structural Equation Models with unobservable variables and measurment error. *Journal of Marketing Research*, 18, (2) 39-50.
[68] Nunnally, J. C. and Bernstein, I. H. 1994. *Psychometric Theory*. Mc Graw-Hill, New York.
[69] Barker, C., Pistrang, N. and Elliott, R. 1994. *Reserach methods in clinical and counseling psychology*. John Wiley, Chichester.
[70] Bagozzi, R. P. and Yi, Y. 1988. On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, (1) 74-94.
[71] Diamantopoulos, A. and Winklhofer, H. M. 2001. Index Construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38, (1) 269-277.
[72] Lohmöller, L. B. 1989. *Latent variable path modeling with Partial Least Squares*. Springer, Heidelberg.
[73] Walker, M. A. 1998. What can I say? Evaluating a spoken Interface to E-mail. In Conference on Human Factors in Computing Systems (CHI) (Los Angeles, 1998), Los Angeles, 35-45.
[74] Mendling, J. and Strembeck, M. 2008. Influence Factors of Understanding Business Process Models. *Proceedings of the 11th International Conference on Business Information Systems* 7, (1) 142-153.

Similarity Determination in Activity Sequences – A Supportive Framework

Jörg Schmidl
Technische Universität München
Boltzmannstraße 3
85748 Garching, Germany
+49 89 322 99 83 10
schmidl@in.tum.de

Holger Wittges
Technische Universität München
Boltzmannstraße 3
85748 Garching, Germany
+49 89 322 99 83 13
wittges@in.tum.de

Helmut Krcmar
Technische Universität München
Boltzmannstraße 3
85748 Garching, Germany
+49 89 289 19 530
krcmar@in.tum.de

ABSTRACT

An increasing number of information systems support their users by helping them in reusing existing knowledge and experience. Often this is done by retrieving similar instances like similar documents, similar process executions or similar persons. While the recommendations use similarity as central concept, the selection of a suitable measure is often done by intuition. This paper introduces a framework that supports the application engineer in selecting and configuring a suitable similarity measure. The requirements of the intended framework are gathered before the architectural implications are detailed. The resulting framework is applied in a case study in which project performance prediction is to be supported by the similarity of the projects' activity sequences. The results show the framework's utility by allowing a comparably simple configuration to yield a considerable support in selecting and configuring a suitable similarity measure.

Keywords

Similarity Measure, Activity Sequence, Similarity Framework, Business Processes

1. INTRODUCTION

Today, support by Information Systems (IS) is omnipresent in organizations. In particular the trend of increasing knowledge intensity of everyday tasks leads to the increased use of IS to support the knowledge worker [7]. IS typically log users' activities for administrative reasons e.g. authorization control or for analytical purposes e.g. web server logs to determine usage behavior. Activity logs are also increasingly used by applications for the support of the end user. These applications provide recommendations based on the activity logs. For example, they deliver similar items to the user depending on his previous searches for other items [19]. Other software recommends the next steps in a process, based on what has been done so far and what others have done in a similar situation [23; 25; 21]. Yet other support systems recommend who to contact in a social network based on the contextual overlap of the two [16].

The common theme among these support systems is their reliance on the concept of similarity to support the user. They recommend similar documents, similar process steps or similar persons, contingent on the current context, which needs to be modeled in an appropriate way.

While all these applications build on the concept of similarity, the interpretation of why two or more objects are to be considered similar depends on the application and its use cases for which the similarity-based application is employed.

The challenge when using similarity-based applications lies in determining a suitable notion of similarity. This is a complex task. There are many approaches stemming from diverse disciplines. They build upon definitions of similarity that are specific to those disciplines. In addition, most similarity measures use a number of parameters to determine how the similarity between two objects is determined. Finding a suitable configuration of a similarity measure is, therefore, a challenge.

Responding to this challenge, this paper introduces a framework that supports the developer of similarity-based applications when faced with the selection and configuration of suitable similarity measures. The framework allows to configure different similarity measures and in a next step to evaluate their appropriateness for the target application. Features of various applications that use similarity determination are taken into account to determine requirements for the framework. The implementation based on these requirements is detailed subsequently. In a case study, we demonstrate the utility of the framework in creating a similarity-based application.

The paper is organized as follows. Section 2 gives an overview of related work. Section 3 derives requirements for a framework to support the selection and configuration of similarity measures for activity sequences. Section 4 presents the paper's core contribution, outlining the framework that meets the requirements as outlined in section 3. Section 5 presents a case study that shows the utility of the framework in a real world setting. The paper concludes with a discussion and recommends further research.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

2. RELATED WORK

This contribution introduces a generic framework that supports the application engineer in selecting and configuring a suitable similarity measure for activity sequences. Its benefits are twofold. Practitioners can use the framework with minimal effort to find suitable similarity measures for their applications. On the other hand it provides a basis for researchers investigating properties of similarity measures for activity sequences.

A similar approach was taken in the process mining discipline where the generic ProM framework supports the development and application of process mining algorithms [6; 30]. The acceptance of this generic framework inspired us to create a similar framework to support further research of similarity measures for activity sequences and their application. In analogy to the framework described in this contribution, ProM acts as incubator for new algorithms and concepts in process mining. Additionally, ProM also operates on activity sequences. However, ProM's central concern is the support of process mining, i.e. the recovery of process models from event logs. Although it has incorporated many additional concepts and algorithms that extend its use to more than only process recovery, the determination of a suitable similarity measure is not in the scope of process mining. This is why our framework makes use of ProM's facilities wherever appropriate for example when accessing data sources or creating process models if this is necessary for the similarity measure. However, our framework's core functions are too different to integrate them into the existing ProM framework as a plug-in

Some of ProM's plug-ins make use of the concept of similarity for example to predict execution times of process instances based on previously executed process instances [28]. This is similar to what we discuss in the case study in chapter 5. However, while our case study aims at finding a well-suited similarity measure underlining the utility of our framework, in [28] the authors only use one kind of similarity measure and describe how to adapt its parameters best.

The challenge to find a suitable similarity measure has also been addressed in another context. In [10] the authors investigated how similar users or similar content can be determined in different social media to increase its usage. In all their experiments the authors used the same similarity measure, but used nine different sources for similarity information investigating their effects on six different social media applications. The results showed that the source of similarity information had significant influence on the perceived quality of the system's suggestions and also that the influence varies with respect to the different applications. In this paper we also stress that similarity measures must be tailored to the application that makes use of them. Additionally, in our framework we also acknowledge the great influence of the initial data and in what format it is collected. However, while in [10] quantifying the influence of different data sources for a concrete application was the goal, we focus on creating a generic framework that could support tasks like the one in [10]. Similarly, in [24] the authors investigate the influence of different similarity measures on recommendations in an online social network. They applied six different similarity measures to recommend potentially interesting sub-communities to their users and investigate the influence of the measure on the quality of the recommendation. In contrast to the contribution at hand, the authors in [24], however, focus on one use case and not on a generic support tool. Also they only use similarity measures that operate on sets, while our framework allows the usage of other similarity features as well.

3. REQUIREMENTS ANALYSIS FOR THE FRAMEWORK

Here, the requirements for a framework to support the selection and configuration of measures for similarity-based applications are documented. Requirements engineering can be done in many different ways [13]. Sources of requirements can be for example domain knowledge, existing systems, users, standards or regulations. In the following we use existing frameworks, source systems, data format standards and applications that use similarity measures as source for requirements elicitation. Each requirement is annotated with a number for reference in the outline of the framework's architecture.

A similarity measure is appropriate if it supports the goals of the target application. Therefore the determination of an appropriate similarity measure consists of two phases: firstly, selecting and configuring a similarity measure and secondly, checking the fitness for the target application. However a prerequisite is having data that is suitable to act as source for similarity information. This step is particularly important, because the selection itself greatly influences the result of the similarity determination [10]. Therefore, a generic framework must have the three components as shown in Figure 1. In the following each identified requirement references its corresponding component as indicated in this figure.

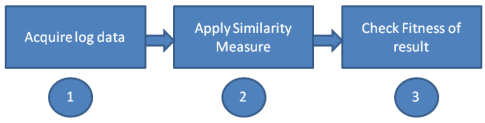


Figure 1 : Steps for determining suitable similarity measure

The requirements for the first component are elicited by inspecting different process aware information systems (PAIS) [8], such as ERP systems, project management systems and personnel management systems, the format of their data and how the log information is interpreted by the application and by persons.

Whenever it is desirable for an application to utilize the similarity of activity sequences, the first step is to retrieve those activity sequences from source application logs. But many applications may serve as sources for information and their data formats also take many forms. Some applications, for example ERP systems, store log information in databases while others such as web servers use files for this purpose, which are accessed through differently. A requirement for the framework, therefore, is:

> The framework should support both – data stored in files and in databases (*R.1a*)

In addition, the format of the data may vary. While data in a database is structured by definition, file-based logs can be stored

in comma separated value (CSV) files, using plain text with or without providing header information about the meaning of each section in a line, which is a still quite common solution. Other file-based logs are stored using some XML dialect with or without providing a schema definition along with it, especially when interoperability is important. Yet other IS use a log structure that is not plain text, is proprietary and needs parts of the IS logic to decode the log. Since all of these data formats are found in applications that can act as data source, the framework should fulfill the following requirement:

> The framework should support the retrieval of data from sources that are structured, semi-structured or use structures in a proprietary format (*R 1b*).

Additionally, the granularity of the log information may differ. Some applications log every user interaction, such as web servers, while others only log certain events including, for example, the change of a status indicator in a project management system. The same is true for context information that goes along with the log entry, which also can differ significantly in its extent. For example, while browsing in an intranet, much context information of an user is typically at hand, while anonymous access to an internet site offers less context information. Therefore, another requirement for the framework is as follows:

> The framework should be flexible enough to handle both rich data sources and to extract or amend less rich data sources (*R.1c*).

IS also differ in their pervasiveness. Some IS log user interaction in the background with little or no user involvement, such as web browsers while they browse through a website, while other IS only write into their logs when explicitly requested by the user, such as for example accounting systems. This influences the granularity and the possibilities for interpreting the log, because in the first case we often need to interpret implicit behavior while in the second case the intention of an user is more explicit and related more strongly to the log entry. Depending on the knowledge about the process that is supported by the IS, it is possible to amend log data with context information. This creates another requirement for the framework:

> The framework needs to be agnostic to how the data is captured from a technical point of view, but needs to provide means for amending the data with implicit information (*R.1d*).

Also, the kinds of stored data differ. Some applications store an event, or activity respectively, in their logs, i.e. what has happened. Others store data that reflects the situation after an activity has been performed, i.e. the result of what has happened. A web browser for example might store the event "page index.html has been requested", while a project management system might store the status "project budget is (now) 100k €", but not the event itself that increased the budget to this amount. Another requirement for the framework, therefore, is as follows:

> The framework should have the capability to transform log information containing status snapshots into log information containing status changes (*R.1e*).

Independent from the characteristics of the data source itself, more than one application log may contain information for an activity sequences, i.e. the information contained in one application log can augment information from another log. For example, a project management system could contain the execution history of a project, while in a separate accounting system, information about consumed budget is kept. This is why the framework should fulfill another requirement:

> The framework should allow the flexible and iterative enrichment of log data from multiple sources (*R.1f*).

Being able to import data from arbitrary data sources and being able to transform them in a suitable format forms the basis for the second component of the framework in Figure 1. It allows to apply similarity measures to the input data. The requirements for the second component were elicited by reviewing the properties of thirteen similarity measures found in literature, extracting their common features and deriving requirements from their common features. The measures were used in a wide range of disciplines such as protein function prediction in biology, comparison of Web Service definitions in computer science and overlap calculation in graph theory, to name a few application scenarios.

Before discussing different kinds of similarity measures, it is necessary to take the goal of their use into account. Applications that make use of the similarity of activity sequences can have different target functions. For example, in project controlling it is often relevant to assess the likeliness of success. This could be done by determining similar projects that have been completed already, taking their success as an indicator for the currently running project. In that case, the goal is to make a good estimation about project success. Another example with a different target function can be found in product recommendation engines where users are presented with similar products that overlap with their peer's preferences. In this case, the goal is to leverage the cross-selling potential. Different target functions have different definitions of when a similarity measure works well on a set of activity sequences and when it does not. It is often appropriate to adapt a similarity measure to suit its intended support for a goal, using supervised learning techniques [33]. The framework's similarity measure component should, therefore, fulfill the following requirement:

> The framework should have the capability to label a training set of activity sequences with an indicator of its utility in relation to the target application's goal (*R.2a*).

There are a number of different ways to determine the similarity between two entities. For this reason the framework needs to be flexible enough to support each different way. In a first instance, an entity can be described by certain flat attributes, for example a project is described by the number of project members and the total budget. In that case, the two entities can be compared according to their attribute values, where the comparison can be done with different algorithms depending on for example the data types or data ranges. This is why the framework should fulfill the following requirement:

> The framework should be able to support similarity measures that operate on input entities that are described by attribute-value pairs (*R 2b*).

In a second instance, an entity can have structured components, for example a project is described by the activities that have been performed during its execution. In that case, the two entities can be compared according to the overlap of the same constituting parts, i.e. the same activities. The framework should, therefore, fulfill another requirement:

> The framework should be able to support similarity measures that operate on structured input entities by for example comparing the overlap of components (*R.2c*).

Additionally, the constituting parts can themselves have attributes, for example each activity in a project can have a specific person that is responsible. Therefore, the comparison of entities can be based on constituting components, acknowledging the difference in attributes as well. Essentially, this is an extension of the requirement described before, where the constituting components, were treated as flat structures and were compared for equality. The framework should fulfill the derived requirement:

> The framework should be able to support similarity measures that operate on structured input entities where each structured component is (additionally) described by attribute-value pairs (*R.2d*).

Finally, it is possible to take the relationship between the constituting parts into account. The relationship represents the temporal or logical order of the constituting parts and may also reflect interleaving of those activities. Addressing this fact, the framework should fulfill the following requirement:

> The framework should be able to support similarity measures that operate on the structure of its entities, i.e. that use structural properties of the input data for similarity determination (*R2.e)*.

Additionally, many similarity measures use one or more parameters to configure the computation of similarity. The framework must, therefore, fulfill an additional requirement:

> The framework should offer the capacity to process parameters for each similarity measure that determine its behavior (*R.2f*).

Each similarity measure typically focuses on one or at least a small set of properties of the input object. However, it is possible that the desired notion of similarity is best reflected by a combination of different properties. In this case, the simultaneous application of different similarity measures is necessary. This poses another requirement to the framework:

> The framework should allow for a compounded calculation of similarity using different measures (*R.2g*).

As indicated above, the structural properties of activity sequences can be used for the similarity determination. Yet, each activity sequence itself has a linear structure by definition. To find out about the dependencies between activities, a model of possible sequences indicating their relationship is required. In many cases, explicit models of activity sequences are not available because they are too expensive to create or because the activity sequences are too flexible to render a model useful. Nevertheless, if the usage of structural properties is deemed necessary, there needs to be a way to at least recover an implicit model for the activity sequences. It would have to be reconstructed from the IS logs and would then indicate the process "as it is lived". In terms of similarity determination, it can be used to deduce structural properties of an otherwise linear activity sequence. The framework should fulfill another requirement:

> The framework should provide a possibility to create a (process) model using the activity sequences that are available (*R.2h*).

We have reasoned above, that similarity measures only have a purpose with respect to their target application. Therefore, our framework should facilitate the selection and configuration of an appropriate measure. To find out about the utility of the selected measure and its configuration, the results have to be seen in the light of the application that they will be used for. The framework needs to be integrated into the target implementation or needs to be integrated into a suitable representation thereof to show its utility. Especially if supervised learning techniques are used for the selection and configuration of similarity measures, the feedback of the application about how well-suited the similarity measure's results are for the intended use is pivotal. Therefore, the framework needs to fulfill a requirement that intentionally covers a broad spectrum of interpretations to encompass arbitrary similarity-based applications:

> The framework should support the integration of similarity-based applications or suitable representations thereof that consume the results of the similarity determination and give the framework feedback about the quality of the results (*R.3*).

4. PROPOSED ARCHITECTURE

The requirements discussed above informed the design of the framework. The logical structuring in three different components (see Figure 1) proved useful for the elicitation of requirements. For the implementation of the necessary functionality it turned out that the second component can be split into three modules: one that supports classifications for supervised learning, one that can mine a process model from input logs for the support of structural similarity measures and one for the application of similarity measures itself. Therefore the architecture features a modular design with five main modules (see Figure 2). The framework handles the flow of action by instantiating one or more plug-ins for each module and passing on the control subsequently. However, it is not mandatory to use all modules, i.e. classification of activity sequences is only necessary when supervised learning should be supported and the creation of an activity sequences' model is only necessary if structural properties should be used in the similarity determination.

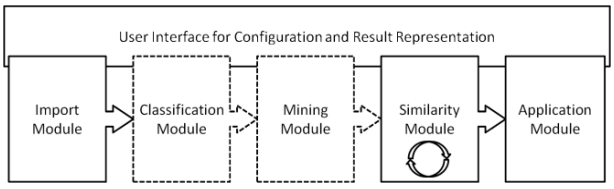


Figure 2 : The framework's modules

The framework was implemented using Java, because it is a very common programming language that many application engineers can work with. Wherever appropriate, existing applications and frameworks were integrated into the framework directly. This is true for parts of the import module that builds upon the ProMImport application and the process mining module ProM, that relieved us from the effort of implementing process mining algorithms. If the functionality of the incorporated applications did not entirely suit our needs, we extended them to meet our set of requirements. In this section, the framework's five modules are detailed. The reasons for the design choices in each module are explained by reference to the requirements in chapter 3.

4.1 Importing Data into the Framework

The requirements analysis showed that activity sequences are frequently stored in different locations, i.e. databases or files, and have different formats. Therefore, importing data into the framework for similarity determination cannot be done with one single import routine but rather must be done by offering an interface that supports as much flexibility as possible. A plug-in that implements this interface then offers the application-specific ability to extract activity sequences.

Extracting information from logs is a common challenge. An existing tool, the ProMImport Framework [9], served as a good basis for the import module, although its extraction mechanisms primarily focus on the support of process mining. For its use in the framework, ProMImport had to be extended. The ability to store the extracted data before displaying it and the ability to connect the output of one import plug-in to the next import plug-in were added. In this way, requirements *R.1a* and *R.1b* are addressed by delegating the specifics of the data extraction to plug-ins while offering a generic interface.

The requirements analysis also shows that relevant information for activity sequences could be spread across different logs that logically complement the information contained in each. The ProMImport Framework had no support for aggregating information found in different kinds of sources for one combined import result. Therefore, the concept of chaining importer plug-ins has been implemented into the framework. While ProMImport would display the results of the import directly, the framework's import module can deliver the results to another importer plug-in to augment the results, yielding, in the end, one integrated import result. This fulfills requirement *R.1f*, which states that information could be spread across different sources.

Offering the ability to connect arbitrary data sources also requires that a suitable data format is defined within the framework that can be used for intermediate storage. As the requirement analysis indicated, logs can have varying expressiveness concerning the granularity of logged activities and concerning each activity's context information. The internal data format must be designed in such a way that the transformation of source data into this internal format is without loss of expressiveness or limits the loss to a minimum. Data formats that are crafted like this can be found in the WFMC's specifications of the Common Workflow Audit Data (CWAD) [32] and in the MXML format [29]. Both data formats are abstract enough to represent the contents of different application logs. Being tailored to the needs of a workflow system, the CWAD format has a considerable number of attributes that only apply in a workflow context. The MXML format on the other hand, abstracts from workflows and therefore allows a more straightforward transformation of arbitrary data. Therefore, MXML is adopted as the internal data format for the framework because it is flexible enough to handle logs with varying granularity and is abstract enough to handle logs with varying contextual data. This fulfills requirements *R.1c* and *R.1d*, which state that the granularity of data can vary.

The chosen data format conceptually stores sequences of events, which is also true for the CWAD format and many others. On the other hand the results of the requirements analysis pointed out that there are also logs that do not store activities as events but rather by storing the results of the activities. Consequently, there needs to be a mechanism to transform logs containing data states into event sequences. While this problem has been addressed in theoretical computer sciences [15], there is no actionable implication for an implementation as would be necessary for the framework. For that reason, the framework incorporates a configurable, XML-based application for this purpose. After one-time configuration, it automatically selects defined portions of the source data that contains sequences of states and transforms them into event sequences during data import. The events are created using activity names that are configured before importing the data. This fulfills requirement *R.1e*. In addition, for each activity attributes can be assigned and their values can be calculated using basic arithmetic and string operations. This fulfills requirement *R.1f*.

4.2 Classifying Activity Sequences

The requirement analysis specifies that a label must be assigned indicating the utility with respect to the desired target, if supervised learning is to be used. In principle, two different ways to allow for labeling are possible: automated or manual. In one case, activity sequences are labeled according to one or more rules that are created by a domain specialist. For example, the result of project executions as good, mediocre or bad could be automatically determined using the budget-to-spending ratio of each project as basis for a rule. However, this automated approach has a significant disadvantage. If there was a rule available that perfectly labels this type of activity sequences, then this rule is at the same time a perfectly suitable similarity measure and there would not be any need for using the framework in the first place. However, the more complex activity sequences are, the less likely it is that a person knows according to which measures an activity sequence should be evaluated. This person nevertheless may often be able to indicate the result tacitly without knowing how to derive this judgment formally. Therefore, the framework uses the other choice, namely the manual labeling approach. In this case, a person classifies a training set of the activity sequences and stores the results in a

csv file. This file is then used as an input source for the framework to automatically classify the training set of activity sequence. As outlined in the requirements analysis, labeling is not always necessary and is implemented as an optional step in the framework. Having the labeling module fulfills requirement *R.2a*.

4.3 Generating a Model of the Activity Sequences

The requirements analysis shows that many applications do not have an explicit model for the execution of activity sequences, but its users may follow implicit models that for example stem from corporate rules or from the technically possible interaction via a graphical user interface. If the implicit model can be explicated in a possibly only approximated model, it can be used to extract structural properties of an otherwise linear activity sequence. Inductively creating a model from instances contained in a log is the main concern of the process mining discipline [1; 3]. There are many algorithms available to mine a model from instances. The application of these algorithms is facilitated by the ProM framework [6; 30] that has many of them integrated as plug-ins already. However, ProM returns its mined models typically as event-process-chains (EPC) or as petri-nets [18] that perfectly serve the purpose of modeling the execution semantics. In the context of determining similarity, the execution semantics do not play a large role though, which allows for the simplification of petri-nets and EPCs into simple graphs that only consist of edges and vertices. This also creates the ability to use similarity measures that work with simple graphs.

The transformation is done as follows (see Figure 3). Whenever the ProM framework returns a petri-net[1], the framework needs to transform it by creating one node for each of the petri-net's transitions (blocks in the figure). Those nodes are connected to one another by inspecting which transitions are connected in the petri-net, where the term connected is interpreted as follows. Two transitions are connected if there is exactly one place (circles in the figure) in between them. If the model has explicit routing nodes (not shown in the figure; would be *XOR*, *OR* or *AND*, with the obvious semantics), then two transitions are connected if there is a sequence of zero or more routing nodes and one place in between them, but no other transition. Additionally the resulting graph is extended with explicit *Start* and *End* nodes that are implicit in petri-nets. Using this definition and this way of transforming the petri-net, it does not matter if one uses explicit routing nodes or implicit routing by means of petri-net firing semantics. Both will be transformed into the same graph. An example of implicit routing can be found in the petri-net on the left part of Figure 3. In that example Task 3 must always be executed in accordance to petri-net firing semantics.

[1] In the context of process mining and also similarity determination EPCs can be transformed into petri-nets without loss of relevant information

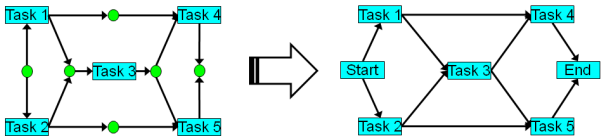


Figure 3 : Transformation of a Petri-net to a graph

The transformation is adequate. The argument is as follows. Since essentially places are removed from the petri-net, along with the routing nodes, if there are any, only the connection between transitions are left that are now considered nodes. It is, however, not obvious anymore if those connections are to represent exclusive execution or parallel execution. It is no longer known which transitions are prerequisites of certain nodes and, therefore, explicit information about the execution semantics is lost. However, this is not a problem for the intended use. The transformed model is solely used to identify the path that was taken by a process instance so far. Therefore, there is no need to know which connections previously modeled exclusive or parallel execution. The connections that are used by the instance were obviously parallel, the ones that have not been used but could have been according to the graph are of no interest, but obviously have represented OR/XOR split alternatives. It is also not a problem that there is no information about which nodes are prerequisites of another node, since the further proceeding of the process instance is of no interest. Utilizing the existing ProM framework but adapting its output fulfills requirement *R.2h*.

4.4 Determining the Similarity between Activity Sequences

The requirements analysis indicates that the framework must offer an interface for the creation of its own similarity measures. Nevertheless, it is desirable to have a reasonable number of algorithms available in the framework to make it useful from the start. As the algorithms differ with respect to the parts of the activity sequence they use for the computation, the framework should offer at least one algorithm for each kind of similarity measure. This guarantees that other similarity measures that operate on the same kind of input data can be integrated into the framework. The framework has 13 similarity measures integrated in its initial phase, that can work with the general properties (requirement *R.2b*) of an activity sequence, the overlap of activities (requirement *R.2c*), taking into account the activities attributes if necessary (requirement *R.2d*) and also taking into account structural properties (requirement *R.2e*).

Within the framework, the main task of each similarity measure is to determine a similarity matrix, i.e. it must create a matrix with as many rows and columns as there are activity sequences with each entry containing the degree of similarity between the respective combinations of activity sequences. This very general representation of a similarity measure's result allows applications to extract the relevant information flexibly. This fulfills requirement 5 which states that applications building upon the similarity of activity sequences can have diverse needs with respect to the similarity measure. A downside is, that this way of storing a similarity measure's result is not the most

efficient way considering computational cost and space requirements.

The requirements analysis also reveals that similarity measures frequently need configuration to a certain degree. The framework provides the means to properly configure a similarity measure. Each similarity measure plug-in is requested to publish its necessary parameters to the GUI component and can then process them as needed. This fulfills requirement *R.2f*.

Further, the requirement analysis shows that cases must be supported in which one similarity measure is not enough. A combined result of different similarity measures might fit the application's needs better than a single measure could. The framework supports this kind of configuration. Internally, each measure computes one similarity matrix. The matrices are then combined to yield one similarity matrix. The combination is done by also allowing for a weighting between the measures. This fulfills requirement *R.2g*.

4.5 Applying the Selected Similarity Measure in the Target Application

The framework is designed to support any application that builds upon the use of activity sequence similarity. This creates the need to offer different configuration mechanisms. One way of configuring the similarity application is the use of incrementable parameters. The parameters are set up with a maximum, a minimum and an initial value, along with a step size. The application iteratively performs its task and changes the incrementable parameter as indicated by the step size, until the upper or lower limit is reached. This functionality is indicated by the circle below the similarity component in Figure 2. The usage of the parameter is not controlled by the framework itself, but by the application, while the framework performs the increment. If more than one incrementable parameter is set for the application, the framework ensures that every possible combination is explored.

Additionally, the framework offers the application an interface for interacting with the intermediate results[2]. The concept is as follows: Directly after the intermediate information is created, the application is asked to pre-process the intermediate information. This happens before the determination of weight distribution and before using incrementable parameters. If the application uses this option, it can tell the framework how to pre-process the dataset and by this means adapt it to its needs. It is then provided with the pre-processed data instead of the plain intermediate results. This step is performed as long as the application indicates that it still wants to change the data. This implementation is generic enough to support arbitrary applications but offers enough functionality to still support the application engineer which fulfills requirement *R.3*. An illustration of the utility of this feature is part of the case study in chapter 5.

[2] The imported activity sequences, the potentially mined model, and the potentially created classification are considered as intermediate results.

4.6 General Features

The goal of the framework is not only to relieve the application architect of the task of finding a suitable similarity measure, but also to find a well-suited configuration of the similarity measure. The similarity application module supports this feature. When the usage of more than one measure is desired, the framework can be used to determine the best combination in terms of weighting. The user only needs to specify how fine-grained the search of the best solution should be by providing an increment value. This value is then used to exhaustively search the result space, which is done by iteratively using each weighting combination for the similarity measures. The combined measures' result is determined in the light of the application that builds upon them, which in turn informs the framework how well this combination is suited to its needs.

After each possible iteration that might stem from the presence of incrementable parameters application pre-processing calls or optimal weight determination, the application returns its collected information to the framework. The collected information reflects the respective performance of each possible combination. For this purpose, it uses a multi-dimensional array, where each dimension represents one incrementable parameter, and the array's value represents the parameterization's performance with respect to the application's performance criteria.

To enable the user to visually explore the relationships, the user can select a graph that shows a two-dimensional projection of the resulting multi-dimensional array. The two dimensions of the graph can be determined without limitation.

5. CASE STUDY

The framework's capabilities are investigated in a real life scenario, where the similarity of activity sequences is used to amend the functionality of an existing application. As one instantiation of an IS that benefits from similarity determination, in the case study, a project management system is investigated. The results of the case study inform the company how to make better use of what has been learned in previous projects with only minimal effort. On the other hand the case study shows the framework's ability to scale and support the application engineer.

5.1 Case Study Background

The company in our case study had been using a proprietary project management application that kept track of the status and the customer interaction during project execution for a number of years prior to the case study. It distinguished between nine different statuses a project can have, such as *customer contacted*, *price negotiated*. Additionally in each status, information like assigned employees, estimated project cost and profit and realized cost and profit are stored. Also an SAP system was used for keeping track of the employee's time on different projects.

The system contained a history of 124 projects covering consultancy and prototypical development of applications for customers. Each dataset contained predefined steps that indicated the status of the project, interactions with the customer, the

respective dates of these interactions and information about who is involved in the respective phases of the project.

Up until the case study, the project management system was used for (retrospective) project reporting and for giving the project portfolio manager an up-to-date overview of the status of the projects. There was, however, no actionable support feature, like project progress projection or reusing experiences from previous projects. The project members saw it as a valuable approach to find similar projects using the activity sequences that could be extracted from the logs. The assumption was, that helping a project manager of a currently running project in finding similar projects, would allow him to learn from the experience of similar previously finished projects. This way he would be presented with the likely performance of his current project utilizing the performance of the similar projects as a predictor and contact details of the related projects' managers to ask them for support.

However, the team members could not clearly define "similar" in this context. The project portfolio manager was able to indicate the quality of the projects' processes, which ranged from poor to good, without being able to state which parameters could be used to support his judgment. This is a typical problem in complex decision environments.

The goal, therefore, was to identify similar projects, where relevant information about the different projects was stored in a project management system. Because it was not known which similarity measures can be useful and neither which features to use, the central research question was: Which similarity measures should be used and in what way to support this knowledge management initiative. As the range of possible measures and possible configurations is large and the evaluation of each single measure and configuration is a time-consuming task, the case study lends itself to applying our framework.

5.2 Configuration of the Framework's Modules

The proprietary data within the project management system was stored in an XML dialect specific to the application that could not be imported into the framework using an existing ProMImport plug-in. This is why a new one was developed. Because it was not known which influence the granularity of the log entries would have on the similarity measure's suitability, we created the importer plug-in configurable to this respect. This allowed us to extract two, differently verbose representations. One transformed the data by interpreting the change between nine given high-level status indicators as activities. The other imported data by additionally interpreting more fine-grained interactions like "insert expense type" as activities. Having two differently large sets for the same source information supports the analysis of the effect on similarity measures that is related to the size of activity sequences. Additionally, some accounting related data was not maintained in the project management IS directly, although it is logically connected to it. Hence, in the case study setting, the imported data from the project management system had to be amended with additional data from an SAP system, for which we could reuse parts of PromImport. Implementing the new importer plugin required some effort but did not take longer than a few days. The configuration of the plug-ins however was straight forward and took only a few minutes.

One of the goals for the company was to estimate the performance of a project by utilizing the similarity of its activity sequence with respect to previously finished projects. However, the stakeholders did not know which features were the best ones to use to determine similarity, while knowing how projects as a whole can be evaluated. For this reason, making use of a supervised learning approach is a suitable approach, which justified the use of the framework's classification module. Within this step, each activity sequence was augmented with the performance judgment of the project portfolio manager using a three-valued classification indicating whether a project was positive, negative or neutral. The configuration of the classification module was straight forward and took less than an hour.

In interviews, the stakeholders agreed that the interaction between different activities on the project were related to its later performance, giving rise to the use of structure-oriented similarity measures. Given the complex interactions within a project, the company did not have an explicit interaction model for their project management system. If structure-oriented similarity measures were used, retrieving a model required using the model generating facilities offered by the framework. A limited number of algorithms included in the ProM framework proved useful in this case study. After some experimentation, the α-algorithm [2], the multi-phase algorithm [2] and the genetic-mining-algorithm [17] proved suitable enough for the model determination task.

The stakeholders could not give an informed recommendation on which properties would best support or not support a similarity determination. This is why a diverse set of different measures has been used to determine the most suitable one. As the activities in the logs amounted to changes in the project status and are known in beforehand, it was viable to interpret activities as similar whenever they have the same name. The usage of equivalence classes or the consideration of the activities' attributes was not necessary in this case. Altogether, the case study used nine similarity measures, out of which five neglected the structural properties, while the other four relied on structural properties for the determination of similarity. They included the Dice Coefficient, the Overlap Coefficient [26], a bag of words [14] adaptation to activities, a Term-Frequency-Inverse-Document-Frequency [11] adaptation to activities, the Levenshtein distance [31] for activities, graph isomorphism [27], maximum common sub-graph [5], graph edit distance [20] and random walk kernels [12]. This covered a broad range of different measures which made use of all the functionalities supported by the framework. Each of these measures has its special advantages and disadvantages which is why we expected them to operate differently well depending on the input data. However, while the description and especially the comparison of their properties is a valuable contribution, it is out of scope in this paper. The configuration of the similarity measures took no more than a couple of minutes for each measure.

The configuration of the application module was done as follows: The application that is to benefit from the determination of

activity sequence similarity is intended to estimate the performance of a project using three predefined values for the performance. To create its estimation the application does a classification, which in the case study was done using a k-nearest-neighbor classifier – a common approach for classification. To determine which neighbors, i.e. which projects are "near", the application utilized the similarity measures' results as its basis. The number of nearest neighbors typically has a large influence on the classification results and must therefore be taken into account when searching for suitable configurations. In the framework, it is configured as an incrementable parameter (see section 4.5), i.e. the framework iterates through different combinations of this parameter and tests the results of each configuration separately.

The application used a second incrementable parameter. Because the aggregation of the k-nearest neighbors' class indicator into one single answer can be done in different ways, the desired algorithm can be selected using an incrementable parameter. The application module offered four different ways to do this such as majority vote and using different weighting mechanisms according to distance. The parameter iterated over those four.

The application should later serve the purpose of providing an estimate concerning the future outcome of a current project, i.e. a project that has not ended yet. To evaluate the performance of the similarity measures and their parameter configuration, the available data was split in training and test data. The test data, however, needed special treatment. The available data consisted of finished projects, but for testing the prediction quality, it is necessary to have projects that are not finished yet. For that reason, each activity sequence was first pruned using a value as indicated by an incrementable parameter and then compared to the remaining completed activity sequences to emulate the situation of a currently running project. The pruning was performed to an increasing degree using the third incrementable parameter.

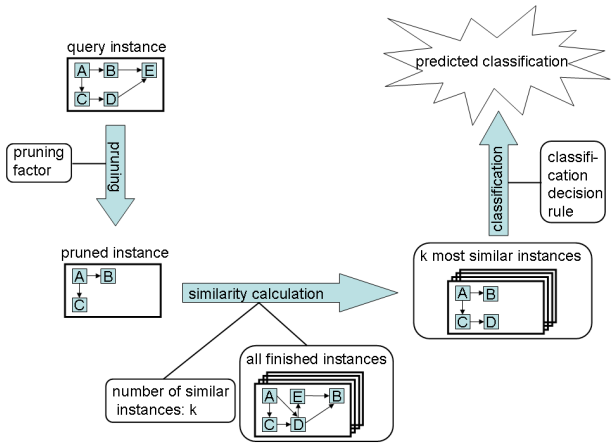


Figure 4 : Illustration of the steps for the use case

Because every project has a performance value, assigned by the project portfolio manager before the test run, it was possible to compare the results of the prediction to the actual performance value for each project. The aggregation of the single results were used to determined the overall suitability of a similarity measure for the task of predicting the performance of a project by using four different indicators that are typically used to evaluate classifiers: precision, recall, accuracy [4] and the F-measure [22]. The results were stored in a multi-dimensional data structure and were selectively displayed in a 2-D graph according to user-defined selection criteria. The steps within the case study to test the quality of different similarity measures are illustrated in Figure 4.The implementation of the application module's plug-in for the case study consumed most time and took a few days. However the program code can be integrated into the target application, therefore, the time would have been necessary anyway.

5.3 Case Study Results

The configuration as detailed in the previous section was used to perform the project performance prediction with 11 different settings for the pruning of a respective activity sequence, which reflects increasingly mature projects in terms of their run time. Also 124 unique values for the k-nearest-neighbor classifier were tested. In each iteration the four fitness indicators for the similarity measure were determined. Altogether seven different classification approaches were used, three of which were using simple heuristics[3]. The heuristics were used to compare the result of the other approaches in the light of reference results. This helped to understand the influence of potential biases in the input data. Most similarity measures outperformed all heuristics which indicated that a potential bias of the data had no significant influence. In each iteration, the data structure consisted of a 124-by-124 matrix – one line and column per activity sequence, corresponding to 15,376 entries, which in turn needed 7,688 computations of similarity values due to symmetry in the matrix. For each similarity measure, there were 11 * 124 application configurations for the 7,688 computations resulting in 10,486,432 similarity results per measure and 94,377,888 in total.

Table 1. Results of different similarity measures on prediction accuracy in the case study

	Small activity log	**Large activity log**
Measure		
Dice Coefficient	76 %	**72 %**
Overlap Coefficient	69 %	68 %
Bag of activities	70 %	70 %
TFIDF	74 %	66 %
Levenshtein	78 %	**72 %**
Graph isomorphism	67 %	33 %
Max. common-sub-graph	**79 %**	?[4]
Graph edit distance	75 %	64 %

[3] Simple heuristics were to always classify as good, bad or neutral

[4] Determination was not possible due to the algorithm's computation complexity in combination with the large dataset.

Random walk	78 %	66 %

For lack of space, not all results can be displayed in this paper. Table 1 shows the highest accuracy values for each similarity measure on both data sets. The Table shows the variation of results that can be experienced when using different measures. The difference in accuracy can be quite significant (10% on small logs, 39% on large logs), where the highest values are on a level, suitable for real life application.

The results of this case study can be seen from different angles. For the company that utilized our framework, knowing the maximum achievable accuracy for project performance prediction was valuable information, as it supports the project managers' interpretation of predictions. Without the framework, the effort for the determination would have been too high and some arbitrary, possibly non-optimal, similarity measure would have been used.

This relates to another result of the case study. We wanted to find out how well the framework could support application engineers and how much effort could be saved. The most time in the case study was spent programming the importer plug-in and the application module plug-in. These two tasks were necessary for the extension of the project management system anyway and both are independent of the similarity measures that were applied. Only the adaptation to the framework's interface caused additional effort. Together these implementation tasks took several days. Afterwards, however, the configuration of each module could be done in a matter of hours. This indicates the framework's value for application engineers, as the implementation and configuration without the framework would have taken much longer.

Another goal was to determine the scalability of the framework. In the case study, the framework extracted large volumes of data from the initial data source. Additionally, it computed process models and performed classifications. And finally, it performed nearly 100 million similarity calculations. With the exception of one similarity measure that is inherently computationally hard (it is NP-complete), the calculations were performed very quickly and none of them took longer than a few hours on standard desktop PC. This indicates the scalability of the framework.

6. CONCLUSION AND OUTLOOK

This contribution was motivated by the observation that many applications make use of the concept of similarity of activity sequences. However, the problem lies in finding the right measure for determining similarity and configuring the measure appropriately. Due to the large number of possible algorithms and configurations, a selection and configuration of a suitable measure should be automated and supported to relieve the domain specialist of routine tasks. The authors, therefore, call for the creation of a framework that supports the application engineer in finding the right measure. The requirements for such a framework are deduced by analyzing the data formats of contemporary information systems, similarity measures that are used in similarity-based applications and frameworks that are used in similar disciplines. Building upon and structuring the requirements, the components of a supportive framework are proposed. It is geared to be as flexible as possible, highlighting five modular components that allow the integration of plug-ins to cater for expendability. The framework's utility is shown in a case study where a suitable similarity measure for the performance prediction of projects is investigated. Utilizing the framework it was possible to successfully automate the computation of almost 100 million similarity values to find a suitable similarity measure. This was a task that did not take more than one person-day in the case study for configuring the framework.

While the framework was shown to be of great use, it was applied only in one case study. Great care has been taken to anticipate the needs of all applications that could potentially benefit from using the framework. To further verify the frameworks utility and also benefit from its potential, we intend to perform more case studies, especially in the area of knowledge management. We will use the framework to find suitable measures for recommendations, this time using persons and their interactions with IS as units of analysis. Another direction for further research lies in determining the properties of different similarity measures with respect to the input data. The case study already gave some interesting insights in possible properties. Those will have to be investigated more thoroughly to derive general recommendations.

7. REFERENCES

[1] Aalst, W.M.P.v.d.; van Dongen, B.F.; Herbst, J.; Maruster, L.; Schimm, G.; Weijters, A.J.M.M. (2003): Workflow Mining: A Survey of Issues and Approaches. In: Data and Knowledge Engineering, Vol. 47, 2, pp. 237-267.

[2] Aalst, W.M.P.v.d.; Weijters, T.; Maruster, L. (2004): Workflow Mining: Discovering Process Models from Event Logs. In: IEEE Transactions on Knowledge & Data Engineering, Vol. 16, 9, pp. 1128-1142.

[3] Agrawal, R.; Gunopulos, D.; Leymann, F. (1998): Mining Process Models from Workflow Logs. In: Sixth International Conference on Extending Database Technology, pp. 469-483.

[4] Baeza-Yates, R.; Ribeiro-Neto, B. Modern information retrieval.

[5] Bunke, H.; Shearer, K. (1998): A graph distance metric based on the maximal common subgraph. In: Pattern recognition letters, Vol. 19, 3-4, pp. 255-259.

[6] Dongen, B.F.v.; de Medeiros, A.K.A.; Verbeek, H.M.W.; Weijters, A.J.M.M.; van der Aalst, W.M.P. (2005): The ProM Framework: A New Era in Process Mining Tool Support. In: ICATPN, pp. 444-454.

[7] Drucker, P.F. (1999): Knowledge-Worker Productivity: THE BIGGEST CHALLENGE. In: California Management Review, Vol. 41, 2, pp. 79-94.

[8] Dumas, M.; van der Aalst, W.; Ter Hofstede, A. (2005): Process-aware information systems: bridging people and software through process technology, Wiley-Blackwell 2005.

[9] Gunther, C.; van der Aalst, W. (2006): A generic import framework for process event logs. In: Applications and Theory of Petri Nets 2005, 26th

International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005, Proceedings, Vol. 4103, pp. 81.
[10] Guy, I.; Jacovi, M.; Perer, A.; Ronen, I.; Uziel, E. (2010): Same places, same things, same people?: mining user similarity on social media, p. 41-50.
[11] Jones, K. (2004): A statistical interpretation of term specificity and its application in retrieval. In: Journal of documentation, Vol. 60, pp. 493-502.
[12] Kondor, R.; Lafferty, J. (2002): Diffusion kernels on graphs and other discrete input spaces, p. 315-322.
[13] Kotonya, G.; Sommerville, I. Requirements Engineering: Processes and Techniques. 1998. John Wiley & Sons.
[14] Lewis, D. (1998): Naive (Bayes) at forty: The independence assumption in information retrieval. In: Applications and Theory of Petri Nets 2005, 26th International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005, Proceedings, Vol. 1398, pp. 4-18.
[15] McCarthy, J. (2002): Actions and other events in situation calculus, p. 615-628.
[16] McDonald, D.; Ackerman, M. (2000): Expertise recommender: a flexible recommendation system and architecture, p. 231-240.
[17] Medeiros, A.K.A.d.; Weijters, A.J.M.M.; Aalst, W.M.P.v.d. (2004): Using Genetic Algorithms to Mine Process Models: Representation, Operators and Results. Eindhoven University of Technology, Eindhoven.
[18] Peterson, J. (1981): Petri Net Theory and the Modeling of Systems. In: PRENTICE-HALL, INC., ENGLEWOOD CLIFFS, NJ 07632, 1981, 290.
[19] Resnick, P.; Varian, H.R. (1997): Recommender Systems. In: Communications of the ACM, Vol. 40, 3, pp. 56-58.
[20] Sanfeliu, A.; Fu, K. (1983): Distance measure between attributed relational graphs for pattern recognition. In: IEEE TRANS. SYS. MAN CYBER., Vol. 13, 3, pp. 353-362.
[21] Schonenberg, H.; Weber, B.; Van Dongen, B.; van der Aalst, W. (2008): Supporting flexible processes through recommendations based on history. In: Business Process Management, pp. 51-66.
[22] Shaw, W.; Burgin, R.; Howell, P. (1997): Performance standards and evaluations in IR test collections: Cluster-based retrieval models. In: Information Processing and management, Vol. 33, 1, pp. 1-14.
[23] Shkundina, R.; Schwarz, S. (2005): A Similarity Measure for Task Contexts. Paper presented at the 6th International Conference on Case-Based Reasoning, ICCBR, August 23-26, 2005, Workshop Proceedings, Chicago, IL, USA, p. 261-270.
[24] Spertus, E.; Sahami, M.; Buyukkokten, O. (2005): Evaluating similarity measures: a large-scale study in the orkut social network, p. 684.
[25] Stahl, A. (2004): Learning of knowledge-intensive similarity measures in a case-based reasoning. PhD, Universität Kaiserslautern 2004.
[26] Tan, P.; Steinbach, M.; Kumar, V. (2005): Introduction to data mining, Pearson Addison Wesley Boston 2005.
[27] Ullmann, J. (1976): An algorithm for subgraph isomorphism. In: Journal of the ACM (JACM), Vol. 23, 1, pp. 31-42.
[28] van der Aalst, W.; Schonenberg, M.; Song, M. (2009): Time Prediction Based on Process Mining. In: BPM Center Report BPM-09-04, BPMcenter. org.
[29] van Dongen, B.; van der Aalst, W. (2005): A meta model for process mining data, p. 309–320.
[30] Verbeek, H.; Buijs, J.; van Dongen, B.; van der Aalst, W. ProM 6: The Process Mining Toolkit. In.
[31] Wagner, R.; Fischer, M. (1974): The string-to-string correction problem. In: Journal of the ACM (JACM), Vol. 21, 1, pp. 168-173.
[32] WFMC (1998): Audit data specification. Technical Report WFMCTC - 1015 Version 1.1, 1998.
[33] Witten, I.; Frank, E. (2002): Data mining: practical machine learning tools and techniques with Java implementations. In: ACM SIGMOD Record, Vol. 31, 1, pp. 76-77.

Towards the Analysis of Information Systems Flexibility: Proposition of a Method

Daniel Wagner
University of Bamberg
Feldkirchenstraße 21
96045 Bamberg, Germany
+49 951 863-2664
daniel.wagner@uni-bamberg.de

Christian Suchan
University of Bamberg
Feldkirchenstraße 21
96045 Bamberg, Germany
+49 951 863-2652
christian.suchan@uni-bamberg.de

Benjamin Leunig
University of Bamberg
Feldkirchenstraße 21
96045 Bamberg, Germany
+49 951 863-2661
benjamin.leunig@uni-bamberg.de

Jochen Frank
University of Bamberg
Feldkirchenstraße 21
96045 Bamberg, Germany
+49 951 863-2653
jochen.frank@uni-bamberg.de

ABSTRACT

The flexibility of information systems (IS) is a topic of growing importance within IS analysis and design. IS have to cope with the increasing structural and behavioral complexity of environmental and business systems. Both increase the need for flexibility of a company's IS. On the one hand, researchers and practitioners agree that IS flexibility is a crucial success factor for the viability of business systems, on the other hand, however, earlier research has come to the conclusion that the concept of flexibility is hard to capture. One of the reasons for this is the difficulty to perform flexibility analyses on IS. This paper investigates different definitions of flexibility and methods for the analysis of IS flexibility regarding their precision, completeness and applicability by using a systems and organizations theoretical foundation. The paper's research goal is to propose a certain way of understanding flexibility as well as a method for analyzing it. Furthermore, "rules of thumb" are introduced to support the development of IS strategies with regard to flexibility aspects. Finally, the applicability of the IS artifact is exemplified by a scenario within the Plant Engineering and Construction industry.

Keywords

Flexibility, Method, Systems Theory, Organizational Theory, Plant Engineering and Construction, Strategic IS Planning

1. PROBLEM SETTING

The flexibility of information systems (IS) is a topic of growing importance within IS *analysis* and *design*. An IS can be considered the *information processing part* of an *environmental system* or a *business system*. Hence, an IS consists of (a) information processing tasks (management processes and servicing processes) and (b) IS task actors (business application systems, managers and other staff) [1]. From a *systems theoretical* perspective, an IS exhibits the characteristics of *structure* (system components and the relations between them) and *behavior* (change of system states in time) [2], [3]. The primary goal of a system is its *viability*, i. e. the successful existence over time [4], [5]. Due to the increasing *structural* and *behavioral complexity* of the *environmental system* and the *business system* [6], [7] the *flexibility potential* of an IS gains importance for IS analysis and design – besides the consideration of *functional* (e. g. production, distribution) and other *non-functional requirements* (e. g. data security, Business-IT-Alignment, IT availability). The IS has to *fit*, in particular with regard to flexibility, into both the environmental system and the business system in order to ensure the *viability* of a business system.

In view of these challenges it is essential for IS managers to get a precise and comprehensive understanding of the concept of *flexibility*. Generally, flexibility is understood as the *ability to adapt to changes* [8] or "capability to respond to environmental changes*"* [9]. These definitions capture the concept of flexibility on an abstract level. CONBOY and FITZGERALD postulate that "the body of research on the definition of flexibility indicates such an interpretation is too simple" [10]. Hence, these definitions lead to many semantic (mis)interpretations of the term by IS managers (problems of *ambiguity* and *imprecision*). It is therefore important to analyze the available definitions of flexibility with regard to their *linguistic* characteristics. In addition, due to the increasing structural and behavioral complexity of IS, IS managers have to cope with increasing demands regarding their functional and cognitive capabilities. As a conclusion, IS managers need an appropriate method to support the analysis and design of IS.

From a systems theoretical perspective the problem setting can be interpreted as an *input output system* $S_{IO} \subseteq IN \times OUT$ (cf. Figure 1), with IN representing the input set and OUT the output set (external view). Furthermore, let be $IN \cap OUT = \varnothing$. The relationship between IN and OUT of S_{IO} is unknown (internal view, "black box") [1], [11]. IN represents the characteristics of the given IS and the environmental system, OUT the need for adjustments of the existing IS to close the gap between current and

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

required IS flexibility. In order to support the analysis of flexibility, we investigate the following research questions:

a) *How is the term flexibility currently understood in literature (esp. IS literature)?* In order to answer this research question, we have to identify the dimensions which constitute the term flexibility. A precise understanding of the term flexibility is a prerequisite for a successful "opening of the black box" of S_{IO} (*conceptual perspective*).
b) *How can the analysis of the current degree of flexibility and the required degree of flexibility of an IS be supported?* This research question shall "fill the black box" of S_{IO} with a method which copes with IS flexibility (*methodological perspective*).

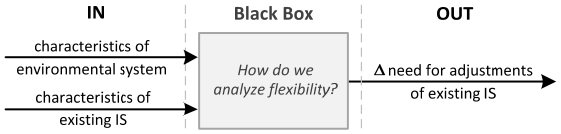


Figure 1: Problem setting interpreted as input output system

The research goal of this paper is to examine the two research questions (a) and (b) and to propose a method to support the analysis of IS flexibility as a conceptual IS artifact. To do so, we use a *systems theoretical* [2], [3] and *organization theoretical* foundation [12]. The paper is organized as follows: section 2 offers an overview of the current understandings of the term flexibility in literature and identifies the research gap. In section 3, we develop a systems theoretical understanding of flexibility. In section 4, a method is introduced representing the IS artifact. The applicability of the IS artifact is shown in section 5, exemplified by a scenario within the Plant Engineering and Construction Industry (PEC industry). Finally, section 6 summarizes the paper, discusses limitations, and gives an outlook on our future research.

The *deductive* research method used in this paper is based on VON GLASERSFELD's *Radical Constructivism* [13] and results in a *method* [14] as a *conceptual IS artifact*. This paper reflects a *design-oriented IS research* and is developed according to the guidelines of HEVNER ET AL. [15], and also [14].

2. RELATED WORK

In order to investigate the research question (a), we perform a literature review according to FETTKE, LIGHT et al. and COOPER et al. [16], [17], [18]. The literature review is aimed at (1st) identifying the *current understanding* of the term flexibility, particularly within IS research, and (2nd) gaining an overview of the *available methods* which support the analysis of flexibility. We investigate literature especially within the period of 1990-2010 rated with "A+" to "B" within the VHB-JOURQUAL 2 ranking (part *IS and information management*). Furthermore, the IS conferences AMCIS, ECIS, ICIS and HICSS of the same period are consulted. Papers are selected if they contain the concept of "flexibility" (with or without post- or pre-fixes) or methods for analyzing flexibility. In addition, typical papers of other domains (e. g. industrial industry, service industry) are consulted in order to provide a broad overview. The *quantitative* results of the literature review are as follows (cf. Table 4, Appendix): 113 papers mention the term "flexibility"; 34 definitions of the term flexibility are further investigated. Certain concepts to define flexibility, like "change" (59 %), "adaption" (35 %) or "environment" (38 %) are used frequently. Others are used rarely, like "structure" (9 %) or "behavior" (6 %).

From a *qualitative* perspective, the analyzed definitions of flexibility can be divided into a main class (a) of *generic* definitions and two subclasses. Subclass (b) provides definitions referring to *objects of flexibility,* whereas subclass (c) provides definitions referring to special *kinds of flexibility* (cf. Figure 2).

The generic definitions in class (a) capture the concept of flexibility on a very abstract level (e. g. "Flexibility is a quality of a system, which allows it to change effectively and recently" [19], "Flexibility of a system is its adaptability to a wide range of possible environments that it may encounter" [8] or "flexibility as an adaptive response to unpredictable situations" [20]).

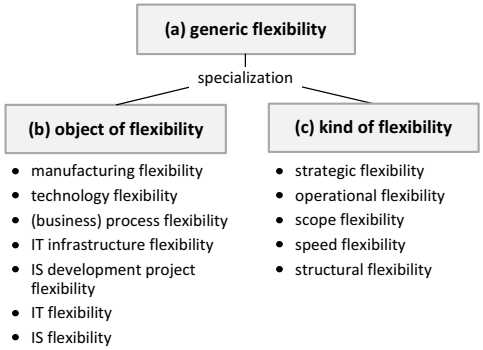


Figure 2: Classification of flexibility definitions

Generic definitions use concepts such as *adaption*, *change*, *reaction* and *variability* to define flexibility (cf. Table 4, Appendix). Due to the multiplicity of different definitions including different meanings (one-to-many relation between concept (syntax) and meaning (semantic), in the narrower sense ambiguity), there is the risk that IS managers misinterpret the flexibility concept, which might lead to lower decision quality. Furthermore, the concept flexibility as well as the concepts used to explain flexibility (e. g. *adaption*, *change* or *response*) exhibit *intrinsic fuzziness* (ordinal scaled values) and *informational fuzziness* (complex concept with the need for decomposition and usage of other metric scaled concepts) [21], [22], [23]. Summing up, the generic definitions may lead to decisions with less decision quality because of *ambiguity* and a *lack of precision*.

The second class of definitions (class (b)) concretizes the generic considerations of flexibility to a (special) object of flexibility like *manufacturing flexibility* [24], *IT infrastructure flexibility* [25], [26], [27] *IT flexibility* [28] or *business process flexibility* [29]. These definitions either use the generic definitions within a specific domain without adding any further details, or enumerate certain aspects of flexibility which are applicable for a special domain. However, these definitions may be incomplete. An IS manager who uses these definitions might miss important aspects while analyzing and designing an IS.

Several authors limit the general understanding to specific kinds of flexibility (class (c)). Examples are: *strategic flexibility* [30], *operational flexibility* [31], *scope flexibility* [32], *speed flexibility* [32] or *structural flexibility* [33].

Apart from the above analysis of different understandings of flexibility, we investigate available methods supporting the analy-

sis of IS flexibility. None of the investigated papers offers such a method. Nevertheless, some authors offer flexibility understandings regarding single aspects:

- DE LEEUW and VOLBERDA [34] implicitly develop a solution to the informational fuzziness mentioned above. They propose an understanding of organizational flexibility and differentiate between flexibility of management and organization tasks. However, a method is not presented.
- SCHOBER and GEBAUER [35] present a formal model to determine the value of IS flexibility. Their approach is based on decision tree analysis (DTA) as well as on real options analysis (ROA) and supports IS managers evaluating IS flexibility.
- Based on a comparison of available IS planning methods, PALANISAMY [36] develops hypotheses representing relationships between user *involvements*, *flexibility* and *IS success*. The author concludes that IS success and organizational flexibility can be achieved by IS flexibility. IS flexibility itself can be influenced by user involvement. This causal chain might support IS managers in preparing IS design decisions.

Other authors (e. g. [37], [38] and [39]) offer abstract methods without explicitly considering flexibility. Summing up, we identify two flexibility problem fields:

- **Flexibility problem field (a) – 'Possible problems due to imprecision, ambiguity and incompleteness of the term flexibility':** The term "flexibility" is often defined in a generic way, resulting in intrinsic and informational fuzziness (imprecision) as well as ambiguity. Furthermore, the completeness of the definitions (at least to a given model or concept) remains uncertain. This raises the imminent danger of focusing on the wrong aspects or leaving out important ones.
- **Flexibility problem field (b) – 'Lack of methodological support':** None of the papers investigated offers a method to support the analysis of flexibility. In contrast to that, we postulate that a support by a method is mandatory in order to reduce the complexity of the task complex 'analysis and design of flexible IS'. Likewise, a stepwise approach simplifies the task execution by the IS manager (decomposition of the task complex). Note that a solution of the problem field (a) is a necessary condition to maximize the decision support of a method.

In the following sections, we investigate possibilities for resolving the flexibility problem fields (a) and (b).

3. DERIVATION OF CONSTRUCTION IDEAS

As a starting point, the flexibility problem fields (a) and (b) are analyzed using an *analysis level hierarchy* to gain *construction ideas* that support the construction of an IS artifact. The analysis level hierarchy consists of four *analysis levels* representing the *level of abstraction* of the concept of flexibility (cf. Figure 3). All hierarchically lower analysis levels *inherit* the characteristics of the hierarchically higher analysis levels (specialization relationship).

In relation to our research questions, the analysis levels represent the following (cf. Figure 3):

- **'Analysis level 3':** This level represents the theoretical foundation, i. e. the theoretical (and also philosophical) assumptions of a flexibility definition. The findings of 'analysis level 3' are the foundation of all further hierarchically lower levels. The theoretical foundation of the generic flexibility definitions remains almost uncertain. This level is missing in the existing flexibility definitions. Only one of the investigated papers explicitly mentions a theoretical foundation. DE LEEUW and VOLBERDA develop their findings from an organizational and control theoretical perspective [34]. Nevertheless, the authors focus on *variety* and *structure*. Hence, *behavior* is missing (cf. section 1 and subsection 3.1).
- **'Analysis level 2':** The theoretical assumptions of 'analysis level 3' are the foundations for 'analysis level 2'. Within this level the term flexibility is elaborated. From the perspective of linguistics, a meaning (semantic aspect) (here: explanation with e. g. adaptability) is assigned to a symbol (syntax aspect) (here: flexibility). The precision and unambiguousness of this relation is an important aspect for further maximizing decision quality (cf. flexibility problem field (a)). The considered object to which the flexibility refers to is still left open.
- **'Analysis level 1':** Within 'analysis level 1' the considered object is limited to a class of objects. The available flexibility definitions focus on special objects of flexibility (e. g. manufacturing flexibility, IT flexibility) and special kinds of flexibility (e. g. speed of change). The completeness of these definitions remains uncertain (cf. flexibility problem field (a)). Furthermore, the missing theoretical foundation of the 'analysis level 3' and existing imprecision and ambiguities located within the 'analysis level 2' lead to additional problems (e. g. incomplete analysis, misinterpretations).
- **'Analysis level 0':** Within this level a concrete object (e. g. an existing enterprise) is located. Existing definitions do not consider this level. We treat this level in section 5 of this paper.

Based on the previous analysis, we postulate that an IS artifact is needed that focuses on 'analysis level 3' and 'analysis level 2'. On the one hand, the IS artifact can be applied to multiple problem classes (high degree of abstraction), on the other hand, the completeness of the investigations are increased simultaneously. The challenge is to tackle these at least partly conflicting goals. In the following, we develop the right hand side of Figure 3.

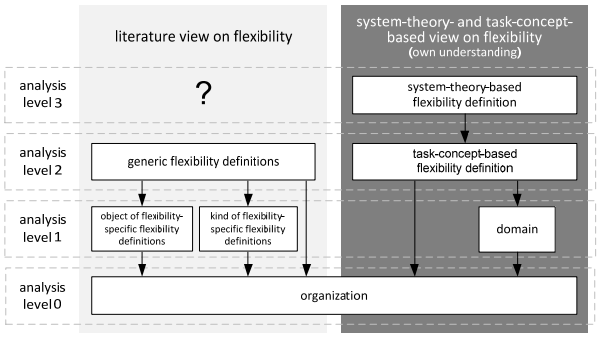


Figure 3: Analysis level hierarchy

3.1 Elaborating 'Analysis Level 3'

The first *construction idea (a)* is the use of the *General System Theory* (GST) by BERTALANFFY as a theoretical foundation [2].

The core concept of GST is the *system*, which has already been introduced in section 1. Furthermore, the methods *decomposition* and *abstraction* simplify the dealing with structure and behavior complexity as origins of the need for IS flexibility. Due to its generality, GST can be applied to multiple problem classes.

In the following, an IS is interpreted formally as a system according to GST. In order to render our argumentation more precise, let $I \neq \emptyset$ be an arbitrary index set and $v = \{V_i: i \in I\}$ a family of non-empty sets. Therefore, the general system S_G is defined as a relation on the sets V_i, $S_G \subseteq \times V_i$, where $i \in I$. The sets V_i within S_G are the system components. The set $R_G \subseteq \{(V_i,V_j):i,j \in I \wedge i \neq j\}$ is the structure of S_G describing the pairwise relationships between system components. The projection $S_G \rightarrow V_i \times V_j$ describes the behavior of the structure element (V_i, V_j) [11], [1].

If we interpret an IS as a general system S_{IS}, the set V_i can be divided into the subsets V_t representing *tasks* and V_{ta} representing *task actors*, where $V_t \cap V_{ta} = \emptyset$. The subsets V_t and V_{ta} are the subsystems S_t and S_{ta} of S_{IS} [2], [3], [11]. By analyzing the structure and behavior of S_{IS}, an IS can be investigated completely.

Both, behavior and structure can be flexible to a certain extent. This extent is either passively generated by IS design or it is explicitly considered at IS' build time. Our approach enables the inclusion of flexibility aspects into IS at build time. Summing up, we introduce an own concept of flexibility on level 3:

Flexibility is the capability of a system to react to or anticipate system or environmental changes by adapting its structure and / or its behavior considering given objectives.

From a systems theoretical perspective, we postulate that this definition is *complete*. The enhancement of the *precision* of the flexibility concept is developed in the following subsection.

3.2 Elaborating 'Analysis Level 2'

The second *construction idea (b)* concerns the usage of KOSIOL's task concept[1] [12] extended by [1]. The task concept's utility has been proven within IS research in numerous cases e. g. [40], [41], [42]. Although already developed in the 70s, the task concept is *robust* with regard to new requirements such as increasing flexibility. An IS consists of information processing tasks (*task level*) and task actors performing these tasks (*task actor level*) (cf. section 1). Every task within the task level is interpreted according to the construction idea. A business process that is part of an IS can be considered as a *network of tasks*. This network of tasks reflects the *structure* of the business process. The sequence of the task executions characterizes the *behavior* of the business process. Modeling languages (e. g. Unified Modeling Language (UML) or Semantic Object Model (SOM)) offer independent models for modeling the structure and behavior of business processes [43]. Ideally, task actors are designed or chosen ("make-or-buy-decision") with respect to the flexibility requirements of the task level (top-down approach). For this reason, we concentrate on the task level.

Every task (e. g. the task 'plant construction') can be characterized from an *external view* and an *internal view* (cf. Figure 4).

[1] Note that the method in section 4 includes KOSIOL's *task analysis*. The *task synthesis* is part of the *design* of an IS strategy, which is not treated in the paper.

The *external view* reveals (1) the *object of the task*, (2) the *goal* and one or more *objectives* of a task as well as (3) one or more *pre-events* and (4) *post-events*. The object of the task includes all *attributes* of the system that are affected by the procedure of the task (e. g. material, building ground). The goal specifies the expected results after the task has been executed (e. g. 'produce plant!'). An objective specifies the degree of goal achievement after the execution of tasks such as *time*, *quality* or *cost constraints* (e. g. 'minimize costs!' and 'maximize security!'). For the current investigations we distinguish between types of goals or objectives (e. g. "produce plant!" or "costs") and instances of the goal or objective (e. g. "*one* plant" or "*minimize* costs!"). Furthermore, one or more pre-events trigger the execution of a task (e. g. 'production order exists'), while one or more post-events are generated after the execution of a task (e. g. 'plant construction is finished'). The *internal view* defines the *procedure* which realizes the goal of the task. The procedure can be further decomposed in *workflow management* and *activities*. The workflow management controls the process of activities, whereas the activities interact with the object of the task by using *sensor* and *actor relationships*. The relationship between workflow management and activities forms a control loop. Furthermore, a task is executed by one or more *task actors* which may be a human for *non-automated* tasks, an application system or machine for *fully automated* tasks or a man-machine system for *partially automated* tasks. The components of the task concept can be analyzed by using both structure and behavior to gain support in analyzing and designing IS.

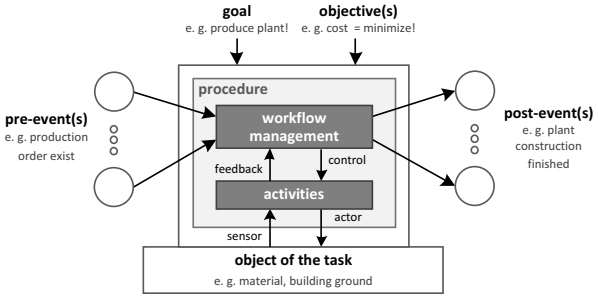


Figure 4: Task concept

The task is analyzed from a structural and behavioral point of view by combining the *construction ideas (a)* and *(b)*, (cf. Table 1).

Table 1: Flexibility of task components from a systems theoretical perspective

task component	flexibility type: structure flexibility (adding or removing of…)	flexibility type: behavior flexibility
pre-event	pre-event(s)	time characteristics of pre-event(s)
goal/ objective	goal and/or objective(s)	variety of goal and/or objective(s)
procedure	parts of the procedure (workflow management or activities)	variety of possible results of the procedure
object of the task	attribute(s)	variety of attribute(s)
post-event	post-event(s)	time characteristics of post-event(s)

The above flexibility types can occur (a) separately, which means that only a *single task component* is flexible ("single task component flexibility"). In addition, also (b) *multiple* task components can be flexible. In this case, *more than one task component* is flexible ("multiple task component flexibility"). In order to simplify the analysis of multiple task component flexibility, some task components can be summarized:

- **'Loose task coupling flexibility'**: Tasks can be *loosely* coupled using *pre-* or *post-events* [44]. Considering $task_A$ and $task_B$, if $task_A$ (e. g. 'production') is loosely coupled with $task_B$ (e. g. 'distribution'), one or more post-events of $task_A$ are identical with one or more pre-events of $task_B$ (post-event $task_A$ (e. g 'product is manufactured') is identical to pre-event of $task_B$). In the case of *structure flexibility*, the type or existence of those events change, whereas in the case of *behavior flexibility* the time of occurrence of the events changes. Two variants of loose task coupling flexibility can be identified: flexibility of one or more pre-events *or* post-events *as well as* one or more pre-events *and* post-events.
- **'Task coordination flexibility'**: Tasks are coordinated *hierarchically* or *non-hierarchically* using goals and objectives. Goals as well as objectives can be *typecasted* (e. g. the goal type is "produce!" or the objective type is "costs") and *instantiated* (e. g. the goal instance is "1,000 pieces" or the objective instance is "minimize!"). Task coordination comprises goals and one or more objectives.

4. INTRODUCTION OF THE METHOD

Based on section 3, a method aimed at supporting IS analysis and design with regard to flexibility is constructed (cf. Figure 5). The method supports the identification of current as well as required IS flexibility. Note that the method *does not* imply a *functional* relationship between IN and OUT of S_{IO} (cf. section 1).

Firstly, business processes are "captured", using a business process modeling language (e. g. BPMN, ARIS, SOM). The business process model abstracts from the complexity of the real system and consists of a network of tasks. *Secondly*, every task within the business process can be examined in detail by determining the components of the task in inside and outside view (cf. Figure 4). Note, that every task (in analogy to a system) can be further *decomposed* into several tasks (in analogy to subsystems). The granularity of the model is sufficient if the IS manager is able to differentiate between *inflexible* and *flexible* tasks.

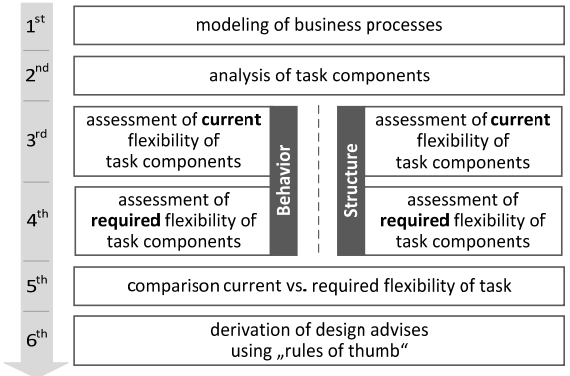


Figure 5: Method as an IS artifact

In the *third* step, the current flexibility of an individual task is assessed. For every component of the task the IS manager has to decide *which* component is *currently* flexible with regard to its *structural* and *behavioral* aspects. Furthermore, the IS manager has to investigate *in what way* the components *are* flexible with regard to *these* aspects. The results can be classified on an ordinal scale using Figure 6 and Table 1. The portfolio consists of the two dimensions *flexibility of inside view* and *flexibility of outside view of the task*. Inside view as well as outside view of the task can be more or less flexible independent of each other (orthogonality of the dimensions). The classification of the task components has to be done twice: once for *structure* and once for *behavior flexibility*. Depending on the classifications of the dimensions, the structure or behavior flexibility is either "low", "medium" or "high".

The results of the structure and behavior flexibility can be aggregated (cf. Figure 7). The fields 1, 2, 4 of the matrix (cf. Figure 6) correspond to "low", 3, 5, 7 to "medium" and 6 as well as 8 to 12 to "high" values of flexibility. The results represent the *current* structure and behavior flexibility of the investigated task.

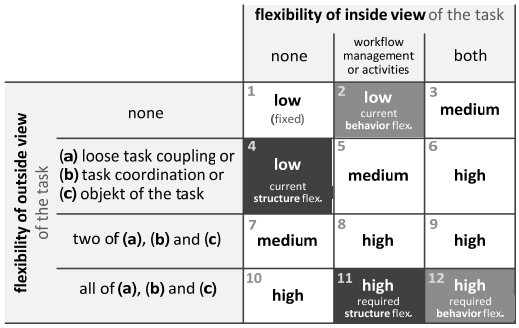


Figure 6: Determination of the structure and behavior flexibility level

In the *fourth* step, the *required* flexibility is determined in analogy to the *third* step described above. Characteristics of the environmental system and the business system (e. g. frequently changing suppliers or business model and plans) constitute the input IN for this assessment. Depending on IN, the IS manager has to answer questions about the *way* in which the IS *should* be flexible. The results are the *current* and the *required* flexibility of a task (output OUT of S_{IO}).

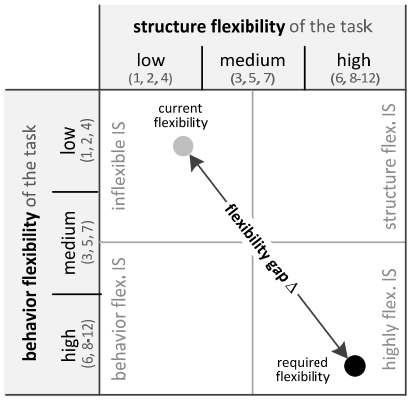


Figure 7: Flexibility level portfolio

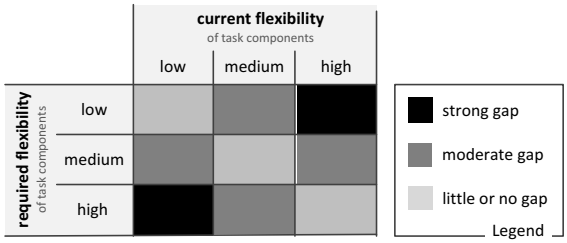


Figure 8: Flexibility gap portfolio

In the *fifth step*, the required structure and behavior flexibility of each task component is evaluated. The results of the "flexibility level portfolio" are interpreted using the 3 × 3 "flexibility gap portfolio" (cf. Figure 8). The location of the examined task components in the portfolio depends on the findings of the *third* and the *fourth* step. Black quadrants suggest *strong gaps*, grey quadrants indicate *moderate gaps*. In particular for task components residing in black quadrants, a flexibility-oriented modification of the IS is recommended. If the levels of current and required flexibility are equal, the organization is in a good position and does not have to make any major changes concerning the analyzed task. If the required flexibility is lower than the current flexibility, the organization may think about lowering the current flexibility rather than providing too much flexibility. On the basis of the steps *one* to *five*, it is possible in step *six* to deduce recommendations for action. These "rules of thumb" might support IS managers in designing the IS strategy and are *complete* with regard to the structure of the task concept.

- **Rule of thumb (a) – 'Flexibility of the goal and objectives of the task':** The inside view and the object of the task is defined depending on its goal *type* (e. g. "construct power plant!") or objective type (e. g. costs). Nevertheless, goals may vary regarding the type (e. g. power plant variants) and instances (e. g. number of power plants) ("variety of goals"). IS managers should be aware of the variety of the goal (types and instances). *Recommendation: Investigate the procedure of the task regarding its current flexibility. Furthermore, a decomposition of the procedure into workflow management and activities (cf. Figure 4) might be necessary. The current flexibility of every decomposition product has to be investigated. If necessary, the procedure has to be changed. It is recommended to evaluate the utility of a Service oriented Architecture (SOA), for example, as a paradigm for IS strategy.*
- **Rule of thumb (b) – 'Flexibility of pre- and post-event(s) of the *task':*** Flexibility requirements often originate from the need to execute tasks in a sequence that differs from the initially planned sequence. Another driving force of flexibility requirements is the uncertainty concerning the time characteristics of pre- and post-events. A flexible IS must be able to handle uncertain points of time of pre- and post-events. *Recommendation: It is recommended to evaluate the utility, for example, of an Event-Driven Architecture (EDA) [45] as a paradigm for IS strategy.*
- **Rule of thumb (c) – 'Flexibility of the procedure of the task':** For both behavior and structure flexibility of the procedure, *building subsystems* of the procedure is recommended. Subsystems can either be determined using an *object-oriented* decomposition [1] of the procedure or a decomposition of the procedure according to the *action principle* [1]. *It is recommended to evaluate, for example, the utility of a SOA as a paradigm for IS strategy in order to support modularized procedures.*
- **Rule of thumb (d) – 'Flexibility of object of the task':** Flexibility within the object of the task means that the attributes of the task can be changed. The IS needs to ensure that all attributes can be accessed by the procedure at the time needed. *Recommendation: It is recommended to evaluate Enterprise Application Integration (EAI), e. g. object integration [1].*

5. APPLICATION OF THE METHOD

The Plant Engineering and Construction industry (PEC industry) is used in this paper to (a) highlight the special requirements of the domain concerning IS and (b) to show the applicability of the IS artifact. For giving evidence, we chose a stepwise research design:

1. We carried out *semi-structured interviews*[2] with experts (n = 5) of the PEC industry from two different companies. As the PEC industry represents an *oligopoly*, it is difficult to increase the sample size. The interviews include questions concerning the understanding and concept of flexibility, the postulated origin of flexibility (process- or IT-driven), the determination of flexibility levels as well as PEC-specific processes.
2. The results (anonymous summaries and models) are discussed and the interview partners committed to them (*consensus theory* as *theory of truth*) in a second round.

The interviews confirmed the study of GALLIERS [46] which revealed that the practitioners' understanding of strategic IS planning differs from the definitions in academic literature. All persons confirmed that the process of evaluating flexibility requirements in practice differs from IS literature (e. g. [37], [47]). Furthermore, none of the interviewees could confirm that the evaluation of flexibility requirements of the business system or the environmental system is part of their strategic IS planning. From this we conclude that the existing concepts for evaluating flexibility requirements of business processes and transforming the results into recommendations for IS strategies are unknown to practitioners or provide insufficient support. Furthermore, the interviewees state that the PEC industry has *high flexibility* requirements compared to other industries. From the point of view of the IS artifact, those requirements can be divided into *structure* and *behavior flexibility requirements*. Behavior flexibility can be considered as the kind of flexibility which can be handled by using *variants,* for example. The structure flexibility of a business system poses bigger challenges to IS managers. The high flexibility requirements basically result from the fact that companies acting in the PEC industry often have to cooperate flexibly with varying partners in order to fulfill the project task. As most of those companies operate globally, it is nearly impossible to cooperate with the same companies in every project, starting with alternating members of the project consortium right up to the countless contractors worldwide that deliver minor products or

[2] We have not included the interview guidelines for reasons of lack of space but will provide them on request.

provide services. The challenges arising from this fact are further increased by the international setting of projects. The mentioned characteristics of the PEC industry lead to difficulties in the IS design. In practice, this often leads to *fragmented* instead of *integrated* IS architectures.

Based on these findings, a scenario is introduced using the *interaction schema* (IAS) as part of SOM [43], [1]. All the following findings are evaluated by practitioners. An IAS represents the *structure* of a business process (cf. Figure 9) and consists of business objects (BO) which encapsulate tasks according to the task concept (cf. Figure 4). The business objects are connected by transactions (TA).

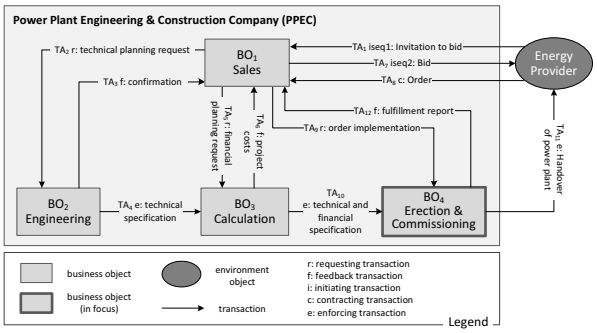


Figure 9. Business process of PPEC industry (structural view)

The IAS shows the interaction of a *power plant engineering and construction company* (PPEC) with an *energy provider* (EP) who purchases a power plant. After invitation to bid (TA_1), the BO_1 *Sales* determines the scope and price of the power plant with the help of BO_2 *Engineering* (TA_2, TA_3) as well as BO_3 *Calculation* (TA_5, TA_6) and submits an offer to the customer (TA_7). *EP* negotiates with *PPEC* through a contracting transaction (TA_8). Afterwards, the BO_1 *Sales* requests the BO_4 *Erection & Commissioning* to implement the order using TA_9 *order implementation*. The BO_4 is hierarchically coordinated by the TA_9 'order implementation' and TA_{12} 'fulfillment report'. Firstly, we use *conventional / available* flexibility definitions (cf. section 2) to derive flexibility requirements on this business process. The following results can be achieved:

1. BO_1 *Sales* must be able *to adapt* to different customer needs.
2. BO_3 *Calculation* must be able to perform *different kinds* of calculations.
3. BO_4 *Erection & Commissioning* must be able to *adapt* to different situations and surrounding conditions on the construction site.

From the interviewees' perspective, only few or no conclusions regarding the design of an IS can be drawn from these generic requirements. In fact, the information gained concerning flexibility requirements is considered too generic. Secondly, in contrast to the previous flexibility assessment, we use the IS artifact developed in section 4. Step (1) of the method is already completed by modeling the business process (cf. Figure 9).

Using the task concept in step *two*, the task 'Erection & Commissioning' can be decomposed into its components (cf. Figure 10).

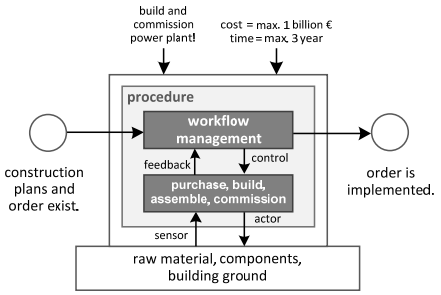


Figure 10. Decomposed task 'Erection & Commissioning'

Based on the components of the task, the current flexibility of the IS is determined (step *three*). According to the flexibility concept, every task component is examined from a behavioral and structural perspective (cf. Table 2).

Table 2. Current IS flexibility

	behavior flexibility	structure flexibility
goal	IS offers flexibility regarding the *size* of the power plant	different *types* of fossil power plants
objective	completion date and/or production cost level	change of objective(s), e. g. higher prioritization of production costs
pre-event	date of request is unknown	not yet supported
procedure	production of variants of a power plant	not yet supported
object of task	different variants of components	not yet supported
post-event	date of completion of task is unknown	not yet supported

Table 3. Required IS flexibility

	behavior flexibility	structure flexibility
goal	***CO_2 emission level of the power plant,*** size of the power plant to be implemented.	***all types of power plants (fossil, solar, wind, nuclear)***
objective	completion date and/or production cost level	change of objective(s), e. g. higher prioritization of the level of production costs
pre-event	date of request is unknown	***engineering and construction services will be offered to other companies, too → task must be available as a service for other companies***
procedure	production of variants of a power plant	***use of new methods and procedures to build power plants***
object of the task	different variants of components	***availability of new construction tools must be considered***
post-event	date of completion of the task is unknown	***post-event must also be provided to external company if the request to provide engineering and commissioning originates from outside PPEC.***

In step *four* of the method, the *required* flexibility is evaluated. Like in the previous step, flexibility requirements are determined from a behavioral and structural perspective. Comparing the current and required flexibility (step *five*), we conclude that there are several flexibility gaps in the current IS. The required flexibility aspects which are not yet provided by the current IS are written in bold, italic letters (cf. Table 3).

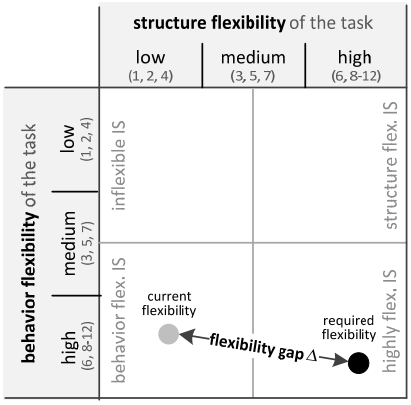


Figure 11: Flexibility level portfolio

We conclude that the flexibility requirements concerning the goals of the task are higher than the flexibility currently provided by the IS. Besides, structure flexibility is not provided by the current IS within pre-events, procedure, object of the task and post-events. The findings of Tables 2 and 3 can be aggregated to the "flexibility level portfolio" (cf. Figure 11) and summarized in a flexibility gap portfolio (cf. Figure 12) in order to provide recommendations.

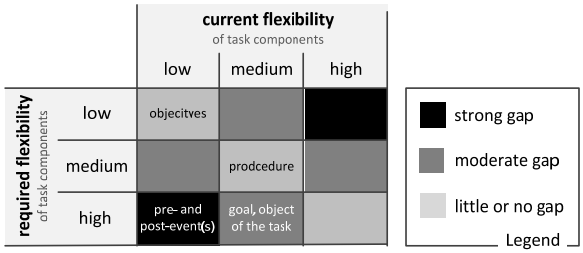


Figure 12: Flexibility gap portfolio 'Engineering & Construction'

Based on the comparison of the current and the required flexibility of the IS, the following (brief) recommendations are made for the design of the IS (step *six*):

- PPEC should implement a SOA. New requirements, like the consideration of CO_2 levels, could be realized by including new services.
- The engineering and construction department would like to offer its services to other power plant engineering and construction companies as well. Therefore, a platform is necessary to interact with external contractors. This could be realized using web technologies.
- The flexibility requirements within the pre- and post-events could be handled by applying event-oriented techniques. By posting events, an EDA abstracts from the origin and destination of messages and thus notably enhances IS' flexibility.

6. CONCLUSION, LIMITATIONS AND FURTHER RESEARCH

This paper provides an analysis of the term flexibility within IS research and develops an own understanding of flexibility based on systems and organizational theory. In addition, we propose a method for determining flexibility gaps ("filling the black box") in order to encourage scientific discourse about IS flexibility. The OUT of S^{IO} (cf. section 1) is determined by offering "rules of thumb". The "rules of thumb" offer input to support the construction of an IS strategy with regard to flexibility. Finally, the *viability* of the method is *deductively* shown using a scenario within the PEC industry. Arguing with POPPER's *Fallibilism* [48], the inherent *hypotheses* of the method – represented by the *construction ideas (a)* and *(b)* – can be *accepted*. Nevertheless, the method still has certain *research limitations*:

- A "semantic gap" exists between the *"rules of thumb"* regarding flexibility and the *construction of IS strategies.* The research goal is to close the gap to increase the utility of the method.
- In order to maximize the support of IS managers, a *software tool* is needed. Ideally, such a tool implements the entire procedure model to reduce time effort and support managing complexity.
- Because of the single scenario, there is the need for one or more extensive case studies which investigate the applicability of the method. Ideally, those scenarios cover different industries. One goal of future research is to increase the evidence of the method.

Although this paper offers only a small contribution towards a deeper understanding of the term *flexibility,* the presented method has the potential to support IS managers in IS analysis and design. However, the research limitations mentioned above are significant. Our research strategy includes further *laboratory experiments* with students and practitioners to identify additional improvements of the method. Furthermore, we carry out an extensive case study within a *medical care center* in order to reduce research limitations.

7. REFERENCES

[1] Ferstl, O. K. and Sinz, E. J. 2008. *Grundlagen der Wirtschaftsinformatik*. 6th Ed., Oldenbourg, München.

[2] Bertalanffy, L. v. 2001. *General system theory*. *Foundations, development, applications*. Braziller, New York.

[3] Klir, G. J. and Valach, M. 1967. *Cybernetic modelling*. Iliffe S.N.T.L., London, Prague.

[4] Ashby, W. R. 1956. *An introduction to cybernetics*. Chapman & Hall, London.

[5] Beer, S. 1995. *Brain of the firm*. The managerial cybernetics of organization. Wiley.

[6] Drucker, P. F. 2007. *The age of discontinuity*. *Guidelines to our changing society*. Transaction Pubs, New Brunswick (USA).

[7] Toffler, A. 1990. *Future shock*. Bantam Books, New York.

[8] Sethi, A. K. and Sethi, S. P. 1990. Flexibility in Manufacturing: A Survey. *The International Journal of Flexible Manufacturing Systems* 2, 4, 289–328.

[9] Lee, G. and Xia, W. 2002. Flexibility of information systems development projects: A conceptual framework. *Americas Conference on Information Systems (AMCIS)*, 1390–1396.

[10] Conboy, K. and Fitzgerald, B. 2004. Toward a Conceptual Framework of Agile Methods: A Study of Agility in Different Disciplines. In *Extreme Programming and Agile Methods - XP/Agile Universe 2004. 4th Conference on Extreme Programming and Agile Methods, Calgary, Canada, August 15-18, 2004*. Proceedings. ACM, 37–44.

[11] Mesarovic, M. D. and Takahara, Y. 1975. *General systems theory. Mathematical foundations*. Academic Press, New York, NY.

[12] Kosiol, E. 1976. *Organisation der Unternehmung*. Gabler, Wiesbaden.

[13] Glasersfeld, E. v. 1997. *Radical constructivism. A way of knowing and learning*. Falmer Press, London.

[14] March, S. T. and Smith, G. F. 1995. Design and natural science research on information technology. *Decision Support Systems* 15, 4, 251–266.

[15] Hevner, A. R., March, S. T., and Park, J. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28, 1, 75–105.

[16] Fettke, P. 2006. State-of-the-Art des State-of-the-Art. *WIRTSCHAFTSINFORMATIK* 48, 4, 257–266.

[17] Light, R. J. and Pillemer, D. B. 1984. *Summing up. The science of reviewing research*. Harvard Univ.Pr., Cambridge, Mass.

[18] Cooper, H. and Hedges, L. V. 1994. Research Synthesis As a Scientific Enterprise. In *The handbook of research synthesis*, H. Cooper and L. V. Hedges, Eds. Russell Sage Foundation, New York, NY, 3–14.

[19] Mandelbaum, M. and Buzacott, J. 1990. Flexibility and decision making. *European Journal of Operational Research* 44, 17–27.

[20] Gupta, Y. P. and Goyal, S. 1989. Flexibility of manufacturing systems: Concepts and measurements. *European Journal of Operational Research* 43, 2, 119–135.

[21] Zadeh, L. A. 1965. Fuzzy Sets. *Information and Control* 8, 338–353.

[22] Dubois, D. and Prade, H. 1980. *Fuzzy sets and systems*. Academic Press, Boston, Mass.

[23] Klir, G. J. and Folger, T. A. 1988. *Fuzzy sets, uncertainty and information*. Prentice Hall, London.

[24] Slack, N. 1983. Flexibility as a Manufacturing Objective. *International Journal of Operations & Production Management* 3, 3, 4–13.

[25] Byrd, T. A. and Turner, D. E. 2000. Measuring the flexibility of information technology infrastructure. *Journal of Management Information Systems* 17, 1, 167–208.

[26] Duncan, N. B. 1995. Capturing Flexibility of Information Technology Infrastructure: A Study of Resource Characteristics and Their Measure. *Journal of Management Information Systems* 12, 2, 37–57.

[27] Kumar, R. L. 2004. A framework for assessing the business value of information technology infrastructures. *Journal of Management Information Systems* 21, 2, 11–32.

[28] Beimborn, D., Franke, J., Wagner, H.-T., and Weitzel, T. 2007. The Impact of Operational Alignment on IT Flexibility - Empirical Evidence from a Survey in the German Banking Industry. In *Proceedings of the 13th Americas Conference on Information Systems*. *AMCIS 2007*. AIS Paper No. 131.

[29] Muenstermann, B., Joachimy, N., and Beimborn, D. 2009. An empirical evaluation of the impact of process standardization on process performance and flexibility. In *Proceedings of the 15th Americas Conference on Information Systems*. *AMCIS 2009*. AIS Paper No. 787.

[30] Hitt, M. A., Keats, B. W., and DeMaiie, S. M. 1998. Navigating in the new competitive landscape: Building strategic flexibility and competitive advantage in the 21st century. *Academy of Management Executive* 12, 4, 22–42.

[31] Kopanaki, E. and Smithson, S. 2003. Examining Organizational Flexibility in an Interorganizational Context. *Americas Conference on Information Systems (AMCIS),* 69, 543–553.

[32] Parthasarthy, R. and Sethi, P. S. 1992. The impact of flexible automation on business strategy and organizational structure. *Academy of Management review* 17, 1, 86–111.

[33] Nelson, K. M. and Ghods, M. 1998. Measuring technology flexibility. *European Journal of Information Systems* 7, 232–240.

[34] Leeuw, A. C. J. de and Volberda, H. W. 1996. On the Concept of Flexibility: A Dual Control Perspective. *Omega* 24, 2, 121–139.

[35] Schober, F. and Gebauer, J. 2009. How Much to Spend on Flexibility? Determining the Value of Information System Flexibility. In *Proceedings of the 15th Americas Conference on Information Systems*. *AMCIS 2009*. AIS Paper No. 193.

[36] Palanisamy, R. 2005. Strategic information systems planning model for building flexibility and success. *Industrial Management & Data Systems* 105, 1, 63–81.

[37] Lederer, A. L. and Salmela, H. 1996. Toward a theory of strategic information systems planning. *The Journal of Strategic Information Systems* 5, 3, 237–253.

[38] Venkatraman, N., Henderson, J. C., and Oldach, S. 1993. Continuous Strategic Alignment: Exploiting Information Technology Capabilities for Competitive Success. *European Management Journal* 11, 2, 139–149.

[39] Langdon, S. 2003. Information systems architecture styles and business interaction patterns: Toward theoretic correspondence. *Information Systems and e-Business Management* 1, 3, 283–304.

[40] Strobel, M. 1998. *Optimierung betrieblicher Systeme auf Basis von Geschäftsprozeßmodellen*. Deutscher Universitäts-Verlag, Wiesbaden.

[41] Schlitt, M. 2004. *Grundlagen und Methoden für Interpretation und Konstruktion von Informationssystemmodellen*. Deutscher Universitäts-Verlag, Wiesbaden.

[42] Suchan, C. 2009. Design of causal coupling patterns supporting causal modeling and flow modeling in the system dynamics methodology. In *Proceedings of the 15th Americas Conference on Information Systems*. *AMCIS 2009*. AIS Paper No. 742.

[43] Ferstl, O. K. and Sinz, E. J. 2006. Modeling of Business Systems Using SOM. In *Handbook on architectures of information systems*, P. Bernus, Ed. International handbooks on information systems. Springer, Berlin, 347–367.

[44] Eckert, S., Suchan, C., Ferstl, O. K., and Schissler, M. 2005. Integration von Anwendungssystemen für die Materialwirtschaft – Anwendung einer Entwicklungsmethodik

im Bereich des Kraftwerkbaus. In *Wirtschaftsinformatik 2005. eEconomy, eGovernment, eSociety*. Physica-Verl., Heidelberg, 667–686.

[45] Taylor, H., Yochem, A., and Phillips, L. 2009. *Event-Driven Architecture. How SOA enables the real-time enterprise*. Addison-Wesley, Upper Saddle River, New Jersey.

[46] Galliers, R. D. 1993. IT Strategies: beyond competitive advantage. *Journal of Strategic Information Systems* 2, 4, 283–291.

[47] Neu, P. 1991. *Strategische Informationssystem-Planung. Konzept und Instrumente*. Springer, Berlin.

[48] Popper, K. 2007. *Logik der Forschung*. Akademie-Verlag, Berlin.

[49] Bahrami, H. 1992. The Emerging Flexible Organization: Perspectives from Silicon Valley. *California Management Review* 34, 4, 33–52.

[50] Groote de, X. 1994. The Flexibility of production processes: A conceptual Framework. *Management Science* 40, 7, 933–945.

[51] Evans, J. S. 1991. Strategic flexibility for high technology maneuvers: A conceptual framework. *Journal of Management Studies* 28, 1, 69–89.

[52] Frost, R. S. 1999. *The growing imperative to adopt "flexibility" as an American principle of war*. Strategic Studies Institute U.S. Army War College, Carlisle Barracks Pa.

[53] Knoll, K. and Jarvenpaa, S. L. 1994. Information technology alignment or 'fit' in highly turbulent environment: The concept of flexibility. In *Proceedings of the 1994 ACM SIGCPR conference*. ACM Press, New York, 1–14.

[54] Lucas, H. C. and Olson Margrethe. 1994. The Impact of Information Technology on Organizational Flexibility. *Journal of Organizational Computing* 4, 2, 155–175.

[55] Monteiro, L. and Macdonald, S. 1996. From efficiency to flexibility: The strategic use of information in airline industry. *Journal of Strategic Information Systems* 5, 3, 169–188.

[56] Patnayakuni, R. and Patnayakuni, N. 2003. Organizational Flexibility and Inventory Flow Integration. In *Americas Conference on Information Systems (AMCIS),* AIS Paper No. 72, 573–580.

[57] Whitworth, B. and Zaic, M. 2003. The WOSP model: Balanced Information system design and evaluation. *Communications of the Association for Information Systems* 12, 258–282.

[58] Shimizu, K. and Hitt, M. A. 2004. Strategic flexibility: Organizational preparedness to reverse ineffective strategic decisions. *Academy of Management Executive* 18, 4, 44–59.

[59] Verganti, R. 1999. Planned Flexibility: Linking Anticipation and Reaction in Product Development Projects. *Journal of Product Innovation Management* 16, 4, 363–376.

[60] Jacome, L. 2007. Evaluating Information Systems Flexibility: a Research Approach to Build a Framework. In *Proceedings of the 13th Americas Conference on Information Systems*. *AMCIS 2007*. AIS Paper No. 202.

[61] Nelson, K. M., Nelson, H. J., and Ghods, M. 1997. Technology Flexibility: Conceptualization, Validation, and Measurement. In Proceedings of the Thirtieth Hawaii International Conference on System Sciences Wailea. IEEE Computer Soc. Press, Los Alamitos, Calif.

8. APPENDIX

Table 4: Mentioned aspects in flexibility definitions

	Origin	kind of flexibility	terms mentioned in literature regarding flexibility (not (!) the way we understand it)									
			reaction	adaptation	variability	(business) processes	un-plannedness	behavior	structure	objectives	change	environment
generic flexibility	[49]	flexibility	x	x			x				x	
	[50]	flexibility									x	x
	[34]	flexibility			x					x		x
	[51]	flexibility		x			x			x		
	[52]	flexibility	x	x							x	
	[20]	flexibility					x					x
	[53]	flexibility									x	x
	[31]	flexibility	x									x
	[31]	flexibility	x				x					x
	[9]	flexibility	x								x	x
	[54]	flexibility					x			x		
	[19]	flexibility	x							x	x	
	[55]	flexibility		x								x
	[56]	flexibility			x	x				x	x	
	[8]	flexibility		x			x				x	x
	[57]	flexibility	x	x			x					x
kind of flexibility	[31]	operational flexibility			x				x		x	
	[32]	scope flexibility			x							
	[32]	speed flexibility								x	x	
	[30]	strategic flexibility	x				x			x	x	x
	[31]	strategic flexibility			x			x	x			
	[58]	strategic flexibility	x							x	x	x
	[31]	structural flexibility		x					x			
	[33]	structural flexibility		x		x					x	
	[59]	structural flexibility					x					
object of flexibility	[29]	business process flexibility	x	x	x					x	x	x
	[33]	process flexibility				x					x	
	[60]	IS flexibility		x		x					x	
	[39]	IS flexibility		x		x					x	
	[9]	ISD project flexibility								x	x	
	[28]	IT flexibility	x			x				x	x	
	[25]	IT infrastructure flexibility			x					x		
	[24]	manufacturing flexibility						x				
	[61]	technology flexibility		x		x				x	x	

How Agile Are You Thinking? – An Exploratory Case Study

Roy Wendler
Technische Universität Dresden
Helmholtzstraße 10
01069 Dresden
+49 (0) 351 463 34005
roy.wendler@tu-dresden.de

André Gräning
Technische Universität Dresden
Helmholtzstraße 10
01069 Dresden
+49 (0) 351 463 33514
andre.graening@tu-dresden.de

ABSTRACT
Agile software development methods reduce project costs and development time by simultaneously enhancing quality. But despite these advantages, agile principles are rarely adopted by the whole organization. In order to gain a deeper understanding of this issue, we conducted an initial exploratory qualitative case study in one medium-sized company. The goal of this study was to find out whether this company is "thinking" agile or traditional. Although we discovered a tendency towards an agile way of thinking, we identified several factors where the way of thinking remained traditional among management as well as employees. Our study reveals that cost-related aspects, a lack of self-responsibility, uncertainty with customer interaction and the urge for comprehensive documentation are obstacles to adopting agile methods beyond the development team. Hence, the results of our study provide useful implications for research and practice by identifying critical problem domains when implementing agile methods at the organizational level.

Keywords
Agile Software Development Methods, Agile Thinking, Exploratory Case Study

1. INTRODUCTION
Agile software development methods and their respective project management methods are a recurring and controversial issue in science and practice. Major benefits include quality improvements because of better teamwork and frequent customer interaction. Furthermore, customer feedback avoids misunderstandings and continuous requirements control, testing and releasing leads to continuous approval of the software by the customer. Many studies have shown that agile methods may reduce project costs and development time by simultaneously enhancing quality (see for example [6, 9, 14, 35]). Nevertheless, there are critical voices, too. Many constraints may be hindering the effectiveness of agile methods [7, 26, 27, 29]. Also, many organizations adopt agile methods, at least partially, without understanding the concept of agility itself [24].

Based on expert interviews with employees of the one software developing and consulting organization, called *SoDeCo* (a pseudonym for the purpose of anonymity), we found that the benefits of agile methods are known but still not used in software development projects at SoDeCo. This situation led to an initial agile project based on Scrum to develop a new e-commerce system for a global company. The use of Scrum was triggered by a few project managers. The experience with this project confirmed the advantages of agile methods. Self-organization was favored by the team and a running prototype was available much earlier as compared to projects using traditional methods. The project members stated that they were able to faster implement new and detailed requirements without losing sight of the overall project goal. Furthermore, the daily Scrum-meetings helped them to better track and predict their progress. Despite of the experienced benefits, the team was not able to convince other project teams to use Scrum and therefore Scrum was not or only partially adopted at SoDeCo.

Interestingly, this turned out to be a phenomenon that is observed quite often as confirmed by Abrahamsson et al. [2] and Ågerfalk et al. [3]. Therefore, questions about the causes for these situations arose. Is a specific organizational culture necessary or is the way of individual and organizational thinking crucial for the use of agile methods? Our assumption is that the ways of thinking (i.e. the attitudes, opinions, knowledge ...) when referring to agility and agile methods are different among several groups within an organization. Potential research needs to include barriers or success factors for the adoption of agile methods at the organizational level. Also, specific needs, dependencies or interactions of organizations, individuals and other departments beyond the development team have to be taken into account [2, 3].

Based on our preliminary findings, the aim of our research is to identify “how agile” the staff of SoDeCo is thinking. Especially, we are going to investigate what characteristics in software development projects lead to a more or less agile way of thinking among project members, project managers and other stakeholders within the organization. Likewise, we want to identify potential obstacles, causing employees or decision makers to show resistance against agile practices and therefore hindering their effective use and further adoption.

To investigate this issue, we conducted an exploratory single case study within SoDeCo to answer the following research questions:

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

(Q 1) How agile is the way of thinking of the organization and its individuals?
(Q 2) Are there differences in the way of thinking between several organizational roles and how do they manifest themselves?
(Q 3) What influencing factors may hinder the adoption of agile software development methods and how can they be overcome?

Extensive research on agile methods was conducted in recent years [2, 3]. Systematic summaries of existing studies are given in [14] and [28]. Nevertheless, there is only little research available that examines the specific cultural and organizational factors negatively influencing the adoption of agile software development methods [26]. In addition there are no studies investigating differences in adoption factors between several roles or departments within an organization.

The remainder of this paper is structured as follows. In the second section, we motivate our research by a literature review and deduce the concept of "Agile Thinking". The third section describes our exploratory case study design and in the following fourth section, the results are qualitatively analyzed and interpreted. The paper closes with conclusions and implications in the fifth section. Furthermore, limitations of the study are addressed and an outlook is given.

2. RESEARCH BACKGROUND

In reaction to the inadequateness of traditional software development methods for many projects with regard to time and cost constraints, bureaucracy of documentation and increasing changes in the business environment, new and more flexible development approaches were developed [1]. In fact, "lightweight" thinking is nothing new. Examples for iterative, incremental or evolutionary development practices can be tracked back till the 1970's. Unfortunately, these practices were not considered to be seriously adopted until the late 1990's [23]. The term "agile" referring to software development became well known through the *Agile Manifesto* which was formulated in 2001 by a group of supporters of alternative development approaches [18]. Today, some of the best known agile approaches among others are Scrum, Extreme Programming, Feature Driven Development and the Crystal Family. A summary of these and other approaches can be found in [7], [14], [19] and [24].

The Agile Manifesto states four key values and twelve principles that underlie all agile software development approaches [5, 11]. Due to the lack of an acknowledged definition until today, the four key values may serve as an explanation while stating the basic characteristics of all agile approaches. An overview of various definitions can be found in [1] and [17]. Many of them refer to the Agile Manifesto as well, which therefore represents the basis for this work, too. The key values of the Agile Manifesto are (1) the concentration on individuals and interactions more than on processes and tools, (2) the delivery of working software instead of focusing on comprehensive documentation, (3) regular customer collaboration over contract negotiation and (4) responding to changes instead of purely following plans [5].

However, despite the benefits of agile development approaches, there are limitations as well. Bleek and Wolf [7] provide a number of indicators against the usage of agile software development methods. Among them are cultural aspects, missing customer commitment, mandatory processes or the fear of responsibility [7]. It becomes evident that a lot of constraints may hinder the effective and efficient use of agile development methods. Nevertheless, more and more organizations report having adopted agile approaches [24, 31]. But with an increasing awareness, the confusion about agility, its meaning and the optimal usage of agile approaches is increasing, too. Therefore, many researchers call upon more studies within this field [2, 3, 14, 15, 24].

In order to be able to investigate the agility of an organization, one has to understand the meaning of agility. Although the Agile Manifesto describes agility in terms of software development, the concept of agility is much older. Agility became well known in the business literature around the 1990's [12]. It was mainly utilized in the fields of management, manufacturing and organizational behavior and emerged out of the concepts *flexibility* and *leanness* [12, 17].

Sharifi and Zhang [34] identified four capabilities an organization has to generate to be agile. These are *Responsiveness*, *Competency*, *Flexibility*, and *Speed*. The authors underline that these capabilities ensure appropriate reactions to changes in the environment [34]. The handling of change as a fundamental prerequisite for agility is confirmed by Conboy [12], who named *Creation* of change, *Proaction* in advance of change, *Reaction* to change, and *Learning* from change as components of agility [12].

These general characteristics of agility can be found in the Agile Manifesto [5], too. The handling of change can be seen in the values of customer collaboration (value 3) and response to change (value 4), thereby competency and responsiveness are reflected in the concentration on individuals (value 1) and flexibility and speed lead to the fast delivery of working software (value 2).

Although the Agile Manifesto covers many aspects of agility, the use of agile methods does not automatically lead to an agile organization. As stated by Abrahamsson et al. [2] and Mangalaraj et al. [26] the success of agile software development projects often sticks to the team level. Mostly it is difficult, if not impossible, to implement agile principles beyond single development teams, because of many constraints and dependencies with regard to the rest of the organization [2, 26].

We assume that the causes for this phenomenon are attitudes and ways of thinking of individuals. Given the fact that methods contain not only isolated practices, but are bound together by a set of values and goals lying behind the principles of the method [4, 12], it is reasonable that these values and goals have to be coexistent in people's minds when working with these methods. This necessity becomes particularly evident when looking at agile software development methods. As stated above, the core of the Agile Manifesto consists of four "values" which have to be shared by every user. It has to be assumed that agile methods may only be adopted at an organizational level when the way of thinking of the whole staff of an organization is congruent with agile values. This assumption is supported by the conceptual framework of agility by Sharifi and Zhang [34]. Besides technology and innovation they define *people* and *organization* as main supporting areas of agile capabilities. Furthermore, Sambamurphy et al. [32] distinguish between operational, *customer* and *partnership* agility [32]. While operational agility is directed to processes, partnership and customer agility deal with relationships

to customers and partners [32, 33]. For the latter two it is not sufficient to be agile in one or two projects only. To be fully agile means that software developers and project staff as well as managers, sales and distribution staff, and all other departments have to share the above stated agile values and principles.

Here only those factors are of interests that have an influence on the selection of the development method and on the realization and the success of the project itself. It is obvious that a more or less agile way of thinking in factors critically affecting the named areas may play an important role for adopting agile methods. Therefore, we reviewed the literature for potential success factors in agile software development projects as a starting point for our study. A comprehensive overview of existing empirical studies on agile development can be found in Dybå and Dingsøyr [14]. We also focused on the studies of Chow and Cao [10] and Misra et al. [28], who identified several success factors for the adoption of agile development methods that fit the purpose of our research. Furthermore, we initiated discussions and conducted interviews with software development project experts of SoDeCo to identify additional factors. The whole set of identified factors is called *Influencing factors* in the following. Table 1 lists all influencing factors used. The factors taken from literature are explained in the respective studies [10, 28]. A short explanation of the influencing factors identified through the expert interviews follows:

- Self-responsibility is considered important for adopting agile approaches including personal characteristics like autonomous execution of tasks, initiative and self-organization.
- Distribution of power refers to issues like power to direct a company or organization. It is assumed that a potential loss of power will make people hold back information and therefore threaten the effectiveness of agile approaches.
- Pricing models focuses on the question, if alternative and flexible pricing models for agile approaches are understood by customers and employees.
- Documentation covers the aspect that detailed and comprehensive project documentation may be seen as a quality indicator in some environments and therefore hinders the acceptance of agile approaches.

Table 1. Identified influencing factors

Framework category	Influencing factor	Source
Individual factors	IF1: Self-responsibility	Expert interviews
	IF2: Competencies	[10, 28]
Team factors	IF3: Communication	[10, 28]
	IF4: Decision processes	[10, 28]
	IF5: Team size	[10, 28]
	IF6: Distribution of power	Expert interviews
Environmental factors	IF7: Pricing models	Expert interviews
	IF8: Customer satisfaction	[10, 28]
	IF9: Customer collaboration	[10, 28]
	IF10: Documentation	Expert interviews

Furthermore, Mangalaraj et al. [26] developed a framework for the acceptance of software process improvements. According to this framework acceptance depends on individual, team, technology, task, and environmental factors [26]. Along our research aim, the framework gives us additional categories to investigate the influence on the way of thinking.

In the context of our research, we exclude technology and task factors, dealing with technical characteristics and supported task types of the development methods. We rather concentrate on the *individual*, *team* and *environmental* factors that are able to cover Agile Thinking. Individual factors contain personal attitudes, knowledge and abilities; team factors contain the social culture within a development team, high status individuals' opinions and majority opinions and attitudes; and environmental factors consider mainly customers and their influence on the use of development methods [26]. As table 1 shows all identified influencing factors are assigned to the categories of [26].

As illustrated in figure 1, the accumulation of the different influencing factors, the underlying categories in relationship to the agile values and the identification of the principles of agility of individual persons at an organizational and individual level is what we call *"Agile Thinking"*.

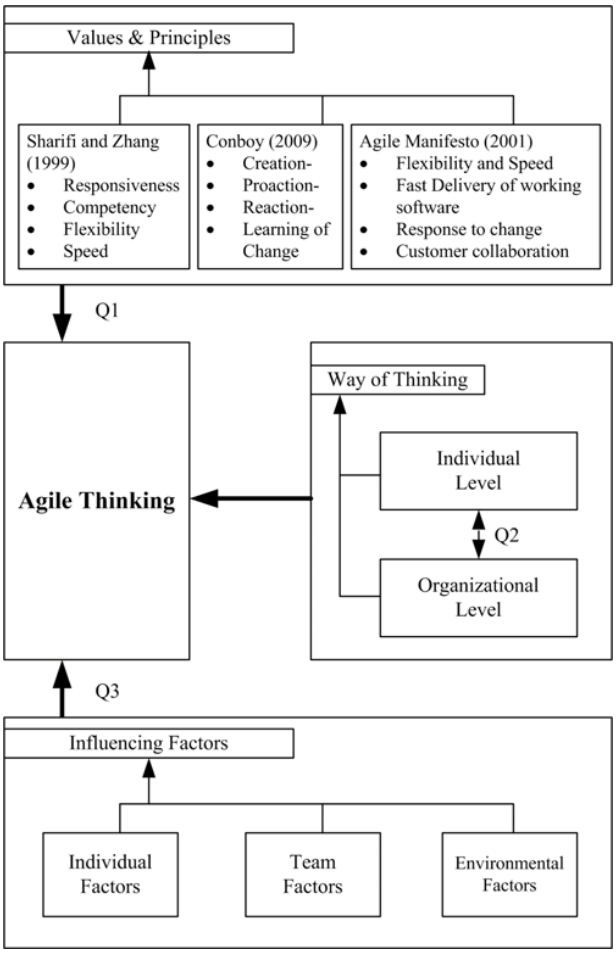


Figure 1. Agile Thinking

Agile Thinking comprises the attitudes and opinions of individuals with regard to agility within their everyday business. It describes the willingness of individuals to act agile – even if this means to change traditional habits and approaches. Ganguly et al. [17] describe an agile enterprise as being able to "adjust to any unexpected or sudden changes in the environment both rapidly and efficiently". Therefore, the attitudes regarding agility of all individuals of an organization strongly affect the agility of the whole enterprise. The idea behind this concept is that individuals covering key positions within an organization may have a great influence in supporting or hindering the adoption of agile methods. The same effect may occur when the majority of the staff of one or more departments demonstrates a more or less agile way of thinking.

3. CASE STUDY DESIGN AND CASE DESCRIPTION

According to Gable [16] and Yin [36] the case study approach is suitable to gain a deep qualitative understanding of problems being investigated [16, 36]. Especially in the rapidly changing field of Information Systems, case studies offer the opportunity to deliver valuable insights into organizational behavior [16]. Furthermore, we intend to identify factors potentially influencing the use and adoption of agile methods. This exploratory character of our research strengthens the applicability of a case study approach, too.

In order to gain a comprehensive overview about and first insights into the phenomenon of Agile Thinking, we chose a single and exploratory case study design. To demarcate the case study, the unit of analysis has to be clearly defined [36]. Therefore, the unit of analysis of our study is the overall software developing and consulting organization (SoDeCo), especially the way of thinking regarding agile methods as demonstrated by individuals and groups within this organization.

Having this in mind we have to make sure to gather appropriate and sufficient data. Although case studies are qualitative research methods, the data collected and analyzed may be both qualitative and quantitative [8, 16, 36]. In fact, combining different types of data to compensate strengths and weaknesses of the single ones is often favored as it provides a fuller picture of the underlying phenomena investigated [8, 20, 21, 36].

To take advantage of the use of qualitative and quantitative data, we decided to use expert interviews and discussions as well as a web-based survey for data collection. Initial ideas for potential factors influencing the way of thinking were gathered qualitatively by conducting interviews with software development and project management experts at SoDeCo. The results were used to develop a set of questions to investigate the way of thinking within the overall organization, including as many departments and functions of SoDeCo as possible. According to Darke et al. [13], time consuming interviews should only be conducted when desired data cannot be obtained in any other way [13]. Because of that, we decided to utilize a web-based questionnaire, which can be considered a useful data collection method within case studies as well [16, 36].

To obtain answers about the way of thinking along the identified influencing factors, we developed a set of statements (items) for every single one of them that represent the characteristics of agile approaches (see appendix). To give an example, one statement for the factor self-responsibility is "Team internal self-organization prevents slack of team members". All statements are formulated in such a way that an agreement implies either an agile or a traditional way of thinking on the part of the respondent. For the example mentioned, an agreement implies the assumption that project members coordinate and execute their tasks independently and consequently avoid slack that could emerge while using inflexible and superior project plans. For assessing the statements we used a five-point-Likert-scale. Respondents had to answer every statement with "agree", "rather agree", "neither/nor", "rather disagree" or "disagree". Additionally, the respondents were given the opportunity to provide further comments to every question in free text forms. Therefore, the results are comparable to those that would have been collected via standardized interviews. To ensure validity and readability, three members of the company and two additional professionals were asked to review the statements and to check their face validity. The feedback was included before starting the study. The invitation to answer the web-based questionnaire was sent to the staff of SoDeCo via an internal email distribution list.

After three weeks, a total of 58 people responded by answering the online questionnaire. After reviewing the raw data and checking the consistency of the 58 responses, 9 had to be excluded, because answering the questionnaire was terminated before completion. Finally, 49 responses were used for further analysis and interpretation. The data collected was then analyzed and interpreted in a qualitative way. To validate and verify the results of the interpretation, we again conducted expert discussions afterwards.

SoDeCo undertakes individual software projects in manufacturing and commerce for large and medium sized companies. The company consists of three divisions and operates in Europe and the US. Mostly, they use traditional software development methods, such as the Waterfall-Model, the V-Model and some customized versions. Nevertheless, some of the staff has experiences with agile software development methods, too. Although the benefits of agile methods were recognized during the initial Scrum-project mentioned in the introduction, the team members found it difficult to convince substantial parts of the rest of the organization to use agile methods in future projects. This problem served as an entry point for our study. We assumed that parts of the organization had attitudes, which hindered the adoption of agile methods. Thus, we intended to further investigate "how agile the organization is thinking". Insights into the characteristics of the organization are provided in table 2.

The majority of respondents to our questionnaire already had experience in software projects, with nearly 80% of them having participated in software projects for more than 12 months (see table 3). As the level of experience of the participants is quite high, we assume the data set to be suited to serve the purpose of our study.

Table 2. Profile of investigated company

Manpower	ca. 180 employees, 3 divisions
Location	Headquarter Germany, 4 locations world wide
Customers	Large and medium sized enterprises world wide
Used traditional methods	Waterfall, V-Model (main part of projects)
Used agile methods	Scrum (single project); Scrum-like fragments in mixed methods (some projects)
Company background	Specialized in individual software development. Planning, implementation and maintenance of developed software. Specialized in IT consulting for semi conductor and energy markets.

Table 3. Experience in software development projects

Item	Value	Number of respondents	Percentage (%)
Experience in software projects	Yes	42	85,7
	No	7	14,9
Duration of participation in software projects	< 12 months	3	6,1
	12-60 months	17	34,7
	60-120 months	11	22,4
	>120 months	11	22,4

4. DATA ANALYSIS AND RESULTS

First, the respondents had to clarify their level of experience with project management and software development methods we listed in the questionnaire. It contained eleven agile methods and ten traditional methods. The results show that the traditional methods are more likely to be known and used than the agile ones. Surprisingly, only very few respondents consider themselves as professionals (5% for agile methods; 8% for traditional methods), although the study took place in a software developing organization. In detail, at least one traditional and agile method is known by all but one. Furthermore, about 70% have used at least one traditional method and about 56% experienced at least one agile method.

Second, related to Q2 we divided the data set for further qualitative analysis into two independent subgroups:

- *decision makers*, including board members, managers and group leaders and
- *employees*, including software developers, consultants and others.

Third, all statements were grouped according to the identified influencing factors (see table 1). The analysis of every statement included the calculation of mean values and standard deviations and a qualitative interpretation of them. Thereby the mean value represents the average attitude throughout the analyzed groups and the standard deviation is an indicator for the concentration of this attitude. A mean value of 3 shows, that the considered group is indifferent as to agile or traditional thinking, whereas mean values of 4 to 5 reveal Agile Thinking and mean values of 1 to 2 represent traditional thinking. A small standard deviation supports this result, whereas a high standard deviation reveals that the group is either divided into different fractions or that the attitudes are equally distributed among the respondents. In addition, we examined the distribution of the answers for every statement. For example, a skewed distribution is an indicator for tendencies. After the preparation of the data set, the qualitative interpretation of the results was done independently by the authors. All disagreements in interpretation and misunderstandings about the data set were solved in several discussions.

A summary of the results is given in figure 2. It shows the mean values per influencing factor and represents the way of thinking related to decision makers and employees. Surprisingly, a comparison of the two subgroups revealed no or only minor differences between decision makers and employees. However, the distribution of the ways of thinking over the influencing factors is varying.

Interestingly, IF6 is the only factor, where a small difference between employees and decision makers is noticeable. Looking at the single items, it seems that employees fear a loss of power after sharing implicit information more than decision makers do. This might be due to the fact that decision makers think more in terms of the overall organization than employees do. However employees' reluctance to share knowledge because of a fear to lose power may severely threaten agile projects – probably more than decision makers' lacking willingness would do.

The only factor revealing a clear agile attitude is IF3 (Communication). This shows that decision makers as well as employees prefer non bureaucratic and flexible communication processes and hence share an Agile Thinking regarding this issue. Furthermore, the respondents of both subgroups show a tendency towards agile over traditional thinking related to the different influencing factors. Therefore we examined the single statements for every factor to gain further insights. This was done by the following procedure:

Every statement was either classified agile, traditional or neutral according to its mean value for the total sample as well as for the two subgroups. Afterwards, we counted the classified statements and calculated the percentages per factor. Additionally, the percentages of respondents were calculated according to agile, traditional or neutral answers. Table 4 summarizes the results of this procedure, structured into two columns based on the percentage of statements and the percentage of respondents per factor. Furthermore, the results for the two subgroups are shown accordingly. As a result, every cell delivers insights into the way of thinking. Noticeable deviations are subsequently qualitatively analyzed in detail and interpreted.

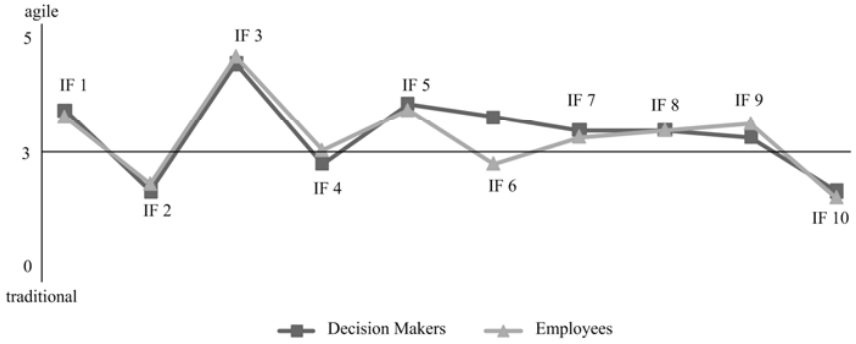


Figure 2. Mean values per influencing factor of the whole sample

Similar to figure 2, the results in table 4 hardly differ between decision makers and employees. Minor differences between the two subgroups are noticeable with influencing factors 2 (Competencies), 5 (Team size), 6 (Distribution of power), 8 (Customer satisfaction), and 9 (Customer collaboration). The differences mainly result from a higher number of neutral thinking respondents in the employee group. With none of the statements a substantial difference between agile thinking in one subgroup and traditional thinking in the other is ascertainable. Despite this observation, interesting results can be derived by investigating the content of the single statements where differences were observed.

There were two influencing factors showing a total agile way of thinking along all statements. These were IF3 (Communication) and IF5 (Team size). Respondents state that regular and intensive communication between team members improves quality and that smaller teams are more effective and flexible with respect to changes under time and cost constraints. Looking at the percentages of the respondents, employees seem to be more indifferent regarding team size (IF5) as opposed to decision makers. This could be an indicator that the actual number of team members is less important as long as communication processes are flexible. Closely connected to this issue is IF2 (Competencies). It addresses the question, whether team members should be generalists or specialists in order to deliver high quality under time and cost constraints. Although decision makers seem to support specialists, a concentration on the neutrally answered statements on employee side can be observed. This means that the respondents prefer neither of the two alternatives, but they state that a software developing team should always contain specialists and generalists to be successful. The number of neutral answers by all respondents supports this issue. This does not hinder the adoption of agile approaches directly, but it indicates a slight tendency to combine agile and traditional approaches.

Differentiated results have to be emphasized for the influencing factor IF1 (Self-responsibility). Most of the statements belonging to this factor suggest an agile way of thinking. They deal with the issues quality, self-organization, motivation and flat hierarchies. In contrast, three statements showed a different way of thinking. It is striking that all of these addressed leadership issues. Respondents showed a tendency to traditional thinking while stating that a successful project needs leadership by superiors, the project management should be supervisory to the project team and important decisions have to be confirmed by management. This leads to the conclusion that the respondents know that characteristics of agile approaches will increase quality and motivation, but they still urge for someone superior, who will be responsible for the project's success or failure. This clearly is an obstacle to adopting agile approaches and is supported by the answers to IF4 (Decision Processes). Here, the respondents are very indifferent – showing an agile way of thinking referring to fast and non bureaucratic decisions on the one hand and asking for confirmation by management on the other hand.

Influencing factors mainly considering customer issues do not deliver a clear picture. The statements of IF8 (Customer satisfaction) and IF9 (Customer collaboration) reveal either very controversial answers or a concentration on the neutral position, although a slight tendency to an agile way of thinking can be observed. In contrast, single statements dealing with cost-related aspects reveal a more traditional way of thinking by decision makers. Furthermore, there are differing opinions on the statements about additional or changing requirements. Some state that changing requirements lead to rising costs and diminish the quality of the products. This fact is clearly hindering the adoption of agile approaches, because changing requirements are one of their key principles. IF10 (Documentation) appears to be a potential customer-related obstacle, too. Nearly all respondents show a traditional or neutral way of thinking, stating that comprehensive documentation is required to prove quality and professionalism. Additionally, IF6 (Distribution of power) may support the assumption of customer-related problems. When people withhold information due to the fear of losing bargaining power, the effective adoption of agile methods is critical.

The problems identified above that are caused by cost-related issues are confirmed by IF7 (Pricing models). There is an agile attitude regarding a preference for flexible pricing mechanisms to improve the quality, because of more flexibility with features. But despite this, there are again very controversial answers for statements dealing with the ability to keep control over costs with flexible pricing mechanisms and addressing whether customers will accept such mechanisms. Similar to IF6, an explanation for this behavior could be the aspiration for pricing sovereignty by the software developer and this again hinders the adoption of agile methods at an organizational level.

Table 4. Way of thinking by statements and respondents per influencing factor

Framework Category	Influencing factor (IF)		Way of thinking (in % of statements per factor)			Way of thinking (in % of respondents per factor)		
			Total	Decision makers	Emplo-yees	Total	Decision makers	Emplo-yees
Individual factors	IF1: Self-responsibility	Agile	67,00	67,00	67,00	61,22	66,00	60,00
		Traditional	11,50	11,50	11,50	18,37	16,00	18,97
		Neutral	11,50	11,50	11,50	20,41	18,00	21,03
	IF2: Compe-tencies	Agile	0,00	0,00	0,00	20,41	15,00	21,79
		Traditional	33,33	83,33	33,33	44,22	46,67	43,59
		Neutral	66,67	16,67	66,67	35,37	38,33	34,62
Team factors	IF3: Communi-cation	Agile	100,00	100,00	100,00	89,80	90,00	89,74
		Traditional	0,00	0,00	0,00	2,04	5,00	1,28
		Neutral	0,00	0,00	0,00	8,16	5,00	8,97
	IF4: Decision processes	Agile	50,00	50,00	50,00	40,82	35,00	42,31
		Traditional	50,00	50,00	50,00	39,80	40,00	39,74
		Neutral	0,00	0,00	0,00	19,39	25,00	17,95
	IF5: Team size	Agile	100,00	100,00	100,00	59,18	66,67	57,26
		Traditional	0,00	0,00	0,00	10,88	13,33	10,26
		Neutral	0,00	0,00	0,00	29,93	20,00	32,48
	IF6: Distribution of power	Agile	50,00	0,00	50,00	35,71	50,00	32,05
		Traditional	50,00	50,00	50,00	33,67	20,00	37,18
		Neutral	0,00	50,00	0,00	30,61	30,00	30,77
Environmental factors	IF7: Pricing Models	Agile	50,00	50,00	50,00	41,84	47,50	40,38
		Traditional	0,00	0,00	0,00	21,94	17,50	23,08
		Neutral	50,00	50,00	50,00	36,22	35,00	36,54
	IF8: Customer satisfaction	Agile	25,00	50,00	25,00	55,10	60,00	53,85
		Traditional	0,00	25,00	0,00	24,49	25,00	24,36
		Neutral	75,00	25,00	75,00	20,41	15,00	21,79
	IF9: Customer collaboration	Agile	33,33	33,33	33,33	57,48	48,33	59,83
		Traditional	0,00	16,67	0,00	23,81	26,67	23,08
		Neutral	66,67	50,00	66,67	18,71	25,00	17,09
	IF10: Documen-tation	Agile	0,00	0,00	0,00	13,27	20,00	11,54
		Traditional	50,00	50,00	50,00	53,06	55,00	52,56
		Neutral	50,00	50,00	50,00	33,67	25,00	35,90

5. CONCLUSIONS AND OUTLOOK

We conducted an exploratory case study in a medium-sized software developing and consulting company. The aim was to identify, how agile this company is thinking with regard to project management in software development projects. Therefore, we asked two independent groups of respondents – decision makers and employees – about their experience with different project management and software development methods. We additionally concentrated on their perceptions along ten influencing factors in order to gain a deeper understanding of their way of thinking.

The case study provided first insights into the concept of Agile Thinking according to the investigated influencing factors. Referring to Q1, we could ascertain that the company has a tendency to agile thinking. Despite of this, there exist several obstacles hindering the implementation and adoption of agile approaches. Furthermore, we have to revise our assumption of

different attitudes between decision makers and employees as a possible reason. With respect to Q2, we detected only minor differences in the way of thinking between different organizational roles. A slight tendency towards a more traditional way of thinking is recognizable for decision makers, but mainly resulting from a large number of employees with a neutral attitude.

According to Q3, we identified the following factors and problem domains, hindering the adoption of agile approaches at an organizational level.

(1) Mainly, cost-related aspects track attention. Most statements which are answered in a traditional, neutral or controversial way of thinking, focused on cost aspects. Related issues are the loss of cost control, no acceptance of flexible pricing models and problems in assessing the cost-related consequences of agile methods.
(2) Another aspect is a clear demand for superiors taking responsibility for the project's results. Decisions have to be confirmed by management. This indicates that a leadership position within projects is appreciated.
(3) Customer satisfaction and collaboration are problem domains, too. Despite the awareness that customer interaction improves the quality, there is a kind of fear of losing bargaining power. Furthermore, there were very contradictory answers in this field. This shows uncertainty about the optimal degree of customer interaction.
(4) A comprehensive documentation is still seen as a quality indicator for customers. Thus, there is the risk that agile methods may be seen as insufficient, only because of their reduced documentation.

Summarizing the results, it is obvious that there are certain potential obstacles, hindering the adoption of agile methods, although they are not different when comparing decision makers with employees. Our results offer insights into problem domains, which have to be addressed explicitly, while implementing agile methods within an agile-inexperienced company. These results have several implications for research, SoDeCo and similar companies. Given the three categories of influencing factors (see figure 1), none of them has a totally hindering or supporting influence on Agile Thinking. Table 5 summarizes the most critical aspects, where Agile Thinking was mainly missing.

Table 5. Gaps in Agile Thinking

Individual factors	• Lack of self-responsibility • Demand for leadership
Team factors	
Environmental Factors	• Cost-related aspects • Fear of sharing knowledge • Fear of losing bargaining power • Comprehensive documentation as quality indicator

The best agile attitudes were found among the Team factors. The factors Communication and Team size showed clear agile ways of thinking, whereas Decision processes and Distribution of power resulted in many neutral answers. The latter two are closely connected to the Individual factors, where the lack of self-responsibility and a demand for leadership were obvious obstacles to agile methods. The most hindering issues were revealed with the Environmental factors. Although the agile principle of customer collaboration and the creation and response to change are known, they are often seen as annoying and uncontrollable. Also, the fear of losing power, because of too knowledgeable customers is an obstacle within the Environmental factors.

As a result, our research reveals the influencing factors, where an agile way of thinking is probably already in place when intending to make use of the advantages of agile methods within a more traditional project organization. These factors, for example Communication (IF3) and Team size (IF5), are especially suitable to start implementing agile approaches, tools or process components, because there will be no resistance among decision makers and employees. The other way round, our results show where traditional ways of thinking are still predominant and how they manifest themselves. Therefore, it is possible to explicitly foster attempts to change these organizational aspects in terms of enabling agile methods.

Hence, future research should focus on approaches to avoid the identified obstacles in the different categories. Some studies are already partially addressing these issues, mainly by investigating the combination of agile and traditional methods, but the results are very different. Karlström and Runeson [22] for example find that Extreme Programming as an agile approach can work well within stage-gate oriented organizations [22]. However, very formalized and bureaucratic organizational cultures hinder the effective execution of agile approaches [25, 30] as confirmed by our study, too. The given obstacles may serve as a starting point to investigate the combination of agile and traditional approaches at an organizational level.

Additionally, a further investigation of the introduced concept of Agile Thinking offers the possibility to gain an insight into an organization. This insight may then serve as an instrument to prove the ability of a company to act and think agile. Further aspects of future research could be to develop methodologies to change the way of thinking towards agile for the identified problem domains. A suitable instrument could be the systems dynamics approach. Another interesting aspect is to investigate the way of thinking of customers and compare the results with ours.

Our survey suffers from some limitations that should be addressed in later studies. Due to its exploratory and qualitative character, our case study was limited to one company as a starting point. Its tentative results require additional substantiation on the basis of a multiple case study design. Based on larger samples testable propositions may be derived in the long run.

6. REFERENCES

[1] Abbas, N., Gravell, A. M., and Wills, G. B. 2008. Historical Roots of Agile Methods: Where Did "Agile Thinking" Come From?. In *Agile Processes in Software Engineering and Extreme Programming, 9th International Conference XP 2008* (Limerick, Ireland, June 10 - 14, 2008). XP2008. Springer, Berlin, Germany, 94-103.

[2] Abrahamsson, P., Conboy, K., and Wang, X. 2009. 'Lots done, more to do': the current state of agile systems development research. *Eur. J. Inform. Syst.* 18, 4 (Aug. 2009), 281-284.

[3] Ågerfalk, P. J., Fitzgerald, B., and Slaughter, S. A. 2009. Flexible and Distributed Information Systems Development:

State of the Art and Research Challenges. *Inform. Syst. Res.* 20, 3 (Sep. 2009), 317-328.

[4] Ågerfalk, P. J. and Wistrand, K. 2003. Systems Development Method Rationale: A Conceptual Framework for Analysis. *5th International Conference on Enterprise Information Systems* (Angers, France, April 23 – 26, 2003). ICEIS.

[5] Agile Manifesto 2001. Manifesto for Agile Software Development. http://www.agilemanifesto.org/. Accessed 02 August 2010.

[6] Balzert, H. 2008. *Lehrbuch der Softwaretechnik: Software-management.* 2nd edition, Spektrum, Heidelberg, Germany.

[7] Bleek, W. G. and Wolf, H. 2008. *Agile Softwareentwicklung: Werte, Konzepte und Methoden.* dpunkt, Heidelberg, Germany.

[8] Bonoma, T. V. 1985. Case Research in Marketing: Opportunities, Problems, and a Process. *J. Marketing Res.* 22, 2 (May 1985), 199-208.

[9] Bose, I. 2008. Lessons Learned from Distributed Agile Software Projects: A Case-Based Analysis. *Commun. AIS* 23, 619-632.

[10] Chow, T. and Cao, D. B. 2008. A survey study of critical success factors in agile software projects. *J. Syst. Software* 81, 6 (Jun. 2008), 961-971.

[11] Cockburn, A. 2003. *Agile Software-Entwicklung*. mitp, Bonn, Germany.

[12] Conboy, K. 2009. Agility from First Principles: Reconstructing the Concept of Agility in Information Systems Development. *Inform. Syst. Res.* 20, 3 (Sep. 2009), 329-354.

[13] Darke, P., Shanks, G., and Broadbent, M. 1998. Successfully completing case study research: combining rigour, relevance and pragmatism. *Inform. Syst. J.* 8, 4 (Oct. 1998), 273-289.

[14] Dybå, T. and Dingsøyr, T. 2008. Empirical studies of agile software development: A systematic review. *Inform. Software Tech.* 50, 9-10 (Aug. 2008), 833-859.

[15] Fernandez, D. J. and Fernandez, J. D. 2008. Agile Project Management – Agilism versus Traditional Approaches. *J. Comput. Inform. Syst.* 49, 2 (Winter 2008/2009), 10-17.

[16] Gable, G. G. 1994. Integrating case study and survey research methods: an example in information systems. *Eur. J. Inform. Syst.* 3, 2 (Jan. 1994), 112-126.

[17] Ganguly A., Nilchiani, R., and Farr, V. F. 2009. Evaluating agility in corporate enterprises. *Int. J. Prod. Econ.* 118, 2 (Apr. 2009), 410-423.

[18] Highsmith, J. and Cockburn, A. 2001. Agile Software Development: The Business of Innovation. *Computer* 34, 9 (Sep. 2001), 120-122.

[19] Hruschka, P., Rupp, C., and Starke, G. 2009. *Agility kompakt: Tipps für erfolgreiche Systementwicklung.* 2nd edition, Spektrum, Heidelberg, Germany.

[20] Jick, T. D. 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Admin. Sci. Quart.* 24, 4 (Dec. 1979), 602-611.

[21] Kaplan, B. and Duchon, D. 1988. Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quart.* 12, 4 (Dec. 1988), 571-586.

[22] Karlström, D. and Runeson, P. 2005. Combining Agile Methods with Stage-Gate Project Management. *IEEE Software* 22, 3 (May/Jun. 2005), 43-49.

[23] Larman, C. and Basili, V. R. 2003. Iterative and Incremental Development: A Brief History. *Computer* 36, 6 (Jun. 2003), 47-56.

[24] Lee, G. and Xia, W. 2010. Toward Agile: An Integrated Analysis of Quantitative and Qualitative Field Data on Software Development Agility. *MIS Quart.* 34, 1 (Mar. 2010), 87-114.

[25] Lindvall, M., Basili, V., Boehm, B., Costa, P., Dangle, K., Shull, F., Tesoriero, R., Williams, L., and Zelkowitz, M. 2002. Empirical Findings in Agile Methods. In *Proceedings of Second XP Universe and First Agile Universe Conference* (Chicago, IL, August 4 - 7, 2002). XP/Agile Universe 2002. Springer, Berlin, Germany, 197-207.

[26] Mangalaraj, G., Mahapatra, R., and Nerur, S. 2009. Acceptance of software process innovations – the case of extreme programming. *Eur. J. Inform. Syst.* 18, 4 (Aug. 2009), 344-354.

[27] McAvoy, J. and Butler, T. 2009. The role of project management in ineffective decision making within Agile software development projects. *Eur. J. Inform. Syst.* 18, 4 (Aug. 2009), 372-383.

[28] Misra, S. C., Kumar, V., and Kumar, U. 2009. Identifying some important success factors in adopting agile software development practices. *J. Syst. Software* 82, 11 (Nov. 2009), 1869-1890.

[29] Müller, T. and Gross, B. 2009. Welcome to Reality! Agile vs. Klassisch. In *Tagungsband 26. Internationales Deutsches Projektmanagement Forum* (Berlin, Germany, October 14 - 15, 2009). PM Forum 2009. GPM, Berlin, Germany, 350-357.

[30] Nerur, S., Mahapatra, R., and Mangalaraj, G. 2005. Challenges of Migrating to Agile Methodologies. *Commun. ACM* 48, 5 (May 2005), 73-78.

[31] Salo, O. and Abrahamsson, P. 2008. Agile methods in European embedded software development organisations: a survey on the actual use and usefulness of Extreme Programming and Scrum. *IET Softw.* 2, 1 (Feb. 2008), 58-64.

[32] Sambamurthy, V., Bharadwaj, A., and Grover, V. 2003. Shaping Agility through Digital Options: Reconceptualising the Role of Information Technology in Contemporary Firms. *MIS Quart.* 27, 2 (Jun. 2003), 237-263.

[33] Seethamraju, R. 2006. Influence of Enterprise Systems on Business Process Agility. In *Global Conference on Emergent Business Phenomena in the Digital Economy* (Tampere, Finland, November 28 – December 3, 2006). ICEB+eBRF.

[34] Sharifi, H. and Zhang, Z. 1999. A methodology for achieving agility in manufacturing organizations: An introduction. *Int. J. Prod. Econ.* 62, 1-2 (May. 1999), 7-22.

[35] Shine Technologies 2003. *Agile Methodologies Survey Results*. Shine Technologies Pty Ltd, http://www.shinetech.com/attachments/104_ShineTechAgileSurvey2003-01-17.pdf. Accessed 02 August 2010.

[36] Yin, R. K. 2009. *Case Study Research: Design and Methods.* 4th edition, SAGE Inc., Thousand Oaks, CA

Appendix

Appendix 1. Statements (items) per influencing factor

Influencing factor	Statements (items)
IF1: Self-responsibility	• Ein erfolgreiches Projektteam benötigt Führung durch Vorgesetzte. • Das Projektmanagement ist dem Projektteam übergeordnet. • Eigenverantwortung der Projektmitarbeiter erhöht die Qualität des Produktes • Eine selbständige Organisation der Projektteams erhöht die Motivation der Mitarbeiter • Teaminterne Selbstorganisation verhindert Leerlaufzeiten der Mitarbeiter.
IF2: Competencies	• Ein Softwareprojekt ist nur erfolgreich, wenn das Team sowohl aus Generalisten als auch Spezialisten besteht. • Einzelne Mitarbeiter sollten entsprechend ihrer Kompetenzen eingesetzt werden. • Generalisten sind Spezialisten vorzuziehen. • Die Entscheidungskompetenzen sollten bei einer Person liegen. • Ein Team von Generalisten ist flexibler. • Durch viele Spezialisten entstehen häufig Leerlaufzeiten der Projektmitarbeiter.
IF3: Communi-cation	• Regelmäßiger Austausch zwischen den Projektmitarbeitern erhöht die Qualität des Projektergebnisses. • Die Kommunikation zwischen den Projektmitarbeitern wird durch komplexe Hierarchien behindert.
IF4: Decision processes	• Wichtige Entscheidungen müssen formal von der Projekt-/Unternehmensleitung bestätigt werden. • Eine flache Projekthierarchie führt zu schnelleren Entscheidungen.
IF5: Team size	• Große Projektteams erzeugen höhere Kosten. • Kleinere Projektteams arbeiten effizienter.
IF6: Distribution of power	• Mitarbeiter neigen dazu, Informationen bei drohendem Machtverlust zurück zu halten. • Die Preisgabe impliziten Wissens führt zu Machtverlust bei einzelnen Mitarbeitern.
IF7: Pricing Models	• Festpreise sind flexiblen Preisbildungen vorzuziehen. • Zusätzliche Kundenwünsche sollten nur realisiert werden, wenn der vereinbarte Preis eingehalten werden kann. • Flexible Preisbildungen werden vom Kunden nicht akzeptiert. • Flexible Preisbildungen erhöhen die Qualität.
IF8: Customer satisfaction	• Die Kundenzufriedenheit ist von der Einhaltung der Kosten abhängig. • Kunden können die Qualität der Ergebnisse nur ungenügend einschätzen. • Kundenzufriedenheit erhöht sich durch regelmäßige Zusammenarbeit während der Projektlaufzeit. • Kunden akzeptieren eine erhöhte Projektdauer bei höherer Qualität. • Bei Zeitproblemen empfiehlt es sich, zusätzliche Mitarbeiter hinzuzuziehen.
IF9: Customer collaboration	• Viele Anforderungen des Kunden sind überflüssig. • Der Kunde muss sich zu Beginn des Projektes über die Anforderungen im Klaren sein. • Der Projekterfolg hängt maßgeblich von der Kooperation des Kunden ab. • Der Kunde ist in der Verantwortung, Teilergebnisse zu beurteilen. • Nachträgliche Änderungen des Kunden beeinträchtigen die Qualität. • Der Kunde ist nicht in der Lage, den Projektfortschritt zu beurteilen.
IF10: Documen-tation	• Der Kunde assoziiert ausführliche Dokumentation des Projektverlaufs und der Projektergebnisse mit Professionalität. • Der Kunde erwartet umfangreiche Dokumente zu Projekt(teil)ergebnissen.

Analysis of Two Theoretical Perspectives on Information Systems Development: Towards an Integrated Perspective

Frank Zickert
Chair of Business Administration, esp. E-Finance and Services Science
Goethe University Frankfurt, Grüneburgplatz 1,
60323 Frankfurt, Germany
+49 179 4885644
mail@frankzickert.de

ABSTRACT
In this paper, we analyze two theoretical perspectives and investigate their explanatory power on information systems development (ISD) projects. Building upon a case study, we illustrate that the perspectives of ISD as an economic transformation process and ISD as complex problem solving address different but complementary ISD phenomena. By integrating both theoretical perspectives, we are able to analyze and predict more ISD phenomena than each of the theories individually. Therefore, the contribution of this paper is twofold. Firstly, it supports researchers in their selection of a theory when addressing ISD phenomena. Secondly, it serves as an example of how researchers can develop a new theoretical perspective to address a phenomenon of interest not covered appropriately by existing theories.

Categories and Subject Descriptors
D.2 [Software Engineering]: Requirements/Specifications.

General Terms
Design, Theory.

Keywords
Theoretical Perspectives, Information Systems Development, Economic Transformation Process, Complex Problem Solving.

1. INTRODUCTION
Research in information systems development (ISD) provides us with numerous theories that explain how ISD works. These theories frame our understanding of phenomena in and around ISD. Two of the most commonly used theories are, for example, ISD as an economic transformation process [33], in which resources are used to transform the requirements of a system into a working code [20], and ISD as complex problem solving [9], in which the solution is sought by generating and evaluating alternatives of the system under construction [41].

Although diversity in theory can be useful to ISD research [44], it confronts the researcher with the problem of deciding on which theory to use for the investigation of a phenomenon of interest [57]. This decision is crucial, since the phenomena that command our attention are linked inextricably to the theories and paradigms we use to understand the world [34]. Consequently, an inappropriate selection of a theory may result in the inability to investigate the phenomenon of interest.

The purpose of this paper is to provide support for researchers in their selection of a theory when addressing ISD phenomena. Therefore, this paper aims at answering the research question of which ISD phenomena can be addressed appropriately by using the theories of ISD as an economic transformation process and ISD as complex problem solving. Moreover, by integrating both theories with each other, this paper aims to extend the scope of ISD phenomena beyond what can be addressed by either theory alone. The empirical basis for the evaluation of the theories is a software development project that we were able to investigate in a large financial institution.

The remainder of this paper is organized as follows. We depict the theories of ISD as an economic transformation process and ISD as complex problem solving in section two. After that, in section three, our case study illustrates the different insights that these theories disclose. Subsequently, in section four, we integrate both theories and return to our case applying the integrated theory. Subsequent to a brief discussion in section five, we conclude by noting benefits and limitations associated with this analysis in section six.

2. TWO THEORIES
In this paper, theories are regarded as lenses through which we see problems and observe phenomena [8]. Following this notion, theories provide explanations of how phenomena are related to the problem and from which predictions can be derived or the problem can be solved [25]. Theories are thus tools that researchers use in order to investigate phenomena of interest. Since they are tools, there is not one single correct theory that implies all others are wrong, but rather any theory can at best be appropriate or inappropriate for the investigation of a specific phenomenon. Table 1 depicts a collection of ISD theories.

This paper focuses on two theories, ISD as an economic transformation process and ISD as complex problem solving. Both theories belong to the functionalist paradigm, in which "the

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

economic reality (translated into quantitative financial goals, and systems performance characteristics) allows system objectives to be derived in an objective, verifiable, and rational way [and where] systems design becomes primarily a technical process" [28]. Thus, both theories share fundamental assumptions about the nature of ISD, such as that in ISD, social order is used to find consensus on a solution that is the rational choice because it satisfies goals [14].

Table 1. Theories on ISD

Theory	Short Description
Economic Transformation Process	The system is transformed from objective goals into subsequent forms, such as requirements and code. [2][5][7]
Complex Problem Solving	The system is a set of parameters for which a configuration must be found that results in the desired system behavior. [9][40]
Knowledge-based	Systems are created using the aggregated knowledge of stakeholders. The team process needs to be coordinated. [23]
Negotiation	The system serves as means to the individual objectives of the stakeholders. The system characteristics are determined by negotiation. [10][42]
Complex Adaptive Systems	The system emerges as a result of the individual behavior of agents and their local optimization processes. [9][29]

The theories of ISD as an economic process and ISD as complex problem solving have been selected in this paper since both are widely used and acknowledged (e.g. [6][7][9] [21][37]).

2.1 Economic Transformation Process

The theory of ISD as an economic transformation process builds upon the economic theory of the firm that provides a formal description of the relationship between the quantity of outputs produced and the input resources employed. In the ISD process, input factors including labor (the programming team) and capital (tools and techniques) are transformed into outcomes such as new or modified software [7] as depicted in Figure 1.

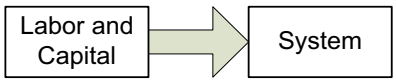


Figure 1. Software development project as a collection of transformation activities.

The central assumption underlying this perspective is a direct relationship between the input factors and the outcomes. For example, Banker et al. [7] apply the transformation process perspective in order to assess the effect of code generators or packaged software on productivity of the ISD maintenance process. Although not explicitly mentioned, Agrawal and Chari [2] build upon the notion of ISD as an economic transformation process when investigating the effects of high process maturity on outcomes, such as effort, quality, and cycle time. Anda et al. [5] quantify the impact that variations and reproducibility in the ISD process have on the quality of software projects in terms of delivery within budget and on the quality of the product in terms of functionality, reliability, usability, efficiency, maintainability, and portability.

The main phenomenon of interest of ISD as an economic transformation process is the productivity of the ISD process and related attributes, such as effort or cycle time [6]. Insights about ISD productivity are crucial since the technical ISD process is an engineering task of creating cost effective solutions to practical problems [52]. The purpose of this perspective is thus to support the creation of cost effective solutions.

In order to measure productivity, both inputs and outputs need to be measured. The most important output is represented by the system size, which can be measured by the number of function points [3], [16], a metric of business systems functionality [4], [42] or by the number of source lines of code [58], [46], [12]. Labor, as the most important input factor, is represented by the project effort, which results from the time and number of staff that are needed to build the system [1]. Moreover, both input and output factors are homogenous.

Another important input factor that is missing in this notion are the requirements of the system under construction. Although labor is also required for the elaboration of requirements, there are conceptual differences between the requirements of the system and the labor required for building the system [23]. Requirements correspond to the system under construction [56]. Just as source code, requirements are a representation of the system. Each representation of the system serves a specific purpose, has an intended audience, and has its own language. While the purpose of source code is to run on a computer and developers write it in programming languages, the purpose of requirements is to describe what the system does in its environment [29]. Requirements are socially constructed and negotiated by stakeholders as means to satisfaction of their goals [47], [10] and requirements are written in natural language [32] or specific notation languages, such as KAOS [55] or Problem Frames [50].

In a refined notion of a transformation process, the purpose of ISD is the transformation of an early representation of the system, such as requirements, into a working instance that is represented by compiled and tested source code. Since all representations correspond to the same system, correctness of the transformation can be evaluated by a direct comparison of whether the representations are congruent [19], for instance, do the requirements that describe what the system is supposed to do match with what the source code of the system actually does when it is executed.

On the contrary, other input factors, such as labor, are not actually transformed but rather consumed by transforming one representation of the system into another [13]. Labor and other consumables are thus not added to the system, but these factors refer to the ISD project in which they are consumed.

Another class of input factors comprises tools and techniques, which are neither transformed nor consumed. Tools and techniques are used within the transformation process through which system representations are transformed by using labor. Both the amount of required consumables for a transformation and the quality of a transformed system representation depend on

the employed tools and techniques [59]. For instance, using a complex technique for the formal elaboration of requirements may require more labor than an easy and informal technique does. When using formal techniques however, the quality of the resulting requirements may be improved.

Figure 2 depicts the refined notion of ISD as an economic transformation process, which distinguishes between these three classes of input factors. In fact, this is still an abstract notion of ISD. The ISD process determines which specific activities are accomplished at all, whether they are done sequentially or concurrently, which representations of the system are produced, and at which points consumables are required. The waterfall model [48] serves as a blueprint of an ISD process from the perspective of ISD as an economic transformation process.

The theoretical perspective of ISD as an economic transformation process treats ISD as a black box, which means that there is no further analysis of how the transformation specifically works. On the contrary, since the input factors are homogenous, it is assumed that the transformation is repeatable and therefore predictable. That means that the ISD can be repeated with the same productivity each time it is executed. Consequently, if the ISD productivity has already been assessed, it is possible to forecast required labor for the transformation of specific systems. Cost and effort estimation methods, for example, build upon this assumption when they estimate the labor that is required for the system development based on the system size [11].

The following Table 2 summarizes the major characteristics of the theoretical perspective of ISD as an economic transformation process.

Table 2. ISD as an Economic Transformation Process

Purpose	- Creation of cost effective solutions - Effort estimations
Treats ISD as	- Black Box
Assumptions	- Direct relationship between input factors and outcomes - Input factors are homogenous - The result of transformation is predictable - Transformations are repeatable
Input factors	- Resources/Labor - Process model - System content
Phenomenon of interest	- Productivity of the ISD process

2.2 Complex Problem Solving

Another theoretical perspective on ISD is ISD as complex problem solving [9]. This perspective mainly aims at disclosing what needs to be done in order to find a satisfactory solution for the problem [41].

Marengoa and Dosi [37], for example, find in their investigation of the degree of decentralization in problem solving that decentralized structures are unlikely to generate optimal solutions if the problem is complex. Duimering et al. [21] examine the influence of product requirement ambiguity on the task structures of the development project. Their results highlight the role of communication, coordination, and knowledge as distributed development project teams struggle to resolve ambiguity. Espinosa et al. [22] investigate the effect of familiarity on how long the development team requires in order to find an error free solution to the problem.

The theoretical perspective of ISD as complex problem solving builds upon the notion of a parametric representation, in which the system is regarded as a collection of parameters. The behavior of the system, once it is completed, depends on the set of values that are assigned to the parameters. The objective in ISD is to define values for all parameters of a system in a way that results in the desired behavior of the system [31]. The complexity of finding appropriate values for all parameters originates from interrelations among the parameters [51]. Due to interrelations, whether a specific value for a parameter is valid depends on the value itself and also on values that have been assigned to related parameters.

In ISD, the problem to be solved is represented by requirements that describe what the system has to accomplish [29]. The problem is solved if all requirements are met. Requirements engineering (RE) methods, such as KAOS, support the elaboration and verification of requirements in a way that assures the requirements appropriately address the superordinate problem [18]. Therefore, requirements can be regarded as the parameters of the problem.

Subsequently, in ISD, specifications that describe how the system works are designed in order to accomplish the requirements. The specifications therefore serve as values for the parameters. Other RE methods, such as problem frames, support the correct derivation of specifications from requirements and therefore aim at assuring that only valid values are assigned to the parameters [50].

Solving a problem requires assigning valid values to all parameters. The assignment is not carried out randomly but follows a search procedure that aims at favorable configurations for the values of the parameters. These search procedures are called heuristics [41] and are well covered by literature on artificial intelligence [36], [49]. Heuristics usually converge towards a solution, which means that they do not instantly find the right configuration but start with a configuration and alter it in a way that approaches the final solution. For example, the hill climbing heuristic starts with a random configuration of parameter values and then iteratively changes parameter values. Changes that improve the resulting solution performance are kept, whereas changes that decrease resulting performance are withdrawn. As depicted in Figure 3, the performance is increased

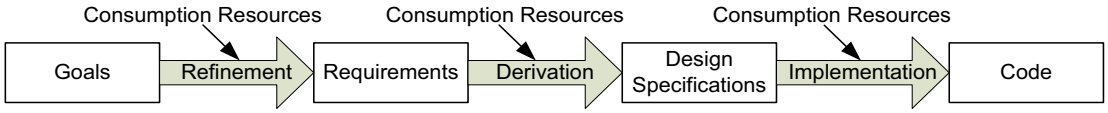


Figure 2. Software development project as a collection of transformation activities.

until an optimum is found, from which each parameter change results in a lower performance. However, depending on the starting point, the hill climbing heuristic may become stuck in local optima that may not achieve the desired performance output. In such cases, in order to find a satisfactory solution, the current path must be left and a completely different must be taken. This is done by backtracking, in which new values are assigned to parameters that have already been set in another way.

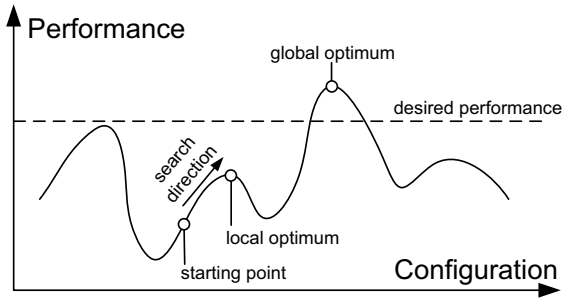


Figure 3. Hill climbing heuristic.

An underlying assumption of the perspective of ISD as complex problem solving is the decomposability of the problem. In order to be able to search for parameter values that solve the problem, the problem first has to be decomposed into a set of parameters. Based on Simon [51], problems often exhibit 'near decomposability', which refers to the idea that there are groups of problem components that have a high degree of interdependence to each other, whereas they are only loosely coupled with other groups of components. These groups appropriately serve as parameters, since they are relatively independent and thus it is easier to find valid values for them [21].

However, since the parameters remain interrelated with each other to a certain extent, the performance of a configuration results from the combination of parameter values, where even small changes in one parameter value can result in significant changes in the overall performance [54], [38]. As a result, in order to achieve satisfactory performance, assigning a value to a parameter may also require other parameters to take specific values. However, if the other parameters already have values that do not correspond to the required ones, some already set values must be changed respectively and reassigned. This reassignment—or backtracking—can also require other parameter values to change. Thus, it may result in cascade effects that require the complete configuration to change [10].

An assumption underlying the theoretical perspective of ISD as complex problem solving is the specificity of problems, where each problem is decomposed into a specific set of parameters [51]. Since each set of parameters exhibits a specific structure with regard to how the parameters are interrelated with each other, there is no general best way of how to solve a problem, but the performance of the applied heuristic depends on its fit to the problem structure [40]. For example, due to its property of getting stuck in local optima, hill climbing is an inappropriate heuristic for solving a problem with many local optima that do not achieve satisfactory performance. Other heuristics that do not get stuck in local optima, such as genetic algorithms, would find better solutions for such problems.

The underlying assumption is that activities are not generally repeatable but that it depends on the specific parameters whether valid values exist. In order to account for the specificity of problems and how they are solved, this perspective treats ISD as a white box. Moreover, since specific values are assigned to specific parameters, how such assignments affect the solution performance is not predictable unless they are given a try.

Table 3 summarizes the major characteristics of the theoretical perspective of ISD as complex problem solving.

Table 3. ISD as Complex Problem Solving

Purpose	- Find a satisfactory solution
Treats ISD as	- White Box
Assumptions	- Each problem is specific - Decomposability of the problem - The effect of an activity on solution performance is unknown unless it is tried - Direct manipulation of parameters and values possible
Input factors	- Problem - Heuristic/the way of how the problem is solved
Phenomenon of interest	- Performance of the system under construction

3. ISD CASE

3.1 Case Study Design

In order to get first hand information about the phenomena that the theoretical perspectives on ISD investigate, we applied a case study on a software development project in a large financial institution.

The observed project involved various stakeholders and affected different systems. Moreover, the project comprised reengineering of an existing system and its integration with another recently built system. The project was selected since the variety of both participating stakeholders and involved systems promised to be fruitful for making a distinction between two different theoretical perspectives in use.

The most important source of data was observations that we made by accompanying the business and technical analysts when requirements and design specifications were elaborated. We spent 103 hours over 40 days with the analysts on the project. During this period, we frequently had discussions with the analysts. Moreover, we were able to also interrogate other stakeholders in the project, such as the retail customer division whose representative acted as internal customer, the project manager, developers, representatives of the vendors, and members of the testing team. Moreover, sources of data also included access to documents, including working versions und reviews to the documents, such as concepts, meeting minutes, and e-mails. In total, this documentation comprised 323 pages.

Following Yin [60], we matched our data to the elements and characteristics of both theories in order to identify patterns in the data that disclose whether the project took place in the way the theories suggest. Based on this, we analyze how the theoretical perspectives explain the observations and which kind of insights each theoretical perspective supports.

As proposed by Miles and Huberman [39] we conducted data gathering and analysis concurrently so that we were able to capture all information that we found necessary for matching patterns to the theories.

3.2 Description of the Case

The project started in January 2009 and was completed in August 2009. In the project, a front-end system had to be integrated with a recently built payment processing system and therefore required reengineering. Previously, payment orders once entered at the front-end were transferred to a legacy processing system. Since the legacy processing system was planned to be deactivated, orders needed to be transferred to the new processing system instead. Moreover, since the new processing system required different data and a different payment order format than the old processing system, the order entry at the front-end had to be changed completely, wherefore it was decided to reengineer the whole system.

The applied ISD process generally adhered to the waterfall model [48]. At the beginning of the project, the business analyst collected the objectives of the retail customer division representative, which served as a basis for the elaboration and formulation of the requirements that had to be met to satisfy the objectives. All requirements were collected in a requirements document.

Since the retail customer division representative had many issues concerning the functionality of the front-end, the analyst elaborated various requirements. In order to achieve these requirements, significant changes in the front-end design were made. For example, it was requested that payment orders had to be already checked for correctness at the front-end. Since this involved verification, whether entered bank codes are valid, access to a complete list of all allowed bank codes was required.

The business analyst evaluated different alternatives for the requirement of integrating the front-end with the processing system. Using reasoning of lower maintenance effort for the future front-end system, he selected a direct interface between the front-end system and the processing system

Based on the requirements document, two technical analysts derived the software design specifications and prepared the software design specification document. While the business analyst is part of an IT department that is aligned with the retail customer division, the technical analysts are assigned to specific IT systems. Thus, in this project, there was a separate technical analyst involved for each affected system, one for the front-end system and one for the processing system.

Reengineering the front-end system not only aimed at providing the new functionality but also was intended to straighten its design. Since the front-end system already had run for several years and had undergone frequent changes, its design was quite tangled. Consequently, in order to obtain a system that is not bound to legacy structures, reengineering the front-end system started from scratch [26], only adhering to the given requirements but ignoring any constraints given by the existing systems. This however, also included leaving unconsidered the constraints given by the processing system.

Starting from scratch, the technical analyst responsible for the front-end system addressed the requirement of a direct connection between the front-end and the processing system by specifying a web-service interface. The integration of the front-end and the processing system proceeded after the front-end design was completed. Subsequent to reviewing the specified front-end design, the technical analyst, who was responsible for the processing system, rejected the implementation of a web-service interface at the processing system side, since it would not be implementable within the given constraints in time and budget. Instead, he suggested a file transfer. However, the architecture of the front-end system did not support file transfers in the suggested way. The inability to provide a web-service on the one hand and the inability to transfer requested files on the other hand not only required rework of the already specified front-end system design, but it also rendered unworkable the requirement of the direct connection between both systems. Therefore, the requirement of an indirect connection replaced the direct connection requirement, although it implied higher maintenance costs.

Subsequent to the resolution of this issue and rework of the front-end design, the on-site developers and the external vendor developed the software code based on the software requirements specifications document. Subsequently, the testing team performed the software tests. Despite some minor bug-fixings, neither code development nor testing disclosed any problems that required considerable rework.

3.3 From an Economic Transformation Process Perspective

The most significant observation regarding the project at hand from the theoretical perspective of ISD as an economic transformation process is provided by the organizational policy that specifies all activities and their outcomes in the project. According to this policy, each activity has to have a described output that is the input for the next activity. For example, in the requirements analysis phase, requirements had to be elaborated by the business analyst and had to be written down in natural language. A template for the requirements document had to be used, which provides a document structure and the required contents. This requirements document served as input for the design specification phase, in which the technical analysts derived design specifications from the requirements. A software design specifications document had to be produced, whose content was also pre-structured by a template that had to be used. The design specifications were handed over to the developers and the external vendor, who prepared the source code, which was finally handed over to the testing team.

Each of these produced outputs referred to the system under construction and is thus a representation of this system. The process model described which activities the employees had to accomplish and which tools (e. g. templates) they had to use. The required labor for accomplishing the activities was gathered

using the organizational accounting tool, which every employee used to charge the spent working time to a project.

Altogether, these factors do not only support the assessment of the ISD productivity in this project, they also allow identification of productivity drivers. For example, the required formalism in the activities consumed a significant amount of labor and thus negatively affected productivity. Although the analysts had delineated requirements and design specifications using self-made models or descriptions, they had to spend about the same amount of work filling out the required template documents. Other factors that affected the productivity were the number of requirements that needed to be transformed into the working system and the number of required attempts for the correct transformation. The business analyst, for instance, elaborated different alternative integration requirements of the connection between the front-end and the processing system. Each of these alternatives needed to be elaborated and described and thus required labor that reduced productivity. Reworking the integration requirement after the web-service had been rejected is another example of a factor that negatively affected productivity.

However, although this perspective allows the identification of factors that affect productivity, such as required rework, it does not explain why rework occurred. Building upon the assumption that both input and output of an activity are homogenous, the investigation of any specific input or output is unsupported from this theoretical perspective.

3.4 From a Complex Problem Solving Perspective

Despite general adherence to the given organizational policy, the project at hand was not accomplished in a unidirectional and straightforward manner, but could be characterized as a continuous search, in which different alternatives were evaluated in order to find a solution that exhibited the requested performance characteristics.

At the beginning of the project, for example, the business analyst considered different alternative requirements before he was able to determine that a direct interface between the front-end system and the processing system is the requirement with the best performance attributes since it resulted in low maintenance cost. In fact, however, this of all requirements turned out to be inappropriate for a satisfactory solution, because it was not accomplishable. The designed web-service could not be integrated with the processing system within the given constraints in time and budget and the file transfer that would have worked with the processing system did not work with the front-end system.

In this situation, the integration requirement of a direct connection served as a parameter that comprised a dependency between the interfaces at the front-end and the processing system. Because of this dependency, a design specification that represents a value of this parameter had to work with both systems. However, although assigning a value that worked with both systems to this parameter was impossible in this situation, the selection of this parameter was not per se false. In fact, two design specifications could have achieved the requirement and thus depicted valid values for this parameter. What made this requirement unworkable was the dependency, due to which both the front-end and the processing system had to share the same value. While the web-service specification did not work with the processing system, the file transfer specification did not work with the front-end system. Thus, the inappropriateness of the requirement of a direct connection was not disclosed until design specifications were derived from it.

In order to solve the problem, despite the inconsistency among the required values for the direct connection requirement, the dependency between the values had to be resolved. Backtracking the direct connection requirement and replacing it with the requirement of an indirect connection decoupled both systems from each other and therefore enabled solving the problem. Without having had the chance to withdraw the requirement of the direct connection, the problem would not have been solvable. It would have resulted in failure of the project. Although the requirement of the indirect connection created other dependencies, such as the interfaces to a routing system, these new dependencies did not result in any problems with regard to finding appropriate design specifications as values.

This perspective offers insights about which specific activities and decisions in ISD were required in order to find a satisfactory solution. It provides an explanation of why specific requirements and design specifications had to be reworked in our case. For example, it discloses that the inconsistency between required values for the requirement of the direct connection inhibited solving the problem.

However, this perspective does not put the decisions made for solving the problem into an ISD context that explains why inconsistencies occurred at all, for example, whether the reason for the inconsistency was the ISD process, insufficient resources, or the problem of building a system.

4. TOWARDS AN INTEGRATED PERSPECTIVE

4.1 Theoretical Perspective

While both the theoretical perspectives of ISD as an economic transformation process and ISD as complex problem solving support the addressing of different phenomena of interest, both perspectives also have limitations with regard to which aspects they are able to explain. While the transformation process perspective sets input factors, such as attributes to the ISD process, into relation with the produced output and therefore discloses factors affecting productivity, it does not give underlying reasons of why the factors matter. The complex problem solving perspective, on the contrary, allows investigating the structures underlying ISD and therefore provides insights into why specific problems occur in a project. However, it does not put these problems into relation to attributes of the ISD process, and therefore, it fails to provide measures on how to improve ISD.

Since the phenomena of interest that the theoretical perspectives address are complementary, an integrated perspective that combines both theories may address phenomena of interest beyond the phenomena addressed by either theory alone. Moreover, both theories share the same fundamental assumptions

about the world, because they both build upon the functionalist paradigm.

The integrated perspective regards ISD as a collection of solution space transformation activities. The solution space contains all potential solutions to the problem, regardless of whether their performances are satisfactory or not [10]. Like the complex problem solving perspective, the integrated perspective builds upon the notion of a parametric representation of the system under construction, in which the configuration of the parameter values results in the behavior of the system once it is built. Therefore, the solution space contains all configurations of parameter values.

However, in contrast to the complex problem solving perspective, in which ISD takes place as a conscious search for the parameters and their values, in the integrated perspective, accomplished activities unconsciously determine the parameters and their values, as is explained in the following.

Seen from the perspective of ISD as complex problem solving, parameters and their values are directly manipulated and therefore the configurations whose performance is sought to be evaluated are known. Although the performance of a configuration is unknown unless it is evaluated, heuristics calculate configurations worth consideration based on the performance of already evaluated configurations. For example, the genetic algorithm heuristic generates promising configurations by recombining parts of configurations with good performance [36].

On the contrary, seen from the integrated perspective of ISD as a collection of solution space transformation activities, only activities are consciously selected, whereas the configuration of parameters and their values results from the activities in an unpredictable way. That means, not only the performance of a configuration but also the specific configuration is unknown unless the activity that results in the respective configuration is accomplished. As a result, it is impossible to employ a heuristic because it is impossible to generate specific configurations selectively. Therefore, it is not a heuristic but the current situation in the ISD project that supports decisions on which activities to execute and which resources to employ in order to solve the given problem.

In this regard, the integrated perspective is similar to the perspective of ISD as an economic transformation process. There are specific activities in ISD that are executed in order to build the system and each activity requires resources—most importantly labor. However, while in the notion of the economic transformation perspective, activities directly transform the content of the final solution in a predictable way, in the notion of the integrated perspective, activities transform the current configuration in an unpredictable way.

Since the current configuration determines which other configurations can be achieved by performing further activities, one needs to distinguish between the actual solution space that only contains solutions that are achievable from the current configuration and the overall solution space that contains all configurations.

The actual solution space evolves over time. With each activity, it approaches a solution which, however, is unknown both in terms of its configuration and its performance. Therefore, whether the solution exhibits satisfactory performance is unknown, too. Since neither the configuration nor its performance are predictable, although both depend on the activities, the employed resources, and the specific problem, ISD is not directed in any way, neither in terms of conscious problem solving, nor in terms of simply transforming the content of the system under construction. The phenomenon of interest of the integrated perspective therefore is to investigate why ISD is successful or fails at all.

Table 4. ISD as a Collection of Solution Space Transformation Activities

Purpose	- Investigations of the structures underlying ISD and putting them into relation with general input factors
Treats ISD as	- White Box
Assumptions	- Each problem is specific - Decomposability of the problem - Activities are repeatable, but their outcome is not predictable because it depends on the content and the employed resources
Input factors	- Resources/Labor - Composition of activities - System content
Phenomenon of interest	- Reasons for ISD success or failure

By investigating how the actual solution space evolves in a project, this perspective allows tracing back problems, such as inconsistencies, to their origins. This perspective discloses whether the origin of success or failure in a specific case is the process model, the employed resources, or an unsolvable problem. Table 4 summarizes the major characteristics of the theoretical perspective of ISD as a collection of solution space transformation activities.

4.2 THE CASE REVISITED

The most significant characteristic of the observed case, which supports the notion of ISD as a collection of solution space transformation activities, is that the solution space was unknown. At no time, did decision makers consciously take into account how many or which solutions the actual solution space comprised. However, we will particularly consider the actual solution space in the following, when applying the integrated perspective on the case.

The major problem in the project at hand became evident when the front-end system was integrated with the processing system. In this situation, the actual solution space contained no valid solution. Although there were two considered solutions, the web-service interface as suggested by the technical analyst who was responsible for the front-end system design and the file transfer suggested by the technical analyst responsible for the processing system design, no solution worked with both systems. Therefore, both solutions were invalid, leaving no valid solution in the actual solution space.

In order to look into the cause for this “emptiness” of the actual solution space that resulted in backtracking and thus rework, related activities are analyzed. The design of the web-service was

the proximate activity, in which the technical analyst's task comprised addressing the requirement of a direct connection when reengineering the front-end system from scratch. The analyst successfully accomplished the task since the designed web-service appropriately addressed the direct connection requirement. It represented a valid solution for the given sub-problem.

Since the analyst successfully accomplished the design task, the reason that caused the empty actual solution space is not the analyst's fault but rather the activity itself. Particularly, the conscious neglect of the dependency to the interface of the existing processing system is questionable, because it delayed discovery of the empty actual solution space until the integration of both systems. However, since the front-end system turned out not to be able to support the file transfer as required from the processing system, even an early consideration of the dependency would not have resulted in anything but an empty actual solution space. Since neither insufficient nor incapable resources nor the neglect of the dependency caused the empty actual solution space, it must have been already empty prior to the derivation of design specifications.

Nevertheless, the division of labor affected the amount of accrued rework. Early consideration of the interdependency between the front-end and processing system interfaces would have disclosed earlier that the actual solution space was empty. It would have been recognized before a significant amount of work was spent on the complete front-end design specification. Thus, although the activity setting in this situation did not cause rework, it determined its extent. The recommendation therefore is to take into account all dependencies early.

In order to further investigate the cause of rework in this project, the activity, in which the parameters were set, needs to be analyzed. The business analyst set the parameters when elaborating the requirements at the beginning of the project. This activity aimed at requirements that can be met and, if met, satisfy the stakeholders' objectives. Although the first elaborated requirement of a direct connection could not be met, the activity was generally accomplishable as the second elaboration of the indirect connection requirement discloses. Therefore, the actual solution space at this time contained at least one valid and satisfactory solution that, however, was not selected right away.

Nevertheless, the selected requirement of a direct connection was a rational decision, because it was the best choice reflecting the information available to the business analyst at the time [17]. Firstly, the requirement of a direct connection best satisfied the objectives, because it also resulted in lower maintenance cost than the indirect connection requirement. And secondly, the information about the requirement of a direct connection to be unworkable did not emerge until the design was specified. When the business analyst first elaborated the direct connection requirement, the resulting actual solution space contained two seemingly valid solutions: the web-service and the file transfer. The decision would have been irrational only if some feasible arrangement for recognizing and achieving a preferred outcome existed, but that outcome was not obtained [35].

Since the problem of selecting an appropriate requirement was solvable and the decisions were rational in the given context, the activity that set the context for the decision needs to be critically analyzed. The given process model arranged for the final elaboration of requirements before their viability was checked further. As a result, information required in order to not only make rational but also beneficial decisions was unavailable when decisions had to be made. Therefore, the insights of this theoretical perspective recommend an ISD process that assures all relevant information be available when decisions need to be made. Concurrent requirement elaboration and design specification would make available information about requirement viability early and therefore could improve the quality of decision making [45]. However, since design specifications are built upon not yet finalized requirements, rework would occur if requirements turned out to not completely address the stakeholder objectives [53].

Altogether, the integrated perspective of ISD as a collection of solution space transformation activities suggests that good decisions are not necessarily those that best satisfy goals, but those that also allow further problem solving. In two situations within the observed project, the decisions that aimed at achieving the best solution resulted in severe consequences. Firstly, although the selected direct connection requirement would have implied lower maintenance cost, it resulted in an empty actual solution space and therefore in an unworkable situation that caused rework of the requirement and all design specifications building upon it. Secondly, although the chosen web-service interface would have implied a straightened design, it resulted in a large extent of rework.

However, this does not imply that goal satisfaction should not be a major factor for decision-making. It rather implies that the effect that decisions have on the actual solution space also needs to be included in the decision-making. For example, the requirement of the direct connection had the disadvantageous effect of coupling the interfaces between front-end and processing system and therefore increased complexity of accomplishing the activity [15]. However, much work is needed in order to assess the effect that decisions have on the actual solution space.

5. DISCUSSION

Theories allow knowledge to be accumulated in a systematic manner and this accumulated body of knowledge enlightens professional practice [25]. Therefore, the primary interest of scientific research is to add to the body of knowledge by the creation, refinement, and validity assessment of theories. However, since theories in the body of knowledge also serve as utilities from and through which IS research is accomplished [27], the researcher must be aware of the nature of the applied theories. Theories are only valid in a context that is determined by basic assumptions about the world and specific assumptions about the phenomenon of interest [28]. These assumptions must be considered when applying theories. Otherwise, findings may be misinterpreted or even void. Therefore, a critical eye on theories in the body of knowledge is required in order to not rely on serendipity when selecting a theory. Researchers need to be aware of the assumptions and beliefs that they employ in their day-to-day activities [28]. Therefore, further analyses are required in order to structure the body of knowledge in a way that makes it comprehensible and usable for subsequent research.

6. CONCLUDING REMARKS

In this paper, we investigated which phenomena of interest two widely used theoretical perspectives address in the area of ISD support. By building upon an analysis of the perspectives and a case study of a software development project conducted in a large financial institution, this paper has three findings.

Firstly, ISD productivity is the main phenomenon of interest of the theoretical perspective of ISD as an economic transformation process. While this perspective allows identifying factors affecting ISD productivity, such as rework, it does not explain the rationale underlying these factors, since it treats ISD as a black box. Therefore, it does not disclose measures positively influencing the factors, for example, measures reducing rework.

Secondly, the performance of the system under construction is the main phenomenon of interest of the perspective of ISD as complex problem solving. Since this perspective treats ISD as a white box, it supports investigations of how decisions in the ISD process affect performance. For example, it discloses that backtracking is vital for finding satisfactory solutions. However, this perspective does not put the decisions made for solving the problem into the specific ISD context. Therefore, it does not support conclusions on whether decisions, such as to backtrack, are reasoned with the ISD process, insufficient resources, or the problem of building a system.

Thirdly, an integrated perspective that combines both ISD as an economic transformation process and ISD as a complex problem solving, supports addressing the underlying reasons for ISD success or failure. In our case, the integrated perspective discloses that the applied process caused rework and determined its extent. Based on the insights that the integrated perspective provides, measures positively influencing ISD success can be identified, for example, making information about the consequences of decisions available as early as possible.

Our findings about which ISD problems can be addressed by using which theoretical perspectives provide support for researchers in their selection of a theoretical perspective when investigating ISD problems. Moreover, by integrating two theories, the paper serves as an example of how researchers can prepare a theoretical lens that is suited for the investigation of a phenomenon of interest that is not appropriately addressed by one single perspective.

However, this analysis has some limitations that future work needs to address. Firstly, the scope of this analysis is limited to the evaluation of two theoretical perspectives on ISD within the functionalist paradigm. There are in fact other theoretical perspectives within this or within other paradigms, which still have to be critically analyzed. Future work needs to evaluate these theories in order to create a framework that researchers can use when selecting a theory.

Secondly, although this analysis of the theoretical perspectives also provides some insights on ISD, it has to be noticed that these insights build upon a single case. In fact, we do not claim to have gathered any statistically generalized insights but rather analytical ones. In this paper, the insights on ISD illustrate which kind of insights the theoretical perspectives can provide. Nevertheless, the insights on ISD seem interesting and therefore deserve further scientific investigation.

7. Acknowledgements

This work was developed as part of a research project of the E-Finance Lab at Goethe University Frankfurt. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the E-Finance Lab or its supporting partners. The authors are indebted to the participating universities and gratefully acknowledge all support of their industry partners.

8. REFERENCES

[1] Abdel-Hamid, T. and Madnick, S. 1991. *Software Project Dynamics: An Integrated Approach*. Prentice-Hall, Englewood Cliffs, NJ.

[2] Agrawal, M. and Chari, K. 2007. Software Effort, Quality, and Cycle Time: A Study of CMM Level 5 Projects. *IEEE Transactions on Software Engineering* 33, 3, 145-156.

[3] Albrecht, A. 1979. Measuring Application Development Productivity. *Proceedings of IBM Applications Development Symposium*, 83-92.

[4] Albrecht, A. and Gaffney, J. 1983. Software Function, Source Lines of Code, and Development Effort Prediction: A Software Science Validation. *IEEE Transactions on Software Engineering* 9, 6, 639-648.

[5] Anda, B., Sjøberg, D., and Mockus, A. 2009. Variability and Reproducibility in Software Engineering: A Study of Four Companies that Developed the Same System. *IEEE Transactions on Software Engineering* 35, 3, 407-429.

[6] Banker, R., Datar, S., and Kemerer, C. 1991. A Model to Evaluate Variables Impacting the Productivity of Software Maintenance Projects. *Management Science* 37, 1, 1-18.

[7] Banker, R., Davis, G., and Slaughter, S. 1998. Software Development Practices, Software Complexity, and Software Maintenance Performance: A Field Study. *Management Science* 44, 4, 433-450.

[8] Benbasat, I. and Weber, R. 1996. Research Commentary: Rethinking 'Diversity' in Information Systems Research. *Information Systems Research* 7, 4, 389-399.

[9] Benbya, H. and McKelvey, B. 2006. Toward a Complexity Theory of Information Systems Development. *Information Technology & People* 19, 1, 12-34.

[10] Bergman, M., King, J., and Lyytinen, K. 2002. Large-Scale Requirements Analysis Revisited: The need for Understanding the Political Ecology of Requirements Engineering. *Requirements Engineering* 7, 3, 152-171.

[11] Boehm, B., Abts, C., and Chulani, S. 2000. Software Development Cost Estimation Approaches – A Survey. *Annals of Software Engineering* 10, 177-205.

[12] Boehm, B. 1981. *Software Engineering Economics*. Prentice-Hall, Englewood Cliffs, NJ.

[13] Boehm, B. 2007. Improving software productivity. In: *Software Engineering*, R. Selby, Ed. John Wiley and Sons, Hoboken, NJ, 151-178.

[14] Burrell, G. and Morgan, G. 1979. *Sociological Paradigms and Organizational Analysis*. Heinemann, London.

[15] Campbell, D. 1988. Task Complexity: A Review and Analysis. *The Academy of Management Review* 13, 1, 40-52.

[16] Cheung, Y., Willis, R., and Milne, B. 1999. Software Benchmarks Using Function Point Analysis. *Benchmarking: An International Journal* 6, 269-279.

[17] Cyert, R. and March, J. 1963. *A Behavioral Theory of the Firm*. Blackwell, Cambridge.

[18] Dardenne, A., Fickas, S., and van Lamsweerde, A. 1991. Goal-directed concept acquisition in requirements elicitation. *Proceedings of the 6th international workshop on Software specification and design*, Como, Italy, 14-21.

[19] Darimont, R. and van Lamsweerde, A. 1996. Formal refinement patterns for goal-driven requirements elaboration. *Fourth ACM SIGSOFT Symposium on the Foundations of Software Engineering*, San Francisco, 179-190.

[20] Davis, A. 1993. *Software Requirements: Objects, Functions, and States*. Prenctice-Hall, Englewood Cliffs, NJ.

[21] Duimering, P., Ran, B., Derbentseva, N., and Poile, C. 2006. The Effects of Ambiguity on Project Task Structure in New Product Development. *Knowledge and Process Management* 13, 4, 239-251.

[22] Espinosa, J., Slaughter, S., Kraut, R., and Herbsleb, J. 2007. Familiarity, Complexity, and Team Performance in Geographically Distributed Software Development. *Organization Science* 18, 4, 613-630.

[23] Faraj, S. and Sproull, L. 2000. Coordinating Expertise in Software Development Teams. *Management Science* 46, 12, 1554-1568.

[24] Glinz, M. 2007. On Non-functional Requirements. *Proceedings of the IEEE Joint International Conference on Requirements Engineering (RE'07)*, 21-26.

[25] Gregor, S. 2006. The nature of theory in IS. *MIS Quarterly* 30, 611-642.

[26] Hammer, M. and Champy, J. 1993. *Reengineering the Corporation*. Harper Collins, New York, NY.

[27] Hevner, A., March, S., Park, J., and Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28, 75-105.

[28] Hirschheim, R. and Klein, H. 1989. Four Paradigms of Information Systems Development. *Communications of the ACM* 32, 10, 1199-1216.

[29] Holland, J. 1996. *Hidden Order, How Adaption Builds Complexity*. Basic Books, New York, NY.

[30] Jackson, M. 1995. *Software Rrequirements and Specifications*. Adison-Wesley, New York, NY.

[31] Kauffman, S. 1995. *At Home in the Universe*. Oxford University Press, New York, NY.

[32] Kotonya, G. and Sommerville, I. 1998. *Requirements Engineering: Processes and Techniques*. John Wiley and Sons, New York , NY.

[33] Kriebel, C. and Raviv, A. 1980. An Economics Approach to Modeling the Productivity of Computer Systems. *Management Science* 26, 3, 297-311.

[34] Kuhn, T. 1970. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, IL.

[35] Liebowitz, S. and Margolis, S. 1995. Path Dependence, Lock-in, and History. *Journal of Law, Economics, and Organization* 11, 1, 205-226.

[36] Luger, G. 2002. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Addison Wesley, Reading, MA.

[37] Marengoa, L. and Dosi, G. 2005. Division of Labor, Organizational Coordination and Market Mechanisms in Collective Problem-solving. *Journal of Economic Behavior & Organization* 58, 303-326.

[38] McCann, J. and Ferry, D. 1979. An Approach for Assessing and Managing Inter-Unit Interdependence. *The Academy of Management Review* 4, 1, 113–119.

[39] Miles, M. and Huberman, 1994. *A Qualitative Data Analysis*. Sage Publications, Thousand Oaks, CA.

[40] Newell, A. and Simon, H. 1972. *Human Problem Solving.* Prentice Hall, Englewood Cliffs, NJ.

[41] Newell, A. and Simon, H. 1976. Computer Science as Empirical Enquiry: Symbols and Search. *Communications of the ACM* 19, 113–126.

[42] Ovaska, P, Rossi, M., and Smolander, K. 2005. Filtering, Negotiating and Shifting in the Understanding of Information System Requirements. *Scandinavian Journal of Information Systems* 17, 1, 31-66.

[43] Perry, W. 1986. The Best Data Processing Measures. *System Development* 6,6, 4-6.

[44] Pfeffer, J. 1993. Barriers to the Advance of Organizational Science: Paradigm Development as a Dependent Variable. *Academy of Management Review* 18, 4, 599-620.

[45] Pich, M., Loch, C., and de Meyer, A. 2002. On Uncertainty, Ambiguity, and Complexity in Project Management. *Management Science* 48, 8, 1008-1023.

[46] Putnam, L. 1978. General Empirical Solution to the Macro Software Sizing and Estimating Problem. *IEEE Transactions on Software Engineering* 4, 345-361.

[47] Robinson, W. 1990. Negotiation Behavior During Requirements Specification. *Proceedings of the 12th international conference on Software engineering*, Nice, 268-276.

[48] Royce, W. 1970. Managing The Development of Large Software Systems. *Proceedings of IEEE WESCON* 26, 1-9.

[49] Russel, S. and Norvig, P. 2003. *Artificial Intelligence, A Modern Approach*. Prentice Hall, Upper Saddle River, NJ.

[50] Seater, R., Jackson, D., and Gheyi, R. 2007. Requirement Progression in Problem Frames: Deriving Specifications from Requirements. *Requirements Engineering* 12, 2, 77-102.

[51] Simon, H. 1996. *The Sciences of the Artificial*. MIT Press, Cambridge, MA.

[52] Shaw, M. 1990. Prospect for an Engineering Discipline of Software. *IEEE Software* 7, 15-24.

[53] Terwiesch, C. and Loch, C. 1999. Measuring the Effectiveness of Overlapping Development Activities. *Management Science* 45, 4, 455-465.

[54] Thompson, J. 1967. *Organizations in Action*. McGraw-Hill, New York, NY.

[55] van Lamsweerde, A., Darimont, R., and Massonet, P. 1995. Goal-directed Elaboration of Requirements for a Meeting Scheduler: Problems and Lessons Learnt. *Proceedings of the 2nd IEEE International Symposium on Requirements Engineering*, 194-203.

[56] van Lamsweerde, A. 2001. Goal-oriented Requirements Engineering: A Guided Tour. *Invited Paper for RE'01 - 5th IEEE International Symposium on Requirements Engineering*, Toronto, 249-263.

[57] Walsham, G. 2006. Doing Interpretive Research. *European Journal of Information Systems* 15, 320-330.

[58] Waltson, C. and Felix, C. 1977. A Method of Programming Measurement and Estimation. *IBM Systems Journal* 16, 1, 54-73.

[59] Wirth, N. 1995. A Plea for Lean Software. *IEEE Computer* 28, 2, 64-68.

[60] Yin, R. 1994. *Case Study Research: Design and Methods*. Sage Publications, Thousand Oaks, CA.

Track 5: Adoption, Usage and Impact of Information Systems

Jens Dibbern
Institute of Information Systems, University of Bern, Engehaldenstrasse 8, CH-3012 Bern, Switzerland
jens.dibbern@iwi.unibe.ch

Armin Heinzl
Information Systems, University of Mannheim, Schloss, D-68131 Mannheim, Germany
heinzl@uni-mannheim.de

Wynne W. Chin
Department of Decision & Information Sciences, University of Houston, 4800 Calhoun Rd., Houston, Texas, 77204, Unites States
wchin@uh.edu

Editorial

This track contributes to a better understanding of the various factors, processes, concepts, and methods that influence the adoption and usage of information systems as a precondition for achieving the desired impacts of information systems (IS) on the individual, group, and organization level. The papers of this track may be categorized into three groups.

The first group of papers focuses on the adoption of IS. As such, the focus is set on emergent Internet technologies which are increasingly used by individuals in order to facilitate social interaction with others. Such social interaction over the Internet takes place in private life as well as in organizational settings. The boundaries between such settings are often blurred and hence overlaps exist in the patterns explaining differences in the adoption and usage behavior of individuals. The papers falling under this category consider user heterogeneity by examining gender, age, and cultural differences between users. Moreover, the technology-enabled shift from offline to online social networking as well as the role of online social networking within organizations are examined.

The second group of papers focuses on the implementation of IS, which often takes place in the form of IS projects. The studies range from enhancing our understanding of the evolution of risks in IS projects to examining particular cultural challenges in IS offshore projects, and elaborating on the key factors that influence the adoption of project management methodologies. Moreover, the specific challenges arising in the course of implementing real-time communication and collaboration technologies are examined.

Finally, a third group of studies focuses on particular emergent IS concepts, such as electronic negotiation systems and service oriented architectures, discussing their underlying design characteristics and examining their organizational impacts, respectively.

Associate Editors

- Roman Beck, Johann Wolfgang Goethe University of Frankfurt, Germany
- Paola Bielli, University of Bocconi, Italy
- Mike Gallivan, Georgia State University, USA
- Oliver Hinz, Johann Wolfgang Goethe University of Frankfurt, Germany
- Norman Johnson, University of Houston, USA
- Jan Marco Leimeister, University of Kassel, Germany
- Yan Li, ESSEC Business School, France
- Thomas Myrach, University of Bern, Switzerland
- Nancy Pouloudi, Athens University of Economics and Business , Greece
- René Riedl, University of Linz, Austria
- Andrew Schwarz, Louisiana State University, USA
- Jason Thatcher, Clemson University, USA
- Tim Weitzel, University of Bamberg, Germany

Conflict Resolution Support in Electronic Negotiations

Alexander Dannenmann
University of Hohenheim
Information System I
D-70593 Stuttgart
+49 (0) 711 - 459 23834
alexander.dannenmann
@uni-hohenheim.de

Mareike Schoop
University of Hohenheim
Information System I
D-70593 Stuttgart
+49 (0) 711 - 459 23345
m.schoop
@uni-hohenheim.de

ABSTRACT
Negotiation Support Systems offer a sophisticated support of electronic negotiations. During a negotiation process, different types of conflicts can occur. Up to a certain level, they can be constructive, afterwards they become destructive. Such escalating conflicts should be handled and resolved. In this paper, the potential and challenges of conflict resolution support for such type of negotiations are discussed. Based on a state-of-the-art in electronic negotiation support, the application of the concepts of moderation, consultation, and mediation for conflict management and conflict prevention during an asynchronous electronic negotiation process is discussed.

General Terms
Management, Performance, Economics, Experimentation, Human Factors.

Keywords
Electronic negotiation, conflict management, negotiation support system, moderation, consultation, mediation

1. Introduction
In general, parties negotiate in an iterative communication and decision process. Their motivation is that they want to reach a goal that they cannot achieve alone. During the negotiation process the parties act through exchanging information, offers, and counteroffers to find an agreement [3]. Thus, a conflict is the reason for every negotiation. Electronic negotiations are a specific form of negotiations and their use in a business context has become more and more important during the last years [29]. Compared to face-to-face negotiations, electronic negotiations can offer a multiplicity of advantages which can lead to different economic effects such as cost and time saving [28]. On the other hand, there are several challenges to deal with, e.g. the electronic medium, distributed locations.

During a negotiation, different kinds of conflicts can occur. Whilst parties would not negotiate without the initial conflict, escalating conflicts can lead to a rejection of negotiation which in turn can cause additional costs (e.g. costs of litigation). For example, it has been reported that German companies go to court about 500 times per year with a value of claim of over 500.000€ each leading to more than 1.000.000.000€ costing of litigation [33]. There are several alternative options for dispute solution. Additionally, conflicts are influenced by different factors and can change during a negotiation [17]. In early stages of a conflict, moderation or consultation can help parties to find a joint solution. In later stages, the negotiators can ask a neutral third party called a mediator for assistance. Mediation is a structure process in which a third party with not decision making authority supports the conflict parties to find a new win-win solution [18–20, 23]. Surprisingly, only around 600 mediations between German companies take place every year. 75% of mediations are successful [33]. Comparing the costs, mediation is a real alternative to conflict resolution by a court. Mediators can act much quicker than a judge who must follow standardised processes. As a consequence, those who correctly solve conflicts will save costs and be prepared for success [20].

The conflict behaviour within electronic negotiations is different to that in face-to-face negotiations. Based on the restrictions and potential of the medium, parties focus on some selected aspects. Communication plays a more important role whilst gestures, mimics, and tone of voice which can increase or decrease conflicts are missing. Thus it depends to a certain degree on the conflict behaviour and the negotiation strategy whether the conflict potential in electronic negotiations is higher than in normal face-to-face negotiations [32].

Although there are differences in the conflict behaviour, the conflict process in electronic negotiations is in points similar to that in face-to-face negotiations (cf. figure 1).

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

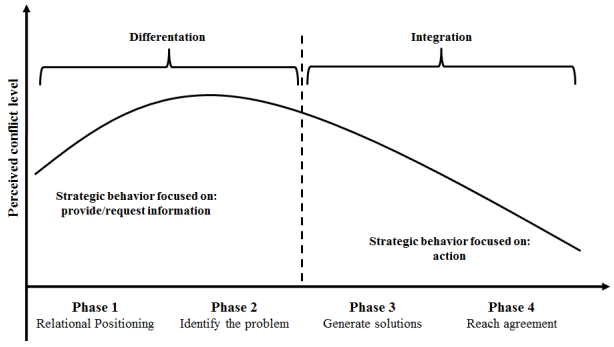


Figure 1: Negotiation behavior over time in a face-to-face setting [32]

In electronic negotiations, we have a differentiation and an integration phase. In the first phase, the parties focus on providing and collecting information. During the process, they exchange information and start to go into the integrative phase, where the main focus is on action, i.e. on exchanging offers with the objective to generate solutions [32].

In this idealised process, there are different conflict stages. Conflicts develop as the consequence of the negotiators behaviour. During the first period, the perceived level of conflict increases before latent conflicts become stronger and start to manifest. At the end of this period, the participants have, on the one hand, a high level of conflict while, on the other hand, they have created a basis for a mutual understanding based on the exchanged information. They continue with the integration phase, communicate more openly and start to generate solutions. Normally, the level of conflict will decrease with this process and the negotiators will come to an agreement at the end. Taken together, the parties first have a high perceived degree of conflict before they move closer together to find an optimal solution for both sides [32]. Figure 1 shows an approach focusing on the initial conflict - the prime reason for the negotiation - but fails to consider dynamic conflicts occurring during the process. There can be points during the negotiation, where the conflict starts to escalate and as a consequence will lead to a rejection of the negotiation.

The main objective of this paper is to introduce and discuss a framework for conflict resolution management in electronic negotiations to enable agreements and to prevent parties from breaking off their negotiation process without agreement. To this end, the following questions need to be addressed: Are the conflict resolution methods used in face-to-face conflict situations useable in the virtual world? What are the requirements and consequences for the conflict resolution support of the electronic negotiation process to overcome impasses and to help the parties finding an integrative agreement? The paper will first introduce the concepts relevant in electronic negotiations (section 2) before discussing diagnosis and methods in conflict management (section 3). Section 4 will then present the framework of conflict resolution in electronic negotiations. The paper will be concluded with a summary and a final discussion of the work.

2. Electronic Negotiations

2.1 Definition and classification

Negotiations that are conducted by means of information technology and that provide rules for communication and/or decision support that are enabled only through the use of information technology are called electronic negotiations [43]. This definition includes different forms of electronic negotiations ranging from fully automated ones conducted by negotiation agents over semi-automated electronic auctions to negotiation support in which the decision are taken by the human negotiator who is supported in the various negotiation processes.

A negotiation support system (NSS) is software which implements models and procedures, has communication and coordination facilities, and is designed to support two or more parties and/or a third party in their negotiation activities [26]. In contrast, for example, to email, the system supports the parties in different ways, namely by means of communication support, decision support and document management [38–41]. The level of involvement an NSS can offer has different levels. Kersten et al. [26] introduced the following three stages, cf. table 1.

Table 1: Involvement categorisation of NSSs (based on [26])

Stage	Involvement	Functions
1	Passive	- Support of interaction between users located in different places (= Communication Support) - Mathematical and Statistical calculation of utilities (= Decision Support) - Different visualization technologies to represent data
2	Active	- Support in evaluating, formulating and problems solving - Offering concession range and possible new counter-/offers (= Facilitation-Mediation Systems)
3	Pro-Active	- Same functions than active systems, extend with the knowledge of an artificial intelligence. The system supports the negotiator more active, based on a monitoring of the process and the activities. (= Intervention-Mediation System)

The main differences between active and pro-active NSSs are the role of the user and the knowledge base of the system. In active systems, the negotiator asks for an advice, the system uses the given data, makes an analysis, and visualises it. It thus depends on the activities of the negotiator and its counterpart. Pro-active systems also analyse the process but make inferences and intervene without a request [26].

There already is a conceptual awareness of supporting negotiators in different ways. However, most NSSs do not offer a holistic support. For this reason it is necessary to discuss their components next.

2.2 State-of-the-Art in Negotiation Support Systems

Existing systems such as *Negoisst* [39], *SmartSettle* [44], and *Inspire* [24, 25] offer different support features. Inspire and SmartSettle are firmly rooted in the decision support school. They provide advice limited to a quantitative support to get closer to a Pareto optimal agreement. Support for the other parts of the negotiation process is limited; *Negoisst* is based on a holistic approach offering an integrated sophisticated support for all parts of an electronic negotiation. In particular, *Negoisst* provides decision support, communication support, and document management. For this reason, *Negoisst* is the most sophisticated system and has been used in various negotiation studies [13, 28, 42]. Therefore, Negoisst will be in the focus of the following discussions.

Decision Support is the basic for a quantitative analysis. It rates the offers and counteroffers based on the preferences of the user. The used methods can be static and dynamic; preferences can be fixed or fuzzy. If users are not sure about their preferences, they can use indirect preference elicitation methods such as a conjoint analysis or hybrid methods in case of a change of preferences during the process. Based on the preference elicitation, the negotiator will get an individual linear-additive utility function. Additionally, so called utility graphs can visualize the utility history of all offers and counteroffers made up to now [35, 39].

Communication Support offers support on all semiotic levels (i.e., syntactic level, semantic level, pragmatic level). The main objective is to reduce the disadvantages of the electronic medium and to ensure effective and efficient communication. Negotiation protocols structure and coordinate the message exchange on the syntactic level. To create a common understanding of the negotiation issues, the semantic level is relevant. An ontology underlies the communicative exchanges in Negoisst. The negotiation issues can be integrated into the natural language text preventing misunderstandings about the concepts under negotiations. All concepts are defined and the definitions can be shown to the negotiation partners if need be. Thus, a common background is created. This approach is called semantic enrichment. The pragmatic level deals with the intentions of negotiation partners. Each message in Negoisst is classified by the author using a message type based on the protocol. This shows, e.g. whether a message is meant as an informal enquiry or as a formal request, thereby enabling a clear understanding of how a message is to be interpreted by the recipient. We call this pragmatic enrichment of the messages. Negoisst distinguishes between formal (offer, counteroffer, accept and reject) and informal (questions and clarifications) message types. Formal messages are binding and are linked to a contract. Informal messages help to clarify open points which do not belong to the core negotiation [13, 40, 41].

Document Management offers autonomous dynamic contract generation based on each offer/counteroffer. Based on the pragmatic enrichment and the negotiation ontology, the NSS creates a binding contract. The main objective is to create trust between the parties, because each offer/counteroffer can be accepted and has a final contract as consequence. In addition, contract templates fit the different needs of negotiation contexts. Based on the ontology, an inference machine can simulate the consequences of contract violation [42].

There have been some preliminary discussion about introducing conflict resolution support to sophisticated negotiation support systems [12, 47]. Similar to the focus of decision support in most NSSs, a focus on quantitative analysis and advices is noticeable in these preliminary sketches. However, it is vital to consider also qualitative aspects for conflict resolution. The main challenge is thus to create a holistic concept, including the already established methods of conflict management. The following chapter will give a brief overview over the basic theory of conflict management.

3. Conflict Management

Conflicts are an integral part of any cohabitation, but they don't have to be destructive. De Fleur et al. [9] differentiate between constructive and destructive conflicts, Lewicki et al. [30] between functional and dysfunctional conflicts. Since conflicts are often perceived as a battle situation, easily an internal conflict dynamics unfold and a peaceful, constructive and non-violent solution is not longer possible. Not every conflict can be handled in the same way. They can result of different interests, cognitive abilities, norms and values of the society or objectives [5]. Before a conflict can be handled it is necessary to diagnose and classify it.

3.1 Aspects of conflict diagnosis

The level of escalation in general can be can be influenced by 5 aspects [17]:

1) The relationship and positions between the parties (including formal and informal positions to each other, character constellations, etc).
2) The parties by themselves can be individuals, groups, organizations, etc.
3) Their attitude to the conflict (including aspects like: Do they know how to solve conflicts? Does the party itself thinks, that the conflict can be solved? What are their expectations of a solution? Etc).
4) The specific conflict points, the so called conflict issues, which are brought into.
5) The conflict escalation (= process) by itself, that means: When was the first appearance? How did the intensification proceed? Etc.

All of these aspects are very important for a conflict diagnosis. A holistic diagnosis starts with the recognition of conflicts and its mechanism. The mechanism by itself keeps the level of escalation or pushes it further.

Normally all dimensions have to be put into an overall context to diagnosis the degree of escalation in a conflict. There are several escalation models in the theory. A common used one is the 9-level-escalation developed by Glasl [17]. It divides the escalation in 9 different stages, namely 1) hardening, 2) debate, 3) action instead of words, 4) images and coalition, 5) loss of face, 6) threats, 7) partial destruction, 8) fragmentation and 9) elimination.

In the first stages, there is a polarisation and debate between the parties. During the next stages, the conflict gets more intensive and the parties start to make threats and are afraid of a loss of face. During the first stages, conflicts can be functional. As

discussed, even in successful negotiations there are different positions and opinions. The parties have to exchange information to achieve a mutual understanding. With rising escalation, the conflicts are getting dysfunctional and after stage 5 to 6, they cannot be resolved without the help of a third party any more.

A diagnosis can be of different levels of details. A very detailed diagnosis cannot be guaranteed due to lack of time or missing, distorted, or masked information. On the other hand, it is also not always relevant to include all aspects. In the context of electronic negotiations, the access to interpersonal factors is particularly difficult. The negotiators act in offset locations and communicate over time. For a possible third party, it is far harder to make a complete conflict diagnosis or to influence relevant aspects. In the following, therefore, not all aspects will be considered. Only the conflict escalation (=process), the issues and the relationship are in the focus. Especially the issues play an important role. These are represented in *Negoisst* by specific agenda items and their characteristics and can be quantifiable.

3.2 Common Conflict Resolution Methods

After the diagnosis it is possible to intervene and try to deescalate the conflict. Based on the model of Glasl, there are several conflict resolution methods which can help to resolve the dispute.

3.2.1 Moderation

On levels 1-3, moderation can be used. A moderator tries to solve problems of interaction and questions to the content and the process immediately. This kind of intervention is very useful for conflicts on the first two levels and for simpler conflicts on the third level. To identify the moment in which a moderator no longer has the necessary effect and a consultant would be better, is very difficult. It is a grey area; the cross-over between both methods is fluent. Characteristic for the first escalation levels is the creation of a common understanding and the clarification of the issues. Misunderstandings can be solved by explaining unclear terms and definitions. A moderator can help to structure polarizing issues and less critical issues. Moderation can force the discussion of less critical issues and create awareness for a joint objective at the heart of the negotiation. Interventions can also be behaviour-oriented or focus on the clarification of tasks, roles and functions. Characteristic for moderation is its passiveness and restriction as an adviser. A moderator can make interventions to a certain point of time, but has no power to force the parties to accept the advice. The effect of moderation is short-term. The current conflict development between parties will be aligned and structured.

3.2.2 Consultation

Consultation - on level 3 to 5 in the escalation model - is active, constructive and integral dispute resolution. Active means that the consultant helps the participants to deal with their problems. This does not exclude the consultant helping only one party. The main objective is to support the parties in resolving their conflict alone. First of all the parties need to understand the mechanism of the conflict and that they can influence it (this is meant by "constructive"). The consultation should be perceived as helpful, motivating and not judging. Parties can activate the consultant by asking for an advice.

The objectives are similar to those in moderation with just one difference: Through conflict management by a third party, negotiators have acquired the knowledge and ability to cope largely independently during future conflicts. Interventions within the consultation focus on socio-psychological aspects. This mainly means unconscious blockades, attitudes and patterns. Consultation will help the parties to control their emotions, thoughts and intentions and break out of the spiral of escalation. The conflict potential is analysed and the reduction of conflict attitude is aimed for. Interventions should prevent deadlock situations and increase the flexibility of the parties. Further, the parties should reflect their personal identity, self-perception and the images others have of them [17].

It is very important to point out one aspect: A consultant will not start to create a bilateral interaction between both parties at the same time.

3.2.3 Mediation

From levels 5 to 7, conflicts cannot be resolved alone; a cooperative conflict resolution is not possible any more. The negotiators just agree on one point, namely to prevent further damage by ending the negotiation. Apart from the multiplicity of definitions, mediation can be simply defined as assisted negotiation through a third party [19, 37]. This definition does not include the procedures, methods or tools of mediation. This abstraction offers the opportunity to investigate all the different perspectives of mediation in combination with the new information technologies. Mediation can be used to assist the parties in their negotiation, not to negotiate with the parties. It is a communicative process between all parties with the objective for the parties to generate a solution themselves. Mediation has the following principles which are essential for a complete process [19]:

- The participation is voluntary; each party (also the mediator) can stop the participation at every point.
- The advice is not binding; the mediator has no decision-making authority.
- There is no relationship between the mediator and the negotiators. The third party is neutral and as a consequence independent.
- There are no secrets between the parties. All information has to be disclosed.
- Everybody knows the whole concept, the objectives and the process. Parties have mediation awareness.

The mediation process is similar to the negotiation and has the same sequences. The focus of mediation is on the negotiation outcome. A negotiation problem should be transferred into a successful and integrative agreement. The aim is not only to stop a fight but also to secure an agreement [27]. The impact of mediation on deep rooted conflict attitudes is a side effect, not the main goal. This is also the difference to consultation which focuses on the conflict awareness and the acquisition of conflict prevention. Parties can find a solution by themselves. Mediation uses similar methods but has the objective of finding an agreement.

Compared to mediation, an arbitrator can help to find a solution on conflict levels 6 to 8. In this case the final advice of the arbitrator is binding. This missing flexibility is a fundamental reason for excluding arbitration from further considerations. Traditionally, alternative dispute resolution (ADR) is an alternative to court cases and as a consequence to a binding advice

[14–16]. Mediation with a trusted neutral party is usually a better solution to impasse than being forced to go to court [33].

In this paper the focus is on the first three methods of conflict resolution, namely moderation, consultation and mediation as arbitration limits the flexibility of the negotiators.

4. Conflict Resolution in Electronic Negotiations

To create a concept for conflict resolution support within electronic negotiations, it is necessary to analyse the main challenges and restrictions given by the medium.

To combine conflict resolution methods with an electronic medium is an interdisciplinary task and it is necessary to introduce the definition of a socio-technical system. Electronic conflict resolution is based on two components. The technical system is characterized by software and hardware which enables the communication through technology. The social system includes the relationships between the negotiators, their roles and interaction rules. Both systems influence each other. The way how parties communicate with each other influences their relationship and the interaction rules. As a consequence, it also influences the conflicts between them and vice versa [31]. Last but not least it is obvious that the success of such a support needs a perfect balance between both systems. For further consideration it is necessary to underline the two perspectives.

It is necessary to discuss the fit of technical systems into the different methods and how they can support the conflict resolution process in different ways. Which NSS component supports which method in which way? The other view focuses on the consequences for communication. Fuzzy communication and decision-making can influence the parties in negative ways. Possible tools should not be over-formal or be based on logical models alone. Rather, psychological concepts are also necessary to keep electronic conflict resolution in line.

Apart from the already mentioned considerations, the idea to transfer established conflict resolution methods into an online context is not new. Starting with the growth of the internet, the interest of online dispute resolution has also increased in different ways [1, 2, 21, 36, 37, 45, 46]. In contrast to methods such as moderation and consultation, mediation is already an option of conflict resolution which is popular and has intensively been discussed for more than a decade. Several researchers started to discuss the opportunities and challenges of a so-called online mediation. This research area is wide and can be summarised by three topics: 1) analytical computer support of the (human) mediator, 2) electronic medium for the interaction between mediator and the parties, 3) partial or full replacement of a human mediator by an electronic environment.

Up to now, online mediation is widely discussed, but empirical evidence of its efficiency is rare. One reason could be the missing acceptance within the mediation community [34]. There is no common classification of mediation systems. Many authors focus on discussing the potential of transferring tools and techniques into electronic environment. Most platforms offer a web-based interaction tool enabling parties at different places to come together and to find a solution with the help of a human mediator as third party. These kinds of systems can be synchronous or asynchronous. The degree of computer-support is limited to offering a forum, safe message exchange or several groupware functions [4, 8, 10].

Mediation functions can also be more active. The *Negotiator Assistant* of Druckman [11] has the objective to transfer the research-based knowledge on flexibility and to implement it into a diagnostic tool which monitors the process of an active negotiation. The analysis includes all dimensions of the conflict diagnosis. The process based on 3 functions: 1) diagnosis (questionnaire), 2) analysis (graphical grid) and 3) advice (qualitative suggestions for an improvement). Negotiator Assistant shows whether the parties can expect a fair outcome, a conflict on both sides or more on one side, an impasse etc. Whilst the system includes the main functionalities of mediation it depends on the input of the user and does not provide an autonomous diagnosis.

4.1 Strategy for conflict resolution support

For conflict resolution support in NSSs, it is necessary to formulate an overall strategy. The acceptance of interventions will depend on the point of time the interventions will occur. The success will be higher when the interventions intensify over time. We follow Glasl [17] who concludes that it does not matter whether an intervention is suitable or not. It is necessary to introduce the principles of conflict resolution to the parties, keep them in line during the intervention and evaluate the results with them. These three steps can be formulated as 1) Preparation, 2) Intervention and 3) Reflection. Preparation has the objective to create acceptance and process understanding at the negotiator. It is indispensable that the parties agree on an intervention and that they are prepared for their own input into the process. Preparation clarifies the roles, rights and responsibilities [17]. As a consequence this means that the parties need detailed briefings of the NSS and especially of the basic components and the additive functions. They have to accept that the system offers multi-level support in the form of interventions. Only if they agree to such an approach can the socio-technical system offer balanced support.

Intervention would include special strategies like moderation, consultation and mediation. Each method would include a different sub-process, but they build up on each other. A typology for intervention and its scope will be given in the next chapter.

Reflection has the objective to consolidate the negotiators and help them to understand the outcome of the whole process. The effect of intervention will be internalised and can be recalled in similar situations. In chapter 3.2.2 it was already mentioned that parties get used to the process and can resolve conflicts in future without the help of a consultant. Reflection as a part of the holistic strategy includes an overview of all interventions and their results.

4.2 Typology of interventions

In chapter 3.2, the main methods of conflict resolution have been explained. Based on these considerations, we will formulate a 3-stage intervention typology which fits the theoretical aspects of conflict management (cf. sections 3.1 and 3.2). This approach is user-driven in that .the user determines himself/herself at which point which kind of help is required. In addition, we use the nine step escalation model of Glasl [17]. It shows the general intervention methods related to the level of escalation. The reader should bear in mind that we deal with conflict management in *electronic* negotiations. Therefore, some cues are missing such as body language, tone of voice, gestures etc. (as explained in section 1). This can both escalate a conflict (e.g. when a partner assumes a reaction on the partner's side) and deescalate a conflict (e.g. because threatening behaviour is less obvious, a partner can

think before replying etc.). Therefore, it is important that each partner decides on the individual perceived level of conflict.

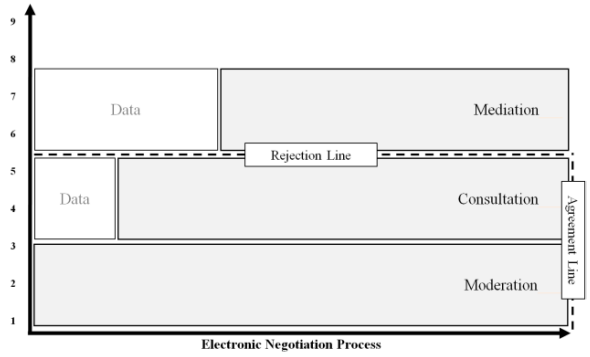


Figure 2: Multi-level intervention during the electronic negotiation

It will now be discussed how the intervention methods of moderation, consultation, and mediation will relate to the different aspects of conflict. Normally these methods would intervene in different intensity on all aspects. We will restrict the intervention on selected points - marked in the following table 2 with an "x"- and categorise them with the considerations discussed in chapter 2.1 (cf.[26]).

Table 2: Scope of interventions in electronic negotiations

		Intervention			
		Moderation	Consultation		Mediation
			I	II	
Conflict Aspects	Relationship				x
	Issues		x	x	x
	Parties				
	Behavior			x	x
	Process	x	x	x	x
Information exchange		No	No	Yes	Yes
Involvement		Passive	Active		Pro-Active

Moderation

Moderation has been described as a method that helps the participants to structure the negotiation process. *Negoisst* offers sophisticated communication support. This kind of support already moderates the negotiation, because it constrains the users in several interaction rules. On the first level, a negotiation protocol structures the message exchange. Negotiators know that they can exchange messages in an alternating manner. Thus interruptions as in face-to-face negotiations or chats are prevented. Additionally, one party cannot flood the other party with streams of messages. A moderator also helps the participants to reduce misunderstandings. *Negoisst* offers semantic and pragmatic enrichment of messages with the main objective to reduce misunderstandings. Messages need a type declaration before sending, so the recipient will know whether the message is meant as an offer, request, question, etc. Additionally ontologies can help to create a mutual understanding. Ontologies are formally ordered representations of a set of concepts and relations between them in a particular subject area [22]. By explaining the concept through the NSS, parties avoid wrong interpretations. So we can confirm that an intervention in form of moderation is already given in *Negoisst*.

Apart from moderation, the perceived level of conflict can still continue to rise. In this case, as shown in figure 2, there is still some range before the rejection line, where the parties can negotiate without reaching a point where they need the help of a neutral third party.

Consultation

In this case, a possible consulting component can start to support parties. The consultation process can be divided into three steps: 1) Diagnosis, 2) Analysis and 3) Advice. The request for a consultation will be done by the parties themselves. We divide the consultation into two aspects which are very important for further approaches. The first two steps of the consultation approach will focus on the aspect that no exchange of information between the negotiators will occur. This happens in stage 1 and 2:

Diagnosis and Analysis

First of all we have to clarify the scope of the consultation: the intervention will focus on the negotiation issues. During an escalation, the flexibility - represented through the issues - decreases. In NSS, the issues are represented through the agenda items. This means, that the main points of negotiation and their characteristics are known by the parties. When they exchange offers/counteroffers they attach specific agenda items which mostly represent their preferences. Based on the preference elicitation, the system generates utility values for each offer/counteroffer. If a party makes few concessions during the negotiation, this can lead to frustration on the other side. This behaviour can be caused of different reasons [17]:

- It can be part of the negotiation strategy
- The other party already cannot make further concessions.
- Issues can be linked with each other
- They have a special (unknown) importance.

All this can lead to a fixed and extreme perception on the issues. In NSSs, the negotiators have two ways to transfer their attitude, opinion and willingness for a successful and integrative agreement: First they exchange offers/counteroffers in form of a specified agenda list (represented through issue values). Secondly they can use written communication to add arguments (represented through words) to their agenda. The objective of a scope on the issues would be to show the participants their own concession history compared to the negotiation history and the concession done by the counterpart. This one sided data analysis offers the opportunity to break up hardenings and increase flexibility – one of the objectives of consultation mentioned in chapter 3.2.2. – based on reflection.

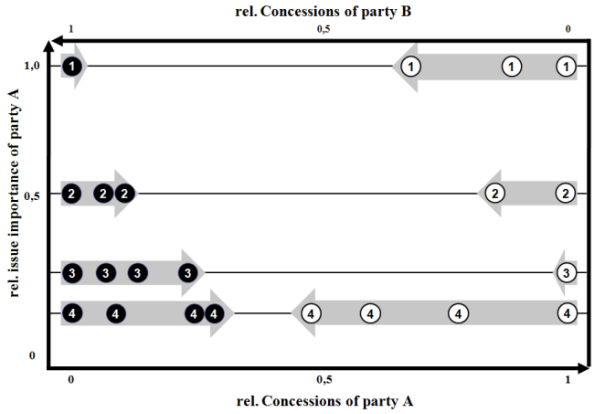


Figure 3: Relative concession history of party A after 8 messages

To this point no data has been revealed between the negotiators. Based on the relative concessions and the relative importance, a graphical preparation of the concession history can be given up to now. In this simple example we have static preferences and make a diagnosis after 8 exchanged messages (4 offers/counteroffers on each side). We compare the relative concessions in each issue with the concessions the counterpart made. Figure 3 shows that party A made no concessions for issue 1, while the most concessions can be reported on both sides for issue 4. Compared to that, party B made several concessions for issue 1 (the most important one for party A), but no concessions for issue 3. Based on this development it would be our objective to score the issues automatically to a flexibility degree and estimate the consequences of an issue change for the counterpart. As a consequence we could score the results and transfer it into a portfolio. Figure 4 shows the exemplary portfolio. It offers party A now following interpretation: Issue 1 is a strategic issue. It is from high importance for party A, possible concessions should be done carefully. Issue 3 is integrative, because it seems that issue 3 is of great importance for the counterpart. Cause of their concession history, issue 2 and 4 can be rated as problematic and unproblematic.

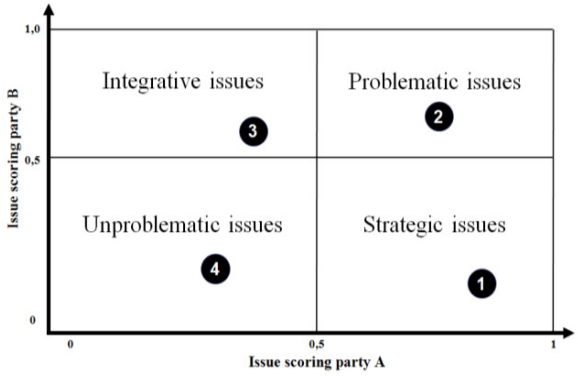


Figure 4: Issue scoring and visualization from the view of party A.

The main challenge during the analysis is to animate the negotiators to make an interpretation themselves. The scoring does not include the total utility development! Initially one party asks for an analysis and the system offers one. The objective of this procedure is that the negotiator can interpret the given data and is not forced to accept a given advice.

Advice

Before we discuss the possibilities of giving an advice it is necessary to point out, that we focus again on the issues. Consequently, a quantitative advice in form of suggestions of possible new offers/counteroffers is offered. It is also possible to provide communication advice such as "Have you told your partner that issue xyz is very important to you?" or preference advice such as "Are you sure, that issue n is so important for you?"

Bargaining advices are already part of some NSSs. Software agents monitor and analyses the negotiation process to consider a possible new concession based on the last bargaining step [6, 7].

Vetschera [47] introduces an alternative approach called "Analytical Concession/Advising Technology Model (AC-AT)" which exactly fits the needs of staged and holistic conflict resolution approach: The concept focuses on the entire bargaining process rather than on a single bargaining step of one negotiator [47]. It also assumes that negotiators communicate via offers/counteroffers and that they represent negotiation issues. Based on the issue values it uses an optimization model to predict a negotiation path which is similar to the negotiators concession path. A negotiator can thus preserve the current strategy. Afterwards the approach starts to generate a new offer which is inside of the concession cone. The AC-AT offers the opportunity to choose the degree of "toughness" for the next offer generation. Therefore, the negotiator will get not only an advice but can also choose the cognitive effect of the advice. This suggests more freedom on the negotiator's side which as a result raises the acceptance of the advice and with it the consultation component.

Up to now, the consultation component does not reveal information to the other negotiator However, the AC-AT needs the utility values of the opponent to generate the advice. This implies, that before such an advice can be given, both parties must agree to unmask their preferences to the consultant (not to the counterpart!). So, as an example, party A's perceived level of conflict continues to rise. Even the analysis of the concessions and the joint reflection does not help to increase the flexibility. Party A will ask now for an advice. The system replies that an advice would be possible but requires publication of the preferences to the consultant. So party A has to wait until party B agrees to the publication. As a consequence, both parties can use the consultant to generate advices. Additionally party B – whose perceived level of conflict is probably not as high as party A's – gets an awareness of the counterpart's frustration. This knowledge might help to review one's own conflict behaviour, another consultation aspect mentioned in table 2.

By definition, the exchange of information (here information about preferences) is also one aspect of mediation. In this case, the private information will only be exchanged on the access level of the consultation component. The counterpart will not see the preferences.

Mediation

If consultation does not help to deescalate the conflict and lead to an agreement, one of the parties will reach a point of terminating the negotiation without agreement. At this point, they could ask for active mediation or a pro-active mediation component would

step into the negotiation process before the negotiation reaches such an escalation point. Both options set high requirements for the system, the mediation process. Negotiator Assistant has been combined with *Negoisst* for mediation advice [12]. Empirical experiments (without the considerations to consultation made in this paper) demonstrated that there is a traceable influence of electronic mediation on negotiators behaviour and leads to fewer rejections in negotiations [12]. At the same time, it was shown that the negotiators expect more than a mere self/reflection and general diagnosis from a mediation tool. As mentioned before, mediation is assisted negotiation and as a result a very communicative process. To support negotiators on this level means also to support the communication. When the breaking up of hardenings (lack of flexibility) does not lead to the favoured result, the conflicts between the parties have another source. Aspects such as relationships and conflict behaviour complete the already treated conflict aspects "issues" and "process". As a consequence, the mediator has to improve the communication quality within the electronic negotiation. To research the transfer of possible mediation functions to electronic negotiations, we suggest a staged research classification of mediation in asynchronous NSS:

Stage 1:

Negotiators can request mediation. A human mediator steps into the electronic negotiation and executes the mediation process. *Negoisst* will offer an extended role model, negotiation (=mediation) protocol and new message types to support the whole mediation process by itself. The participants and the mediator have to start at the beginning of the process and complete each stage of it.

Stage 2:

The process is almost the same but *Negoisst* automates parts of the mediation process: The closure of mediation contract can be automatically generated by the document management support. An inventory of issues is already provided. The NSS knows the main points of conflicts and can visualize them in a flexibility graph as described. Afterwards, the mediator can start the treatment of issues. The considerations of a consulting advice based on the AC-AT also take in the development and evaluation of alternatives stage effect. A mediator can use these analytical functions to diagnosis and identifying possible solution ranges and integrate them in the interaction. Already given decision support functions can help the participants to evaluate the new options. In a final step, the group designs the solutions and hopefully finds an agreement.

This form of a sophisticated computer-supported mediation should be compared to the results of research in stage 1 with the objective to create on the one hand knowledge about the acceptance and impact of electronic mediation and possible improvements, on the other hand it would be possible to get data – and as a result a knowledge base – of conflict processes and their communicative characteristics.

Stage 3:

In a next step, the human mediator would be replaced and the concept to full computer mediation would be introduced. In this case, a qualitative analysis of communication quality is indispensable. Given concepts of text-mining and natural language perspective already offer the potential to monitor communication processes. In combination with the communication quality model developed by Duckek [13], it is possible to get a live monitoring of communication quality within electronic negotiations. The opportunities in this case are not only the diagnosis and analysis of written messages; also the qualitative advice given by the system should fit the individual circumstances of a negotiator. There will be no replacement of human mediators and no acceptance of such functions until the final qualitative advice does not exactly fit the cognitive needs. A challenge which even experienced human mediators cannot always cope.

5. Summary and Conclusion

It was the objective of this paper to show, that several methods of conflict resolution methods for face-to-face negotiations exist, and that they offer multidimensional advantages (e.g. costs and time saving, better relationships, etc.). Furthermore, we introduced a framework of conflict resolution support in electronic negotiations.

Conflicts occur in any type of negotiation. There are new challenges and new opportunities imposed by the electronic medium when it comes to conflict resolution support in electronic negotiations. We first gave a brief theoretical overview of current Negotiation Support Systems (NSSs) and their classification within electronic negotiations. Following the main characteristics and a description of the components communication support, decision support and document management. The discussion of the current state of the art discloses that already a few years ago different researchers had an idea of more "involvement" by the NSS in the negotiation process. But holistic realisations are still rare due to the difficulties of transferring a very communicative and staged process into an electronic context, which is characterised by several restrictions. To get a better understanding of this "communicative and staged process", we introduced the basic concepts of conflict management. The last one defines aspects, which influences the level of escalation, such as conflict issues, conflict process, conflict attitude, the parties involved in the conflict and their relationships. Existing escalation models offer the opportunity to intervene at a specific level. In early stages, moderation can be useful to keep the conflict process in line and to reduce misunderstandings. If conflicts continue rising, consultation can be deescalating. It focuses on hardenings and tries to increase the flexibility of negotiators. If this still does not help, the negotiators can ask for the help of a neutral third party. This process is called mediation and can be understand as an assisted negotiation. All this methods have their characteristics and use different methods to deescalate conflicts.

First of all, our considerations focused on a holistic approach. It is not only important to choose the right method, but rather to educate the parties additionally about the process and its possible impact. Only if negotiators accept this type of conflict resolution, the actual intervention can be successful. Last but not least the focus of our considerations was still on the specific intervention methods. Before formulating an approach, we discussed the current state of the art in existing online conflict resolution methods. One finding was that especially the online-mediation already enjoys high popularity within different research areas. Regarding the possible area of application within NSS the implementations get rare.

Based on the idea, that interventions have to build up on each other, we started to formulate and characterize a 3-stage intervention model for NSS, characterized by following aspects:

Scope of intervention, degree of involvement, exchange of information, quantitative and/or qualitative advice. The intervention starts with moderation on low conflict levels, continues with consultation and ends in mediation. Due the discussion we discussed that the basic concept of NSS already matches the requirements of moderation. As a consequence, negotiators using *Negoisst* to negotiate electronically and asynchronously already use several concepts to structure the negotiation and to reduce misunderstandings (negotiation protocol, semantic and pragmatic enrichments, ontology based agenda items …).

The considerations to a consultation component are characterised by the point of intervention and the possible exchange of information. We assume, that a consultation is requested by the negotiators. The intervention by itself will be structured in diagnosis, analysis and advice. Especially for a possible advice we differentiate between information exchange or no information exchange between the involved parties. In the first two stages the collected data will be analyzed and presented to the users. The objective is to provoke self-reflection and solve hardenings without giving any advice until the user requests one. If this happens, we assume to offer quantitative advice in form of possible bargaining steps without information exchange (TIT-for-TAT strategy) or with information exchange. The last option would base on the AC-AT concept, which offers multiple advantages in the prediction of possible concession in combination with the individual "toughness" of a negotiator. The mutual information exchange (in form of preferences) additional has the effect, that the counterpart gets an awareness of the frustration of his partner.

Last but not least the participants can use the option to go into mediation together. Existing literature shows, that there are no studies of the acceptance and consequences of an online mediation executed through a human mediator. Therefore we suggest first of all integrating a human mediator into the negotiation process and analysis the outcome. Based on this finding it is possible to add (and replace) functions and process steps of the mediation to make it more efficient. Along these two steps it is possible to build up a knowledge base for a full computer-mediated negotiation. As a consequence this kind of support would be pro-active, what means, that the software monitors the ongoing negotiation (text-mining, concession analysis, etc.) and step into the process at a certain point where the perceived conflict level of one or more parties is significant for an intervention.

We have shown that our 3-stage model provides the potential for effective and efficient conflict resolution support in electronic negotiations resulting in more agreements and fewer unsuccessful terminations of negotiation processes.

6. References

[1] Adler, P. 2005. eMediation - Können wir die MediatorIn einsparen? Überlegungen zum Einsatz des Internets und der EDV. In *Wirtschaftsmediation zwischen Theorie und Praxis*, M. Gruber, Ed. Schriften zur Rechtspolitik 24. NWV Neuer Wiss. Verl., Wien, Graz, 23–40.

[2] Behr, N. 2005. *Konfliktlösung im Internet*. Freie Univ., Diss.--Berlin, 2004. Beiträge zum Informationsrecht 12. Duncker & Humblot, Berlin.

[3] Bichler, M., Kersten, G., and Strecker, S. 2003. Towards a Structured Design of Electronic Negotiations. *Group Decision and Negotiation* 12, 4, 311–335.

[4] binary-objects. 2010. *Die Discourse Machine*. www.discourse-machine.de/. Accessed 10 August 2010.

[5] Chang, M. K. and Woo, C. 1994. A Speech-Act-Based Negotiation Protocol: Design, implementation, and Test Use. *ACM Transactions on Information Systems* 12, 4, 360–382.

[6] Chen, E., Kersten, G., and Vahidov, R. 2004. An e-marketplace for agent-supported commerce negotiations. *InterNeg Research Papers,* 08/03.

[7] Chen, E., Kersten, G., and Vahidov, R. 2005. Agent-supported Negotiations in the E-marketplace. *International Journal of Electronic Business* 3, 1, 28–49.

[8] Daube, J. 2010. *umgangskalender.de*. https://www.umgangskalender.de/. Accessed 10 August 2010.

[9] DeFleur, M., Kearney, P., Plax, T., DeFleur, M., DeFleur, M. H., and DeFleur, M. L. 1. 2005. *Fundamentals of Human Communication*. *Social science in everyday life*. McGraw-Hill; McGraw Hill, Boston.

[10] Die Verbraucher Initiative e.V. 2010. *ombudsmann.de - die neutrale Schlichtungsstelle für den Online-Handel*. http://www.ombudsmann.de/. Accessed 10 August 2010.

[11] Druckman, D., Druckman, J., and Arai, T. 2004. e-Mediation: Evaluating the Impacts of an Electronic Mediator on Negotiating Behavior. *Group Decision and Negotiation* 13, 6, 481–511.

[12] Druckman, D., Koeszegi, S., Schoop, M., van der Wijst, P., Vetschera, R., Dannenmann, A., Duckek, K., Filzmoser, M., Gettinger, J., Mitterhofer, R., and Reiser, A. 2010. *Acceptance and Evaluation of Automated Mediation in e-Negotiation*. Proceedings of the 11th Group Decision and Negotiation Conference in Delft, Netherlands. The Center of Collaboration Science, Omaha.

[13] Duckek, K. 2010. *Ökonomische Relevanz von Kommunikationsqualität in elektronischen Verhandlungen*. Univ., Diss.--Zugl.: Hohenheim, 2009. Gabler Research. Betriebswirtschaftlicher Verlag Gabler; Gabler, Wiesbaden.

[14] Duffy, K. G., Grosch, J. W., and Olczak, P. V. 1991. *Community mediation. A handbook for practitioners and researchers*. Guilford Press, New York.

[15] Folberg, J. and Taylor, A. 1990. *Mediation. A comprehensive guide to resolving conflicts without litigation*. The Jossey-Bass social and behavioral science series. Jossey-Bass, San Francisco, Calif.

[16] Friedman, G. J. and Himmelstein, J., Eds. 1996. *Die Scheidungs-Mediation. Anleitungen zu einer fairen Trennung*. rororo Sachbuch 9944. Rowohlt, Reinbek bei Hamburg.

[17] Glasl, F. 2004. *Konfliktmanagement. Ein Handbuch für Führungskräfte, Beraterinnen und Berater*. Organisationsentwicklung in der Praxis 2. Haupt [u.a.], Bern.

[18] Haft, F., Schlieffen, K. von, Bamberger, H. G., and Haft-Schlieffen, Eds. 2009. *Handbuch Mediation*. *Verhandlungstechnik, Strategien, Einsatzgebiete*. Beck, München.

[19] Hauser, C. 2002. *Eine ökonomische Theorie der Mediation*. Univ., Diss--Fribourg, 2002. Luzerner Beiträge zur Betriebs- und Regionalökonomie 9. Rüegger, Chur.

[20] Hertel, A. von. 2005. *Professionelle Konfliktlösung*. *Führen mit Mediationskompetenz*. Campus-Verl., Frankfurt/Main.
[21] Hörnle, J. 2003. Online Dispute Resolution: The Emperor's New Clothes? *International Review of Computers and Technology* 17, 1, 27–37.
[22] Horrocks, I. and Tessaris, S. 2002. Querying the Semantic Web: A Formal Approach. In *First International Semantic Web Conference*. Springer, 177–191.
[23] Kals, E. and Webers, T. 2001. Wirtschaftsmediation als alternative Konfliktlösung. *Wirtschaftspsychologie,* 2, 10–16.
[24] Kersten, G. 2002. The Science and Engineering of E-negotiation: Review of the Emerging Field. *InterNeg Research Papers,* 05/02.
[25] Kersten, G. 2004. E-negotiation systems: Interaction of people and technologies to resolve conflicts. *InterNeg Research Papers,* 08/04.
[26] Kersten, G. and Lai, H. 2007. Negotiation Support and E-negotiation Systems: An Overview. *Group Decision and Negotiation* 16, 553–586.
[27] Kim, N. H., Wall, J., Sohn, D.-W., and Kim, J. 1993. Community and Industrial Mediation in South Korea. *Journal of Conflict Resolution* 37, 2, 361–381.
[28] Köhne, F. 2007. *Electronic Negotiation Support Systems and Their Role in Business Communication*. Vdm Verlag Dr. Müller, Saarbrücken.
[29] Köhne, F., Schoop, M., and Staskiewicz, D. 2005. An Empirical Investigation of the Acceptance of Electronic Negotiation Support Systems. In *Proceedings of 13th European Conference on Information Systems in Regensburg, Germany*.
[30] Lewicki, R. J., Barry, B., and Saunders, D. M. 2010. *Negotiation*. McGraw-Hill/Irwin, Boston.
[31] Märker, O. and Trénel, M. 2003. Online-Mediation: Konfliktvermittlung in neuem Kleid - eine Einführung. In *Online-Mediation*. *Neue Medien in der Konfliktvermittlung ; mit Beispielen aus Politik und Wirtschaft*, O. Märker and M. Trénel, Eds. ed. sigma, Berlin, 7–20.
[32] Pesendorfer, E.-M., Graf, A., and Koeszegi, S. 2007. Relationship in electronic negotiations: Tracking behavior over time. *Zeitschrift für Betriebswirtschaft* 77, 12, 1315–1338.
[33] Peter, B. 2002. *Wirtschaftsmediation: Preiswert streiten*. http://www.zeit.de/2002/44/200244_wirtschaftsmedia.xml. Accessed 15 June 2010.
[34] Pichler, J. 2005. *Empirie zu Online Mediation - eMediation (ODR) und eMediation im eCommerce (ODReC) und weitere Überlegungen*. Accessed 19 March 2010.
[35] Reiser, A. and Schoop, M. 2010. *The Use of Dynamic Preference Elicitation for Negotiations with Incomplete or Missing Information*. Proceedings of the 11th Group Decision and Negotiation Conference in Delft, Netherlands. The Center of Collaboration Science, Omaha.
[36] Rohde-Liebenau, B. 2005. Online Mediation. How It Can Contribute to Justice. *Médiation en Europe – Èchanges sur les pratiques* 2, 212–227.
[37] Rule, C. 2002. *Online dispute resolution for business*. *B2B, e-commerce, consumer, employment, insurance, and other commercial conflicts*. Jossey-Bass, San Francisco, CA.
[38] Schoop, M. 1. A Language-Action Approach to Electronic Negotiations. *Journal of Systems, Signs and Action* 2005, 1 (1), 62–79.
[39] Schoop, M., Jertila, A., and List, T. 2003. A Negotiation Support System for Electronic Business-to-Business Negotiations in E-Commerce. *Data and Knowledge Engineering* 47, 3, 371–401.
[40] Schoop, M., Köhne, F., and Ostertag, K. 2010. Communication Quality in Business Negotiations. *Group Decision and Negotiation* 19, 2, 193–209.
[41] Schoop, M., Köhne, F., and Staskiewicz, D. 2004. An Integrated Decision and Communication Perspective on Electronic Negotiation Support Systems: Challenges and Solutions. *Decision Systems* 13, 4, 375–398.
[42] Staskiewicz, D. 2009. *Document-centred electronic negotiations*. Univ., Diss.--Hohenheim, 2009. Informatik. Verl. Dr. Hut, München.
[43] Ströbel, M. and Weinhardt, C. 2003. The Montreal Taxonomy for Electronic Negotiations. *Group Decision and Negotiation* 12, 2, 143–164.
[44] Thiessen, E. and Soberg, E. 2003. Smartsettle described with the Montreal Taxonomy. *Group Decision and Negotiation* 12, 165–170.
[45] Trénel, M. 2001. Online-Mediation: Nutzen, Software und Strategien. *Wirtschaftspsychologie,* 2, 45–50.
[46] Turel, O., Yuan, Y., and Rose, J. 2007. Antecedents of attitude towards online mediation. *Group Decision and Negotiation* 16, 6, 539–552.
[47] Vetschera, R., Filzmoser, M., and Mitterhofer, R. 2010. *Analytical Concession-Advisor Technology (ac-at)*. Proceedings of the 11th Group Decision and Negotiation Conference in Delft, Netherlands. The Center of Collaboration Science, Omaha.

The Impact of National Culture on Control in IS Offshoring Projects

Jakob Heumann
University of Erlangen-Nuremberg
Lange Gasse 20
90403 Nuremberg
Germany
+49 911 5302-901
jakob.heumann@wiso.uni-erlangen.de

Dr. Martin Wiener
University of Erlangen-Nuremberg
Lange Gasse 20
90403 Nuremberg
Germany
+49 911 5302-859
martin.wiener@wiso.uni-erlangen.de

Dr. Ulrich Remus
University of Canterbury
Private Bag 4800
Christchurch
New Zealand
+64 3 364-2618
ulrich.remus@canterbury.ac.nz

ABSTRACT
To our best knowledge, this is the first quantitative study on the impact of the supplier's national culture on the client's choice of control modes in IS offshoring projects. A survey-based field study was conducted, using a client-supplier matched pair as the unit of analysis. This approach allowed for the examination of the direct control relationship within 46 unique matched pairs. The study results offer empirical evidence that the supplier's national culture (i.e., power distance and time perception) affects the client's choice of controls in IS offshoring projects. However, the supplier's cultural background seems to play a less important role than suggested by prior research. These results (1) adapt previous research to the IS offshoring context, (2) enhance prior findings by establishing a more detailed understanding about the cultural influence on the exercise of controls as well as by confirming the significance of project size–an influencing factor that has previously shown mixed results, and (3) incorporate new constructs and measures in developing an integrated model that should be broadly applicable to other IS project contexts.

Keywords
IS offshoring, control theory, managerial control, national culture, cultural dimensions, matched pair survey, partial least squares.

1. INTRODUCTION
IS offshoring (ISO), defined as the relocation of IS services to a captive or third party organization in a foreign, mostly low-wage country [42] continues to be an important global trend [25]. ISO promises many benefits, such as cost reduction, access to highly skilled professionals, and time-to-market reduction (e.g., [42]).

Despite the manifold benefits, companies' ISO experiences have not been consistently positive and often ISO projects fail [1]. These project failures can often be traced back to national cultural differences between client and supplier [38; 44]. This is not surprising as ISO projects involve actors from different countries and cultures, working together in complex, intensive, and dynamic activities that require close cooperation and coordination [27]. In particular, many risks associated with ISO projects, such as blocked knowledge transfer, differences in the interpretation of processes, barriers between individuals, and lack of acceptance of foreign behaviors, they all may result from cultural distance [10].

One powerful approach for managing client-supplier relationships in ISO projects is exercising control [27; 28], which refers to any attempt to motivate individuals to behave in a manner consistent with organizational objectives [21]. Because ISO "entails complex issues of geographical, cultural, and lingual differences", Rustagi et al. ([45], p. 139) stress the need for research on control in ISO. Here, especially the client's control over the supplier is an important instrument to ensure project success [9]. However, the use of control is complicated by cultural differences which impact the coordination of the supplier employees as well as the cooperation with them [59]. In a recent study it was found that espoused cultural values, such as norms, values, and beliefs, affect control choices [28]. In another study, Beck et al. [3] found that formal project management and control mechanisms are mainly driven by the cultural intelligence of the client's ISO project manager. High cultural intelligence might lead to a better understanding of the controllee's cultural values and thus enables her/him to better select and execute suitable modes of control [3]. It is thus important to examine how the supplier's national culture affects the client's exercise of controls in ISO projects.

Previous literature has already acknowledged the important role of national culture in ISO in general (e.g., [38]). However, two gaps are still obvious: First, IS outsourcing and ISO research is still primarily based on anecdotal evidence using qualitative (interpretive) case studies as main research method [57]. Thus, there is still a need to validate and complement these important findings by quantitative analyses. Second, so far there has been no empirical study which has examined the influence of national culture on the choice of control modes. Although Narayanaswamy and Henry [37] proposed an initial set of propositions regarding the relationship between three of Hofstede's cultural dimensions and control modes used in offshore-outsourced IS development projects, they did not empirically test these propositions. This study seeks to fill these gaps by examining how the client's selection of controls relates to the supplier's national culture. We address this question by developing a research model which

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

integrates control theory with prior literature on ISO and national culture. To test our model, we perform a survey-based field study using a client-supplier matched pair as the unit of analysis.

2. LITERATURE REVIEW

2.1 Culture Theory

In this study, we adopt the value-based definition proposed by Hofstede [18]. He defines culture as "the collective programming of the mind that distinguishes the members of one group or category of people from another" (p. 11). This definition describes culture as a set of value patterns that are shared by individuals and influence how they behave (e.g., [18; 55]). Values refer to relationships among abstract categories that are characterized by strong affective components and imply a preference for a certain type of action [22]. They provide individuals with fundamental assumptions about how things are. Based on these assumptions, researchers attempt to generalize the patterns of different cultures into several dimensions. The defined cultural dimensions provide a framework to measure and compare the cultural differences from one country (or group) to another [43].

ISO is situated within a complex and multi-leveled socio-cultural context, which comprises not only the national (societal) level but also organizational, professional (functional), team, or individual levels [22]. Thus, national culture may not be the only type of culture which influences the choice of control. However, cultural differences on the national level are presumed to constitute a predominant factor influencing ISO project control [33]. Wilkins and Ouchi [58] argue, for instance, that "the learning of organizational 'culture' [is] neither as deep nor as immutable as the anthropological metaphor would suggest" (p. 479). We believe that this rationale also applies to the professional and team level of culture.

Even though culture is seen a collective phenomenon and, thus, irreducible to the individual level of analysis it can only manifest itself through the individual [50]. We therefore argue "that individuals espouse national cultural values to different degrees", treating these values as an individual difference variable ([48], p. 680). This perspective allows us to analyze cultural differences on different levels without presuming an aggregated type of culture on the national level. By disaggregating the monolithic national culture construct into its espoused value dimensions it is possible to use it as individual difference construct in individual-level research such as control theory. As a result, it is possible to capture the nuances and distinct effects of the relationship between each cultural value and control, which otherwise treated as monolithic construct could have been shown as culturally invariant [48].

2.2 Control Theory

Our study adopts a behavioral view of control. This view implies that the controller takes some action in order to regulate or adjust the behavior of the controllee [26], and draws upon organization and agency theories consistent with prior studies in IS (e.g., [9; 26; 27; 28; 29]), organization design (e.g., [13]), and marketing (e.g., [21]).

A control situation typically involves an individual exercising control (the controller) and a target of control (the controllee) [28]. However, this distinction sometimes becomes fuzzy [9], in particular in an ISO context. For instance, the controller and the controllee may not be individuals but teams of individuals representing their organizational unit or organization respectively. Furthermore, in an ISO project the supplier project manager may be controlled by the client and, in turn, may control the supplier project team members. However, for the specific focus of this study, the distinction between controller (in terms of an individual in the client organization) and controllee (in terms of an individual in the supplier organization) remains largely valid.

The behavioral view of control presumes that the controller uses certain mechanisms to exercise four modes of control, which may broadly be divided into formal and informal controls [27].

There are two modes of formal control: behavior and outcome control [13]. In behavior control, the controller seeks to influence the process to achieve the desired outputs by explicitly prescribing specific rules and procedures, monitoring their implementation, and rewarding the controllee based on the extent to which the implementation complies with these rules and procedures [26]. This is achieved through the use of mechanisms that either specify appropriate behaviors, or allow for the evaluation of the controllee's behavior [27]. In outcome control, only the outputs (both interim and final) are measured and evaluated. Here, the controller explicitly defines specific goals and rewards the controllee for meeting these goals [13; 27]. Outcome control is exercised through mechanisms that specify or measure desired outcomes [9].

Informal control modes are clan and self-control. Clan control refers to mechanisms that minimize the differences between the controller and controllee's objectives [13] by "promulgating common values, beliefs, and philosophy within a clan, which is defined as a group of individuals who are dependent on one another and who share a set of common goals" ([27], p. 217). According to this definition, it is questionable whether the clan control construct can be applied to ISO projects as the client-supplier relationship is assumed to be adversarial [32]. Thus, we adopt a different interpretation of clan control "refer[ring] to a situation in which the traditional relationship is replaced by a scenario where the two organizations perceive themselves as having a common, shared goal" ([9], p. 293). Unlike clan control, self-control is a function of intrinsic motivation [36] as well as individual standards and objectives [21]. Even though controllees control themselves by their own actions (e.g., setting their own goals) [26], the controller can use control mechanisms to assist and promote the exercise of self-control by the controllee.

Controllers often use the four control modes in combination, creating a portfolio of controls [21; 27]. Within a portfolio, each control mode can itself be implemented through multiple control mechanisms [27]. The choice of controls is influenced by different factors in the project, stakeholder, and global contexts [27; 28]. Factors related to the global context include priority differences among stakeholders from different countries, as well as geographic, time zone, and cultural differences. In this context, it is the cultural factors influencing the choice of controls that are still not well understood. These relationships are discussed in more detail in the following section.

3. RESEARCH MODEL AND HYPOTHESES

Our research model considers the relationship between the controller's choice of controls and the controllee's national culture. The model draws on five cultural dimensions which can be used to define national culture. Here, however, we do not focus on the cultural characteristics of specific nations. Instead, we attempt to understand how the characteristics of the underlying

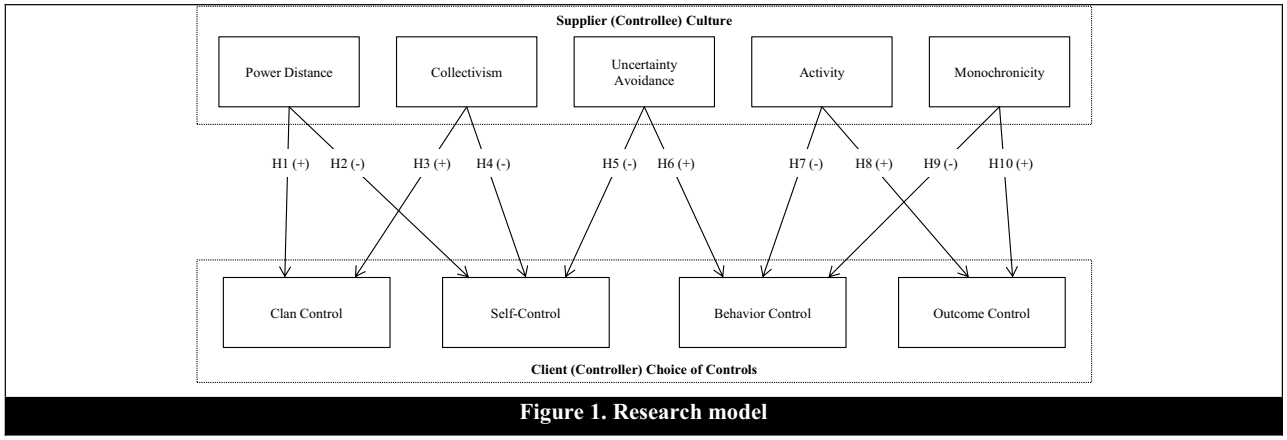


Figure 1. Research model

espoused cultural values interact with ISO project control [14; 48]. Figure 1 shows the different constructs and hypotheses.

According to Carmel [6], cultural dimensions are useful in modeling culture-related issues in globally distributed projects. From the variety of dimensional models existing at the national level, the following five dimensions were selected: *power distance*, *individualism/collectivism*, *uncertainty avoidance* [18], *activity/passivity* [35; 52], and *mono-/polychronic time perception* [16]. The selected dimensions were evaluated in terms of their suitability to explain cultural characteristics that may affect the choice of different control modes in ISO projects. For example Hofstede's masculinity/femininity dimension was dropped as it seems to be highly correlated to age [14]. Some of these dimensions have already been applied to ISO-related research, including power distance [59], individualism/collectivism [59], uncertainty avoidance [49], and activity/passivity [59]. In addition, the model was enhanced by the dimension mono-/ polychronicity because different views about timelines, deadlines, work rhythms, and/or punctuality may impose challenges to the coordination (and control) of globally distributed projects [46].

3.1 Power Distance

Power distance is defined as the extent to which the less powerful members of institutions and organizations within a society expect and accept that power is distributed unequally. In high power distance cultures superiors make decisions without consultation with subordinates. Employees are fearful of disagreeing with their superiors and expect to be told what to do. For example, it was found that in rather high power distance cultures like India, offshore team members have difficulty in saying no [38; 59]. In contrast, in cultures that are low in power distance, relationships between superiors and subordinates are more participative and egalitarian and subordinates are likely to contradict their superiors directly and do not expect to be told what to do [17]. They also participate more in decision making activities and prefer a consultative relationship with their superiors [17].

In particular, in high power distance cultures the controllee feels less comfortable in debating and contradicting. They tend to pay more attention to the opinions of others and thus tend to be more attuned to social norms [48], typically facilitated by clan control. Here, clan control mechanisms, such as rituals, ceremonies and socialization might mitigate the difference in objectives between controller and controllee [9], triggering compliance, identification and internalization effects [23], although this might be difficult to achieve unless they are part of a long-term alliance [9].

H1: *The higher the controllee's power distance the greater the exercise of clan control.*

On the other hand, as individuals feel self-motivated and more productive when there is less intervention by the managers, self-control is also likely to be used. This calls for less formal control substituted by self-control, defined as controllee-driven noncontrolling [51]. This is reinforced by the issue that in offshore projects, control of behavior is more difficult and is often facilitated and supplemented indirectly by means of self-control [51]. In this situation the controller encourages the controllee to use self-control or even exerts informal social pressure to use clan control [26; 9]. The controllee, then in his role as controller is encouraged to use the more difficult formal controls, such as outcome or behavioral control on his (supplier) team [9].

H2: *The lower the controllee's power distance the greater the exercise of self-control.*

3.2 Individualism/Collectivism

Individualism is defined as the extent to which people prefer to act as individuals rather than as members of groups [18]. In individualistic cultures the ties between individuals are loose. They value personal time and personal accomplishments. In contrast, in collectivistic cultures group goals and interests are more important than individual desires, and people are integrated into strong, cohesive groups.

In a study on collectivistic and individualistic work groups, Earley [12] found that collectivistic individuals show higher performance when working in an in-group (i.e., a group they identify with), as compared to collectivistic individuals who work by themselves or as part of an out-group (i.e., a group they do not identify with). In collectivistic societies the focus seems to be more on how well subordinates follow prescribed processes instead of assessing merely the outcomes [54]. In particular, social norms, duties, and obligations guide team members' behavior, and group (clan) goals seem to have priority [54]. Another aspect is that collectivistic employees view their relationship with the employer in moral terms. They tend to have a strong sense of identity with and loyalty to their organization. Consequently, they will strive to achieve outcomes that are in the organization's best interest and will do so with little expectation of personal gain. The controllee team will assume joint responsibility and/or receive joint recognition for actions taken or decisions made [55]. They also

see self-development occurring through harmony and reciprocity in interpersonal relations and contributing to the welfare of other group members [56]. Hence, implementing control through a process of socialization and promoting interpersonal dynamics to create shared beliefs will harmonize the values and beliefs among the team members [37]. Often, collectivistic cultures prefer training and other learning opportunities [54], all indications that clan control would be an appropriate choice of control [39].

H3: *The higher the controllee's collectivism the greater the exercise of clan control.*

On the other side, it is expected that in highly collectivistic cultures individuals will keep individuals and organization's interests and goals in line because they expect personal reward and recognition for their decisions [55]. Thus, providing individuals with autonomy will help them to monitor their own progress towards achieving common goals. This would favor the use of self-control, defined as a function of individual objectives and standards and intrinsic motivation [26; 9]. As motivated in "Power Distance", the encouragement of self-control might indirectly help to implement more difficult formal control modes, such as behavior control on the supplier side.

H4: *The lower the controllee's collectivism the greater the exercise of self-control.*

3.3 Uncertainty Avoidance

This dimension describes the extent to which the members of a culture feel threatened by uncertain or unknown situations. Individuals with low uncertainty avoidance believe that problems can be solved without formal rules [37]. They do not seem to be dependent on experts and prefer a less structured and rule-oriented environment [17]. They prefer rules only in situations of absolute necessity. Thus, providing a high degree of autonomy by means of self-control will increase project performance [37].

H5: *The lower the controllee's uncertainty avoidance the greater the exercise of self-control.*

On the other hand, a culture high in uncertainty avoidance would exhibit rule orientation, prefer employment stability, and exhibit stress when trying to explain, mitigate and minimize the uncertainty that is inherent to life [48]. Often, controllees with high uncertainty avoidance are dependent on experts and superiors for answers and feel secure with defined behaviors [37], thus feeling more comfortable with tight behavior controls. In addition, high uncertainty avoidance is associated with close communication, posing questions, feedback and reassurance, increasing the controller's understanding of appropriate behaviors as controllees might be more willing to reveal their actual work behaviors. The understanding of behavior is in turn crucial for introducing behavioral controls [37] as it is associated with higher behavior measurability. It has been found that, in particular high behavior observability facilitates behavior control [26; 27]. This finding was recently refined by Kirsch et al. [29] who concluded that high behavior observability is associated with the use of either behavior control (if the controller is knowledgeable) or clan control (if the controller has low knowledge). In particular, as the controller's technical knowledge is likely to be higher in outsourced/offshored projects there is a tendency for the controller to use more behavior control [51].

H6: *The higher the controllee's uncertainty avoidance the greater the exercise of behavior control.*

3.4 Activity/Passivity

The activity-passivity dimension is defined as the "extent to which individuals in a culture see themselves as doers (active shapers of the world) or beers (passive reactors to the world)" ([35], p. 178). In passive cultures people change themselves to fit into the environment. They are more cooperative, emphasize the experience of living, and are especially concerned with getting along with others [53]. In a recent study it was found that offshore teams from low activity cultures, such as India, prefer to have precisely described software specification and were described as "rather passive reactors to pre-specified tasks and methodologies" ([59], p. 249). Moreover, the Indian developers didn't appreciate open team meetings in which they were urged to actively participate in discussions related to wider project-relevant topics [59]. This would rather exclude clan and self-control as viable control options and instead suggest more directive forms of management, i.e., guiding the controllee through the process [52].

H7: *The lower the controllee's activity the greater the exercise of behavior control.*

In contrast, in active cultures individuals try to change the environment to fit them; active cultures are more competitive, action-oriented, and emphasize self-fulfillment. Furthermore, in more active cultures autonomy to complete requested tasks is emphasized as well as fitting to goal-oriented work environments, suggesting the use of outcome control and more liberal methods of management to be effective [52].

H8: *The higher the controllee's activity the greater the exercise of outcome control.*

3.5 Mono-/Polychronicity

Individuals with polychronic time perception are involved in many different activities with different people at the same time [24]. They view time commitment only as an objective to achieve when possible and make changes to plans when needed [16]. They feel that getting to know their counterparts and building a relationship is more important than adhering to a preset schedule. Here, monitoring the process may be required to assure compliance with project schedules [9].

H9: *The less monochronic the controllee's time perception the greater the exercise of behavior control.*

In contrast, individuals with monochronic time perception do only one thing at a time, take time and deadlines seriously, and adhere to preset schedules. For them, time is structured, linear, and sequential [16]. They set agendas for meetings and adhere to preset schedules. They schedule negotiations in ways that create psychological pressure in having to arrive at a decision by a certain date [16].

H10: *The more monochronic the controllee's time perception the greater the exercise of outcome control.*

4. RESEARCH METHODOLOGY

To test the hypothesized relationships in our research model, we developed matched-pair survey instruments. Most prior research in IS outsourcing/offshoring has collected data from either clients or vendors [45]. However, Koh et al. [31] found that stakeholder perspectives differ in such arrangements. Consequently, we collected data from both clients and suppliers. Clients were surveyed on their use of different control modes within the examined project and general project characteristics. Suppliers responded to items about their national culture. In addition, clients

and suppliers provided information about their position and professional experience. The use of two questionnaires significantly reduces the risks of common source bias [30].

4.1 Data Collection

To ensure the quality of the survey data, projects and respondents had to satisfy three criteria for inclusion in the sample. First, ISO projects either had to be completed for not more than twelve months, or had to be underway for at least three months and already reached one milestone. This ensured that included projects had progressed to a reasonable maturity [45] and that significant activities had recently occurred [30], increasing the reliability of the participants' perceptions and answers. Second, projects had to allow access to both a client representative and her/his supplier counterpart. For each selected project, completed survey instruments from one matching pair were required. Third, the client and supplier members of a dyad must have had operated in their roles for at least two months. Establishing this criterion ensured that the dyads have had adequate time to work with each other and to develop a relationship [45]. Furthermore, all survey participants held key positions in their respective organizations being responsible for managing the client-supplier relationship–a major influence on ISO success [59].

A website (http://survey.international-outsourcing.de) was launched to host the survey instruments, accelerate communications to respondents, and improve accuracy and efficiency in data collection and analysis. Next to the online questionnaire, we also prepared a paper version of our questionnaire to eliminate coverage error [47].

We used a convenience sample to collect the survey data. To identify appropriate ISO projects and respondents, we contacted management executives of client and offshore supplier firms by e-mail followed by a personal phone call. The executives were professional acquaintances of one of the authors. This was a key criterion for selecting these sites because it enhanced our ability to ensure the appropriateness of the ultimate respondents [45]. If an executive agreed to participate, she/he was asked to nominate suitable projects and respondents and solicit the participation of the executive of the counterpart organization. The use of this "known sponsor approach" [40] often resulted in immediate legitimacy and credibility of the research team and study. The client and/or vendor executive then forwarded a personalized e-mail with the study invitation to each potential respondent within the nominated project(s). This e-mail contained the URL address and a link to the website where the survey was available, the name of the questionnaire to be filled in, and the unique matched pair ID which was used to join the data records of the paired client and supplier representatives during data analysis. The e-mail also guaranteed the anonymity and confidentiality of the respondents, and clearly specified the goals of the study, the potential benefits to the participants, and the required level of participation. As an a priori strategy to minimize non-response error and its impact on the validity of inferences, we used Dillman's [11] Tailored Design Method.

Of the 18 executives who were initially requested to participate in the study, 14 agreed, for a response rate of 78 percent. Follow-up communications with the four non-participating executives did not reveal any trends or reasons that would point toward a non-response bias. A total of 96 client and supplier project team members were asked to participate in our study. In all, 94 respondents (46 client and 48 supplier representatives) filled in the questionnaire, resulting in a response rate of 98 percent. In order to form one data record for each matched pair, the matching client and supplier data records were joined based on the included ID. Two non-paired data records were dropped from the analysis, resulting in a sample size of 46 unique matched pairs. A comparison of the data of early returned questionnaires with that of later returned ones showed no indication of non-response error. Furthermore, the wide range of responses to our survey items indicates a lower risk of non-response bias [45].

4.2 Instrument Development

Two survey instruments were developed for this study, one for collecting data about the dependent variables (choice of controls) from the clients, and one for collecting data about the independent variables (national culture) from the suppliers. Generally accepted guidelines were followed in developing these instruments. All latent variables were measured with multiple items. Scale items were derived from prior research: To measure the four modes of control, we adopted Kirsch et al.'s [29] items for behavior, outcome and clan control, and adapted Brief and Aldag [5], Choudhury and Sabherwal [9], and Kirsch et al.'s [29] items for self-control. Measures for the cultural dimensions power distance, individualism/collectivism, and uncertainty avoidance were adopted from Hofstede's "Values Survey Module" [19]. However, scale items for the constructs activity and monochronicity were newly developed since we were not able to identify suitable measuring instruments. Although some scholars have conceptualized the cultural dimensions associated with these constructs, relatively few have attempted to measure them directly. The new items used in this study reflect and measure key concepts of activity and monochronicity, and are grounded in the work of Lytle et al. [35] and Triandis [52], and Hall and Hall [16], respectively. The three activity items assessed the controllee's individual initiative to complete tasks and find solutions to recurring issues, and her/his willingness to accept challenging tasks. The two monochronicity items involved the adherence to preset plans and time targets. All constructs were measured reflectively. Except for the demographic items, all items were rated on five-point Likert scales.

Prior research noted the effect of project size [27] and controllee's domain-specific knowledge [9; 26] on control choices. Hence, we included project size and supplier ISO knowledge as control variables in our analysis. Project size was measured by requesting client managers to indicate the amount of person months needed to execute the ISO project. ISO knowledge was estimated by asking (supplier) respondents to indicate their number of years of experience in the ISO field on a three-point Likert scale. The project volume and the ISO experience were used as proxies for project size and supplier ISO knowledge, respectively.

In March 2010, we conducted a pretest with five IS practitioners and four academics with expertise in ISO and survey methods. Furthermore, two experienced IS faculty members reviewed the client and supplier questionnaire and provided comments for improvement. Following the pretest, we selected a large-scale ISO arrangement as the site for the pilot study. This arrangement involved a multinational client organization with annual revenues of more than ten billion US dollars that has offshored IS services to an Indian vendor. A total of eleven respondents participated in the pilot study, eight client and three supplier representatives. The pilot resulted in clarification of the unit of analysis: The client-supplier pair rather than the ISO project. In addition, a power distance measure was added, the wording of some measures was

slightly changed, and the degree of data anonymity and confidentiality was further emphasized. Respondents in the pilot study were not in the main sample.

5. DATA ANALYSIS AND RESULTS

Before testing the research model, we provide descriptive statistics for our sample. Over a five-month period, we collected data from a total of 36 projects from 16 client organizations. All of these organizations operate from German-speaking countries (12, 3, and 1 from Germany, Switzerland, and Austria, respectively). 14 of them are large for-profit firms and two small or medium-sized enterprises (SMEs).

26 projects dealt with "applications development/testing"; five projects were characterized as "applications management", and two as "IT infrastructure management (managed services)". The project volume ranged from very small (less than 24 person months) to very large (600 or more person months), with a slight preponderance of larger projects. 33 projects involved large-scale supplier organizations, while three projects involved SME suppliers. 20 projects were executed with independent supplier firms (third party vendors and global IT service providers), 15 with a subsidiary of the client firm, and one with a joint venture. One project involved a nearshore supplier (Slovakia). In contrast, more than 90 percent of the projects were offshored to India.

63 percent of the client representatives stated that they had more than five years of experience in the IS field, while 96 percent declared having more than one year of experience in the ISO field. Almost 70 percent of the supplier representatives indicated having more than five years of experience in both the IS and ISO field.

To test the research model, we transformed it into a structural equation model, using the software *SmartPLS*. Partial least squares (PLS) has the ability to handle relatively small sample sizes [30], making it an appropriate choice. Basically, our data analysis followed a two-stage process as suggested by Chin [7] and Hulland [20]: First, we assessed the reliability and validity of the measurement model. Second, we tested the structural model and its hypotheses, and analyzed the effect of control variables.

5.1 Assessment of the Measurement Model

Five criteria need to be examined to determine the adequacy of the measurement model. These criteria are discussed below.

(1) *Content validity:* Content validity indicates to what extent the variables of a measurement model belong to the domain of the construct [4]. This was assured by selecting well established measures from prior research (wherever feasible), consulting experts in a pretest, and conducting a pilot study.

(2) *Item reliability:* Item reliability specifies which part of an item's variance can be explained by the underlying construct. If item loadings within the PLS model are lower than 0.40 they should be eliminated [20]. Loadings were analyzed using the PLS path weighting scheme. A number of items were below the 0.40 threshold. On the part of the independent variables, similar to Srite and Karahanna [48] we encountered difficulties in some of the original culture items adopted from Hofstede's "Values Survey Module" [19]. The analysis indicated problems with three power distance items, two collectivism items, two uncertainty avoidance items, and one activity item. On the part of the dependent variables, the generated item loadings showed problems with three outcome control items, one clan control item, and one self-control item. These items were removed from the model. PLS analysis was then run again reporting high loadings for all culture-related and control-related items (above 0.60 and 0.71, respectively), except for one behavior control item (0.53). A generally accepted rule of thumb is that item loadings should be greater than 0.70 [2]. However, in exploratory work loadings of 0.50 are still acceptable [8]. Thus, all items can be considered significant.

(3) *Construct reliability:* Construct reliability (or internal consistency) indicates how well a construct is measured by its items. It can be assessed with the composite reliability measure [15]. As seen in the "Fornell" column in Table 1, all constructs exceed the recommended cut-off of 0.6 [15], and are thus reliable.

(4) *Convergent validity:* Table 1 displays the correlations analysis of the independent variables, the dependent variables, and the two control variables. The boldface diagonal cells are the square root of the average variance extracted (AVE), which is a measure of the variance shared between a construct and its items. Each variable has an AVE of at least 0.5, establishing convergent validity for all scales [15].

(5) *Discriminant validity:* A necessary condition for discriminant validity is that a latent variable shares more variance with its assigned items than with any other latent variable [15]. The off-diagonal cells in Table 1 show the correlations between the

Table 1. Correlations between constructs

Construct	Fornell	PD	CO	UA	AC	MO	BC	OC	CC	SC	PS	IK
Power Distance (PD)	0.77	**0.80**										
Collectivism (CO)	0.76	-0.18	**0.79**									
Uncertainty Avoidance (UA)	0.75	0.40	-0.05	**0.78**								
Activity (AC)	0.78	-0.04	-0.23	0.05	**0.81**							
Monochronicity (MO)	0.67	0.23	-0.04	0.38	0.30	**0.71**						
Behavior Control (BC)	0.65	0.26	-0.06	0.06	0.16	0.41	**0.71**					
Outcome Control (OC)	0.83	-0.29	-0.29	-0.13	0.37	0.14	0.13	**0.84**				
Clan Control (CC)	0.80	0.39	0.12	-0.01	0.13	-0.02	0.38	-0.05	**0.76**			
Self-Control (SC)	0.83	-0.29	-0.16	-0.24	0.14	0.00	0.05	0.22	0.04	**0.79**		
Project Size (PS)	1.00	-0.09	-0.01	-0.05	0.02	-0.33	-0.42	-0.07	0.02	0.03	**1.00**	
ISO Knowledge (IK)	1.00	-0.11	-0.10	-0.13	0.09	-0.08	-0.13	0.06	-0.38	-0.14	0.10	**1.00**

constructs. The diagonal values are significantly greater than the off-diagonal values in the corresponding rows and columns, indicating discriminant validity for all scales [20]. Additionally, each within-construct item loads highly on the construct it is supposed to measure, and cross-loadings are lower than the within-construct item loadings.

5.2 Assessment of the Structural Model

The assessment of the inner model involves estimating the path coefficients and the R^2-values. Path coefficients specify the strengths of the relationships between the independent and dependent variables, while the R^2-value is a measure of the predictive power of a model for the dependent variables [30]. A bootstrap resampling method (1.000 re-samples) was used to determine the significance of the paths within the structural model. The sample size of 46 matched pairs exceeded the recommended minimum of 30 data records, which is ten times the largest number of independent variables influencing dependent variables in the structural model [2].

H1 and H2 pertain to power distance. As expected, power distance is significantly related with the exercise of clan control ($b = 0.387$; $t = 2.426$; $p < 0.05$) and self-control ($b = -0.287$; $t = 1.743$; $p < 0.10$). Both paths have effects in the predicted directions, supporting H1 and H2. Collectivism does not significantly affect the choice of informal controls (clan and self-control). Thus, H3 and H4 are not supported. H5 suggests a negative relationship between uncertainty avoidance and the use of self-control. H6 proposes a positive effect of uncertainty avoidance on the exercise of behavior control. However, both hypotheses are not significant. H7, which hypothesizes a negative relationship between activity and behavior control, as well as H8, which assumes a positive effect between activity and outcome control, are not significant, either. The latter path almost reached the critical t-value of 1.66 ($b= 0.368$, $t = 1.484$), and might therefore be worthwhile for further investigation in future research. Contrary to expectations, monochronicity has a significant and positive relationship with behavior control ($b = 0.310$; $t = 1.751$; $p < 0.10$). This finding is in the opposite direction of the relationship hypothesized (H9) and suggests that if the controllee's monochronic time perception is high, the controller's exercise of behavior control is also high, and vice versa. Finally, monochronicity is not significantly associated with the use of outcome control. Thus, H10 is not supported. Table 2 gives a detailed overview of the hypotheses test results.

Approximately 30 percent ($R^2 = 0.298$) of the variance in clan control, 19 percent ($R^2 = 0.192$) of the variance in self-control, 28 percent ($R^2 = 0.278$) of the variance in behavior control, and 15 percent ($R^2 = 0.145$) of the variance in outcome control are explained by the model. Ranging from 0.287 to 0.387, all path coefficients of the supported hypotheses clearly exceeded the suggested minimum value of significance at 0.20 [7]. Therefore, the fit of the overall model is deemed to be good [30].

The controllees' ISO knowledge and project size were included in the model as control variables. For both variables, one relationship with a dependent construct was found to be significant: ISO knowledge lowers the exercise of clan control ($b = -0.333$; $t = 2.463$; $p < 0.05$); project size is negatively associated with the use of behavior control ($b = -0.310$; $t = 1.756$; $p < 0.10$).

6. DISCUSSION

Before discussing the study results and their implications, some of the key limitations have to be mentioned. First, only a moderate sample size of 46 matched pairs was achieved. This sample size could potentially limit the power of the statistical techniques. Second, the findings of this study may be specific to ISO arrangements between Germany and India as the majority of the participating controllees were Indian, while most controllers were German. Third, the extent or amount to which the four control modes were exercised was not examined. Forth, this study only provides insight into the client's choice of controls. Hence, it does not examine the mechanisms used internally by the supplier. Finally, there are also limitations specific to measuring cultural values on the individual level that apply to all culture studies. As such, there is a concern that some of these cultural values are subtle and implicit and as such cannot easily be reported [48].

Before discussing the cultural value dimensions showing positive effects on control modes, we provide a brief discussion of the non significant relationships.

Collectivism, uncertainty avoidance and activity don't show significant relationships to any of the four control modes. Please note that the path from activity to outcome control is just below the significance threshold of 1.66. Obviously, ISO project managers don't pay particular attention to these three dimensions

Table 2. Hypotheses test results

		Hypothesis	Standardized Path Coefficient (b)	t-Value for Path	p-Value (two-tailed)
✓	H1	Power Distance → Clan Control (+)	0.387	2.426	0.05
✓	H2	Power Distance → Self-Control (-)	-0.287	1.743	0.10
X	H3	Collectivism → Clan Control (+)	0.162	1.036	
X	H4	Collectivism → Self-Control (-)	-0.239	1.014	
X	H5	Uncertainty Avoidance → Self-Control (-)	-0.169	0.947	
X	H6	Uncertainty Avoidance → Behavior Control (+)	-0.086	0.524	
X	H7	Activity → Behavior Control (-)	0.088	0.365	
X	H8	Activity → Outcome Control (+)	0.368	1.484	
✓	H9	Monochronicity → Behavior Control (-)	0.310	1.751	0.10
X	H10	Monochronicity → Outcome Control (+)	0.004	0.025	

"✓" indicates significant relationship; "x" indicates not significant relationship

when selecting their portfolio of control. Several explanations might account for this finding. First, it is possible that controllers do not care about the cultural values associated with these dimensions. Second, it might be particularly difficult to identify and observe these three cultural dimensions at the supplier's side. It might well be that these values are hidden as they are more tacit and deeply engrained [48]. Hence, the controller cannot take into account these cultural values when selecting appropriate controls. Finally, there could also be mediation effects between these cultural values and control choices responsible for these non-significant results. For example, behavior measurability could be mediating the relationship between uncertainty avoidance and control. Here, individuals with high uncertainty cultural values are more willing to reveal their actual behavior, a prerequisite for high behavior measurability, which in turn is associated with the use of either behavior control (if the controller is knowledgeable) or clan control (if the controller has low knowledge) [29].

6.1 Power Distance

Power distance was found to have a significant impact on informal control modes. This finding supports our hypotheses such that the higher the controllee's power distance the greater the exercise of clan control and the lower the controllee's power distance the greater the exercise of self-control.

Obviously, in high power distance cultures controllers take into account the controllees' predispositions such as that employees are fearful of disagreeing with their superiors and expect to be told what to do. They thus select clan control mechanisms, such as rituals, ceremonies and socialization to mitigate the difference in objectives between them and the controllees [9]. As controllees are more attuned by social norms [48] it is likely that they are more reachable by clan control.

On the other hand, our finding also confirms that self-control is a feasible option for low power distance cultures and a less favorable option for high power distance cultures. Typically, in high power distance cultures controllees ask for guidance. As a result, controllers are less likely to use control mechanisms that require high levels of autonomy and self-management. This finding is important because it further supports findings from prior literature that already proclaimed the important role of power distance in the context of ISO (e.g., [41; 59]).

In general, our findings emphasize the importance of informal controls with regard to the power distance value, may it be as complementary to formal control, or as dominant control mode.

6.2 Monochronicity

Our results confirmed a significant relationship between monochronicity and behavior control but in the opposite direction as hypothesized, such that the more monochronic the controllee's time perception the greater the exercise of behavior control. This result is counterintuitive. Several explanations are possible. First, the rationale for our hypothesis rested on Choudhury and Sabherwal's [9] finding that in an outsourcing context, tight behavioral controls are preferable in order to meet project schedules. This finding is based on anecdotal evidence from five cases, and thus might be specific to the particular context in these cases. Second, there is also evidence that control that counteract behavior may lead to typical resistance behavior, causing so called "ripple and knock-on effects" [34]. Ripple effects are primary side effects of well-intentioned control efforts, whereas knock-on effects show "secondary impacts of project control efforts, i.e., the impacts of ripple effects, often caused by processes that produce excessive or detrimental concurrence or human factors that amplify the negative effects via channels such as morale. Here, the use of less tight controls (e.g., outcome controls) might make monochronic controllees feel unsecure (ripple effect), decreasing their productivity (knock-on effect). Controllers anticipating these (negative) side effects might thus try to exercise more behavior control for monochronic controllees and, in turn, less behavior control for more polychronic controllees.

6.3 Project Size

Our results show that project size has a significant negative relationship with behavior control, such that the lower the project size the greater the exercise of behavior control. There is a plausible possible for this. Smaller projects are better controllable by means of behavior control as behavior observability is higher compared to larger, more complex projects, thus increasing the option to use behavior control. This extends the findings of Choudhury and Sabherwal [9], who didn't find a significant impact of project size on the choice of controls in outsourcing projects, and it may well be that their anecdotal evidence from five cases didn't allow generalization to a larger population of ISO projects. Our findings also contradict the findings of Jaworski [21], who found that larger projects prefer more formal control. However Jaworski didn't specify the mode of formal control so it is not clear whether this also included behavior control.

6.4 Supplier ISO Knowledge

The path between supplier ISO knowledge and clan control shows a significant negative relationship between these two, such that the lower the supplier's ISO knowledge the greater the use of clan control. This significant effect has not yet been articulated in past research. In general, there is agreement that the choice of controls further depends on the knowledge of the stakeholders [28]. In particular, a knowledgeable controllee makes the controller feel more confident in using self- or outcome control [26]. Even though we used ISO experience (number of years) as proxy for ISO knowledge, our results shed more light into the role of supplier experience (and thus knowledge) for using clan control.

Apparently, for less experienced controllees, controllers tend to rely on clan control, although in distant offshore relationships implementing clan control can be very costly. Exercising clan control by participating in project team meetings requires considerable time and commitment. However, it might well be that clan control is still the only feasible option or supplements well other more formal control mechanisms. On the other hand, if controllees are highly experienced, often these costly clan controls may not be necessary to this extend.

7. CONCLUSIONS AND IMPLICATIONS

Our research aims to contribute to the ISO and control literature in several ways. First, our results enhance prior findings by establishing a more detailed understanding about the influence of the supplier's cultural values on the exercise of control. Second, we were able to confirm the significance of two control variables, such as project size and the supplier's ISO knowledge. In particular project size has so far shown mixed results [27; 9]. Third, we incorporated and successfully applied new measures for self-control and developed new items for the constructs activity and monochronicity.

Our results also have important implications for practice. In general, our results suggest that informal controls are a powerful managerial tool for steering ISO projects. In particular, our view

of self-control (i.e., mechanisms the controller uses to assist and promote the exercise of self-control by the controllee) has interesting implications: Self-control could be used as a means to implement more difficult formal controls for less motivated, difficult to control and dependent controllees. For instance, control of the supplier team members' behavior can be indirectly achieved by means of self-control through the supplier's project manager who acts as controller for her/his supplier team [51]. Furthermore, our results suggest that when ISO client managers select their portfolio of control they should consider the cultural values of their supplier counterparts, may they be "easy to observe" (power distance) or more "hidden" (collectivism, uncertainty avoidance, and activity). Finally, our research sheds new light on the importance of cultural trainings [59]. Trainings focusing on cultural values could effectively improve cultural intelligence, thereby enabling client project managers to better determine culture-specific elements of behavior [3], which in turn is a prerequisite to select appropriate controls and to fine-tune them.

8. REFERENCES

[1] Adelakun, O. and Jennex, M.E. 2003. Success Factors for Offshore Information System Development. *Journal of Information Technology Cases and Applications* 5, 3, 12-29.

[2] Barclay, D., Thompson, R., and Higgins, C. 1995. The Partial Least Squares Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration. *Technology Studies* 2, 2, 285-324.

[3] Beck, R., Gregory, R., and Prifling, M. 2008. Cultural Intelligence and Project Management Interplay in IT Offshore Outsourcing Projects. In *Proceedings of the International Conference on Information Systems* (Paris, France, December).

[4] Bohrnstedt, G.W. 1970. Reliability and Validity Assessment in Attitude Measurement. In *Attitude Measurement*, G.F. Summers, Ed. Rand McNally. Chicago, 80-99.

[5] Brief, A.P. and Aldag, R.J. 1981. The 'Self' in Work Organizations: A Conceptual Review. *Academy of Management Review* 6, 1 (Jan. 1981), 75-88.

[6] Carmel, E. 1999. *Global Software Teams: Collaborating Across Borders and Time Zones*. Prentice Hall PTR. Upper Saddle River, New Jersey.

[7] Chin, W.W. 1998. The Partial Least Squares Approach to Structural Equation Modeling. In *Modern Methods for Business Research*, G.A. Marcoulides, Ed. Lawrence Erlbaum Associates. New Jersey, 295-358.

[8] Chin, W.W. 2010. How to Write Up and Report PLS Analyses. In *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V.E. Vinzi, W.W. Chin, J. Henseler, and H. Wang, Eds. Springer. Heidelberg, 655-690.

[9] Choudhury, V. and Sabherwal, R. 2003. Portfolios of Control in Outsourced Software Development Projects. *Information Systems Research* 14, 3 (Sept. 2003), 291-314.

[10] Dibbern, J., Winkler, J., and Heinzl, A. 2008. Explaining Variations in Client Extra Costs between Software Projects Offshored to India. *MIS Quarterly* 32, 2 (June 2008), 333-366.

[11] Dillman, D.A. 1999. *Mail and Internet Surveys: The Tailored Design Method*. Wiley InterScience. New York, NY.

[12] Earley, P.C. 1993. East meets West meets Mideast: Further Explorations of Collectivistic and Individualistic Work Groups. *Acad. of Manag. J.* 36, 2 (Apr. 1993), 319-348.

[13] Eisenhardt, K.M. 1995. Control: Organizational and Economic Approaches. *Manag. Science* 31, 2 (Feb. 1985), 134-149.

[14] Ford, D.P., Connelly, C.E., and Meister, D.B. 2003. Information Systems Research and Hofstede's Culture's Consequences: An Uneasy and Incomplete Partnership. *IEEE Transactions on Engineering Management* 50, 1 (Feb. 2003), 8-25.

[15] Fornell, C. and Larcker, D.F. 1981. Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research* 18, 3 (Aug. 1981), 328-388.

[16] Hall, E.T. and Hall, M.R. 1990. *Understanding Cultural Differences*. Intercultural Press, Inc. Yarmouth, Maine.

[17] Hofstede, G. 1991. *Culture and Organizations: Software of the Mind*. McGraw Hill. Cambridge, UK.

[18] Hofstede, G. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. 2nd ed. Newbury Park, California.

[19] Hofstede, G.1994. *VSM94*. DOI= http://www.geerthofstede.nl/research--vsm/vsm-94.aspx.

[20] Hulland, J.S. 1999. Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies. *Strategic Management Journal* 20, 4 (Feb. 1999), 195-204.

[21] Jaworski, B.J. 1988. Toward a Theory of Marketing Control: Environmental Context, Control Types, and Consequences. *The Journal of Marketing* 52, 3 (Jul. 1988), 23-39.

[22] Karahanna, E., Evaristo, J.R., and Srite, M. 2006. Levels of Culture and Individual Behavior: An Integrative Perspective. In *Advanced Topics in Global Information Management, Volume 5*, M.G. Hunter and F.B. Tan, Eds. Idea Group Publishing. Hershey, PA, 30-50.

[23] Kelman, H. 1958. Compliance, Identification, and Internalization: Three Processes of Attitude Change. *Journal of Conflict Resolution* 2 (Mar. 1958), 51-60.

[24] Kersten, G., Koeszegi, S.T., and Vetschera, R. 2003. The Effects of Culture in Computer-Mediated Negotiations. *Journal of Information Technology Theory and Application* 5, 2 (Jan. 2003), 1-28.

[25] King, W.R. 2005. Outsourcing and Offshoring: The new IS Paradigm? *Journal of Global Information Technology Management* 8, 2 (Apr. 2005), 1-4.

[26] Kirsch, L.J. 1996. The Management of Complex Tasks in Organizations: Controlling the Systems Development Process. *Organization Science* 7, 1 (Jan. 1996), 1-21.

[27] Kirsch, L.J. 1997. Portfolios of Control Modes and IS Project Management. *Inf. Sys. Research* 8, 3 (Sept. 1997), 215-239.

[28] Kirsch, L.J. 2004. Deploying Common Systems Globally: The Dynamics of Control. *Information Systems Research* 15, 4 (Dec. 2004), 374-395.

[29] Kirsch, L.J., Sambamurthy, V., Ko, D.-G., and Purvis, R.L. 2002. Controlling Information Systems Development Projects: The View from the Client. *Management Science* 48, 4 (Apr. 2002), 484-498.

[30] Ko, D.-G., Kirsch, L.J., and King, W.R. 2005. Antecedents of Knowledge Transfer from Consultants to Clients in Enterprise System Implementations. *MIS Quarterly* 29, 1 (Mar. 2005), 59-85.

[31] Koh, C., Ang, S., and Straub, D.W. 2004. IT Outsourcing Success: A Psychological Contract Perspective. *Information Systems Research* 15, 4 (Dec. 2004), 356-373.

[32] Lacity, M.C. and Hirschheim, R. 1993. *Information Systems Outsourcing: Myths, Metaphores and Realities*. Wiley, Chichester.

[33] Levina, N. and Vaast, E. 2008. Innovating or Doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *MIS Quarterly* 32, 2 (June 2008), 307-332.

[34] Lyneis, J.M. and Ford, D.N. 2007. System Dynamics Applied to Project Management: A Survey, Assessment, and Directions for Future Research. *System Dynamic Review* 23, 2/3 (Oct. 2007), 157-189.

[35] Lytle, A.L., Brett, J.M., Barsness, Z.I., Tinsley, C.H., and Janssens, M. 1995. A Paradigm for Confirmatory Cross-Cultural Research in Organizational Behavior. In *Research in Organizational Behavior* 17, B.M. Staw and L.L. Cummings, Eds. Jai Press. Greenwich, 167-214.

[36] Manz, C.C., Mossholder, K.W., and Luthans, F. 1987. An Integrated Perspective of Self-Control in Organizations. *Administration and Society* 19, 1 (May 1987), 3-24.

[37] Narayanaswamy, R. and Henry, R.M. 2005. Effects of Culture on Control Mechanisms in Offshore Outsourced IT Projects. In *Proceedings of the 2005 ACM SIGMIS CPR conference on Computer personnel research* (Atlanta, USA, April 14 -16, 2005). ACM Press. New York, 139-145.

[38] Nicholson, B. and Sahay, S. 2001. Some Political and Cultural Issues in the Globalisation of Software Development: Case Experience from Britain and India. *Information and Organization* 11, 1 (Jan. 2001), 25-43.

[39] Ouchi, W.G. 1980. Markets, Bureaucracies, and Clans. *Administrative Science Quarterly* 25, 1 (Mar. 1980), 129-141.

[40] Patton, M.Q. 1990. *Qualitative Evaluation and Research Methods*. Sage Publications, 2nd ed. Newbury Park, California.

[41] Prifling, M., Gregory, R., and Beck, R. 2008. Project Management Techniques for Managing Cross-Cultural Differences in IT Offshore Outsourcing. In *AMCIS 2008 Proceedings* (Toronto, Canada, August 14 - 17, 2008). 1-9.

[42] Rao, M.T., Poole, W., Raven, P.V., and Lockwood, D.L. 2006. Trends, Implications, and Responses to Global IT Sourcing: A Field Study. *Journal of Global Information Technology Management* 9, 3 (July 2006), 5-23.

[43] Roberts, C. and Wasti, S.A. 2002. Organizational Individualism and Collectivism: Theoretical Development and an Empirical Test of a Measure. *Journal of Management* 28, 4 (Aug. 2002), 544-566.

[44] Rottman, J. and Lacity, M. 2004. Twenty Practices for Offshore Sourcing. *MIS Quarterly Exec.* 3, 3 (Sept. 2004), 117-130.

[45] Rustagi, S., King, W.R., and Kirsch, L.J. 2008. Predictors of Formal Control Usage in IT Outsourcing Partnerships. *Information Systems Research* 19, 2 (June 2008), 126-143.

[46] Saunders, C., Van Slyke, C., and Vogel, D.R. 2004. My Time or Yours? Managing Time Visions in Global Virtual Teams. *Academy of Management Executive* 18, 1 (Feb. 2004), 19-31.

[47] Schaefer, D. and Dillman, D.A. 1998. Development of a Standard E-mail Methodology: Results of an Experiment. *Public Opinion Quarterly* 62 (Autumn 1998), 378-397.

[48] Srite, M. and Karahanna, E. 2006. The Role of Espoused National Cultural Values in Technology Acceptance. *MIS Quarterly* 30, 3 (Sept. 2006), 679-704.

[49] Straub, D., Keil, M., and Brenner, W. 1997. Testing the Technology Acceptance Model across Cultures: A Three Country Study. *Information Management* 33, 1 (Nov. 1997), 1-11.

[50] Straub, D., Loch, K., Evaristo, R., Karahanna, E., and Srite, M. 2002. Towards a Theory-Based Measurement of Culture. *Journal of Global Information Management* 10, 1 (Jan. 2002), 13-23.

[51] Tiwana, A. and Keil, M. 2010. Control in Internal and Outsourced Software Projects. *Journal of Management Information Systems* 26, 3 (Winter 2009/2010), 9-44.

[52] Triandis, H.C. 1982. Dimensions of Cultural Variations as Parameters of Organizational Theories. *International Studies of Management and Organization* 12, 4, 139-169.

[53] Triandis, H.C. 2008. Culture and Conflict. In *Intercultural Communication: A Reader* 12, L.A. Samovar, R.E. Porter, and E.R. McDaniel, Eds. Boston, 18-27.

[54] Triandis, H.C., Bontempo, R., Villareal, M.J., Asai, M., and Lucca, N. 1988. Individualism and Collectivism: Cross-Cultural Perspectives on Self-Ingroup Relationships. *Journal of Personality and Social Psych.* 54, 2, 323-338.

[55] Trompenaars, F. 1994. *Riding the Waves of Culture*. Irwin, New York.

[56] Wagner, J. 1995. Studies of Individualism-Collectivism: Effects on Cooperation in Groups. *Academy of Management Journal* 38, 1 (Feb. 1995), 152-172.

[57] Wiener, M., Vogel, B., and Amberg, M. Information Systems Offshoring—A Literature Review and Analysis. *Communications of the Association for Information Systems* 27 (Sept. 2010).

[58] Wilkins, A.L. and Ouchi, W.G. 1983. Efficient Cultures: Exploring the Relationship between Culture and Organizational Performance. *Administrative Science Quarterly* 28, 3 (Sept. 1983), 468-481.

[59] Winkler, J.K., Dibbern, J., and Heinzl, A. 2008. The Impact of Cultural Differences in Offshore Outsourcing—Case Study Results from German–Indian Application Development Projects. *Inf. Sys. Frontier* 10, 2 (Feb. 2008), 243-258.

Eine empirische Untersuchung des Wertbeitrages von serviceorientierten Architekturen (SOA)

Nils Joachim Daniel Beimborn Tim Weitzel
Lehrstuhl für Wirtschaftsinformatik, insb. Informationssysteme in Dienstleistungsbereichen
Universität Bamberg, Feldkirchenstr. 21, D-96045 Bamberg, Germany
+49 (0)951 863 – {2876 | 2525 | 2871}
{nils.joachim | daniel.beimborn | tim.weitzel}@uni-bamberg.de

ZUSAMMENFASSUNG
Bisher wurde der Wertbeitrag der Nutzung von serviceorientierten Architekturen (SOA) für Unternehmen weder in der Praxis noch in der Forschung jenseits von Einzelfallbetrachtungen nachgewiesen. Wir untersuchen in einer ersten quantitativen Studie die realisierten Vorteile einer SOA-Nutzung im Hinblick auf Kostenreduktion, unternehmerische Agilität, Datenqualität, Prozess-Monitoring, interne Geschäftsprozessintegration (STP) und unternehmensübergreifende Integration (B2B). Die Analyse von 134 Unternehmensantworten zeigt, dass SOA zu allen sechs Unternehmensvorteilen beiträgt. Außerdem ist erkennbar, dass sich SOA in klassischen Aspekten flexibler IT-Infrastrukturen wie Modularität, Flexibilität und Skalierbarkeit widerspiegelt. Und auch die Realisierung von Unternehmensvorteilen durch SOA kann zu bedeutenden Teilen über die Verbesserung der IT-Flexibilität als Mediator erklärt werden.

Stichworte
Serviceorientierte Architektur, SOA, Wertbeitrag, IT-Nutzen, IT-Architektur, Flexibilität, Modularität, Integration, Empirie, PLS

1. EINLEITUNG
In den vergangenen Jahren sind serviceorientierte Architekturen (SOA) stark in den Fokus der betrieblichen Praxis und der Forschung gerückt [46]. 68% der europäischen und nordamerikanischen Unternehmen gaben in einer aktuellen Studie von Forrester Research an, SOA bereits zu verwenden oder bis Ende 2010 einzuführen [23]. Während die technischen Aspekte von SOA von Beginn an umfassend erforscht wurden, gewinnt die organisationale Sicht auf SOA erst in der jüngeren Literatur an Bedeutung. So wurden wichtige Forschungsbedarfe aus dem organisationalen Themenfeld, wie ökonomische und organisationale Herausforderungen im Zusammenhang mit SOA oder die Identifikation von Faktoren, die die organisationale Adoption von SOA beeinflussen, aufgezeigt [10, 38, 47]. Als Folge werden vermehrt Fallstudien durchgeführt [5, 24, 48] oder auch einzelne der potenziellen Vorteile, die sich durch SOA ergeben (z. B. Informationsaustausch innerhalb von Wertschöpfungsketten, Einfluss auf Joint Ventures oder das Zusammenspiel von SOA im Rahmen des Geschäftsprozessmanagement), in ersten quantitativen Studien untersucht [8, 27, 34, 43]. Es gibt bisher allerdings noch keine Studie, die den Wertbeitrag von SOA, den sog. SOA-Business-Value, umfassend erhoben hat. In einer Literaturrecherche zu SOA haben Viering et al. [46] herausgefunden, dass die Betrachtung der organisatorischen Auswirkungen einer SOA-Einführung mit 19 von 175 identifizierten Artikeln der am wenigsten erforschte Bereich innerhalb der SOA-Forschung ist. Unsere Forschung wird von der folgenden Frage geleitet: Was ist der Wertbeitrag serviceorientierter Architekturen?

Das Schließen dieser Wissenslücke ist auch für die Praxis von großer Bedeutung, da eine SOA-Einführung mit hohem Aufwand verbunden ist und die Unternehmensvorteile (bislang) nicht umfassend in einer Breitenstudie aufgezeigt wurden, sondern nur postuliert werden. Das dies nicht ausreichend ist, wird auch an den Antworten der erwähnten Forrester-Studie deutlich, in der lediglich 20% der Befragten angegeben haben, dass SOA die meisten oder alle erwarteten Vorteile erfüllt hat, während 50% weniger Vorteile als erwartet realisiert haben [23]. Dies zeigt, dass in der Erforschung des SOA-Wertbeitrages ein wichtiges Ziel liegt, zu dem unsere Studie einen Beitrag leistet. Dabei definieren wir den Wertbeitrag von SOA als die Gesamtheit aller quantifizierbaren sowie immateriellen Vorteile, die auf operativer, Management-, strategischer, IT-Infrastruktur- und organisationaler Ebene eines Unternehmens durch die Einführung und Nutzung von SOA zur Unterstützung von Geschäftsprozessen entstehen (vgl. Tabelle 1 für detaillierte Definitionen). Unsere Forschungsergebnisse können Praktikern bei ihrer Beurteilung der Vorteile von SOA (z. B. Reduktion der Kosten, Steigerung der unternehmerischen Agilität, der Datenqualität, des Prozess-Monitorings, der internen Geschäftsprozessintegration (Straight Through Processing (STP)) oder der unternehmensübergreifenden Integration (B2B-Integration)) unterstützen, um die Entscheidung für oder gegen die Einführung von SOA auf eine breite empirische Basis zu stützen.

Die Ergebnisse sind ebenfalls wichtig für die Forschungsbereiche IT-Wertbeitrag und SOA/IT-Architektur. Da wir die Unternehmensvorteile von SOA nicht nur direkt, sondern auch im Kontext IT-architekturaler Eigenschaften untersuchen, ergeben sich Erkenntnisse darüber, wie eine IT-Architektur hinsichtlich ihrer Flexibilität ausgerichtet sein sollte, um den unternehmerischen Mehrwert einer SOA tatsächlich nutzen zu können. Andererseits können zukünftige SOA-fokussierte Forschungsarbeiten von den

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Ergebnissen profitieren, da wir den häufig „unscharfen“ SOA-Begriff, unter dem recht Unterschiedliches verstanden wird, in Bezug zu wohletablierten Aspekten der Flexibilität der IT-Infrastruktur (Modularität, Integration und Skalierbarkeit) setzen.

Im Folgenden geben wir zunächst einen Überblick über die Grundlagen zu SOA, IT-Flexibilität und der Dynamic-Capabilities-Theorie, bevor das Forschungsmodell hergeleitet wird. Danach stellen wir die erhobenen Daten vor, die für die Modellschätzung mit Partial Least Squares (PLS) genutzt werden. Im Anschluss werden die Ergebnisse diskutiert und ihre Bedeutung für Forschung und Praxis aufgezeigt.

2. GRUNDLAGEN

2.1 Serviceorientierte Architekturen

In der Literatur existieren zahlreiche unterschiedliche und unterschiedlich weit reichende Definitionen von SOA. Hierbei sind vor allem Definitionen zu unterscheiden, die SOA rein auf die IT beschränken oder explizit auch Unternehmensaspekte berücksichtigen. Wir beziehen uns daher auf Bieberstein et al., die SOA als umfassenden Architekturansatz definieren, der sowohl IT- als auch Unternehmensaspekte beinhaltet: “A service-oriented architecture is a framework for integrating business processes and supporting IT infrastructure as secure, standardized components – services – that can be reused and combined to address changing business priorities” [12, S. 5]. Diese umfassende SOA-Definition ist für unsere Untersuchung zweckmäßig, da viele Vorteile von SOA erst bei einer ganzheitlichen Betrachtung realisiert und beobachtet werden können.

2.2 IT-Flexibilität

Da ein häufig genanntes Ziel bei SOA-Einführungen die Erhöhung der Flexibilität der IT ist [48], wird deren Rolle für die Realisierung von SOA-Vorteilen explizit berücksichtigt.

Duncan definiert IT-Infrastruktur-Flexibiltiät als “the ability of the IS department to respond quickly and cost-effectively to systems demands, which evolve with changes in business practices or strategies” [19, S. 44]. Byrd und Turner unterscheiden dabei zwischen der Flexibilität des IT-Personals vs. derjenigen der technischen IT-Infrastruktur [13]. Im SOA-Kontext werden wir uns im Folgenden auf die technische IT-Flexibilität beschränken, auch wenn bei SOA-Implementierungen durch den dabei stattfindenden Erfahrungs- und Wissensaufbau die Flexibilität der IT-Mitarbeiter ebenfalls nicht unberührt bleibt.

Nach Duncan [19] hängt die technische IT-Infrastruktur-Flexibilität davon ab, zu welchem Grad IT-Ressourcen gemeinsam nutzbar und wiederverwendbar sind. Basierend auf diesen beiden Aspekten hat sie drei Kriterien der Flexibilität definiert: (1) Konnektivität (ermöglicht die Verbindung von Komponenten); (2) Kompatibilität (ermöglicht verbundenen Komponenten zu interagieren und Informationen auszutauschen); (3) Modularität (Isolierung und Standardisierung von Prozessen in Bezug auf Anwendungsfunktionen und Daten). Byrd und Turner definieren *Modularität* als “the ability to add, modify, and remove any software, hardware, or data components of the infrastructure with ease and with no major overall effect" [13, S. 171]. Eine empirische Untersuchung der drei Flexibilitätsdimensionen durch Byrd und Turner ergab, dass keine ausreichende Separation zwischen Konnektivität und Kompatibilität möglich ist, was zur Zusammenfassung dieser beiden Aspekte zum Konzept der *Integration* führte [13].

Chanopas et al. [14] haben diese Arbeiten erweitert und führen in ihrem Modell zur technischen IT-Flexibilität neben den Aspekten der Modularität, Integration (Konnektivität + Kompatibilität) fünf weitere Faktoren der IT-Flexibilität (Skalierbarkeit, Kontinuität, Schnelligkeit, Zukunftsfähigkeit und Einfachheit der Nutzung) ein. Von diesen Faktoren, ist die *Skalierbarkeit* (“the degree to which hardware/software can be scaled and upgraded on existing infrastructure” [14, S. 645]) im SOA-Kontext ein wichtiger zusätzlicher Faktor gegenüber dem bisherigen IT-Flexibilitätskonzept. So ist SOA durch die Verwendung eines Enterprise Service Bus (ESB) zur Integration von Anwendungen skalierbarer als Punkt-zu-Punkt-Verbindungen. Bei letzteren wird das Handling bei steigender Zahl von zu integrierenden Systemen schnell sehr komplex [35].

2.3 Dynamic-Capabilities-Theorie

Das im Folgenden entwickelte Forschungsmodell basiert auf der Dynamic-Capabilities-Theorie (DCT), welche wiederum auf der Resource-based View (RBV) fußt. Die RBV definiert, was die Ressourcen eines Unternehmens strategisch wertvoll macht [3, 36]: “sources of sustained competitive advantage are firm resources that are valuable, rare, imperfectly imitable, and non-substitutable” [3, S. 101]. IT im Sinne einer Ressource besteht aus technischen, personellen und organisatorischen IT-Ressourcen, die zu einem nachhaltigen Wettbewerbsvorteil führen können [11, 31]. Einer der Kritikpunkte an der RBV ist, dass nur eine statische Sicht auf die Auswahl der Ressourcen gegeben wird, jedoch nicht beschrieben wird, wie ein Unternehmen diese Ressourcen entwickeln und integrieren kann, um einen Wettbewerbsvorteil zu erlangen [29, 30]. Die DCT behebt dieses Defizit und definiert dynamische Fähigkeiten bzw. “dynamic capabilities” als “ability to integrate, build, and reconfigure internal and external competencies to address rapidly-changing environments” [45, S. 516]. Demnach ist die Agilität eines Unternehmens Resultat einer dynamischen Fähigkeit. Kombiniert man die Entwicklung von Fähigkeiten mit der Auswahl der richtigen Ressourcen [30], so lässt sich argumentieren, dass eine Komplementarität zwischen der Ressource SOA und der IT-Flexibilität als Fähigkeit besteht. Basierend auf dieser Annahme führt die gemeinsame Nutzung zu strategischen (und anderen) Unternehmensvorteilen.

3. FORSCHUNGSMODELL

Ausgehend von den theoretischen Grundlagen wird im Folgenden das Forschungsmodell (vgl. Abbildung 1) hergeleitet. Als erstes wird der direkte Wertbeitrag von SOA anhand verschiedener Kategorien verargumentiert. In einem zweiten Schritt wird dann der Einfluss von SOA auf die technische IT-Infrastruktur-Flexibilität hergeleitet sowie deren Einfluss auf den Wertbeitrag dargestellt.

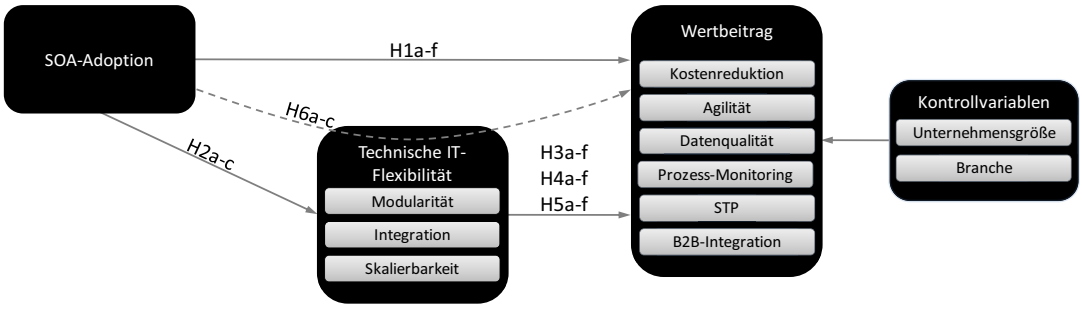


Abbildung 1. Forschungsmodell.

Tabelle 1. Einordnung des Wertbeitrages von SOA in das Rahmenwerk nach Shang und Seddon [41]

Dimension	Definition [41, S. 277-279]	Kosten-reduktion	Agilität	Datenqualität	Prozess-Monitoring	STP	B2B-Integration
Operativ	"streamline processes and automate transactions provides business benefits by speeding up processes, substituting labour and increasing operation volumes"	X				X	X
Management	"allocation and control of the firm's resources, monitoring of operations and supporting of business strategic decisions"			X	X		
Strategisch	"attainment of sustained IT-based competitive advantage"		X	X			
IT-Infrastruktur	"sharable and reusable IT resources that provide a foundation for present and future business applications"	X	X				
Organisational	"build integrated processes, improve employee communication, foster the development of a 'common vision' and user empowerment, support customer services and facilitate a flattening of organizational structure"					X	X
X: Die Dimension aus dem Rahmenwerk ist durch die jeweilige Kategorie abgedeckt.							

3.1 Wertbeitrag von SOA

In der Literatur werden verschiedene potenzielle Vorteile von SOA aufgeführt, die prinzipiell durch die Nutzung von SOA realisiert werden können. Dies sowie die Tatsache, dass sich solche Vorteile häufig nicht direkt in finanziellen Kennzahlen widerspiegeln [41], machen eine multidimensionale Messung des SOA-Wertbeitrages notwendig. Eine Literaturanalyse hat ergeben, dass sich die unterschiedlichen Vorteile einer SOA in sechs Kategorien unterteilen lassen, die alle eine relevante Bedeutung für den Geschäftswert einer SOA-Einführung besitzen. So können durch die Einführung von SOA die IT-Kosten im Unternehmen beispielsweise durch die Wiederverwendung von Services und die damit verbundenen Effizienzsteigerungen reduziert werden [22, 25, 48]. Es wird aber auch die Agilität des Unternehmens erhöht, da es schneller auf Änderungen im Umfeld reagieren kann und bspw. neue Produkte schneller auf den Markt bringen kann [5, 22, 25, 43, 48]. Ebenso soll SOA die Datenqualität durch eine bessere Systemintegration und Datenkonsolidierung [6, 22, 48] sowie die Qualität der Geschäftsprozesse aufgrund durchgehender Integration (STP) erhöhen, was sich auch in der Möglichkeit einer besseren Prozessüberwachung (Prozess-Monitoring) äußert [6, 22]. Zuletzt wird von SOA auch erwartet, dass die Integration mit anderen Unternehmen (B2B-Integration) leichter und günstiger möglich wird [5, 6, 27, 48].

Mueller et al. [32] haben eine sekundärstatistische Analyse von Fallbeschreibungen zu SOA-Projekten durchgeführt und dabei die Vorteile von SOA entlang von fünf Dimensionen gruppiert. Die verwendeten Dimensionen wurden ursprünglich von Shang und Seddon [41] zur Bewertung von Unternehmenssoftware entwickelt und basieren auf einer umfangreichen Literaturanalyse. Tabelle 1 zeigt, dass die von uns identifizierten sechs Kategorien des Wertbeitrags alle fünf Dimensionen abdecken und somit im Einklang mit den bisherigen Arbeiten stehen.

Hypothese (H1a-f): SOA-Nutzung hat einen positiven Wertbeitrag i.S.v. (a) Kostenreduktion sowie Steigerung (b) der Unternehmensagilität, (c) der Datenqualität, (d) des Prozess-Monitorings, (e) der internen Geschäftsprozessintegration und (f) der unternehmensübergreifenden Integration.

3.2 IT-Flexibilität als Mediator

Wie oben ausgeführt werden Modularität, Integration (Konnektivität + Kompatibilität) sowie Skalierbarkeit als Aspekte der technischen IT-Infrastruktur-Flexibilität betrachtet. Ein Gestaltungziel bei SOA ist, die einzelnen Services modular zu gestalten, um sie leichter integrieren und wiederverwenden zu können [5, 27, 34, 48]. Zusätzlich soll SOA helfen, die Skalierbarkeit der IT-Infrastruktur zu erhöhen, um dadurch schneller auf Änderungen des Nachfragevolumens reagieren zu können [48].

Hypothese (H2a-c): SOA führt zu einer Erhöhung der technischen IT-Infrastruktur-Flexibilität i.S. gesteigerter (a) Modularität (b) Integration und (c) Skalierbarkeit.

Insgesamt hilft eine flexible IT-Infrastruktur durch diese höhere Modularität, Integration und Skalierbarkeit die eigentlichen Unternehmensziele im Sinne eines IT-Wertbeitrags besser zu erreichen. So kann die Flexibilität der IT einem Unternehmen neue

Handlungsoptionen eröffnen [40] und dabei helfen, produktiver zu werden [18, S. 23]. Dies kann dadurch geschehen, dass Kosten eingespart werden, die Datenqualität verbessert wird, die Prozesse durchgehend intern wie unternehmensübergreifend integriert werden oder die Geschäftsprozesse besser überwacht und somit weitere Optionen identifiziert werden können. Folglich leisten alle drei Aspekte der IT-Flexibilität einen positiven Wertbeitrag.

Hypothese (H3a-f): Der Grad der Modularität der technischen IT-Infrastruktur hat einen positiven Wertbeitrag im Sinne einer (a) Kostenreduktion sowie Steigerung (b) der Agilität, (c) der Datenqualität, (d) des Prozess-Monitorings, (e) der internen Geschäftsprozessintegration und (f) der unternehmensübergreifenden Integration.

Hypothese (H4a-f): Der Grad der Integration der technischen IT-Infrastruktur hat einen positiven Wertbeitrag im Sinne einer (a) Kostenreduktion sowie Steigerung (b) der Agilität, (c) der Datenqualität, (d) des Prozess-Monitorings, (e) der internen Geschäftsprozessintegration und (f) der unternehmensübergreifenden Integration.

Hypothese (H5a-f): Die Skalierbarkeit der technischen IT-Infrastruktur hat einen positiven Wertbeitrag im Sinne einer (a) Kostenreduktion sowie Steigerung (b) der Agilität, (c) der Datenqualität, (d) des Prozess-Monitorings, (e) der internen Geschäftsprozessintegration und (f) der unternehmensübergreifenden Integration.

Durch die Verkettung der Hypothesen 2 bis 5 und auf Basis der in Abschnitt 2.3 geführten theoretischen Argumentation kann abgeleitet werden, dass SOA ihren Wertbeitrag vor allem über die Erzeugung von IT-Flexibilität (im Sinne einer „dynamic capability") leistet. Eine explizite Untersuchung dieses sog. Mediationseffekts hilft festzustellen, ob es sich dabei um die alleinige Erklärung für die Existenz eines SOA-Wertbeitrags handelt, oder ob SOA auf andere Art und Weise ebenfalls Nutzen für das Unternehmen stiftet. Dazu formulieren wir abschließend:

Hypothese (H6a-c): Die technische IT-Infrastruktur-Flexibilität (im Sinne von (a) Modularität, (b) Integration und (c) Skalierbarkeit) mediiert vollständig den Zusammenhang zwischen SOA und resultierenden Unternehmensvorteilen (d.h. positiven Wertbeitrag)[1].

4. VORGEHEN UND ERGEBNISSE

4.1 Datenerhebung

Zur Evaluierung des Forschungsmodells wurde unter deutschen Dienstleistungsunternehmen (bspw. Finanzdienstleister, Logistik, IT und Kommunikation) in Deutschland eine schriftliche Befragung durchgeführt. Die Dienstleistungsbranche wurde ausgewählt, da dort die IT häufig neben dem Personal den einzig maßgeblichen „Produktionsfaktor" darstellt, während im herstellenden Gewerbe die „physischen" Produktionsanlagen, Rohstoffe etc. eine bedeutende Rolle spielen und damit deutlich umfangreichere Kontingenzfaktoren erzeugen.

Nach dem Erwerb eines Datensatzes mit 1.615 Unternehmen (US SIC Codes 4.000 bis 8.999) wurde zunächst jedes Unternehmen angerufen, um den leitenden IT-Architekten zu identifizieren. Dabei wurden 955 Ansprechpartner ermittelt und diesen zur Datenerhebung postalisch Fragebögen zugesandt. Nach mehreren Erinnerungen per Post, Telefon und E-Mail konnten schließlich 174 Rückläufer empfangen werden (Rücklaufquote 18,2%). Für die nachfolgenden Berechnungen wurden nur diejenigen 134 Fragebögen verwendet, in denen alle 34 für die folgenden Auswertungen notwendigen Fragen beantwortet wurden.

Zum Test des Forschungsmodells und zur Analyse der Daten wurden Partial Least Squares (PLS) (smartPLS 2 M3 [39]) und PASW Statistics 18 verwendet. Fast alle Konstrukte des Forschungsmodells wurden durch reflektive Multi-Item-Messmodelle operationalisiert. Die einzige Ausnahme bildet das Konstrukt der SOA-Adoption, welches formativ („Grad der SOA-Unterstützung fünf verschiedener Geschäftsbereiche", vgl. Appendix) vermessen wurde. Die fünf Items wurden mittels aus einer konfirmatorischen Faktoranalyse abgeleiteter Gewichte zu einem SOA-Adoption-Score verdichtet.

Die Messmodelle wurden, soweit möglich, aus der Literatur abgeleitet. Dies war vor allem bei den Konstrukten der technischen IT-Infrastruktur-Flexibilität (Modularität, Integration und Skalierbarkeit) der Fall. Die Indikatoren zur Messung des SOA-Adoptionsgrades wurden hingegen aufgrund des Fehlens vergleichbarer Untersuchungen neu für diese Studie entwickelt. Für die Kontrollvariablen wurde die *Unternehmensgröße* (Mitarbeiteranzahl des Unternehmens) aus sekundärstatistischen Quellen ermittelt und logarithmiert einbezogen. Auch die *Branchenzugehörigkeit* wurde anhand sekundärstatistischer Quellen ermittelt und durch drei binäre Dummy-Variablen für Finanzdienstleistungen, Informationstechnologie & Kommunikation und Handel & Logistik operationalisiert.

4.2 Datenqualität und Messmodellvalidität

Vor der statistischen Validierung des Messmodells wurden die Daten auf Normalverteiltheit überprüft. Der Kolmogorov-Smirnov-Test belegt, dass die Daten nicht vollständig normalverteilt sind. Dies sowie die begrenzte Stichprobengröße von 134 Datensätzen waren ausschlaggebend für die Verwendung von PLS anstelle kovarianzbasierter Schätzverfahren, die je nach Modellkomplexität Stichproben von mindestens 200 bis 800 Datensätzen benötigen [16]. Einer der Vorteile von PLS ist, dass schwächere Zusammenhänge schon bei kleineren Stichprobengrößen identifiziert werden können [16].

Die Daten wurden zudem hinsichtlich des Verdachts auf *Common Method Bias* (CMB) untersucht. Eine exploratorische Faktoranalyse im Sinne des Harman-Ein-Faktor-Tests ergab, dass die mit dem größten Eigenwert identifizierte Komponente 34,3% der Varianz der Indikatoren erklärt, sodass hier kein Verdacht auf CMB geäußert werden muss. Ergänzend wurden analog zu Liang et al. [28] und basierend auf Podsakoff et al. [37] alle Indikatoren in Single-Item-Konstrukte umgewandelt und von diesen jeweils Beziehungen zu den eigentlichen Modellkonstrukten eingefügt. In das Modell wurde dann ein zusätzlicher latenter Methodenfaktor, der mit einem reflektiven Messmodell aus allen Indikatoren des gesamten Modells operationalisiert wurde, eingefügt und auf die eigentlichen Modellkonstrukte wirkt. Nach diesem Ansatz erklärt der Methodenfaktor eine durchschnittliche extrahierte Varianz (DEV) von gerade einmal 0,008, während die DEV auf Basis der eigentlichen Indikatoren 0,751 beträgt. Das Verhältnis der Varianzen entspricht also 1:94 (Liang et al. [28] schließen in ihrer

[1] Die Kombination der drei IT-Flexibilitätsdimensionen mit den 6 verschiedenen Nutzenkomponenten führt eigentlich zu 18 empirisch zu prüfenden Zusammenhängen, die aber hier zur Übersichtlichkeit in drei grundsätzliche Hypothesen zusammengefasst wurden.

Arbeit schon bei einem Verhältnis von 1:42 CMB klar aus). Insgesamt zeigen beide Tests keine Anzeichen des Vorhandenseins von CMB, auch wenn das bei dieser Art von Studien nie vollständig ausgeschlossen werden kann.

Abschließend wurden die Daten auf das Vorhandensein von *Non-Response Bias* untersucht. Dazu haben wir die Antworten in zwei Hälften eingeteilt und die Daten derjenigen Unternehmen, die zuerst geantwortet haben, mit denen verglichen, die erst nach „Erinnerungen" reagiert haben. Dieses Vorgehen basiert auf der Grundannahme, dass es Gemeinsamkeiten zwischen den Teilnehmern gibt, die nachträglich geantwortet haben, und den Unternehmen, die nicht geantwortet haben [1]. Es existieren keine signifikanten Unterschiede zwischen beiden Gruppen, sodass wir Non-Response Bias ausschließen.

Bevor das eigentliche Forschungsmodell validiert werden kann, ist die Prüfung der Validität und Reliabilität der Messmodelle notwendig. Die Indikatorreliabilitätsprüfung zeigt, dass bis auf die Faktorladung von STP1 alle Werte größer als die verlangten 0,707 [33] sind; die Faktorladung von STP1 ist aber mit 0,663 zumindest deutlich höher als die in [2] geforderten 0,6.

Tabelle 2 zeigt die Einhaltung der Kriterien zur Konstruktvalidität (Konvergenz- und Diskriminanzvalidität). Zum einen sind alle Faktorreliabilitäten (Composite Reliability (C.R.)) deutlich höher als die geforderten 0,7 [33]. Zum anderen ist die durchschnittliche erfasste Varianz (DEV) aller Konstrukte größer als der geforderte Mindestwert von 0,5 [15]. Tabelle 2 zeigt auch eine hinreichende Diskriminanzvalidität, da die Quadratwurzeln der DEV (schattierte Hauptdiagonale) durchgehend größer als die Korrelationen der latenten Variablen in den jeweiligen Zeilen und Spalten sind [21]. Zusätzlich weisen auch die (aus Platzgründen nicht dargestellten) Kreuzladungen immer deutlich niedrigere Werte auf als die Ladungen der Indikatoren auf ihre originären Konstrukte.

Insgesamt erfüllen die verwendeten Daten und Messmodelle also die üblichen Anforderungen.

Tabelle 2. Konvergenz- und Diskriminanzvalidität (Korrelationen der latenten Variablen und Wurzel der DEV (grau hinterlegte Zellen)).

	C.R.	DEV	SOA-Adoption	Modularität	Integration	Skalierbarkeit	Kostenreduktion	Agilität	Datenqualität	Prozess-Monitoring	STP	B2B-Integration
SOA-Adoption	1,000	1,000	1,000									
Modularität	0,874	0,699	0,185	0,836								
Integration	0,902	0,699	0,233	0,548	0,836							
Skalierbarkeit	0,911	0,773	0,145	0,356	0,329	0,879						
Kostenreduktion	0,900	0,643	0,153	0,490	0,470	0,239	0,802					
Agilität	0,921	0,745	0,278	0,445	0,478	0,172	0,281	0,863				
Datenqualität	0,933	0,874	0,152	0,429	0,450	0,453	0,446	0,454	0,935			
Prozess-Monitoring	0,894	0,808	0,345	0,436	0,495	0,296	0,483	0,327	0,399	0,899		
STP	0,850	0,658	0,219	0,452	0,595	0,386	0,598	0,364	0,555	0,579	0,811	
B2B-Integration	0,961	0,891	0,250	0,297	0,476	0,300	0,308	0,262	0,238	0,287	0,334	0,944

4.3 Test der direkten Zusammenhänge

Zur empirischen Überprüfung des Wertbeitrages von SOA wurde zunächst nicht das vollständige Forschungsmodell, wie in Abbildung 1 dargestellt, mit PLS getestet, sondern ein vereinfachtes Modell (nur (H1a-f)) ohne das Konstrukt der technischen IT-Flexibilität (Modularität, Integration und Skalierbarkeit). Die im Rahmen dieser Modellschätzung ermittelten Bestimmtheitsmaße (R^2) sowie die Pfadkoeffizienten (β) dieser direkten Effekte sind in Tabelle 3 dargestellt. Der SOA-Einsatz hat einen moderaten und in allen Fällen deutlich signifikanten Einfluss auf die sechs untersuchten Vorteile auf Unternehmensebene; H1a-f kann angenommen werden. SOA erklärt allerdings nur einen geringen Teil der Varianzen der abhängigen Variablen (3-14%).

Tabelle 3. Bestimmtheitsmaße (R^2) und Pfadkoeffizienten (β) der direkten Effekte im direkten Modell

SOA-Adoption auf…		R^2	β
H1a	Kostenreduktion	0,041	**0,193*****
H1b	Agilität	0,081	**0,278*****
H1c	Datenqualität	0,031	**0,161****
H1d	Prozess-Monitoring	0,142	**0,381*****
H1e	STP	0,059	**0,244*****
H1f	B2B-Integration	0,062	**0,237*****
Anmerkung: *** p ≤ .01; ** p ≤ .05; * p ≤ .1 Wir berichten die einseitigen Signifikanzniveaus.			

In Tabelle 4 sind die Ergebnisse des PLS-Tests des vollständigen Forschungsmodells (H1-H5), wie in Abbildung 1 illustriert, aufgeführt. Es fällt auf, dass die Bestimmtheitsmaße (R^2) der sechs latenten Wertbeitragsvariablen durch Einbeziehung der technischen IT-Infrastruktur-Flexibilität auf 27,6% bis 41,1% ansteigen, sodass das vollständige Modell einen großen Anteil der Varianz in diesen Variablen erklärt.

Die (hoch-)signifikanten Pfadkoeffizienten von SOA-Adoption zur Modularität (0,185), Integration (0,233) und Skalierbarkeit (0,145) belegen den positiven Beitrag von SOA für die IT-Infrastruktur-Flexibilität (H2a-c bestätigt). Angesichts der niedrigen R^2 (2,1% bis 5,5%) liefert SOA jedoch nur einen geringen Erklärbeitrag für diese IT-Flexibilitätsdimensionen.

Generell lässt sich auch beobachten, dass die Pfade von SOA-Adoption zu den Vorteilen auf Unternehmensebene durch Einführung der Flexibilitätskonstrukte wesentlich schwächer ausfallen (vgl. Mediationsuntersuchung im Folgeabschnitt).

Weiterhin können alle vermuteten positiven Effekte einer höheren Modularität bis auf denjenigen zur B2B-Integration bestätigt werden (H3a-e bestätigt und H3f nicht bestätigt). Darüber hinaus wird die große Bedeutung der Integration für die Erlangung der sechs Unternehmensvorteile deutlich. Integration hat durchgehend hochsignifikante positive Einflüsse auf alle sechs Vorteile mit Pfadkoeffizienten von 0,245 bis 0,450 im Falle von STP (H4a-f bestätigt). Demgegenüber kann die Skalierbarkeit zwar zu vier der

betrachten Vorteile signifikant beitragen (H5c-f bestätigt), allerdings deutlich schwächer – und im Fall von Kostenreduktion und Agilitätssteigerung gar nicht (H5a-b nicht bestätigt).

Tabelle 4. Bestimmtheitsmaße (R^2) und Pfadkoeffizienten (β) der direkten Effekte im vollständigen Modell

	Determinanten:	Modularität	Integration	Skalierbarkeit	Kostenreduktion	Agilität	Datenqualität	Prozess-Monitoring	STP	B2B-Integration
H1a-f H2a-c	SOA-Adoption (β)	**0,185*** **	**0,233*** **	**0,145** **	0,027	**0,148** **	0,021	**0,228*** **	0,061	**0,117** **
H3a-f	Modularität (β)				**0,323*** **	**0,265*** **	**0,181** **	**0,192** **	**0,130***	0,009
H4a-f	Integration (β)				**0,276*** **	**0,318*** **	**0,245*** **	**0,305*** **	**0,450*** **	**0,398*** **
H5a-f	Skalierbarkeit (β)				0,028	-0,044	**0,304*** **	**0,094***	**0,184** **	**0,154** **
Kontroll-faktoren	Unternehmensgröße (β)				-0,017	**0,083***	-0,024	-0,019	0,016	**0,095***
	Branche (β):									
	Finanzdienstleistungen				-0,012	**-0,116** **	**0,142** **	**-0,130** **	**-0,068***	**0,127** **
	IT und Kommunikation				-0,031	-0,012	**-0,177** **	-0,021	0,016	0,010
	Handel und Logistik				0,044	0,032	0,020	0,042	**0,079***	0,069
	R^2:	0,034	0,055	0,021	0,300	0,309	0,330	0,341	0,411	0,276
	R^2 (nur Kontrollvariablen):				0,117	0,022	0,047	0,005	0,013	0,055
Signifikanzniveaus: *** $p \leq .01$; ** $p \leq .05$; * $p \leq .1$ (einseitige Signifikanzniveaus)										

Tabelle 5. Effektstärken (f^2) (: mittelstarker Effekt; *: geringer Effekt (nach [15]))**

	Determinanten:	Modularität	Integration	Skalierbarkeit	Kostenreduktion	Agilität	Datenqualität	Prozess-Monitoring	STP	B2B-Integration
H1a-f H2a-c	SOA-Adoption	**0,04***	**0,06***	**0,02***	0,01	0,01	**0,05***	**0,10***	**0,05***	**0,06***
H3a-f	Modularität				**0,10***	**0,06***	**0,03***	**0,04***	**0,02***	0,00
H4a-f	Integration				**0,07***	**0,10***	**0,07***	**0,10***	**0,22** **	**0,14***
H5a-f	Skalierbarkeit				0,01	0,00	**0,13***	0,00	**0,04***	**0,04***

Die Berücksichtigung der Kontrollvariablen[2] zeigt, dass nur marginale Teile der Varianzen (R^2 von 0,5% bis 11,7%) der Unternehmensvorteile in einem Modell, welches nur die Kontrollvariablen als Einflussfaktoren beinhaltet, erklärt werden, sodass der Erklärbeitrag von SOA und IT-Flexibilität bedeutsam ist. Die Unternehmensgröße korreliert nur in zwei der sechs Fälle schwach signifikant ($p \leq 0,1$) mit den Unternehmensvorteilen. Zusammenfassend kann davon ausgegangen werden, dass die Unternehmensgröße, wenn überhaupt, nur einen marginalen Zusammenhang mit den untersuchten Unternehmensvorteilen aufweist. Mit Blick auf die Branchenzugehörigkeit weichen nur Finanzdienstleister (bei fünf der sechs Unternehmensvorteile) etwas von den anderen Unternehmen der Dienstleistungsbranche ab.

Zusätzlich zu den Bestimmtheitsmaßen und Pfadkoeffizienten sind in Tabelle 5 die Einzeleffektstärken (f^2), die aus dem Test des vollständigen Modells resultieren, dargestellt. Die Tabelle zeigt, dass die Nutzung von SOA am stärksten auf Prozess-Monitoring wirkt und etwas schwächer auf Datenqualität, STP und B2B-Integration. Modularität wirkt am stärksten auf Kostenreduktion und schwächer auf die vier anderen Vorteile, während sie keinen Effekt auf B2B-Integration hat. Demgegenüber zeigt Integration die größten Effektstärken und dazu auch auf alle der untersuchten Vorteile. Eine besondere Bedeutung hat Integration für STP, B2B-Integration, Agilität und Prozess-Monitoring. Zuletzt werden ein schwächerer Effekt von Skalierbarkeit auf Datenqualität sowie marginale Effekte auf STP und B2B-Integration aufgezeigt.

4.4 Post-hoc-Analyse der Mediationseffekte

Die starke Abschwächung der Pfadkoeffizienten zwischen SOA und den Wertbeitragsdimensionen beim Vergleich der Werte in Tabelle 3 (vereinfachtes Modell mit H1) und Tabelle 4 (komplettes Modell) impliziert die Existenz von Mediationseffekten. Daher führen wir im Folgenden Mediationstests für jeden der drei Aspekte der technischen IT-Infrastruktur-Flexibilität (Modularität, Integration und Skalierbarkeit) durch. Dies ermöglicht eine Einschätzung darüber, ob der Wertbeitrag, der aus der Nutzung von SOA resultiert, vollständig über die drei Aspekte der technischen IT-Infrastruktur-Flexibilität erklärt werden kann (H6). Dazu wurde als Ausgangslage das reduzierte Forschungsmodell, dessen Ergebnisse in Tabelle 3 dargestellt sind, herangezogen und jeweils nur einer der drei potenziellen Mediatoren eingefügt. Anhand dieses Vorgehens lassen sich die Veränderungen der Pfadkoeffizienten von SOA zu den Unternehmensvorteilen auf einen der drei Aspekte zurückführen und es kann ermittelt werden, ob und wie stark die jeweilige Flexibilitätsdimension die Beziehung mediiert und dadurch erklärt [4]. Zur leichteren Nachvollziehbarkeit des Mediationstests zeigt

[2] Die drei binären Dummy-Variablen der Branchenzugehörigkeit wurden nicht gleichzeitig, sondern in separaten PLS-Modellen verwendet, da PLS keine freie Korrelation zwischen Variablen zulässt, aber ein Unternehmen immer nur einer der drei Branchen angehört. Daher sind alle dargestellten Ergebnisse dem PLS-Test entnommen, der nur die Unternehmensgröße, nicht jedoch eine der Branchenvariablen enthält.

Tabelle 6 in der ersten Zeile noch einmal die Pfadkoeffizienten des ursprünglichen Modells aus Tabelle 3 und in den Folgezeilen dann jeweils die Pfadkoeffizienten, wie sie sich bei Berücksichtigung jeweils eines der Mediatoren (d. h. Dimensionen der IT-Flexibilität) ergeben. In den untersten Zeilen mit den Testergebnissen zur Überprüfung der Mediationshypothesen 6a-c sind jeweils die Ergebnisse der z-Tests nach Sobel [42] aufgeführt, welche zur Überprüfung der statistischen Signifikanz des Vorhandenseins der Mediationseffekte genutzt wurden. Als Ergänzung ist die Stärke der mediierenden Effekte in Form des Variance Accounted for (VAF) [42] aufgeführt.

Tabelle 6. Pfadkoeffizienten (β) der direkten Effekte mit und ohne die jeweiligen Mediatoren, z-Werte nach Sobel sowie VAF-Werte

	Beitrag von SOA für:	Modularität	Integration	Skalierbarkeit	Kostenreduktion	Agilität	Datenqualität	Prozess-Monitoring	STP	B2B-Integration
	ohne Mediatoren (ß)				**0,193*** **	**0,278*** **	**0,161** **	**0,381*** **	**0,244*** **	**0,237*** **
	mit Modularität (ß)	**0,188*** **			0,063	**0,190*** **	0,079	**0291*** **	**0,139** **	**0,183*** **
	mit Integration (ß)		**0,125** **		0,051	**0,161*** **	0,052	**0,253*** **	**0,084***	**0,125** **
	mit Skalierbarkeit (ß)			**0,145** **	**0,148** **	**0,255*** **	**0,100***	**0,321*** **	**0,178*** **	**0,199*** **
H6a	Modularität			Sobel-z:	**2,426*** **	**2,266** **	**2,309** **	**2,317** **	**2,356*** **	**1,978** **
				VAF:	**1,00**	**0,22**	**1,00**	**0,16**	**0,25**	**0,18**
H6b	Integration			Sobel-z:	**2,548*** **	**2,510*** **	**2,471*** **	**2,513*** **	**2,648*** **	**2,482*** **
				VAF:	**1,00**	**0,28**	**1,00**	**0,22**	**0,36**	**0,31**
H6c	Skalierbarkeit			Sobel-z:	**1,554***	1,173	**1,713** **	**1,559***	**1,690** **	**1,576***
				VAF:	**0,16**	0,07	**0,28**	**0,08**	**0,18**	**0,15**
Anmerkung: *** p ≤ .01; ** p ≤ .05; * p ≤ .1 (einseitige Signifikanzniveaus)										

Durch die Berücksichtigung von Modularität bzw. Integration werden die ursprünglich signifikanten Pfade von SOA-Nutzung auf Kostenreduktion sowie auf Datenqualität insignifikant. Für diese beiden Beziehungen kann also von einer vollständigen Mediation sowohl durch Modularität als auch durch Integration gesprochen werden[3]. Dies bedeutet, dass sich der Wertbeitrag von SOA hinsichtlich der Dimensionen Kostenreduktion und Datenqualität vollständig dadurch erklären lässt, dass SOA die IT-Flexibilität im Sinne der Erhöhung des Modularitäts- und Integrationsgrades steigert. Für den Wertbeitrag von SOA hinsichtlich der anderen vier Dimensionen fungieren Modularität und Integration dagegen nur als teilweise Mediatoren (Erklärbeitrag (VAF) zwischen 16% und 36%), d. h. es verbleiben substanzielle Resteffekte von 64%-84%, die anderweitig zu erklären sind.

Die Skalierbarkeit als dritte Dimension der IT-Flexibilität spielt generell eine geringere Rolle. Sie tritt in keinem Fall als vollständiger Mediator auf und erklärt nur zwischen 7% und 28% des SOA-Wertbeitrags, sodass insbesondere im Fall von Agilität im statistischen Sinne nicht von einem Mediationseffekt gesprochen werden kann. Insgesamt belegt die Mediationsanalyse somit die Bedeutung der technischen IT-Infrastruktur-Flexibilität als Mediator bzw. Erklärkomponente für den SOA-Wertbeitrag. Sie zeigt aber auch substanziellen Raum für andere Erklärfaktoren auf, die nicht Bestandteil des Forschungsmodells waren, aber im folgenden Kapitel diskutiert werden.

5. DISKUSSION DER ERGEBNISSE

SOA leistet einen signifikanten Wertbeitrag:

Insgesamt zeigen die Ergebnisse, dass die Nutzung serviceorientierter Architekturen (SOA) einen signifikanten Wertbeitrag im Sinne von Kostenreduktion, Verbesserung der unternehmerischen Agilität, der Datenqualität, des Prozess-Monitorings sowie der internen und unternehmensübergreifenden Prozessintegration leistet. Vor allem der starke Zusammenhang zwischen SOA und dem Prozess-Monitoring fällt auf, was sich dadurch erklären lässt, dass Unternehmen häufig die Einführung von SOA mit einem umfassenden Geschäftsprozessmanagement verbinden [8]. Die niedrigen Bestimmtheitsmaße (R^2) belegen jedoch, dass SOA alleine nur einen eher geringen Teil dieser Unternehmensvorteile zu erklären vermag. Dabei ist natürlich zu berücksichtigen, dass die verschiedenen Vorteile in unterschiedlichem Maß durch zahlreiche andere organisatorische und auch IT-basierte Faktoren beeinflusst werden. Der Aufbau der IT-Architektur kann bspw. nur einen bestimmten Teil der gesamten IT-Kosten erklären.

SOA äußert sich in klassischen Aspekten der technischen IT-Infrastruktur-Flexibilität:

Das Forschungsmodell hat dementsprechend auch untersucht, ob und inwieweit SOA die *technischen* Flexibilitätsaspekte (als dem häufig erstgenannten SOA-Ziel) beeinflusst. Die Ergebnisse in Tabelle 4 zeigen, dass sich die Nutzung von SOA in der Tat positiv auf klassische Aspekte der Flexibilität der technischen IT-Infrastruktur (Modularität, Integration und Skalierbarkeit) auswirkt. Zusätzlich kann unter Berücksichtigung dieser drei wichtigen Aspekte auch der Erklärbeitrag (Bestimmtheitsmaß R^2) der sechs Erfolgsgrößen stark gesteigert werden, sodass dem Modell hohe Erklärkraft in Bezug auf Erreichung dieser Unternehmensvorteile durch die IT-Architektur beigemessen werden kann. Ebenso wird deutlich, dass der Integrationsaspekt der IT-Flexibilität eine überaus hohe Bedeutung für die Erlangung aller sechs Vorteile besitzt (vgl. Tabelle 4) und sogar die höchste Bedeutung für Agilität, Prozess-Monitoring (zusammen mit der Nutzung von SOA), STP und B2B-Integration besitzt (vgl. Tabelle 5). Einen fast ebenso starken Einfluss besitzt die Modularität, die auf fünf der sechs Vorteile wirkt, allerdings nicht zur B2B-Integration beiträgt (vgl. Tabelle 4). Eine Erklärung hierfür ist, dass für die unternehmensübergreifende Integration nicht nur der Grad der Modularität der eigenen IT eine Rolle spielt, sondern vor allem auch die Modularität der IT der anderen Unternehmen. Die Modu-

[3] Die entsprechenden VAF-Werte sind dementsprechend anhand der Empfehlungen von Shrout und Bolger auf 1,00 gesetzt [42].

larität der eigenen IT ist dagegen am bedeutendsten für die Erreichung von Kostenreduktionszielen (vgl. Tabelle 5). Insgesamt zeigt sich allerdings auch, dass SOA nur einen relativ geringen Teil der technischen IT-Infrastruktur-Flexibilität erklärt, sodass die Ergebnisse vor dem Hintergrund der hoch signifikanten Pfade belegen, dass SOA zwar *ein* mögliches Mittel zur Erlangung von IT-Flexibilität ist, Unternehmen jedoch auch sehr gut ohne SOA flexible IT-Infrastrukturen und die entsprechenden Wertbeiträge realisieren können. Dieses Bild mag sich jedoch mit einer Weiterverbreitung des SOA-Konzepts in den nächsten Jahren verändern.

Der SOA-Wertbeitrag lässt sich durch technische IT-Infrastruktur-Flexibilität als Mediator erklären:

Die Analyse der Rolle der technischen IT-Infrastruktur-Flexibilität als Mediator des Wertbeitrages von SOA bestätigt insgesamt die Hypothese, dass SOA vor allem über die klassischen Aspekte der technischen IT-Infrastruktur-Flexibilität (Modularität, Integration und Skalierbarkeit) einen positiven Wertbeitrag leistet (vgl.
Tabelle 6). So wurde in 17 von 18 geprüften Zusammenhängen die IT-Flexibilität als zumindest *ein* substantieller Erklärfaktor für den SOA-Wertbeitrag identifiziert, in vieren davon sogar als der alleinige. Es bleibt nun zu explorieren, welche anderen Faktoren den Zusammenhang von SOA und den Zieldimensionen erklären könnten. Ein mögliches Argument ist die Bedeutung der erhöhten Wiederverwendung von Funktionalitäten und damit Erzielung von Synergiepotenzialen. Ein anderes Argument stellt das IT-Business-Alignment [7] dar, welches durch die Implementierung einer SOA signifikant verbessert werden könnte, da in der Konzeptions- und Implementierungsphase gerade im SOA-Kontext die enge Zusammenarbeit und Abstimmung von Fachbereichen und IT-Entwicklung notwendig ist. Diese Abstimmung führt dazu, dass die IT-Architektur besser auf die Geschäftsbedürfnisse ausgerichtet wird und so der Wertbeitrag der IT zunimmt – unabhängig davon, ob sie flexibler wird oder nicht.

Einschränkungen der Studie:

Da es bisher keine empirischen Studien gibt, die die Vorteile der Nutzung von SOA betrachtet haben, mussten einige Messinstrumente neu entwickelt oder angepasst werden. Die Überprüfung der Messmodellgüte zeigt jedoch, dass dies insgesamt gut gelungen scheint. Anhand unserer Literaturanalyse haben wir uns auf die Untersuchung von sechs Vorteilen für Unternehmen beschränkt. Es besteht natürlich darüber hinaus die Möglichkeit, dass SOA auch auf andere - hier nicht untersuchte - Aspekte positive oder negative Auswirkungen hat. Außerdem haben wir nur Unternehmen in einer Branche befragt Dies führt zu einer zuverlässigeren Interpretation der Ergebnisse, da nicht so viele Einflüsse von heterogenen Branchen auftreten, geht aber zulasten der Generalisierbarkeit der Ergebnisse. Weiterhin wurden nur die IT-Architekten befragt. Diese haben zwar den besten Überblick über den Stand und die Art der SOA-Implementierung und generellen IT-Architektur ihres Unternehmens, Antworten zum Wertbeitrag von SOA könnten allerdings im Vergleich zur Realität zu positiv ausfallen. Generell stehen den Vorteilen von empirischen Breitenstudien (bspw. eine möglichst breite Datenbasis, die nicht so sehr von Einzelfällen beeinflusst wird) auch Nachteile gegenüber. Dabei ist am gravierendsten, dass keine Detailbetrachtung erfolgen kann, wie bspw. in Fallstudien, sondern immer eine gewisse Aggregation und Vereinfachung stattfindet. So zeigen die Ergebnisse zwar, dass SOA einen signifikanten Einfluss auf die untersuchten Wertbeiträge hat, der Erklärbeitrag jedoch verhältnismäßig gering ausfällt. Die gleichzeitige Anwendung von Fallstudien für eine umfassendere Beantwortung der Forschungsfrage ist also wünschenswert und dementsprechend auch Teil unserer laufenden Forschungsarbeiten.

6. ZUSAMMENFASSUNG UND AUSBLICK

In diesem Beitrag konnten wir die in der Literatur gerne postulierten, aber bisher höchstens in Form von Fallstudien evaluierten Vorteile, die durch die Nutzung von SOA entstehen, auf Basis einer breiten Umfrage unter deutschen Dienstleistungsunternehmen empirisch überprüfen. Die Nutzung von SOA zur Unterstützung von Geschäftsprozessen führt demnach zu einer Kostenreduktion, Steigerung der unternehmerischen Agilität, der Datenqualität, des Prozess-Monitorings, der internen Integration von Geschäftsprozessen sowie der überbetrieblichen Integration und trägt so zu den Dynamic Capabilities eines Unternehmens bei.

Auch konnten wir mit unserer Forschung zeigen, dass sich das vage und häufig mehrdeutig verstandene Konzept serviceorientierter Architekturen in „klassischen“ Aspekten flexibler IT-Infrastrukturen, wie Modularität, Flexibilität und Skalierbarkeit, manifestiert. Diese haben einen positiven Einfluss auf die untersuchten Unternehmensvorteile und stellen als „Mediatoren“ auch einen oder tlw. sogar den alleinigen Erklärfaktor für die Erzielung eines Wertbeitrages aus serviceorientierten Architekturen dar.

Mit dieser Studie bieten wir der Forschungsgemeinschaft sowohl eine erste empirische Validierung des SOA-Wertbeitrags als auch einen Vorschlag zur Klassifikation und empirischen Messung der aus SOA resultierenden Vorteile. Wir haben theoretisch argumentiert und empirisch gezeigt, dass die IT-Flexibilität je nach Zieldimension der alleinige oder auch nur ein wenig bedeutsamer Erklärfaktor für den Einfluss von SOA auf die jeweilige Zieldimension ist. Aufgabe zukünftiger Forschung ist es, komplementäre Erklärfaktoren, wie bspw. SOA-Governance [26] oder Prozessstandardisierung [9], zu untersuchen, um ein vollständiges Bild über den SOA-Wertbeitrag zu erhalten und IT-Managern klare Richtlinien vorzugeben, wie eine SOA auszurichten ist, um die beabsichtigten SOA-Ziele zu erreichen.

7. DANKSAGUNG

Die Autoren danken den Partnern des E-Finance Lab für ihre finanzielle Unterstützung.

8. REFERENZEN

[1] Armstrong, J. S. und Overton, T. S. 1977. Estimating Nonresponse Bias in Mail Surveys. *Journal of Marketing Research (JMR)*. 14, 3, 396-402.

[2] Bagozzi, R. P. und Yi, Y. 1988. On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*. 16, 1, 74-94.

[3] Barney, J. B. 1991. Firm resources and sustained competitive advantage. *Journal of Management*. 17, 99-120.

[4] Baron, R. M. und Kenney, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. of Personality and Social Psychology*. 51, 6, 1173-1182.

[5] Baskerville, R., Cavallari, M., Hjort-Madsen, K., Pries-Heje, J., Sorrentino, M., und Virili, F. 2005. Extensible Architectures: The Strategic Value of Service-Oriented

Architecture in Banking. In *Proceedings of the 13th European Conference on Information Systems*, Regensburg.
[6] Becker, A., Buxmann, P., und Widjaja, T. 2009. Value Potential and Challenges of Service-Oriented Architectures - A User and Vendor Perspective. In *Proceedings of the 17th European Conference on Information Systems*, Verona, Italy.
[7] Beimborn, D. und Joachim, N. 2009. Proposing the Relationship between IT Business Alignment and the Business Value of Service-Oriented Architectures in Financial Firms. In *Enterprise Applications and Services in the Finance Industry*, D. Kundisch, D. J. Veit, T. Weitzel, und C. Weinhardt, Hg. Lecture Notes in Business Information Processing. Springer, Berlin, 78-93.
[8] Beimborn, D. und Joachim, N. 2010. The joint impact of service-oriented architectures and business process management on business process quality: an empirical evaluation and comparison. *Information Systems and E-Business Management*. DOI = http://dx.doi.org/10.1007/s10257-010-0129-1.
[9] Beimborn, D., Joachim, N., und Muenstermann, B. 2009. Impact of Service-oriented Architectures (SOA) on Business Process Standardization - Proposing a Research Model. In *Proceedings of the 17th European Conference on Information Systems*, S. Newell, E. Whitley, N. Pouloudi, J. Wareham, und L. Mathiassen, Hg., Verona, Italy.
[10] Beimborn, D., Joachim, N., und Weitzel, T. 2008. Drivers and Inhibitors of SOA Business Value – Conceptualizing a Research Model. In *Proceedings of the 14th Americas Conference on Information Systems*, Toronto, ON, Canada.
[11] Bharadwaj, A. S. 2000. A resource-based Perspective on Information Technology Capability and Firm Performance: an empirical Investigation. *MIS Quarterly*. 24, 1, 169-196.
[12] Bieberstein, N., Bose, S., Fiammante, M., Jones, K., und Shah, R. 2005. *Service-Oriented Architecture (SOA) Compass: Business Value, Planning, and Enterprise Roadmap*. IBM Press, Upper Saddle River, NJ.
[13] Byrd, T. A. und Turner, D. E. 2000. Measuring the Flexibility of Information Technology Infrastructure: Exploratory Analysis of a Construct. *Journal of Management Information Systems*. 17, 1, 167-208.
[14] Chanopas, A., Krairit, D., und Khang, D. B. 2006. Managing Information Technology Infrastructure: A New Flexibility Framework. *Management Research News*. 29, 10, 632-651.
[15] Chin, W. W. 1998. The Partial Least Square Approach to Structural Equation Modeling. In *Modern Methods for Business Research*, G. A. Marcoulides, Hg. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 295-336.
[16] Chin, W. W. und Newsted, P. R. 1999. Structural equation modeling analysis with small samples using partial least squares. In *Statistical strategies for small sample research*, R. H. Hoyle, Hg. Sage, Thousand Oaks, 307-341.
[17] Chung, S. H., Byrd, T. A., Lewis, B. R., und Ford, F. N. 2005. An empirical study of the relationships between IT infrastructure flexibility, mass customization, and business performance. *SIGMIS Database*. 36, 3, 26-44.
[18] Dedrick, J., Gurbaxani, V., und Kraemer, K. L. 2003. Information Technology and Economic Performance: A Critical Review of the Empirical Evidence. *ACM Computing Surveys*. 35, 1, 1-28.
[19] Duncan, N. B. 1995. Capturing Flexibility of Information Technology Infrastructure: A Study of Resource Characteristics and their Measure. *Journal of Management Information Systems*. 12, 2, 37-57.
[20] Gable, G. G., Sedera, D., und Chan, T. 2008. Re-conceptualizing Information System Success: The IS-Impact Measurement Model. *Journal of the AIS*. 9, 7, 1-32.
[21] Gefen, D., Straub, D. W., und Boudreau, M.-C. 2000. Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the AIS*. 4, 1-77.
[22] Haines, M. N. und Haseman, W. D. 2009. Service-Oriented Architecture Adoption Patterns. In *Proceedings of the 42nd Hawaii International Conference on System Sciences*.
[23] Heffner, R. 2009. Insights For CIOs: SOA And Beyond. *Forrester Research*.
[24] Hirschheim, R., Welke, R., und Schwarz, A. 2010. Service-Oriented Architecture: Myths, Realities, and a Maturity Model. *MIS Quarterly Executive*. 9, 1, 37-48.
[25] Janssen, M. 2008. Exploring the service-oriented enterprise: Drawing lessons from a case study. In *Proceedings of the 41st Hawaii International Conference on System Sciences* IEEE Computer Society, Waikoloa, Big Island, HI, USA.
[26] Joachim, N., Beimborn, D., und Weitzel, T. 2011. SOA-Governance für effektive serviceorientierte Architekturen – Eine empirische Studie in der deutschen Dienstleistungswirtschaft. In *10. Internationale Tagung Wirtschaftsinformatik*, Zürich.
[27] Kumar, S., Dakshinamoorthy, V., und Krishnan, M. S. 2007. SOA and Information Sharing in Supply Chain: "How" Information is Shared Matters! In *Proceedings of the 28th International Conference on Information Systems*, Montreal.
[28] Liang, H., Saraf, N., Hu, Q., und Xue, Y. 2007. Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. *MIS Quarterly*. 31, 1, 59-87.
[29] Mahoney, J. T. und Pandian, R. J. 1992. The Resource-Based View Within the Conversation of Strategic Management. *Strategic Management J.*. 13, 5, 363-380.
[30] Makadok, R. 2001. Toward a Synthesis of the Resource-Based and Dynamic-Capability Views of Rent Creation. *Strategic Management Journal*. 22, 5, 387-401.
[31] Melville, N., Kraemer, K. L., und Gurbaxani, V. 2004. Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly*. 28, 2, 283-322.
[32] Mueller, B., Viering, G., Legner, C., und Riempp, G. 2010. Understanding the Economic Potential of Service-Oriented Architecture. *Journal of Management Information Systems*. 26, 4, 145-180.
[33] Nunnally, J. C. 1978. *Psychometric theory.* McGraw-Hill, New York.
[34] Oh, L.-B., Leong, Y.-X., Teo, H.-H., und Ravichandran, T. 2007. Service-oriented Architecture and Organizational Integration: An Empirical Study of IT-Enabled Sustained Competitive Advantage. In *Proceedings of the 28th International Conference on Information Systems*, Montreal.
[35] Papazoglou, M. P. und Heuvel, W.-J. 2007. Service Oriented Architectures: Approaches, Technologies and Research Issues. *The VLDB Journal*. 16, 3, 389-415.
[36] Penrose, E. T. 1959. *The theory of the growth of the firm.* Oxford University Press, Oxford, UK.
[37] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., und Podsakoff, N. P. 2003. Common Method Bias in Behavioral

Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*. 88, 5, 879-903.
[38] Ren, M. und Lyytinen, K. J. 2008. Building Enterprise Architecture Agility and Sustenance with SOA. *Communications of the AIS*. 22, 75-86.
[39] Ringle, C. M., Wende, S., und Will, A. 2007. SmartPLS 2.0 M3. Hamburg.
[40] Sambamurthy, V., Bharadwaj, A., und Grover, V. 2003. Shaping Agility through digital Options: reconceptualizing the Role of Information Technology in contemporary Firms. *MIS Quarterly*. 27, 2, 237-263.
[41] Shang, S. und Seddon, P. B. 2002. Assessing and managing the benefits of enterprise systems: the business manager's perspective. *Information Systems Journal*. 12, 4, 271-299.
[42] Shrout, P. und Bolger, N. 2002. Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations. *Psychological Methods*. 7, 4, 422-445.
[43] Tafti, A., Mithas, S., und Krishnan, M. S. 2008. The Effects of Information Technology and Service-Oriented Architectures on Joint Venture Value. In *Proceedings of the 29th International Conference on Information Systems*.
[44] Tallon, P. P. 2008. Inside the adaptive enterprise: an information technology capabilities perspective on business process agility. *Information Technology and Management*. 9, 1, 21-36.
[45] Teece, D. J., Pisano, G., und Shuen, A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal*. 18, 7, 509-533.
[46] Viering, G., Legner, C., und Ahlemann, F. 2009. The (Lacking) Business Perspective on SOA - Critical Themes in SOA Research. In *9. Internationale Tagung Wirtschaftsinformatik*, H. R. Hansen, D. Karagiannis, und H.-G. Fill, Hg., Wien, 45-54.
[47] Vitharana, P., Bhaskaran, K., Jain, H., Wang, H. J., und Zhao, J. L. 2007. Service-Oriented Enterprises and Architectures: State of the Art and Research Opportunities. In *Proceedings of the 13th Americas Conference on Information Systems*, Keystone, USA.
[48] Yoon, T. und Carter, P. E. 2007. Investigating the Antecedents and Benefits of SOA Implementation: A Multi-Case Study Approach. In *Proceedings of the 13th Americas Conference on Information Systems*, Keystone, USA.

ANHANG: INDIKATOREN DES MESSMODELLS

Konstrukt [Quellen]	Label	Indikator (Bewertung auf 7-stufiger Likert-Skala („trifft gar nicht zu“ bis „trifft voll zu“) sofern nicht anders angegeben)	Ladung
SOA-Adoption	SOA	In welchem Ausmaß sind die Prozesse jeweils durch eine serviceorientierte Architektur (SOA) unterstützt? Produktion, Operations; Beschaffung, Anbindung von Partnerunternehmen; Entwicklung von Produkten/Dienstleistungen (F&E); Marketing, Vertrieb, Customer Relations; Querschnittsfunktionen (Rewe., HR etc.) (jew. 5-er Likert-Skala von „keine SOA“ bis „nur SOA“. Die 5 Indikatoren wurden per Faktoranalyse aggregiert.)	1.000
Modularität [17, 44]	MOD1	Wir können unseren Applikationen ohne größere Probleme neue Funktionalitäten hinzufügen.	0,832***
	MOD2	Durch Austausch oder Veränderung einzelner Komponenten wird die restliche IT-Infrastruktur nicht beeinträchtigt.	0,858***
	MOD3	Unsere Applikationen setzen sich aus klar abgegrenzten Modulen zusammen.	0,818***
Integration [13, 14]	ITG1	Der Austausch von Daten zwischen verschiedenen Applikationen ist leicht möglich.	0,881***
	ITG2	Daten einer Applikation lassen sich leicht in einer anderen Applikation nutzen.	0,888***
	ITG3	Wir können leicht konsolidierte Sichten auf alle einen Kunden betreffenden Daten erzeugen.	0,838***
	ITG4	Zusätzliche Datenformate (EDI, XML) können wir leicht in unsere Applikationen integrieren.	0,727***
Skalierbarkeit [14, 20]	SKA1	Unsere IT-Infrastruktur kann Spitzen in den Transaktionsvolumina leicht kompensieren.	0,859***
	SKA2	Die IT-Infrastruktur stellt genügend Kapazität bereit, um Zusatzaufträge leicht abzuwickeln.	0,892***
	SKA3	Die Performance unserer IT-Infrastruktur erfüllt unsere Geschäftsbedarfe vollständig.	0,887***
Kosten-reduktion	KOR1	Alle Potenziale zur Optimierung unserer IT-Betriebskosten wurden ausgeschöpft.	0,732***
	KOR2	Alle Potenziale zur Reduzierung der Softwareentwicklungskosten wurden ausgeschöpft.	0,826***
	KOR3	Alle Potenziale zur Optimierung unserer IT-Management-Kosten wurden ausgeschöpft.	0,789***
	KOR4	Die Funktionalitäten unserer Applikationslandschaft sind frei von überflüssigen Redundanzen.	0,834***
	KOR5	Geschäftsprozessabläufe sind in unserem Unternehmen nicht mehrfach implementiert.	0,822***
Agilität [17, 44]		Unsere IT ermöglicht es unserem Unternehmen, …	
	AGI1	… Produkte/Dienstleistungen flexibel für einzelne Kunden anzupassen.	0,734***
	AGI2	… unser Produkt-/Dienstleistungsangebot schneller als unsere Wettbewerber zu ändern.	0,902***
	AGI3	… eine kürzere Time-to-Market als unsere Wettbewerber zu realisieren.	0,924***
	AGI4	… schnell und flexibel auf Änderungen der Kundennachfrage zu reagieren.	0,880***
Datenqualität [20]	DAQ1	Die durch unsere Applikationen bereitgestellten Daten sind vollständig und aktuell.	0,931***
	DAQ2	Die durch unsere Applikationen bereitgestellten Daten sind fehlerfrei und konsistent.	0,938***
Prozess-Monitoring	PRM1	Es liegen stets aktuelle Geschäftsprozesskennzahlen (z. B. Durchlaufzeiten, Fehler) vor.	0,905***
	PRM2	Die Transparenz der Geschäftsprozesse erleichtert die Erfüllung von Compliance-Vorgaben.	0,893***
STP	STP1	Unsere Anwender müssen dieselben Daten häufig mehrfach eingeben.	0,663***
	STP2	Alle Applikationen sind, soweit für Geschäftsprozesse sinnvoll, miteinander integriert.	0,875***
	STP3	Die Geschäftsprozesse unserer Fachbereiche sind gut miteinander integriert.	0,877***
B2B-Integration [34]	B2B1	Die Geschäftsprozesse unseres Unternehmens sind gut mit denen unserer Partner integriert.	0,951***
	B2B2	Daten werden effizient zwischen unseren Geschäftspartnern und uns ausgetauscht.	0,946***
	B2B3	Insgesamt ist die Integration zwischen unseren Geschäftspartnern und uns hoch.	0,935***
Unternehmens-größe	UNG	Logarithmus der gesamten Mitarbeiteranzahl des Unternehmens. (Quelle: Sekundärstatistik)	
Branche	BRA1 BRA2 BRA3	Drei Branchen-Binärvariablen: Finanzdienstleistungen, Informationstechnologie & Kommunikation sowie Handel & Logistik (Quelle: Sekundärstatistik)	
Anmerkung: *** $p \leq .01$ (einseitige Signifikanzniveaus)			

When to manage risks in IS projects: An exploratory analysis of longitudinal risk reports

Stefan Hoermann
Technsiche Universitaet Muenchen (TUM)
Boltzmannstr. 3
85748 Garching b. Muenchen
+49 89 289 19546
stefan.hoermann@in.tum.de

Michael Schermann
Technsiche Universitaet Muenchen (TUM)
Boltzmannstr. 3
85748 Garching b. Muenchen
+49 89 289 19507
michael.schermann@in.tum.de

Helmut Krcmar
Technsiche Universitaet Muenchen (TUM)
Boltzmannstr. 3
85748 Garching b. Muenchen
+49 89 289 19532
krcmar@in.tum.de

ABSTRACT
Research attributes the mixed performance of IS projects to a poor understanding of risks and thus limited capabilities to manage such risks. In line with others, we argue that the poor understanding of risks is partly due to the fact, that current research almost exclusively concentrates on *which* risks are important in IS projects. In contrast to this static view, we focus on the temporal aspect of project risks, i.e., we explore *when* risks become more or less important during a project. In doing so, we analyze an archive of risk reports of completed enterprise software projects. Project managers regularly issued the risk reports to communicate the status of the particular project. Our findings are as follows: First, risk exposure and thus the perceived importance of risk types does vary over project phases. Second, the volatility of risk exposure varies over risk types and project phases. Third, risks of various origin exhibit synchronous changes in risk exposure over time. From a research perspective, these findings substantiate the need for a temporal perspective on IS project risks. Thus, we suggest augmenting the predominant static view on project risks to help project managers in focusing their scarce resources. From a practical perspective, we highlight the benefits of regularly performing risk management throughout projects and constantly analyzing the project portfolio. In sum, we provide a first time, descriptive and exploratory view on variations in project risk assessments over time.

Keywords
Risk management, project failure, software project management.

1. INTRODUCTION
Both, researchers and practitioners agree on the challenging nature of managing IS projects. Since the beginning of the IS discipline, researchers continuously report remarkably high failure rates for IS projects (e.g. Alter and Ginzberg, 1978; Zmud, 1980). Despite the breadth and depth of research results on effective project management and the widespread use of tools, methods, and standards designed for supporting project managers, today's IS projects do not seem to be any more successful. Contemporary studies still report failure rates of 33% (Sauer et al., 2007).

A major research stream on IS project management attributes the low performance of IS projects to a poor understanding of related risks and limited capabilities to manage risks in IS projects (e.g. Iversen et al., 2004; Ropponen and Lyytinen, 1997). Following fundamental definitions of risk in reference disciplines (Knight, 2002; March and Shapira, 1987), IS researchers commonly define project risks as events with a perceived probability of occurrence and a perceived negative impact on project objectives (Alter and Sherer, 2004; Boehm, 1991; Charette, 1996; Heemstra and Kusters, 1996). The product of probability and impact is called risk exposure (RE) and denotes the perceived importance of a risk at the time of assessment. Managing risks requires first to identify, understand, and prioritize risks. Following this, the project manager and other stakeholders plan, implement, and monitor actions to control or mitigate risks. Although names and number of phases of risk management vary across authors, the first phase is usually called risk assessment or risk analysis while the latter is called risk control (Boehm, 1991; Heemstra and Kusters, 1996).

Being pivotal to effectively controlling risks in IS projects, many IS researchers focus on the capabilities required for assessing risks (Tiwana and Keil, 2006). Research on ranking and classifying risks establishes the variety of risks in IS projects and subsequently help project managers identify and prioritize risks more effectively (e.g. Barki et al., 1993; Boehm, 1991; Kappelman et al., 2006; Keil et al., 1998; Moynihan, 1997; Schmidt et al., 2001). Other researchers focus on understanding project risks by proposing frameworks of dimensions and domains of projects risks and their effect on IS project performance (e.g. Han and Huang, 2007; Jiang and Klein, 2000; Nidumolu, 1995; Sauer et al., 2007; Wallace et al., 2004). Furthermore, research is available on the effects of risk control activities and contingency factors of IS project risk management and their effect on IS project performance (Barki et al., 2001; Ropponen and Lyytinen, 2000).

In this paper, we focus on the temporal aspect of project risks. While still being relatively unexplored, extant literature argues that

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

understanding how risks change over time is pivotal for progress in managing IS risks effectively and efficiently (Alter and Ginzberg, 1978; Gemino et al., 2008; Pinto and Prescott, 1988; Somers and Nelson, 2004). Hence, we argue that managing IS project risks successfully, i.e., initiating the appropriate measures, depends on the temporal nature of risk and the appropriate point in time for action. This argument is reinforced by the fact that resources for project risk management are frequently in short supply. Understanding the temporal characteristics of project risks would help IS professionals allocate those resources more precisely. Hence, our research question is: *Do IS project risks evolve over time?* Our research goal is to establish a descriptive and exploratory view on the temporal aspect of IS project risks. To do this, we analyze continuous risk reports from 111 enterprise software projects. Our analysis suggests three findings: First, risk exposure and thus the perceived importance of risk types does vary over project phases. Second, the volatility of risk exposure varies over risk types and project phases. Third, risks of various origin exhibit synchronous changes in risk exposure over time. In sum, we provide a first illustration on how risk assessments of project managers vary over time.

The remainder of the paper is structured as follows: In the next section, we analyze extant research on dimensions of IS project risks. In particular, we review existing results on temporal aspects of IS project risks. Extant literature suggests that risks evolve in distinctive ways and that understanding temporal patterns may provide useful insights for both IS researchers and IS practitioners. Next, we analyze an archive of risk assessments by project managers of a leading multinational enterprise software company. Since our goal is to provide a first descriptive and exploratory perspective on temporal patterns of IS project risk types, we employ cluster analysis based on variations in the perceived importance of risk types along the project course. We derive nine clusters with distinct patterns representing changing risk perceptions of project managers. Next, we discuss the characteristics and implications of the patterns. Finally, we describe the potential limitations of our results and recommend future areas of research.

2. THEORETICAL BACKGROUND

2.1 Dimensions of IS Project Risks

IS researchers agree that IS project risks are multidimensional. The checklists mentioned in the introduction are frequently extended by classifying the risks into various dimensions. McFarlan (1981) for instance, suggests three dimensions of IS project risks: project size, project structure and experience with the technology. To quantify IS project risks, Barki et al. (1993) conduct a comprehensive literature review resulting in 35 risks and employs factor analysis to derive five dimensions of IS project risk which elaborate on McFarlan's (1981) dimensions: technological newness, application size, lack of expertise, application complexity, and organizational environment. Schmidt et al. (2001) elicit 53 risks using a Delphi study approach and group them into 14 dimensions: Corporate environment, sponsorship/ownership, relationship management, project management, scope, requirements, funding, scheduling, development process, personnel, staffing, technology, external dependencies, and planning. The risks and dimensions identified by Schmidt et al. (2001) do not only comprise all risks identified in prior studies but also extend these suggesting that new risks have emerged over time.

In another attempt to answer the question of dimensionality, Wallace et al. (2004) generate an extensive list of risks found in academic literature and articles written by practitioners. They also come up with six dimensions of IS project risk: Planning and control, team, complexity, requirements, user, and organizational environment. These dimensions can be mapped to three domains: Project management (planning and control, team), the technical subsystem (complexity, requirements), and the social subsystem (user, organizational environment). Tesch et al. (2007) reinvestigate the risk dimensions identified by Schmidt et al. (2001) and find significant similarities among them. In line with the results of Wallace et al. (2004), the authors reduce the number of dimensions back to six: sponsorship/ownership, funding and scheduling, personnel and staffing, scope, requirements, and relationship management. Sherer et al. (2004) critically reflect on existing approaches to classifying IS project risks and propose a work system framework, which integrates risks and work practices, participants, information, technology, products and services, customers, environment, infrastructure, and strategy of a work system. Table 1 gives an overview on the dimensions identified in these studies.

Table 1. Overview on Dimensions of IS Project Risks

McFarlan (1981)	(1) Project size, (2) Experience with technology, (3) Project structure
Barki et al. (1993)	(1) Technological newness, (2) Application size, (3) Lack of expertise, (4) Technical complexity, (5) Organizational environment
Schmidt et al. (2001)	(1) Corporate environment, (2) Sponsorship/ownership, (3) Relationship management, (4) Project management, (5) Scope, (6) Requirements, (7) Funding, (8) Scheduling, (9) Development process, (10) Personnel, (11) Staffing, (12) Technology, (13) External dependencies, (14) Planning
Wallace et al. (2004)	(1) Project management, (2) Technical subsystem, (3) Social subsystem
Tesch et al. (2007)	(1) Sponsorship/ownership, (2) Funding and scheduling, (3) Personnel and staffing, (4) Scope, (5) Requirements, (6) Relationship management
Sherer et al. (2004)	(1) Environment, (2) Strategies, (3) Infrastructure, (4) Customers, (5) Products and services, (6) Work practices, (7) Participants, (8) Information, (9) Technology

While it is arguable, whether or not these dimensions are exhaustive, all of them are derived in a rather intuitive manner and are based on the domain of origin of the respective risks.

The literature mentioned above has considerably extended our understanding of IS project risks and supports project managers in identifying potential threats to their project goals and formulating 'more specific risk management strategies' (Wallace, 2004). However, in addition to the knowledge which risks appear in IS projects, the question of when they appear and how they evolve is also of substantial interest to IS project managers and researchers. Alter et al. (2004) discuss several potential limitations of extant research on IS project risk, one of them being the 'frequent

omission of the temporal nature of risk'. As the authors state, risks are likely to have different temporal patterns, i.e., not only might their importance vary over the project life cycle but also the points of time at which they occur.

2.2 Temporal Aspects of IS Project Risks

In an early study, Alter et al. (1978) address the temporal aspect of IS project risks and suggest that linking risks to project phases and consequently adapting project risk management increases the likelihood of successful IS projects. The authors identify eight risks and allocate them to seven project phases depending on when their effects become apparent. The identified risks include: 'non-existent or unwilling users', 'multiple users and designers', 'disappearing users, designers or maintainers', 'inability to specify the purpose or usage pattern in advance', 'lack or loss of support', 'lack of prior experience with similar systems', 'inability to predict and cushion the impact on all parties', and 'technical problems or cost-effectiveness issues'. Alter et al. (1978) map all of these risks to one of the first four project phases and propose several risk-reducing strategies.

Sherer et al. (2004) pick up this approach and allocate 228 risks identified in the IS literature to the work system life cycle developed by Alter (2002). The lifecycle describes how work systems evolve over time and consists of the four phases: 'operation and maintenance', 'initiation', 'development', and 'implementation'. It provides a useful and comprehensible model for classifying risks in the context of a work system.

In a more recent study, Gemino et al. (2008) introduce a temporal model of IS project performance that classifies IS project risks into a priori risks and emergent risks. While a priori risks are associated to either structural elements of the project or knowledge resources available to the project team, emergent risks denote deficiencies in organizational support or result from the volatility of IS projects. A project manager may estimate a priori risks before the start of the project; emergent risks become apparent not until particular project phases. Using structural equation modeling the authors show that their model offers an improved explanatory power over traditional models of performance, partly resulting from the temporal perspective on IS project risks.

2.3 Research Gap

Looking at extant work on IS project risks, we see two issues. One is the limited value of present classifications when it comes to managing risks: On the one hand, a broad variety of classifications exist, indicating that little agreement has been established on the scope and scale of IS project risks. On the other hand, extant classifications largely build on the domains of IS project risks. While such classifications reduce the complexity of establishing a thorough and systematic overall risk inventory for a given project, they do not support project managers in managing the life cycle of IS projects (Pinto and Prescott, 1988; Somers and Nelson, 2004).

Second, extant literature agrees on the potential of exploring the temporal aspect for developing a deeper understanding of IS project risks. Existing studies provide a basis by suggesting first classifications such as the differentiation of a priori risks and emergent risks (Alter and Sherer, 2004; Gemino et al., 2008). Other studies conceptually allocate risks to different phases of a work system life cycle (Sherer and Alter, 2004). However, to the best of the authors' knowledge, an empirical investigation of the temporal nature of IS project risks which draws on risk archives is not yet available.

3. RESEARCH DESIGN

3.1 Overview

In the following, we explore the temporal aspect of IS project risk types based on a risk management archive from the multinational enterprise software company BETA. The archive consists of a large set of risk assessments done by project managers at BETA during operational project risk management. Our data set covers 111 software projects between 2004 and 2007. The focus of the projects is implementing, customizing, and updating enterprise software for medium to large customers across various industries. Studying longitudinal archival data allows us to reconstruct the temporal aspect of risks in more detail than it would be possible with sectional ex-post interviews or surveys.

In order to answer the research question mentioned above we proceed as follows: We first describe how the data was collected and prepared for analysis. In the subsequent data analysis phase, we substantiate the central assumption of our research by combining the research design of Alter et al. (1978) and Schmidt et al. (2001). Schmidt et al. (2001) rank IS project risks according to their perceived importance (i.e., their risk exposure) while Alter et al. (1978) allocate the risks to different project phases. In sum, we first analyze the perceived importance of risk types in particular project phases. To do so, we: (1) Integrate the temporal aspect by applying a five-phase process model of IS projects, (2) map risk assessments according to their occurrence in the project to the five project phases, (3) calculate the mean risk exposure per risk type in each project phase, and (4) rank the risk types according to their mean risk exposure in each project phase.

Since the risk exposure varies across project phases, we then examine the archive for patterns in the temporal profiles of risks. We first calculate the changes in the mean risk exposure from project phase to project phase for each risk type, and then cluster the risk types according to similar changes in the mean risk exposure. Finally, we present and discuss the results of our analysis.

3.2 Data Collection and Preparation

Project risk management at BETA follows a common approach: First, risks are identified and assessed. Then actions for controlling the risks are planned, implemented and monitored. The risk reviews take place once before and several times during a project. They are conducted by the project manager and partly by the project team. Depending on the project value and its strategic importance, a central risk management unit assists the process. Risk identification is supported by a check list containing a subset of altogether more than 300 questions which help the project manager identify risks that might occur during the project. Project managers at BETA can chose between 45 different predefined types of risks (see Table 3) which largely match the risks identified by Schmidt et al. (2001). We choose the singular risk as unit of analysis to avoid any influences from particular project types within the project portfolio of BETA. In addition to the type of risk, project managers also assess the risks in terms of their probability of occurrence (from 0 to 1) and their impact (from 0-'Insignificant' to 5-'Catastrophic'). The product of the perceived probability of occurrence and the perceived impact yields the risk exposure of a risk at the time of assessment. Eventually, further quantitative information (such as the expected financial loss or the impact and probability effects of the responses) and qualitative information (such as the condition, the indicator, or the consequence) is recorded for each risk.

Table 2 shows the basic statistics for the three key variables 'Impact', 'Probability' and 'Risk Exposure'. In line with Boehm (1991) and others, we argue that the risk exposure is a suitable construct for illustrating the perceived importance of a given risk at the time of assessment.

Table 2. Descriptive Statistics for Key Variables

Variable	Mean	Min	Max	Std. Dev.
Impact (I)	2,59	0	5	1,25
Probability (P)	0,46	0	0,99	0,22
Risk exposure (PxI)	1,23	0	4,95	0,89

N: 3119

The data generated during the risk reviews are stored in spreadsheet files called risk registers. For each risk review conducted during the life cycle of a project one risk register file is created. In total 1548 files representing 1548 risk reviews were available for our study. Thereof we were able to analyze 1222 files comprising 5066 risk assessments from 111 projects. The remaining 326 files were either corrupt or we were not able to identify the according project and/or customer. Where an automated extraction did not work, we manually extracted the data to ensure high data quality.

Assuming that projects with less than three risk reviews were likely to be still under way at the point of data collection and thus no final conclusion could have been drawn on a risk type's temporal pattern, we excluded 1622 risk assessments from those projects from our analysis. After further adjusting for incomplete records, 3119 of the 5066 risk assessments from 44 projects were retained for analysis. Table 3 provides an overview of the risk types assessed by BETA's project managers, including their frequency, their mean risk exposure and their standard deviation.

Table 3. Risk Ranking According to Risk Exposure

Rank	Risk	N	Mean	Std. Dev.
1	Inadequate Technical Infrastructure	32	2,14	1,44
2	Customer Expectations	109	1,76	0,88
3	Core Development Dependencies	77	1,61	0,79
4	Complex System Architecture	86	1,53	1,01
5	Post Go Live Approach Not Defined	135	1,51	0,89
6	No Ramp-Up	74	1,41	0,95
7	Non-T&M Payment Terms	176	1,36	1,02
8	Customer Inability to Undertake Project	134	1,35	0,92
9	Risk Tolerance	75	1,34	0,83
10	Expected Performance Issues	131	1,34	0,92
11	Functionality Gaps	135	1,33	0,96
12	Implementation and Development Interdependencies	52	1,32	0,75
13	Unrealistic Budget	125	1,31	0,89
14	Non-Conducive Political Environment	79	1,31	1,22
15	Complex Data Conversion	75	1,25	0,73
16	Low Project Priority	106	1,25	0,74
17	No Comparable Installations	102	1,24	0,86
18	Customer Financial Obligations	29	1,23	0,81
19	No Implementation Strategy	40	1,20	0,88
20	No Steering Committee	25	1,19	0,88
21	Undocumented Third Party Services	115	1,18	0,78
22	High Number of Interfaces	88	1,17	0,97
23	Unclear Customer Objectives	113	1,15	0,80
24	Unclear Roles	45	1,14	0,71
25	High Impact on Processes	122	1,13	0,75
26	Unclear Critical Success Factors	77	1,11	1,01
27	Ongoing Escalation Events	56	1,10	0,91
28	Weak Business Commitment	34	1,09	0,74
29	Requirements Not Understood	75	1,08	0,76
30	Implementation Partner Unknown	17	1,00	0,83
31	Production Downtime Impact	133	0,96	0,75
32	Hardware Partner Not Involved	43	0,95	0,77
33	No Quality Assurance or Risk Management	31	0,94	0,71
34	Unclear Governance Model	34	0,93	0,58
35	Language of Development Project	5	0,92	1,51
36	Incomplete Contract Requirements	42	0,86	0,82
37	No Change Management Approach	58	0,83	0,62
38	No Risk Sharing Agreements	42	0,83	0,67
39	High Customer Visibility	95	0,82	0,64
40	Industry Specific Solutions	40	0,77	0,77
41	Inexperienced Project Lead	33	0,73	0,53
42	Penalties and Royalties	9	0,68	0,65
43	Solution Uncertainties	9	0,44	0,61
44	Internal and External Decision Makers	4	0,28	0,21
45	Development Methodology	2	0,25	0,21

3.3 Data Analysis

In order to investigate how the perceived importance of risk types changes over time, we determine the point of time of each risk assessment and assign the assessment to a particular project phase. As our data set does not contain an assessment date but only the number of each individual assessment as well as the total number of assessments for each project (e.g., risk review 3 of 10), we calculate the proportionate project progress at each risk review relative to the total number of project risk reviews (e.g., 30%) and map it to one of five project phases (e.g., 30% to project phase 2) depicted in Figure 1. The mapping procedure is necessary in order to be able to compare risk type assessments on a common temporal basis (as projects have different numbers of risk reviews).

Phase models for enterprise software implementations follow a seven phase approach comprising the phases of 'System Selection', 'Planning', 'Analysis', 'Design', 'Realization', 'Implementation', and 'Operations' (Hansmann and Neumann, 2005). Due to the fact that our data reflect projects from BETA only and during the phase 'Operations' no risk reviews take place, we do not consider system selection and operations in our phase model. The resulting five phase model reflects BETA's approach of conducting projects.

Second, for each project phase we average the risk exposure of each risk type and subsequently rank the risk types by declining risk exposure. In ranking risk types by importance we follow extant research on IS project risks (e.g. Boehm 1991; Kappelman et al. 2006; Schmidt et al. 2001). Table 4 shows the ten most important risk types by project phase. To gain further insights concerning their domain of origin, all risk types are additionally assigned to one of the three domains (project management, technical subsystem, and social subsystem) suggested by Wallace et al. (2004).

Table 4. Top 10 Risk Types by Project Phase

#	Phase 1 "Bid and Planning"	Phase 2 "Analysis"	Phase 3 "Design"	Phase 4 "Realization"	Phase 5 "Implementation"
1	Inadequate Technical Infrastructure (T)	Inadequate Technical Infrastructure (T)	Inadequate Technical Infrastructure (T)	Inadequate Technical Infrastructure (T)	Customer Financial Obligations (S)
2	No Implementation Strategy (P)	No Steering Committee (S)	Low Project Priority (S)	Post Go Live Approach Not Defined (P)	Customer Expectations (S)
3	Customer Expectations (S)	Core Development Dependencies (T)	No Steering Committee (S)	Penalties and Royalties (S)	Complex System Architecture (T)
4	Core Development Dependencies (T)	Post Go Live Approach Not Defined (P)	Customer Expectations (S)	Weak Business Commitment (S)	Expected Performance Issues (T)
5	Non-Conducive Political Environm.(S)	Risk Tolerance (S)	Complex System Architecture (T)	Complex System Architecture (T)	Customer Inability to Undertake Project (S)
6	Post Go Live Approach Not Defined (P)	No Ramp-Up (T)	Core Development Dependencies (T)	Non-T&M Payment Terms (S)	Unrealistic Budget (P)
7	No Ramp-Up (T)	Customer Expectations (S)	Ongoing Escalation Events (S)	Implementation and Dev. Interdep. (T)	Post Go Live Approach Not Defined (P)
8	Non-T&M Payment Terms (S)	Complex System Architecture (T)	Unrealistic Budget (P)	Core Development Dependencies (T)	Implementation Partner Unknown (P)
9	Expected Performance Issues (T)	No Comparable Installations (T)	Functionality Gaps (T)	Unrealistic Budget (P)	Core Development Dependencies (T)
10	Complex System Architecture (T)	Customer Inability to Undertake Project (S)	Customer Inability to Undertake Project (S)	Complex Data Conversion (T)	High Number of Interfaces (T)

P: Project Management Risk, T: Technical Subsystem Risk, S: Social Subsystem Risk (Wallace et al., 2004)

Table 4 reveals two interesting aspects. First, a broad spectrum of risk types occurs, i.e., among the most important risk types are technical, social as well as project management risks. Second, the perceived importance of risk types varies across the projects' life cycle. Although it is surprising to see that many of the most important risk types are of a technical nature (e.g., 'Inadequate Technical Infrastructure', 'Core Development Dependencies', or 'Complex System Architecture') which contrasts the results of much of the existing literature on IS project risks (e.g., Schmidt et al. (2001) or Kappelman et al. (2006)), we focus on the variation in perceived importance over time.

The question arises whether or not patterns in the variations can be identified. For instance, Table 4 indicates that some risk types appear to be important at the beginning of a project but diminish in later phases, such as the risk of having 'No Implementation

Strategy' or having a 'Non-Conducive Political Environment'. Instead, a 'Low Project Priority' and 'Weak Business Commitment' seem to be issues that arise in the middle of a project. In contrast, risk types such as 'Financial Customer Obligations' or 'Implementation Partner Unknown' seem to materialize at the end of a project. In order to derive a classification based on the temporal risk exposure profile, we employ cluster analysis using PASW Statistics 17.0. Since we aim at grouping risk types with similar temporal profiles of risk exposure rather than grouping types with similar absolute risk exposures, we cluster the risk types based on the change in their mean risk exposure from project phase to project phase. Having five project phases results in four clustering variables which all measure the change in risk exposure from one phase to another. To determine the similarity between risk types or rather their temporal patterns we use the squared Euclidean distance as it is known to be very robust (Hair et al., 2008).

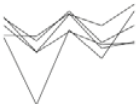
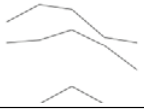
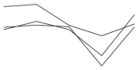
Following the recommendations by Punj et al. (1983), we first identify outliers by using the Single-Linkage (Nearest-Neighbor) approach. The resulting dendogram suggests that seven of the 45 risk types, namely 'Hardware Partner Not Involved', 'Inadequate Technical Infrastructure', 'Language of Development Project', 'No Implementation Strategy', 'No Steering Committee', 'Implementation Partner Unknown', and 'Penalties and Royalties' have quite dissimilar patterns of risk exposure and thus are hard to classify. Consequently, these risk types are initially not included in our analysis.

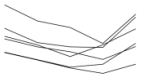
After having identified outliers, we employ the Ward approach to derive the clusters. The elbow check as proposed by Ketchen et al. (1996) indicates that a solution with nine clusters of risk types is the best, since the heterogeneity measure increases disproportionately when moving to a ten cluster solution. The clusters stay relatively stable when using other fusion algorithms, such as the complete linkage algorithm. Six out of nine clusters are identical, the other three show only minor differences. In order to check the validity of the derived clusters we graph the mean risk exposure for each risk type against the five project phases (see Table 5). The high similarity of the graphs suggests that the cluster analysis works well. Where the visual analysis indicates a better solution, we manually re-allocate the risk types to the respective clusters. Furthermore, after re-inspecting the outliers identified above, we are able to assign the risk types 'Hardware Partner Not Involved' and 'Inadequate Technical Infrastructure' to cluster 4 as well as 'Implementation Partner Unknown' to cluster 2.

4. RESULTS AND DISCUSSION

Table 5 depicts the derived clusters. In sum, 41 risk types can be allocated to nine clusters that show distinct risk exposure characteristics across the project phases.

Table 5. Derived Risk Clusters

Cluster	Risk types (Domain of Origin)	Visualization	Temporal Characteristics
1	Complex System Architecture (T) Customer Financial Obligations (S) Solution Uncertainties (T)		Remain constant initially Dramatically gain importance towards project end
2	Low Project Priority (S) Implementation Partner Unknown (P) Ongoing Escalation Events (S) Unclear Critical Success Factors (P) Unrealistic Budget (P)		Vary considerably in importance over time Gain importance towards project end
3	Inexperienced Project Lead (P) No Quality Assurance or Risk Management (S) Post Go Live Approach Not Defined (P) Risk Tolerance (S)		Peak just after project start Lose importance thereafter Re-gain importance towards project end
4	Inadequate Technical Infrastructure (T) Internal and External Decision Makers (S) Hardware Partner Not Involved (P) Weak Business Commitment (S)		Lose importance initially Peak just before project end Lose importance towards project end
5	Development Methodology (P) High Customer Visibility (S) Undocumented Third Party Services (S)		Gain importance after project start Peak in the middle Lose importance towards project end
6	Core Development Dependencies (T) Customer Inability to Undertake Project (S) Functionality Gaps (T)		Lose importance before project end Re-gain importance towards project end

7	Implementation and Development Interdependencies (T) Incomplete Contract Requirements (P) No Comparable Installations (T) No Ramp-Up (T) No Risk Sharing Agreements (P) Production Downtime Impact (T) Unclear Customer Objectives (T) Unclear Governance Model (S)		Peak just after project start Lose importance thereafter Remain comparatively constant until project end
8	Customer Expectations (S) Expected Performance Issues (T) High Number of Interfaces (T) Industry Specific Solutions (T) No Change Management Approach (P) Requirements Not Understood (T)		Lose importance until just before project end Re-gain importance towards project end
9	Complex Data Conversion (T) High Impact on Processes (S) Non-Conducive Political Environment (S) Non-T&M Payment Terms (S) Unclear Roles (P)		Remain comparatively constant over time Tend to lose importance towards project end

T: Technical Subsystem Risk, S: Social Subsystem Risk, P: Project Management (Wallace et al., 2004)

Looking at Table 5, we deem several aspects worth highlighting: First, risk exposure varies across project phases. We see that some risk types reach the highest level of importance in the later phases or at the end of the project while others are rather important in the middle or in the beginning. For instance, project managers perceive the risk type 'Customer Financial Obligations' as stable throughout the project. However, at the end of the project the perceived importance rises drastically. In contrast, comparable drastic changes occur regularly in the perception of the risk 'Low Project Priority'. Other risk types such as 'Complex Data Conversion' slowly decline over time without any major changes in perception (see Figure 2a). This substantiates the suggestions by other researchers that time is an important aspect of IS project risks and has to be considered when managing them (Alter and Ginzberg, 1978; Gemino et al., 2008; Sherer and Alter, 2004). Furthermore, the varying risk exposure across project phases challenges extant research on identifying the most important risk types in IS projects that does not take into account this temporal change. Our data highlights that existing risk rankings fail to acknowledge the practice of structuring projects into project phases (e.g. Barki et al., 1993; Boehm, 1991; Kappelman et al., 2006; Schmidt et al., 2001; Tiwana and Keil, 2006). Risk perception and thus risk management activities change from phase to phase. In addition, literature suggests that risks related to project management and the social subsystem play the most important role in IS project risk management, while risks related to the technical subsystem are of lower importance (Kappelman et al., 2006; Schmidt et al., 2001). In contrast, we see a high importance of technical risk types throughout the project phases (see Table 4). This substantiates the notion of different types of project having different risk profiles, e.g. software implementation projects may be subject to different set of risks then software development projects. Overall, our data does not substantiate any ranking of different risk domains as the perceived importance of domains also varies over time.

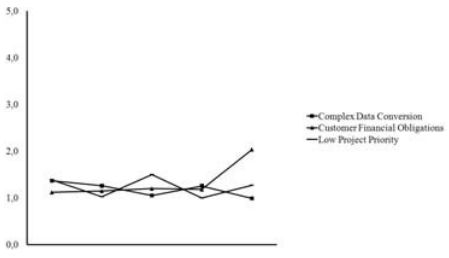


Figure 2a. Varying Risk Exposure

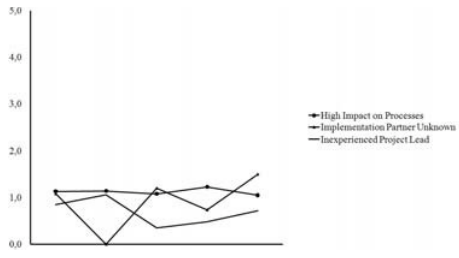


Figure 2b. Heterogeneous Degrees of Volatility

Second, we can observe heterogeneous degrees of volatility of risk exposure across risk types and project phases (see Figure 2b). Frequency and extent of changes in risk assessments vary. For example, the risk type 'Implementation Partner Unknown' varies

considerably from phase to phase with regard to its risk exposure. While being relatively important at the beginning, it becomes almost negligible in the second phase, regains importance thereafter, declines again and drastically peaks at the end. In contrast, the risk type 'High Impact on Processes' remains comparatively stable at a high level of importance. The risk type 'Inexperienced Project Lead' rises at the beginning, declines drastically towards the middle and slowly regains importance. This heterogeneity of risk exposure patterns illustrates the high dynamics of IS projects with respect to shifting business objectives and technical change. Hence, our data substantiates the work by Sitkin et al. (1995), who show that risk perception is largely a function of the changing problem frame underlying project managers' behavior. The changes in risk assessments also implicate that classifications of IS project risk types based on the perceived importance cannot remain stable over time. For instance, risk types will move across the dimensions of relative importance and controllability proposed by Keil et al. (1998).

Third, the cluster analysis suggests distinct temporal patterns of risk exposure, which indicates synchronous changes in risk assessments. For instance, cluster 1 consists of risk types with different levels of risk exposure that remain steady throughout the project and drastically gain importance towards the end (see Figure 3a). In contrast, cluster 9 comprises risk types of similar risk exposure levels which slowly decline to a particular level of risk exposure (see Figure 3b). Interestingly, the clusters contain risk types from all three risk domains. For example, cluster 4 includes the risk types 'Inadequate Technical Infrastructure', 'Internal and External Decision Makers', 'Weak Business Commitment', as well as 'Hardware Partner Not Involved' (see Figure 3c). While the first risk type is of technical nature, the second and the third risk type belong to the social subsystem. The last risk type stems from the project management domain. We agree that classifying risks according to their domain of origin fosters the systematic identification of risks. However, our clusters question the value of this kind of classification for focusing on the most important risks as proposed e.g., by Barki et al. (1993) and Schmidt et al. (2001). Furthermore, the synchronicity of risk exposure graphs within the clusters supports the notion of dependencies between risk types. While Wallace et al. (2004) show particular dependencies between risks of different domains, our clusters suggest common underlying causes that result in synchronous changes of risk perceptions within one cluster. For instance, one possible underlying cause for cluster 6 ('Core Development Dependencies', 'Customer Inability to Undertake Project', and 'Functionality Gaps') could be a software package based on new technology, which is still partly under development resulting in core development dependencies and gaps in functionality. Furthermore – as the technology is new – the customer does not have the capability or skill set to integrate it into the organization's infrastructure.

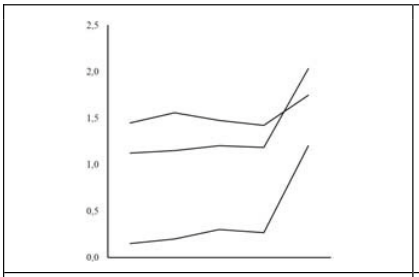


Figure 3a. Risk Cluster 1

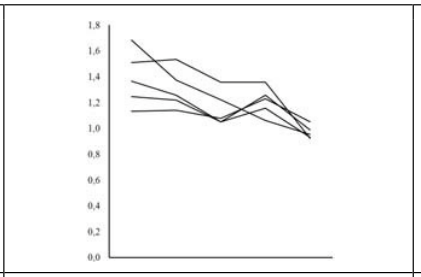


Figure 3b. Risk Cluster 9

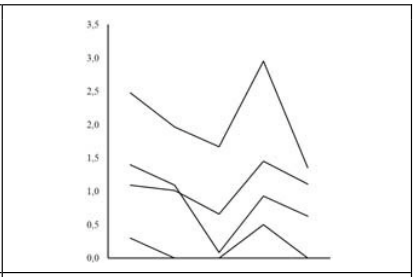


Figure 3c. Risk Cluster 4

5. IMPLICATIONS

In this paper, we present three results: First, risk exposure and thus the perceived importance of risk types does vary over project phases. Second, the degree of volatility of risk exposure varies over risk types and project phases. Third, temporal patterns of risk exposure can be identified. Despite the initial state of our research, we see several implications for IS researchers as well as for IS practitioners.

On the on hand, for IS professionals the identified variations in risk exposure highlight the importance of constantly performing risk management activities throughout a project's life cycle as new risks may emerge in later project phases (Gemino et al., 2008) or already identified risk types may vary in importance. Risk management activities may have to be adapted accordingly. In this regard, our results may help IS practitioners be more aware of these possible variations and employ their resources in a more efficient and effective way.

Furthermore, our results suggest that static lists of important IS project risks are of limited value in practical risk management, since they do not provide effective guidance for a given project phase. In addition, the notion that risk types not only vary with regard to risk exposure but also with regard to risk exposure volatility may be of value for IS practitioners. For instance, the volatility of risk exposure may serve as an indicator to what extent risk types are predictable and/or controllable. As a consequence, these highly volatile risk types may deserve more attention from project managers than risk types that tend to be more stable. In this regard, our results which are based on the analysis of a comprehensive portfolio of enterprise software implementation projects may also prove useful for a company's central project risk management unit: By comparing a project manager's individual set of risk types for a certain project phase to the portfolio's set of risk types for the same project phase, the central risk management unit is able to give some guidance as to which risk types typically require the attention of project managers in that phase. Finally, the results of our cluster analysis suggest that risk types in IS projects can be grouped according to their variation in risk exposure over

time. In this context, we speculate that synchronous changes in risk assessments may have a common underlying cause. This notion of risk archetypes may prove useful for IS professionals as in a concrete project context project managers may be able to identify and manage root causes of risks instead of symptoms.

On the other hand, IS researchers may benefit from a better understanding of the temporal aspect of IS project risks. We extend existing research on the temporal aspects of IS project risks by providing more detailed insights concerning the evolution of risks over time. While extant research (in most cases implicitly) acknowledges that risk exposure varies over time, our data does not only substantiate this thought but also proposes different volatilities in risk exposure. Furthermore, our results show that risks in IS projects may not only be classified into a priori and emerging risk factors but also into more granular temporal patterns. The derived risk clusters may provide a starting point for more sophisticated cause-and-effect models of IS project risks.

6. LIMITATIONS

Our study is subject to several limitations. First, because we analyze the risk archive of one company only, there may be issues concerning the representativeness of our results. BETA's organizational context or the particular nature of its projects may result in specific risk assessments which are not comparable to other companies or other IS projects. We especially consider the specific nature of the analyzed projects an issue. As IS projects are heterogeneous (e.g., small internal development projects vs. implementations of large enterprise software systems) their risk profiles are likely to vary.

Second, our results depend on the quality of the analyzed archival data. Some researchers suggest that risk management is often seen as a burden which creates 'extra work and expense' (Verner and Evanco, 2005). Thus, the possibility exists that risk managers do not carefully maintain the risk registers but rather fill in dummy data just to fulfill the requirements. There is no indication however, that the data is maintained in a careless way. Instead, the comprehensiveness of the free text comments in the risk registers indicate that risk assessment is done properly. Furthermore, other authors explicitly highlight the value of comprehensive archival data (e.g. Ropponen and Lyytinen, 1997). Especially for investigating temporal aspects of risks, longitudinal archival data may be better suited than surveys or interviews as they allow for reconstructing chronological events in much more detail. Moreover, possible bias evoked by the researcher is ruled out when analyzing archival data.

A third limitation concerns the possibility that our research approach is impeded from a methodological point of view: First, the approach of mapping risk assessments to project phases, which is necessary due to the different number of risk reviews per project, is problematic for two reasons: (1), the number and configuration of our clusters depends on the number of project phases as the mean risk exposure per phase changes. Even though BETA typically follows a five phase approach when implementing enterprise software systems, we cannot be sure, that this holds true for all projects investigated. (2), as no exact risk assessment date is available we can only approximate the mapping between risk assessments and project phases which adds to uncertainty. Second, the results of cluster analyses are traditionally prone to criticism as the final number and configuration of clusters depend on a series of choices to be made by the researchers and thus are often considered subjective. This potential issue is aggravated by the manual re-adjustment of clusters described above. However, the argument we want to make does to a large extent not depend on the correct number and configuration of clusters but rather on the finding that the importance of risks (as measured by their mean risk exposure) moves in comparable patterns.

7. SUMMARY AND CONCLUSION

The purpose of our study is to explore how the perceived importance of IS project risks evolves over time. While much research is available on the domains of risks, little is known about their temporal nature. Gemino et al. (2008) explicitly suggest further investigating the temporal perspective. Based on a review of extant research in this field, we investigate a large archive of risk assessments recorded during the operational project risk management process in enterprise software projects. We employ a five-phase process model in order to investigate variations in risk assessments/importance over project phases. Using cluster analysis, we establish a descriptive and exploratory view on temporal patterns of risk types. In doing so, we provide a first illustration of how risk assessments vary over time.

Our results are relevant to both IS researchers and IS professionals. Extending prior studies on risks in IS projects, we shed more light on temporal aspects and thus help better understand and manage IS project risks. Future research will focus on explaining the variations in risk exposure and identifying dependencies between risk types. In particular, we will explore underlying risk archetypes that result in aligned risk assessments of diverse risk types and domains. To do so, we will follow the guidance provided by van de Ven and Huber (1990). Additionally, we will present our results to the project managers of BETA to identify further candidates for risk archetypes.

8. REFERENCES

[1] Alter, S. (2002) *The work system method for understanding information systems and information systems research,* Communications of the Association for Information Systems, 9 (6), pp. 90-104.

[2] Alter, S. and Ginzberg, M. (1978) *Managing uncertainty in MIS implementation,* Sloan Management Review, 20 (1), pp. 23-31.

[3] Alter, S. and Sherer, S. (2004) *A general, but readily adaptable model of information system risk,* Communications of the Association for Information Systems, 14 (2), pp. 1-28.

[4] Barki, H., Rivard, S. and Talbot, J. (1993) *Toward an assessment of software development risk,* Journal of Management Information Systems, 10 (2), pp. 203-225.

[5] Barki, H., Rivard, S. and Talbot, J. (2001) *An integrative contingency model of software project risk management,* Journal of Management Information Systems, 17 (4), pp. 37-69.

[6] Boehm, B. (1991) *Software risk management: principles and practices,* IEEE Software, 8 (1), pp. 32-41.

[7] Charette, R. (1996) *The mechanics of managing IT risk,* Journal of Information Technology, 11 (4), pp. 373-378.

[8] Gemino, A., Reich, B. and Sauer, C. (2008) *A temporal model of information technology project performance,* Journal of Management Information Systems, 24 (3), pp. 9-44.

[9] Hair, J., Black, W., Babin, B., Anderson, R. and Tatham, R. (2008) *Multivariate data analysis,* Prentice Hall, Upper Saddle River.

[10] Han, W.-M. and Huang, S.-J. (2007) *An empirical analysis of risk components and performance on software projects,* Journal of Systems and Software, 80 (1), pp. 42-50.
[11] Hansmann, H. and Neumann, S. (2005) *Prozessorientierte Einfuehrung von ERP-Systemen,* In ProzessmanagementSpringer, Berlin, pp. 329-372.
[12] Heemstra, F. and Kusters, R. (1996) *Dealing with risk: a practical approach,* Journal of Information Technology, 11 (4), pp. 333-346.
[13] Iversen, J., Mathiassen, L. and Nielsen, P. A. (2004) *Managing risk software process improvement: An action research approach,* MIS Quarterly, 28 (3), pp. 395-433.
[14] Jiang, J. and Klein, G. (2000) *Software development risks to project effectiveness,* The Journal of Systems and Software, 52 (1), pp. 3-10.
[15] Kappelman, L., McKeeman, R. and Zhang, L. (2006) *Early warning signs of IT project failure: the dominant dozen,* IT Project Management, 23 (4), pp. 31-37.
[16] Keil, M., Cule, P., Schmidt, R. and Lyytinen, K. (1998) *A framework for identifying software project risks,* Communications of the ACM, 41 (11), pp. 76-83.
[17] Ketchen, D. and Shook, C. (1996) *The application of cluster analysis,* Strategic Management Journal, 17 (6), pp. 441-458.
[18] Knight, F. H. (2002) *Risk, Uncertainty and Profit,* BeardBooks, Washington, DC, USA.
[19] March, J. and Shapira, Z. (1987) *Managerial perspectives on risk and risk taking,* Management Science, 33 (11), pp. 1404-1419.
[20] McFarlan, W. (1981) *Portfolio approach to information systems,* Harvard Business Review, 59 (5), pp. 142-151.
[21] Moynihan, T. (1997) *How experienced project managers assess risk,* IEEE Software, 14 (3), pp. 35-41.
[22] Nidumolu, S. (1995) *The effect of coordination and uncertainty on software project performance: Residual performance risk as an intervening variable,* Information Systems Research, 6 (3), pp. 191-219.
[23] Pinto, J. and Prescott, J. (1988) *Variations in critical success factors over the stages in the project life cycle,* Journal of Management, 14 (1), pp. 5-18.
[24] Punj, G. and Stewart, D. (1983) *Cluster analysis in marketing research: Review and suggestions for application,* Journal of Marketing Research, 20 (2), pp. 134-148.
[25] Ropponen, J. and Lyytinen, K. (1997) *Can software risk management improve system development: an exploratory study,* European Journal of Information Systems, 6 (1), pp. 41-50.
[26] Ropponen, J. and Lyytinen, K. (2000) *Components of software development risk: how to address them,* IEEE Transactions on Software Engineering, 26 (2), pp. 98-112.
[27] Sauer, C., Gemino, A. and Reich, B. (2007) *The impact of size and volatility on IT project performance,* Communications of the ACM, 50 (11), pp. 79-84.
[28] Schmidt, R., Lyytinen, K., Keil, M. and Cule, P. (2001) *Identifying software project risks: an international delphi study,* Journal of Management Information Systems, 17 (4), pp. 5-36.
[29] Sherer, S. and Alter, S. (2004) *Information system risks and risk factors: Are they mostly about information systems,* Communications of the Association for Information Systems, 14 (2), pp. 29-64.
[30] Sitkin, S. and Weingart, L. (1995) *Determinants of risky decision-making behavior: a test of the mediating role of risk perceptions and propensity,* Academy of Management Review, 38 (6), pp. 1573-1592.
[31] Somers, T. and Nelson, K. (2004) *A taxonomy of players and activities across the ERP project life cycle,* Information & Management, 41 (3), pp. 257-278.
[32] Tesch, D., Kloppenborg, T. and Erolick, M. (2007) *IT project risk factors: the project management professionals perspective,* Journal of Computer Information Systems, 47 (4), pp. 61-69.
[33] Tiwana, A. and Keil, M. (2006) *Functionality risk in information systems development: an empirical investigation,* IEEE Transactions on Engineering Management, 53 (3), pp. 412-425.
[34] van de Ven, A. and Huber, G. (1990) *Longitudinal field research methods for studying processes of organizational change,* Organization Science, 1 (3), pp. 213-219.
[35] Verner, J. and Evanco, W. (2005) *In-house software development: What project management practices lead to success,* IEEE Software, 22 (1), pp. 86-93.
[36] Wallace, L. (2004) *Software project risks and their effect on outcomes,* Communications of the ACM, 47 (4), pp. 68-73.
[37] Wallace, L., Keil, M. and Rai, A. (2004) *How software project risk affects project performance: an investigation of the dimensions of risk and an exploratory model,* Decision Sciences, 35 (2), pp. 289-321.
[38] Zmud, R. (1980) *Management of large software development efforts,* MIS Quarterly, 4 (2), pp. 45-56.

Using NFriendConnector to Extend Facebook to the Real World

Felix Köbler
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr. 3
85748 Garching
koebler@in.tum.de

Suparna Goswami
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr. 3
85748 Garching
goswami@in.tum.de

Philip Koene
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr. 3
85748 Garching
koene@in.tum.de

Jan Marco Leimeister
Fachgebiet Wirtschaftsinformatik
Universität Kassel
Nora-Platiel-Str. 4
34127 Kassel
leimeister@uni-kassel.de

Helmut Krcmar
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr. 3
85748 Garching
krcmar@in.tum.de

ABSTRACT
This study presents a novel approach towards establishing online social connections using Facebook and NFC-enabled mobile phones. There is increasing convergence between users' real life social networks and their online social networks, with online connections following actual social acquaintance and interactions. Accordingly there is a need to provide users with means of accessing and establishing online connections in social networking platforms such as Facebook as and when they interact with other people in their real lives. The NFriendConnector is a prototype application which fulfills this functionality. This paper describes the design and development of the prototype application. The expectation confirmation theory is used to analyze the extent to which the NFriendConnector fulfills this inherent need among users and how this influences their intention to adopt and use the prototype. The proposed research model is tested in an experimental setup.

Keywords
Social networking, mobile prototype design, mobile user interaction, expectation confirmation theory, laboratory experiment

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

1. INTRODUCTION
Online social networking has witnessed explosive growth over the past few years. This is evidenced by a nearly ten-fold growth in the number of active users on Facebook from 30 million in July 2007 to about 250 million in August 2009 [17]. Users may have various motivations for joining the different online social networking sites such as LinkedIn, Friendster, MySpace and Facebook, but the primary activity on these sites is that of forming social connections with other users. These connections can lead to potential benefits such as facilitating different tasks within the group, reducing misbehavior, creating online communities of support and generating different types of social capital [11, 13].

Significant research has tried to understand and categorize what motivates people to participate in various online social networking sites. Findings suggest that in previous generations of online communities, relationships were typically initiated online based on shared interests, and were likely to transcend into users' offline worlds [4, 17, 18], often resulting in face-to-face meetings. However, in the case of Facebook, users' main motivation is often to intensify and solidify relationships that started offline [10]. Accordingly, it has been proposed that people use Facebook more for the purpose of "social searching" (which refers to looking up offline contacts and acquaintances, finding out more about them, and connecting to them) as opposed to "social browsing" (which refers to finding people or groups online in order to establish connection with them) [10]. For this group of users, i.e., the social searchers, while Facebook provides the means for maintaining and strengthening of the social ties that were initially established in offline settings, users have to spend significant effort to map their real life ties in Facebook.

Facebook connections are typically established by sending a friend request. Therefore, in order to establish a connection with a person one has met in some kind of offline social setting such as in a common class, at a party, or in a conference, one has to go back to their Facebook account and search for this other person's profile. Once found, a request to become friends has to be sent to the other person, which then has to be accepted by him/her. To establish connections in this form, the person initiating the

connection has to incur certain costs in the form of having to search for the other person's profile, waiting for the other person to respond, and above all having access to a computer terminal with Internet connectivity. Thus there is still a significant gap between the initiation of contact in a real life social setting and that in an online social network.

The NFriendConnector prototype intends to fill this gap. Given the ubiquitousness of mobile phone usage in today's world, we believe that mobile phones have a significant potential of lowering the burden involved in establishing social networking connections, especially for people who frequently use Facebook to establish online connections with their offline contacts. Once the social networking profile information is present on mobile phones, it can further enrich the actual face-to-face interaction experience by enabling users to view, share and match information regarding themselves and others.

This paper presents the NFriendConnector application prototype and describes different aspects of its design and development. The prototype integrates a number of functions from the Facebook platform into a Near Field Communication (NFC) enabled mobile device, by making use of Facebook's Application Programming Interface (API) and Facebook users' profile data. It enables users to map their offline interactions as and when they take place, into their Facebook profiles and also experience a more interesting social interaction complemented by their Facebook profile information, since it allows them to share and compare it. The NFriendConnector prototype application additionally allows them to access other Facebook functionalities such as posting a new status message. Users' perceptions regarding the prototype and their intention to adopt and use it are assessed in a laboratory experiment using the expectation confirmation theory as the theoretical framework. Contributions from this research and its implications are discussed.

2. PRIOR WORK ON MOBILE SOCIAL NETWORKING

The widespread popularity of online social networking has resulted in researchers and system designers investigating various issues related to Internet-enabled social networking, and mechanisms for designing applications and interfaces that make online social networking more meaningful and attractive to its end-users. Accordingly, there is an emerging stream of research that has started investigating the role of mobile devices such as mobile phones, in online social networking. This is evidenced by a number of research projects targeting various aspects of mobile social networking and prototype development efforts. It is likely that commercial applications will follow suit. In fact, analysts predict that mobile social networking users are likely to make up around 18% of all mobile users, the equivalent of 950 million users worldwide by 2012 [3].

Existing research on mobile social networking explore the usage of Bluetooth, GPS, Wi-Fi and similar technologies for proximity and location detection, often called "Social Proximity Applications" [19]. One of the first projects to explore these aspects of mobile social networking was MIT's "Social Serendipity" [6], followed by others like "Dodgeball" [24] and "MobiClique" [21]. Other applications such as "DigiDress" [19] and "Sensor" [20] differentiate on certain aspects of "Social Serendipity", but share the same motivations and basic design principles. The essential idea behind these projects is to inform members with a mobile social network regarding the location and proximity of existing acquaintances in this network or other network users with whom their profile data has a match, so that they may become new acquaintances if they want to. This is achieved by a periodic Bluetooth scan of the environment and reporting all new discovered Bluetooth IDs to a central server, which then detects if the ID is part of the mobile social network and checks the corresponding profile for connections or a positive profile match. Apart from the detection of an acquaintances´ location and proximity, these prototypes usually also support basic communication functions like message transfer [6, 21, 24]. Most of these prototype systems however, are based on small, proprietary social networks, with only a selected few like "MobiClique" and "Cityware" leveraging the large user base of existing online social networks [9, 21], as is done in the case of NFriendConnector and Facebook.

An important concern regarding mobile social networking, especially for applications making use of Bluetooth sensing is that of privacy. The implications of a user´s name and details being made available to a stranger without the user´s knowledge [6] or tracking user´s daily routines and preferences through their location, as is done in the "Cityware" [9] mobile social network are likely to raise perceived threats towards privacy in the potential client base. In contrast to the above described mobile social networking application prototypes, the NFriendConnector does not make use of physical proximity, location or other sensory data for automatic profile matching or generation of other information. The NFriendConnector enables seamless integration of users' offline social interactions with their online social networks using a mobile phone. With NFriendConnector an interaction and exchange of information between users has to be initiated by the users themselves by establishing a near physical contact between the two mobile devices on which the NFriendConnector is running, rather than the system doing things in the background without the users' active participation or knowledge. Since the user has full control over the exchange of information that happens with other users, the NFriendConnector overcomes the privacy concerns that are associated with other mobile social networking applications.

3. NFRIENDCONNECTOR PROTOTYPE APPLICATION

The NFriendConnector prototype makes use of Near Field Technology (NFC) and the Representational State Transfer (REST) application programming interface (API) provided by Facebook along with the profile data of registered Facebook users. NFC is a short range high frequency wireless communication technology that allows data exchange between devices that are about 4 inches apart [18]. It is a simple extension of the ISO/IEC 14443 [7] proximity-card standard (such as contactless card, Radio-Frequency Identification (RFID), etc.) that combines the smart card and the reader into the same device. NFC is primarily aimed for being used in mobile phones. The application environment for the prototype builds on cellular phones equipped with NFC-reader devices, mobile Internet broadband connectivity enabled through Enhanced Data Rates for GSM Evolution (EDGE) or Universal Mobile Telecommunications System (UMTS) technologies and sufficient display measurements.

3.1 Use Case

The simplest use case for the prototype is as follows: two individuals who are Facebook users meet in real-life in a club, a café, a common course, a conference, or some similar setting where social interaction takes place between them. Both individuals possess NFC-enabled mobile phones (as described above) with mobile Internet connectivity and the NFriendConnector application installed. The application enables them to exchange their Facebook profiles over the NFC-interface by establishing a physical contact between their cell phones. This allows them to view and compare each other's Facebook profiles. Based on this interaction, the two individuals discover common interests and a willingness to maintain the newly developed contact and keep in touch. NFriendConnector allows them now to establish a Facebook friend connection, and therefore map each other onto their existing Facebook network, and also to generate a status message.

3.2 Prototype Navigation

The NFriendConnector application prototype can be run in the stand-alone mode (without an active mobile Internet connection) and the connected mode (with and active mobile Internet connection). Accordingly, the features provided by the prototype are classified into client-enabled and connection-enabled. Client-enabled features can be executed in the stand-alone mode, the application accesses Facebook data which has been previously stored on the mobile device. Connection-enabled features can only be executed in the connected mode. It should be noted that the client-enabled features can also be accessed in the connected mode. The user can start the application in either of these modes and then later switch from one mode to the other if so required. The application starts with a splash screen which leads the user to the "Mode options" menu as displayed in Figure 1.

Figure 1: "Mode Options" menu (left) and "Profile Options" menu (right) of NFriendConnector

For both connected and stand-alone modes, the user can initiate a NFC-communication by touching another NFC-enabled mobile phone on which the NFriendConnector is also running. This will present the user with the "Profile options" menu (as in Figure 1) to access more features described below. Table 1 outlines the platform requirements, modes and features of the prototype.

Table 1: Requirements, Modes and Features of NFriendConnector

NFriendConnector		
Requirements	NFC-enabled mobile device	
	Existing Facebook account with profile information	
		Available mobile Internet connection
Modes	Stand-alone	Connected
Features	View Profile	
	Match Profile	
	Save Profile	
		Add as friend
		Make status message

3.3 Client-enabled Features

3.3.1 View Profile

The "View Profile" feature displays Facebook profile data of the other individual, with whom the social interaction is taking place. The mobile phones belonging to the two Facebook users exchange the profile data through a successfully established NFC-connection. For the purpose of the prototype, only the profile picture, profile name, and the Facebook profile fields hometown, interests, movies and music are displayed as a scroll down /side scrolling list (as shown in Figure 2). It is however technically feasible to display all existing profile fields.

Figure 2: "View Profile" (left) and "Match Profile" (right) functionality of NFriendConnector

3.3.2 Save Profile

With the "Save Profile" feature, the user is able to save the Facebook profile data of the other person with whom an exchange of profile information has just happened. This allows the user to

build an individual contact list of Facebook users on his/her mobile phone that can later be browsed even without an active mobile Internet connection.

3.3.3 Match Profile

The "Match Profile" feature makes use of the users' own Facebook profile data stored in the mobile phone, and the newly obtained profile information of the other person to notify if there is a match of interest, dislikes, hobbies, etc. These could be partner-/relationship-seeking, hobby-seeking (e.g., soccer club), or other matching configurations, devised by the user. For the purposes of the prototype application all existing profile fields were matched (as shown in Figure 2). It is also possible to set threshold configurations of certain variables within the profile data (such as matching relationship status, distance of home towns, count of matching interests, etc.) and calculate an output that signals the user of a positive or a negative match. This option is yet to be implemented in the NFriendConnector prototype.

3.4 Connection-enabled Features

The following features can only be used when the NFriendConnector is logged into Facebook through a constant mobile Internet connection.

3.4.1 Add friend

This feature uses the unique Facebook profile ID of the other individual to identify the individual on the Facebook platform and visit that individual's Facebook profile via the mobile web browser installed on the cell phone. The mobile web browser can be launched directly from the NFriendConnector prototype. This provides users with quick access to the Facebook function "add as a friend" in a real life social setting (Figure 3). After executing the desired Facebook functions (such as making a friend-connection request) the user can simply close the mobile web browser to return to the NFriendConnector prototype application.

Figure 3: "Add friend" (left) and "Make status message" (right) functionality of NFriendConnector

3.4.2 Make status message

The NFriendConnector prototype enables users to generate automated status messages that get displayed on their respective Facebook profile status wall. For the purposes of the prototype application, a simple example for such an automated status message is implemented into NFriendConnector. For instance the message states "<individual1> is in <semantic location> with <individual2>" (as seen in Figure 3). The semantic location (in this case "Monaco, France") can be obtained by the GPS receiver or cell identification mechanism of the mobile phone.

The user could also define a current task (e.g., "studying"), which is then mapped onto the status message: "<individual1> is studying with <individual2>" or type in the status message directly through the keyboard of the cell phone. Furthermore, the automated status messages could be enriched by location-based information retrieved from the Global Positioning System (GPS) receiver in combination with other applications such as Google Maps [12].

3.5 Implementation Details

For the purpose of this study, the NFriendConnector is implemented on the Nokia 6212 classic NFC-enabled mobile phone [20]. It is implemented as a Java J2ME midlet that can facilitate communication with the mobile phone on one hand, and gain access to the functions and data of Facebook on the other hand. The communication with the Facebook platform is done by sending and interpreting RESTful HTTP-requests over the mobile Internet connection to the Facebook API REST server [21]. The NFC-reader device of the Nokia 6212 phone is controlled by the J2ME midlet through an API provided by Nokia [1].

Two different communication patterns are possible in the NFriendConnector prototype. Users can download and store their Facebook profile data directly in their mobile phones. Upon contact with another mobile phone on which the NFriendConnector is running, this data can be sent through the NFC-interface to the other phone. Following this exchange, the "View Profile", "Save Profile" and "Match Profile" features can be executed. The advantage of this form of communication is that no mobile Internet connectivity is required for executing the above features. The disadvantage is the possibly large traffic over the NFC-interface, especially when profile pictures are transferred. Also the Facebook profile data on the mobile phone gets outdated without regular updates.

In the other conceivable communication routine, only the unique Facebook user ID is transferred over the NFC-interface. The rest of the profile data is then downloaded on a need basis for executing certain features. This method guarantees up-to-date profile data and low data traffic sent via the NFC-interface, however it also requires an active mobile Internet connection.

For the purpose of the prototype, a balance between the two approaches was chosen (refer to Figure 4). The user can download and update his locally stored Facebook profile upon application startup by logging into Facebook through the application. Only the information fields of the profile data, along with the unique Facebook user ID are stored on the device. This entire information is then exchanged as an XML file through the established NFC-connection with another mobile phone. Multimedia content of the profile (such as profile picture) is stored locally on an user's own mobile phone but not transmitted through the NFC-interface, in order to avoid very high NFC traffic and lengthening of the exchange process. If needed, multimedia content can be

downloaded by the receiving device via a mobile Internet connection.

Facebook's REST API server provides access to profile data without using a browser interface. Facebook method calls can be made over the (mobile) Internet connection by sending HTTP GET or POST requests to this REST server. Most of the features of the NFriendConnector prototype are implemented in this manner.

Figure 4: NFriendConnector communication routine

However, certain Facebook functionality such as logging in, or adding a friend cannot be accessed via the Facebook REST API server [21]. For executing these features, the NFriendConnector automatically opens a web browser and logs the user through the Facebook web interface from the J2ME midlet. The mobile web browser can be automatically set to a specific page, for instance the confirmation page for sending a friend request. The other user can therefore, quickly confirm the friend request and return to the NFriendConnector application by closing the mobile browser. This round-about way of accessing certain Facebook functions through a J2ME midlet is due to poor support for mobile devices currently provided by Facebook APIs. This is however, likely to change in the near future due to recent Facebook initiatives like "Facebook Platform for Mobile" [22].

4. PROTOTYPE EVALUATION

Previous research has indicated that in Facebook, users place a higher priority on being able to establish connections with people they have met in offline settings, than in establishing purely online connections. However, currently Facebook does not provide the functionality that supports this particular pattern of usage. The NFriendConnector addresses this by enabling users to establish a Facebook connection immediately after meeting someone of interest, provided they have an NFC-enabled mobile phone. It is expected that a prototype that satisfies this need of the users will result in increased satisfaction. The extent, to which users find the functionality provided by the NFriendConnector useful and are satisfied with it are examined in an experimental setup. The other purpose of the experiment is to assess the extent to which users are willing to adopt the prototype.

4.1 Theoretical Basis – Expectation Confirmation

We draw on the expectation-confirmation theory (ECT) [15] in general and the expectation confirmation model [2] in particular to explain how a confirmation of users' expectations from the system influences their satisfaction with the system. Satisfaction with an information system and intention to use it are recognized indicators of users' experience of using the system. According to ECT (Figure 5), consumers form an initial expectation regarding a particular product or service before using it. After initial consumption they assess the perceived performance based on their original expectation to determine the extent to which their expectation is confirmed. Confirmation determines satisfaction, which in turn determines repurchase or reuse intention. While initial expectation is formed prior to consumption (in time t_1), assessment of perceived performance, confirmation, satisfaction and repurchase intention are post-consumption variables (measured during time t_2).

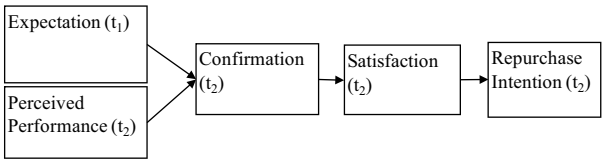


Figure 5: Expectation-confirmation theory

ECT has been used and to analyze repurchase decisions [16, 23], and more recently, to study information technology continuance behavior [2, 14]. The underlying reasoning is that users' decisions to continue using a system is similar to that of consumers' repurchase decisions. Accordingly, the expectation confirmation model (ECM) was proposed [2] by adapting the ECT in the context of technology use and continuance. The theory was extended by drawing from other theories of technology acceptance and use such as the technology acceptance model (TAM) [5]. Instead of examining both pre- and post- use variables, only post-acceptance variables are considered. Post-use expectation is represented by perceived usefulness of the system.

4.2 Research Model and Hypotheses

We apply ECT as the theoretical framework to assess users' experience of using the NFriendConnector because it predicts users' intention to use a system based on their satisfaction after initial usage. Given the setting where users come to know of the prototype and use it for the first time during the course of the

study, we are only interested in examining the post-use variables. This is because while users might have some pre-formed notions or expectations regarding the system before they actually use it for the first time, these expectations are likely to become more defined and concrete only after experiencing and using the prototype.

According to the ECT, confirmation of expectations results in satisfaction. Users evaluate their use experience based on the extent to which their expectations are confirmed and this results in them being satisfied or dissatisfied with the system. Satisfaction is an evaluative response to system use that can be either positive or negative or indifferent [2]. Therefore we hypothesize:

H1. Users' extent of confirmation is positively associated with their satisfaction of NFriendConnector use.

Users may or may not have some initial expectation regarding what the system does and how it can help them in performing their task. However, after having used the system they will form their ex-post expectation regarding the system based on its performance. Perceived usefulness [5] is used to reflect ex-post expectation in this study. Users will form their beliefs regarding the usefulness of the system by combining their evaluation vis-à-vis their initial expectation of the system. The extent to which their pre-use expectations from the system are confirmed after using it will determine the extent to which they perceive the system as being useful. Therefore, we hypothesize:

H2. Users' extent of confirmation is positively associated with perceived usefulness of the NFriendConnector.

Perceived usefulness reflects a belief regarding the usefulness of the system and, beliefs typically result in affect. Satisfaction is an affect that has been found to be significantly related to perceived usefulness in the context of information systems use and continuance [2]. Since trying the NFriendConnector is an instance of information system use that will result in beliefs and subsequent affects we hypothesize:

H3. Perceived usefulness is positively associated with users' satisfaction of NFriendConnector use.

Satisfaction plays an important role in users' decision to repurchase (in consumer behavior) or continue using (in information system continuance). In the context of this study, users encounter the prototype for the first time and therefore their first usage determines the formation of the affect – satisfaction (or the lack of it). Satisfaction has already been validated as a significant predictor of intention to use in TAM [5] and various TAM-based studies (e.g., [5, 8]). Therefore, we hypothesize:

H4. Users' satisfaction with the NFriendConnector is positively associated with their intention to use it.

Previous research has shown that perceived usefulness is a salient belief that influences information system acceptance behaviors or behavioral intentions across a broad range of end-user technologies [5, 8]. Therefore, other than the relationship between perceived usefulness and satisfaction, we also hypothesize a direct link between perceived usefulness and the intention to use the NFriendConnector:

H5. Perceived usefulness is positively associated with the intention to use the NFriendConnector.

Figure 6 represents the research model for the study. All the variables in our study are measured ex-post, i.e., after users have already used the system.

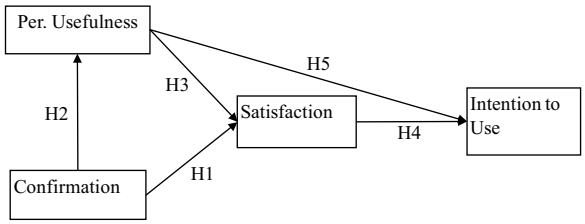


Figure 6: Research Model

4.3 Experiment

An experimental methodology was chosen to test the hypothesized relationships and therefore assess the usability of the NFriendConnector prototype. For this research, laboratory experiment is an appropriate methodology as it allowed us to control factors that are extraneous to the variables of interest. For instance, environmental factors were controlled for by holding the experiment in a common laboratory and using the same NFC-enabled mobile devices (same make and model).

Student volunteers registered in the under-graduate or graduate programs in a large German university were recruited as experimental subjects. The experiment was held over multiple sessions in a designated laboratory. Each experimental session lasted for about 30 minutes. A standard protocol was followed for all sessions. Subjects were provided a brief introductory description about the NFriendConnector prototype and its various features, and the supporting use case. They were then given a demonstration of how they could use the prototype. Following the demonstration, they were provided with NFC-enabled mobile phones which had the prototype installed in it and were asked to evaluate the prototype. Subjects were paired and asked to use NFriendConnector for establishing Facebook connections with each other. The laboratory also had a personal computer terminal, where subjects could login to Facebook and see that the connection had actually been established. Some dummy Facebook accounts were created and subjects were asked to work with these accounts, rather than their own Facebook accounts, to ensure that all subjects worked with similar kinds of Facebook profiles.

Each subject was required to fill two short questionnaires measuring the variables of interest. The first questionnaire had to be filled at the beginning of the experiment before the subjects were introduced to the prototype, and mainly measured demographic variables such as age, educational level, and whether they were users of Facebook or not. The second questionnaire had to be filled after the subjects had used the NFriendConnector prototype. The second questionnaire measured the experimental variables such as Perceived Usefulness, Satisfaction, etc.

Validated instruments from previous research were used to measure the different variables (Table 2). Where necessary, the measurement items were modified to suit the context of this study. All items were measured using 5-point Likert scales, with scale anchors varying from 'Strongly Disagree' to 'Strongly Agree'. For satisfaction, users responded on four different scale anchors:

Very Dissatisfied – Very Satisfied, Very Displeased – Very Pleased, Very Frustrated – Very Contended, Absolutely Terrible – Absolutely Delighted.

Table 2: Measurement Instrument

Variable (Source)	Measurement Items
Confirmation [2]	My expectation of using the NFriendConnector was better than what I had expected. The functionality provided by NFriendConnector was better than what I had expected. Overall, most of my expectations of using the NFriendConnector were confirmed.
Perceived Usefulness [5]	Using NFriendConnector would improve my performance in establishing Facebook / [online social network] connections with friends and acquaintances. Using NFriendConnector would enhance my effectiveness in establishing Facebook / [online social network] connections. Using NFriendConnector would increase my productivity in managing my Facebook / [online social network] connections. I find NFriendConnector to be useful for establishing Facebook / [online social network] connections.
Intention to Use [5]	Assuming that I have access to the NFriendConnector, I intend to use it. Given that the NFriendConnector is available to me, I predict that I will use it.
Satisfaction [2]	How do you feel about your overall experience with the NFriendConnector use?

No time limits were imposed on the subjects, as the main purpose of the experiment was to allow them to get a better understanding of the prototype, and explore its features in addition to performing the task of establishing a Facebook connection with their experimental partner. Most subjects were found to spend about 15 minutes evaluating the prototype.

4.4 Data Analysis

Table 3 reports the demographic details of the experimental sample. A total of 30 subjects participated in the experiment (N = 30). Being a Facebook user was not a pre-condition to participate in the experiment and the data indicates that the sample consisted of equal number of Facebook users and non users. However, subjects were asked whether they had experience with other social networking sites, and the data indicated that all subjects were experienced in online social networking. Subjects were mostly in the age group of 25 to 34, and most of them were pursuing a Master´s degree in the university. Since subjects were volunteers who opted to participate, and we did not try to attain a balance between the genders, the sample turned out to be mostly male.

For validating the hypotheses, we tested the different regression models using SPSS. Summated scales were created for each variable. The primary objective of this experiment is to assess the extent to which subjects would consider adopting and using the prototype if it was made available to them. Accordingly, the first regression model tested the relationship between Perceived Usefulness and Intention to Use, and Satisfaction and Intention to Use. Table 4 reports the results of the regression analysis.

The data indicated that the regression model (model 1) was significant and both perceived usefulness and satisfaction were significantly associated with intention to use the prototype, thus supporting hypotheses H4 and H5. Overall, these two variables explained about 45% of the variance in the intention to use. Model 2 tested the relationship between satisfaction and its antecedents. Both perceived usefulness and confirmation were found to be significant predictors of satisfaction, thus supporting H3 and H1 respectively. Finally, model 3 indicates that the hypothesized relationship between confirmation and perceived usefulness (H2) is not supported.

Table 3: Sample Demographics

Demographic Variable	Categories	Frequency (N=30)
Age	Less than 25 25 – 34 Greater than 35	4 (13.3%) 25 (83.3%) 1 (3.3%)
Gender	Female Male	4 (13.3%) 26 (86.7%)
Educational Status	Bachelor´s Master´s PhD Not specified	4 (13.3%) 16 (53.3%) 5 (16.7%) 5 (16.7%)
Facebook User	Yes No	15 (50%) 15 (50%)
User of other Social Networking Site	Yes No	29 (96.7%) 1 (0.03%)

In addition to testing the hypothesized relationships, we controlled for several extraneous factors that can affect the findings. It was found that the demographic variables such as gender, age and educational background did not affect the responses. Since our experimental sample had equal number of Facebook users and non-users, we also tested whether this affected any of the variables of interest. Subjects Facebook usage status did not affect confirmation, perceived usefulness or their satisfaction with the NFriendConnector prototype application. However, whether or not a person was a Facebook user was significantly associated with their intention to use the prototype application (p=0.043), with Facebook users reporting a higher intention to use the prototype (mean=3.7, s.d=0.82) than Facebook non-users (mean=2.9, s.d=1.21). The implications of these findings are discussed below.

Table 4: Regression Results

Variables	Standardized Coefficients B	Significance	Hypotheses
Model 1: Dependent Variable – Intention to Use			
R^2 = 0.492; Adjusted R^2 = 0.455; F = 13.093; Sig = 0 .000 (N = 30)			
Perceived Usefulness	0.364	0.048 (t=2.067)	H4 supported
Satisfaction	0.413	0.027 (t=2.341)	H5 supported
Model 2: Dependent Variable – Satisfaction			
R2 = 0.506; Adjusted R2 = 0.469; F = 13.809; Sig = 0 .000 (N = 30)			
Perceived Usefulness	0.560	0.000 (t=4.049)	H3 supported

Confirmation	0.339	0.021 (t=2.456)	H1 supported
Model 2: Dependent Variable – Perceived Usefulness			
R2 = 0.041; Adjusted R2 = 0.007; F = 1.208; Sig = 0 .281 (N = 30)			
Confirmation	0.203	0.281 (t=1.099)	H2 not supported

4.5 Findings and Implications

A prototype design exercise can be considered successful when the target group of users find the application useful and are willing to use it. Our results suggest that subjects perceived the NFriendConnector as a useful application. Satisfaction and Perceived Usefulness were found to be significantly associated with users' intention to use the prototype. Satisfaction was also found to be significantly associated with perceived usefulness and confirmation or users' expectation. In addition to validating the usefulness of the NFriendConnector prototype, this study gives empirical evidence to the applicability of the expectation-confirmation model in the context of information system adoption and use. In line with the findings of Bhattacherjee [2], the experimental results indicate that Satisfaction was a stronger predictor of Intention to Use than Perceived Usefulness. Since post-use Satisfaction is grounded in users' first-hand experience with the system, it is believed to be more realistic, unbiased, and less susceptible to change, while Perceived Usefulness could be based solely on cognitive beliefs formed via second-hand information received from various sources.

The NFriendConnector is an innovative application that extends Facebook to its users' real worlds by allowing them to establish Facebook connections and access other Facebook features using their mobile phones. This can be a possible reason behind the insignificant relationship between confirmation and perceived usefulness. Since subjects encounter the application for the first time during the study, it is unlikely that they had pre-formed notions regarding its usefulness and what to expect from it. Accordingly, there was no expectation based confirmation that influenced users' perception regarding the usefulness of the prototype, therefore the insignificant relationship.

Even people who did not have prior experience with Facebook perceived the prototype as useful and were equally satisfied with it (reflected by no significant difference in the mean values of perceived usefulness or satisfaction for the user and non-user groups). This finding has important implications to both researchers and practitioners because it indicates that the current prototype is considered useful by users of other online social networking sites. Therefore, although this prototype was developed predominantly with the usage patterns of Facebook users in mind, the functionality provided by the NFriendConnector is considered meaningful and useful by users of other online social networking sites as well. However, we found that intention to use was significantly associated with whether or not a person was a Facebook user. This is possibly because Facebook non-users felt that the prototype was unlikely to be available to them for use, and therefore they reported lower intentions to use it.

Research and practical developments in the area of online social networking reflect a shift in the behavioral patterns and intentions of users of online social networks to reflect a tighter interlinking between socializing in real lives and the use of SNS. However, there have been few theoretically grounded studies which examine this changing usage pattern, and how providing applications and features to support this pattern, enhances satisfactions and intention to use the system. Our study addresses this gap by examining the use of the NFriendConnector prototype using the ECM as the theoretical framework, and provides empirical validation to the usefulness of such applications and features in the context of the most popular SNS – Facebook.

5. LIMITATIONS

The NFriendConnector prototype application, as well as this study should be interpreted in the context of its limitations. The NFriendConnector prototype demonstrates the extension of online social networking into the real world by integrating the Facebook platform with NFC-enabled mobile devices to provide a real time and ubiquitous access to online social networking. Therefore, the prototype supports only a limited set of Facebook functionality. Moreover, the prototype requires NFC-enabled mobile phones, which might prevent many users from adopting it.

This prototype was developed based on Facebook usage patterns, indicating that for Facebook users the direction of social networking is often from offline to online. This may raise issues regarding the generalizability of the prototype in other social networking platforms. In future research endeavors, experiments can be conducted to assess the perceptions of a different category or users using other social networking platforms (e.g. LinkedIn, Friendster, MySpace).

Conducting a laboratory experiment gives rise to limitations that are inherent to this research methodology. Future studies could assess usability in more natural social settings where users are less likely to feel constrained by an experimental setup. Studies that allow users to use and evaluate the prototype over a longer period of time could be designed to get a better gauge of their feelings and experiences with the prototype. Finally, we used university graduates and student subjects, which might have had some effect on the final results of the experiment. However, this does not really pose a big problem because our sample is generally reflective of the online social networking user profile and most subjects were not just Facebook users but also experienced in using other online social networks.

6. CONCULSION AND FUTURE RESEARCH

Current trends in social media, and mobile applications and devices indicate that there is an increasing demand for ubiquitous access to online social networks, and make it an integral part of people's real lives. In spite of this growing popularity of online social networking, currently, there is little connection between socializing in the real world and online social networking although users often use online social networking platforms like Facebook to map their actual offline contacts. Identifying this gap, we conceptualize, develop and evaluate the NFriendConnector prototype for to enable seamless social networking by extending online social networking into the real world using mobile devices.

The NFriendConnector allows users to make Facebook connections and also access certain other Facebook features, using NFC-enabled mobile phones. Further, social interactions can be

enriched by complementing them with information from their online Facebook profiles to discover mutually shared interests. The experimental evaluation suggests that prospective users perceive the prototype as useful and are satisfied with its functionalities. Further, they indicate an intention to use the prototype if it is available to them. This provides empirical validity to our proposition regarding the importance of providing seamless social networking.

Our experimental results further indicate that even users who do not use Facebook perceive the prototype as useful and are satisfied with it. This indicates that there is a significant potential for developing similar prototypes for other online networking sites. For instance, being able to exchange contact information and establish connections may be particularly useful for online networking sites such as LinkedIn, which is a predominantly professional network. The functionality provided by the NFriendConnector can be thought of as the exchange of "online business cards", where NFC-enabled mobile phones allow users to exchange their contact information and professional interests even without having a mobile Internet connection. Future research could therefore support similar functionalities for other online social networks that are targeted towards different user groups. Finally, more comprehensive, long-term studies should be designed and executed to assess the extent to which different groups of users find the application useful and how its usage influences their online and offline social networking behavior.

7. ACKNOWLEDGMENTS

The NFriendConnector prototype was developed in the context of the research project Mobil50+ (Innovative NFC- und IT-basierte Dienstleistungen für mobiles Leben und Aktivität der Generation 50+). Mobil50+ is funded by the German Federal Ministry of Education and Research (BMBF - FKZ: 01FC08046). It is a joint project of the Technische Universität München and various partners. For further information, see www.projekt-mobil50.de.

8. REFERENCES

[1] Beach, A., M. Gartrell, S. Akkala, J. Elston, J. Kelley, K. Nishimoto, B. Ray, S. Razgulin, K. Sundaresan, B. Surendar, M. Terada, and R. Han 2008. WhozThat? evolving an ecosystem for context-aware mobile social networks. *Network, IEEE.* 22, 4 (2008) 50-55. DOI= http://dx.doi.org/10.1109/MNET.2008.4579771

[2] Bhattacherjee, A. 2001. Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly.* 25, 3 (2001) 351-370. DOI= http://dx.doi.org/10.2307/3250921

[3] Bramson-Boudreau, E. and L. Arathoon 2008. *Analyst Insight: By the end of 2012, 950m user will be accessing social networking sites via mobile devices*. 2008; Available from: http://www.pyramidresearch.com/documents/02.21.08_AI_Mobile%20SNS.pdf.

[4] Cummings, J.N., B. Butler, and R. Kraut 2002. The quality of online social relationships. *Commun. ACM.* 45, 7 (2002) 103-108. DOI= http://dx.doi.org/10.1145/514236.514242

[5] Davis, F.D., R.P. Bagozzi, and P.R. Warshaw 1989. User acceptance of computer technology: a comparison of two theoretical models. *Manage. Sci.* 35, 8 (1989) 982-1003. DOI= http://dx.doi.org/10.1287/mnsc.35.8.982

[6] Eagle, N. and A. Pentland 2005. Social Serendipity: Mobilizing Social Software. *IEEE Pervasive Computing.* 4, 2 (2005) 28-34. DOI= http://dx.doi.org/10.1109/MPRV.2005.37

[7] Fogg, B.J. 2003. *Persuasive Technology - Using Computers to Change What We Think and Do.* Morgan Kaufman, San Francisco, CA.

[8] Karahanna, E., D.W. Straub, and N.L. Chervany 1999. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Q.* 23, 2 (1999) 183-213. DOI= http://dx.doi.org/10.2307/249751

[9] Kostakos, V. and E. O'Neill 2008. *Cityware: Urban computing to bridge online and real-world social networks*. In *Handbook of Research on Urban Informatics: the Practice and Promise of the Real-Time City*, M. Foth, Ed. Information Science Reference, IGI Global, 196-205.

[10] Lampe, C., N. Ellison, and C. Steinfield 2006. A face(book) in the crowd: social Searching vs. social browsing. In *Proc. 20th anniv. conf. on CSCW*. 2006. Banff, Alberta, Canada: ACM. DOI= http://dx.doi.org/10.1145/1180875.1180901

[11] Lampe, C.A.C., N. Ellison, and C. Steinfield 2007. A familiar face(book): profile elements as signals in an online social network. In *Proc. CHI*. 2007. San Jose, California, USA: ACM. DOI= http://dx.doi.org/10.1145/1240624.1240695

[12] Leimeister, J.M. and H. Krcmar 2006. Community-Engineering Systematischer Aufbau und Betrieb Virtueller Communitys im Gesundheitswesen. *WIRTSCHAFTSINFORMATIK.* 48, 6 (2006) 418-429. DOI= http://dx.doi.org/10.1007/s11576-006-0094-y

[13] Leimeister, J.M., K. Schweizer, S. Leimeister, and H. Krcmar 2008. Do virtual communities matter for the social support of patients?: Antecedents and effects of virtual relationships in online communities. *Information Technology & People.* 21, 4 (2008) 350 - 374. DOI= http://dx.doi.org/10.1108/09593840810919671

[14] Lin, C.S., S. Wu, and R.J. Tsai 2005. Integrating perceived playfulness into expectation-confirmation model for web portal context. *Inf. Manage.* 42, 5 (2005) 683-693. DOI= http://dx.doi.org/10.1016/j.im.2004.04.003

[15] Oliver, R.L. 1980. A cognitive model for the antecedents and consequences of satisfaction. *Journal of Marketing Research.* 17, 4 (1980) 460-469.

[16] Oliver, R.L. 1993. Cognitive, Affective, and Attribute Bases of the Satisfaction Response. *Journal of Consumer Research: An Interdisciplinary Quarterly.* 20, 3 (1993) 418-30.

[17] Parks, M.R. and K. Floyd 1996. Making Friends in Cyberspace. *Journal of Communication.* 46, 1 (1996) 80-97. DOI= http://dx.doi.org/10.1111/j.1083-6101.1996.tb00176.x

[18] Parks, M.R. and L.D. Roberts 1998. `Making Moosic': The Development of Personal Relationships on Line and a Comparison to their Off-Line Counterparts. *Journal of*

Social and Personal Relationships. 15, 4 (1998) 517-537. DOI= http://dx.doi.org/10.1177/0265407598154005

[19] Persson, P., J. Blom, and Y. Jung 2005. DigiDress: A Field Trial of an Expressive Social Proximity Application. In *UbiComp 2005: Ubiquitous Computing*, 195-212. DOI= http://dx.doi.org/10.1007/11551201_12

[20] Persson, P. and Y. Jung 2005. Nokia sensor: from research to product. In *Proc. DUX*. 2005. San Francisco, California: AIGA: American Institute of Graphic Arts.

[21] Pietiläinen, A.-K., E. Oliver, J. Lebrun, G. Varghese, and C. Diot 2009. MobiClique: Middleware for Mobile Social Networking. In *Proc. Sec. ACM SIGCOMM Workshop on Online Social Networks*. 2009. Barcelona, Spain: ACM. DOI= http://dx.doi.org/10.1145/1592665.1592678

[22] Richter, A., M. Koch, and J. Krisch 2007. *Social Commerce Eine Analyse des Wandels im E-Commerce.*

[23] Spreng, R.A., S.B. MacKenzie, and R.W. Olshavsky 1996. A Reexamination of the Determinants of Consumer Satisfaction. *The Journal of Marketing.* 60, 3 (1996) 15-32. DOI= http://dx.doi.org/10.2307/1251839

[24] Ziv, N.D. and B. Mulloth 2006. An Exploration on Mobile Social Networking: Dodgeball as a Case in Point. In *Proc. of the ICMB*. 2006: IEEE Computer Society. DOI= http://dx.doi.org/10.1109/ICMB.2006.8

Behind the Curtains of Privacy Calculus on Social Networking Sites: The Study of Germany and the USA

Hanna Krasnova
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Germany
+49 30 2093-5742
krasnovh@wiwi.hu-berlin.de

Natasha F. Veltri
University of Tampa
401 W. Kennedy Blvd.
Tampa, FL 33606, USA
+1 813 257-3970
veltri@ut.edu

ABSTRACT
As social networking sites (SNSs) become increasingly global, the issues of cultural differences in participation patterns become acute. However, current research offers only limited insights into the role of culture behind SNS usage. Aiming to fill this gap, this study adopts a 'privacy calculus' perspective to study the differences between German and American SNS users. Results of structural equation modeling and multi-group analysis reveal distinct variability in the cognitive patterns of American and German subjects. We contribute to the theory by rejecting the universal nature of privacy-calculus processes. From a practical standpoint, our results signal that SNS providers cannot rely on the "proven" means in ensuring user participation when crossing geographic boundaries. When financial means are limited, SNS providers should direct their investments into enhancing platform enjoyment and granting users with more control and, paradoxically, lobbying for more legalistic safeguards of user privacy.

Keywords
Social Networking Sites, Self-Disclosure, Privacy, Trust, Culture.

1. INTRODUCTION
Social Networking Sites (SNSs) have become an inseparable part of daily online routines for millions of people. Web analytics companies, like Alexa.com, report that SNSs top the website traffic charts all over the world, thus, pointing to the ubiquitous nature of social networking (SN). Despite this global popularity, the business of SNS providers is far from being easy. Recruitment of new users as well as retention of existing ones presents a major challenge. Having reached maturity in their home countries, SNS providers increasingly rely on internationalization strategies in their pursuit for growth. For example, Facebook (FB) offers more than 70 translations of its site [10].

This growth strategy, however, is plagued by numerous complexities. When expanding internationally SNS providers have to contend with local rivals. Thus, FB is competing for new members alongside Skyrock in France, Vkontakte in Russia and StudiVZ in Germany. While local SNSs boast first-hand knowledge of their home markets, international platforms like FB are challenged to adopt their platform design, communication strategy and image to national peculiarities.

Given the global nature of users and providers [10] it is crucial to identify, understand and bridge cultural differences in SNS usage. Despite importance of this research question for practice, the literature exploring the moderating influence of culture in the SNS context is largely absent. Filling this gap, our study examines the cultural differences between German and American participants of SNSs. German market is currently dominated by two major players: local StudiVZ and a late entrant FB. In this head-to-head competition for user attention, both platforms have significant advantages on their side: whereas FB is the world leader in SN services, excelling in innovation; StudiVZ has insight into the specific needs of German users.

Building on existing theoretical insights regarding the motivation to use and communicate on SNSs, we propose a theoretical model of SNS participation. Exploring privacy calculus in intercultural setting we examine the moderating effect of culture on construct relationships in our model. Our hypotheses aim to reveal practical insights into the motivational dynamics behind SNS usage. A resulting structural equation model is then tested with American and German FB users. Implications of our results for theory and practice conclude the paper. On the policy side, we expect to offer relevant insights for policy-makers who are interested in protecting the privacy of online users while retaining the unprecedented potential of SNSs to build social capital.

2. RESEARCH MODEL
Krasnova et al. [24] argue that supporting interpersonal communication on the SNS platform is key to user recruitment and retention. Without ongoing communication, network content becomes outdated, leading to decreased user interest and immersion and, as a consequence, lost user loyalty. Beyond

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

ensuring involvement, user *self-disclosure* is also crucial for financial sustainability of SNSs. Even though advertisers are willing to pay for this large database of potential customers, their interest is contingent on active participation. Indeed, freshly updated user content offers advertisers unlimited opportunities for personalization, customer segmentation and market research. Consequently, in order to remain competitive SNS providers must do their best to motivate SNS users to communicate and *self-disclose* on their platform, and do so frequently. This task, however, becomes increasingly complex when operating in a foreign market. As culture determines the way users behave [15], SNS providers need a deeper understanding of the intercultural dynamics of individual *self-disclosure*.

In the past, researchers have applied a variety of theories to explain the factors behind individual adoption of IT. Theories of Reasoned Action, Planned Behaviour, Technology Acceptance Model, Diffusion of Innovations Theory and Social Cognitive Theory have been frequently used to examine usage intensity in a variety of settings. As participation of SNSs is associated with numerous privacy risks, a number of studies advocate the use of the 'privacy calculus' ('PC') perspective when investigating self-disclosure on SNSs (e.g. [24]). In line with this theory, online *self-disclosure* is a product of partially conflicting beliefs, such as *expected benefits*, *privacy concerns* and *trusting beliefs* [8]. Based on this view, Krasnova and Veltri [25] propose an extended model of 'PC' on SNSs particularly adopted for the purposes of intercultural research. Even though their work provides a sound comparative analysis of the means of the model-relevant constructs for German and American SNS users in their sample, the authors leave validation of their empirical model for future work. We pick up this recommendation and adopt their model for the purposes of our study. As our study is dictated by practical considerations, only constructs of immediate relevance for SNS providers and policy-makers are included in the model (Figure 1). We integrate beliefs regarding *Enjoyment, Privacy Concerns,* and *Trust in SNS Provider* as three independent forces defining the dynamics of 'PC' on SNSs. In order to supply SNS providers and policy-makers with insights on how these components of 'PC' can be leveraged in different cultures, we extend our model with three practice- and policy- relevant antecedents: *Legal Assurance, Perceived Control* and *Knowledge*. The reasoning behind the hypothesized relationships as well as their interaction with various cultural dimensions is explained in the following sections.

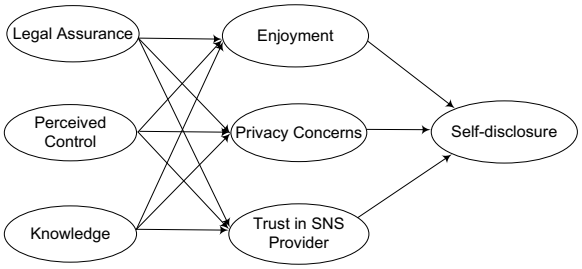


Figure 1. Research Model.

3. PRIVACY CALCULUS IN AN INTERCULTURAL SETTING

Rooted in values, beliefs and traditions, national culture permeates the way people interpret and behave in various situations. It is therefore natural to expect that cultural norms will influence such daily activities as SN online [28].

Even though a multitude of studies address the differences in culture, a framework by Hofstede [15] has received widespread acceptance. Based on Hofstede's national culture indices there are several similarities as well as differences between the German and US cultures. Whereas both cultures exhibit *low power distance* (PDI), *low long-term orientation* (LTO) and *high masculinity* (MAS), they have significant differences in the dimensions of *individualism* (IDV) and *uncertainty avoidance* (UAI). As Germans are significantly *lower* on IDV scale, they are likely to be more interdependent, group-oriented, show higher loyalty to other people and institutions, as well as suppress their emotions and behaviour. On the other hand, *high-IDV* Americans are characterized by higher self-reliance, competitiveness, hedonic attitudes and emotional distance from in-groups [47]. Furthermore, due to *higher UAI*, Germans are expected to be more risk-averse, feel threatened by uncertain conditions, exhibit strong faith in institutions and prefer the tried and tested ways over new methods [9]. Overall, numerous studies confirm the leading role of IDV and UAI dimensions as major explanatory variables in a variety of settings (e.g. online shopping) [6, 28]. Hence, in the context of our study we pay particular attention to these dimensions, when discussing moderating effects of culture on the relationships embedded in our model.

4. RESEARCH HYPOTHESES

Before proceeding, it is important to note that even though subsequent analysis is based on a thorough examination of literature, complexity of cultural influence makes our hypotheses only exploratory in nature.

Enjoyment: A multitude of benefits, such as self-presentation, convenience of social interaction, relationship building, may motivate users to self-disclose on a SNS [24]. Despite their diversity, all these motives have one common denominator: enjoyment. Indeed, based on the findings from social psychology, pleasure can be viewed as a "*consequence of gratification of a motive*" ([40] p. 183). For example, Muniz and O'Guinn [34] demonstrate that conversations in the Internet communities are a pleasurable experience for participants. Furthermore, Rosen and Sherman [43] describe SNSs as purely hedonic platforms. Taking into account the leisure-oriented character of SNS platforms, we integrate *enjoyment* as a *positive* determinant of *self-disclosure* on SNSs as depicted in Figure 1.

Overall, SNS providers have long ago acknowledged the role of pleasant experiences as a crucial part of their competitive strategies. Evidently, the satisfaction of users' desire for fun was the main objective of FB when introducing the News Feed or opening up its platform to third-party developers. As our study aims to deliver practical insights, exploring the role of *enjoyment* in motivating user communication from an intercultural standpoint is important.

Overall, the impact of cultural dimensions on the relationship between *enjoyment* and *self-disclosure* is characterized by high complexity. On the one hand, we expect Germans, as a more *collectivistic* culture, to value in-group interaction and thus enjoy disclosing their information on SNSs as part of group communication process. On the other hand, pleasure-seeking and

hedonism constitute one of the major traits of *individualistic* cultures (like the USA), where people are also more likely to align their behaviour with their own needs and priorities [7, 47].

Besides, it is important to note that both cultures exhibit very low levels of LTO, which implies strong preference for short-term gains as opposed to delayed gratification and forward thinking [15]. As disclosure of information is usually associated with quick reaction from others (e.g. commenting / liking one's status), which in most cases leads to a pleasurable experience, it is plausible to assume that German and American users will be equally encouraged to self-disclose as a result of their anticipated benefits. We therefore hypothesize that:

Hypothesis 1a: *The positive relationship between perceived Enjoyment and Self-disclosure* ***will not differ significantly*** *for US and German SNS users.*

Privacy Concerns: It is a widespread belief that participation on SNSs is associated with significant privacy risks. Indeed, a seemingly innocent piece of information published online may prove to be a career-killer several years later. Employers, law enforcement, tax authorities, creditors, and military recruitment agencies are increasingly relying on SNSs to investigate their clientele. Just recently, a medical insurance company stopped sick-leave payments for their depressed client based on the FB photos depicting her as "happy" [2].

In their attempt to capitalize on privacy concerns of SNS users, providers increasingly integrate privacy-related claims into their web-sites and PR campaigns [3]. Whether or not this strategy will pay off in every country is, however, unclear. On the one hand, the relationship between *privacy concerns* and *self-disclosure* is expected to be strongly negative. However, empirical results provide a mixed picture: whereas a German-based study of Krasnova et al. [24] finds user privacy concerns to be a significant impediment to information sharing on SNSs, a study by Acquisti and Gross [1] finds a discrepancy between claimed privacy concerns and disclosure behaviour for the US respondents. These inconsistencies indicate that the strength of the negative relationship between *privacy concerns* and *self-disclosure* is still to be examined in an intercultural setting.

Lim et al. [28] suggest that both UAI and IDV play a significant role in the formation of *risk attitudes*. People from high-UAI countries have lower tolerance for uncertainty, feel more threatened by ambiguous situations, and perceive more anxiety about the future [15]. As a result, they are likely to be more apprehensive about the consequences of their self-disclosure behaviour. In general, empirical evidence supports a positive link between UAI and the impact of risk perception on the subsequent behaviour. For example, Park [35] finds a positive link between UAI/IDV and insurance penetration on a country-level, hinting that risk-averse individuals are more keen on seeking ways to relieve their anxiety. Similarly, strongly reducing their self-disclosure is a likely response to privacy concerns by people from high-UAI cultures like Germany.

Even though IDV may also play a role in the formation of privacy concerns, the role of this cultural dimension is ambiguous. On the one hand, because *individualistic* societies are typically very competitive as well as opportunistic, SNS users may be more conscious about the sensitivity of the information they publish online, as it can be used to damage their reputation or careers. In support of this argument, Dinev et al. [6] find a stronger impact of *Privacy Concerns* on the use of e-commerce for the US as opposed to Italian subjects. On the other hand, *collectivistic* cultures are more likely to be anxious about the consequences of their self-disclosures. Indeed, public mockery of one's postings by outsiders may lead to the so much feared "loss of social face" in the in-group [33].

Investigating interaction of UAI and IDV in the context of risk perceptions, Lim et al. [28] show that for countries with relatively high uncertainty avoidance levels, as is the case for Germany, IDV dimension shows little impact on Internet shopping. Acknowledging the leading role of UAI in defining the role of *risk perceptions* in user behaviour we hypothesize that:

Hypothesis 1b: *The negative relationship between Privacy Concerns and Self-disclosure* ***will be stronger*** *for German users than for US users.*

Trust in SNS Provider: Trust is a central construct whenever relationships between parties involve some degree of risk or uncertainty. Whereas trust does not necessarily lead to the elimination of risk beliefs, it can overrule their negative impact on self-disclosure [8]. This is consistent with the threshold model which suggests that formation of trust is independent of risk beliefs. Once the level of trust has exceeded the threshold of perceived risk, the trustor will engage in a risky behaviour [12]. In this respect, trust in the network may be the key to explaining the dichotomy between expressed attitudes and actual behaviour of the American SNS users [1].

In the SNS context, authors differentiate between *Trust in SNS Provider* and *Trust in SNS Members* (e.g. [24]). Given practical considerations, in this study we concentrate exclusively on *Trust in SNS Provider*, which we conceptualize as a set of individual beliefs regarding provider's benevolence and integrity. Even though studies from various contexts confirm the positive impact of trust on individual willingness to self-disclose (see [24]), the strength of this influence is likely to be defined by cultural norms [25].

Lim et al. [28] argue that particularly IDV and UAI dimensions are likely to affect individual willingness to engage in risky transactions, which constitutes the basis for trusting intentions. People from high-UAI cultures (like Germany) are likely to be more pessimistic about the incentives of companies, including SNS providers [15]. Because they are afraid of becoming vulnerable to the other party in a transaction, they are also less likely to take a "leap of faith" and actually *act* on the basis of trust. On the other hand, people from the low-UAI cultures (like the USA) may place less emphasis on privacy concerns and rather concentrate on collecting evidence of the trustworthiness of the other party. As a result, they are more likely to engage in a risk-taking behaviour on the basis of trust [9].

Furthermore, most authors hold IDV as more favourable for trust-based behaviour. For example, Dinev et al. [6] argue that members of *collectivistic* cultures strongly differentiate between in- and out-group members when developing trust. Hence, it is unlikely that Germans will exhibit higher readiness than Americans to make themselves vulnerable to socially- and

geographically-distant FB (e.g. translate their trusting beliefs into trusting intentions and then into behaviour [9]).

Furthermore, differences in IDV dimension trigger distinct mechanisms in the trust-formation process. Whereas collectivists (1) concentrate on *predictability* of future actions of the trustee, (2) look for cues that the trustee will *act in their best interests (benevolence)* as well as (3) *easily transfer* trust from one to another within their group; individualists primarily (1) *calculate* the costs and benefits of the defection behaviour of the trustee as the basis of trust [9]. Analysing these differences in the trust-formation patterns in the context of online shopping, Lim et al. ([28], p. 549) argue that as *predictive* trust is more difficult to develop "*Internet shopping is more appealing to individualists than to collectivists*". In the SNS context, more collectivistic German users may also find it hard to form sound *predictions* about the future behaviour of the US-based FB. The *transference* process is also complicated by the controversy of messages present in the German society: while some may admire FB for its structured website design, others scorn it for carelessness in handling user privacy [4]. This complexity of the situation is, however, conducive to the individualistic formation of trust on the basis of calculative thinking. Thus, high-IDV American users may figure that SNS providers have more to lose than to gain by violating their privacy: Once privacy abuse comes to surface, SNS provider will have to deal with ruthless media criticism, stalling user numbers, decreased communication, and even expensive lawsuits. Summarising, *trust* is expected to play a higher role in the decision-making process of high-IDV cultures. Therefore, we hypothesize that:

Hypothesis 1c: *The positive relationship between Trust in SNS Provider and Self-disclosure* ***will be stronger*** *for US than for German SNS users*.

As discussed above, three major forces are critical for the individual disclosure online: *platform enjoyment*, *privacy concerns*, and *trust in SNS provider*. From a practical standpoint, however, our analysis will not be complete without understanding of how the development of these 'PC'-relevant perceptions can be managed. On the policy side, *legal assurance* may impact the dynamics of 'PC' decisions. On the managerial level, enabling users with *control* over their privacy as well as informing them about the essence of the adopted information-handling practices (*knowledge*) represent two interrelated strategies of addressing individual perceptions. We therefore integrate them as positive antecedents of our 'PC'-variables as depicted in Figure 1 and discussed below.

Legal Assurance: As perceptions regarding benefits, privacy risks and trusting beliefs are situational, they are likely to be influenced by the *institutional structures* inherent in the environment in which SN services are consumed [32]. Indeed, when engaging into such risky behaviour as self-disclosure, users are likely to rely on *legal structures* - privacy-related laws, formal policies and procedures – which are designed to give them confidence that things will go well.

By and large, IS-researchers agree that sound legal framework may help to create an atmosphere of *trust* on the platform. In particular, when a relationship is associated with numerous risks – as is the case for SNSs - the legal mechanisms may work to create a much needed "*trust infrastructure*". Furthermore, in the absence of obvious means to control the use of personal data by providers, *legal assurances* are likely to be the best solution to *privacy concerns* [30]. Finally, by setting the rules of the 'privacy game', law-makers ensure that users feel at ease releasing their personal information on a SNS. This, in turn, is likely to allow them to gain the benefits of relationship maintenance, self-presentation and, above all, *enjoyment*. Hence, it comes as no surprise that SNS providers are increasingly relying on third party seals to signal their compliance with the required standards.

As perceptions regarding *legal assurance* are likely to motivate the development of *trusting beliefs* and *enjoyment* as well as mitigate *privacy concerns*, the strength of their impact is likely to differ from country to country.

People in risk-averse cultures like Germany are likely to exhibit "*...higher needs for structure (i.e., formal rules and regulations), and stronger faith in institutions (e.g., the government)*" ([28], p. 547). By establishing protective norms and rules, these cultures are trying to minimize their risk-related anxiety [7]. Significant differences in privacy regulation between Germany and the USA provide support for the validity of this argument. While Americans leave numerous privacy aspects to industry self-regulation, Germans have a comprehensive legal framework covering multiple aspects of personal data access, collection and use. We argue that as risk-averse cultures attach greater importance to rules and standards, these legal assurances will also play a bigger role in the formation of beliefs relevant for 'PC'. We hypothesize that:

Hypothesis 2a*: The positive relationship between perceptions regarding Legal Assurances and Enjoyment* ***will be stronger*** *for German than for US SNS users.* ***Hypothesis 2b****: The negative relationship between perceptions regarding Legal Assurances and Privacy Concerns* ***will be stronger*** *for German than for US SNS users.* ***Hypothesis 2c****: The positive relationship between perceptions regarding Legal Assurances and Trust in SNS Provider* ***will be stronger*** *for German than for US SNS users.*

Perceived Control: In the light of privacy discourse, many authors equate the presence of *control* with the notion of privacy. For example, Son and Kim ([45] p. 504) define information privacy as "*individual's ability to control when, how, and to what extent his or her personal information is communicated to others*".

In the context of SNSs, *control* may take two forms: *control over accessibility of personal information* and *control over information use* [23]. However, the latter type of control needs is hard to address under a current revenue model of SNS providers. In fact, whereas business networks like Xing have long ago introduced premium accounts to capitalize of the privacy needs of its members [3], SNS providers offer their users a sole "take-it-or-leave-it" option when it comes to the use of their personal data. This lack of flexibility in choosing an appropriate mode of information secondary use is partly compensated by *accessibility control* options, which give users the possibility to protect their information from the prying eyes of third parties and other users. We therefore concentrate on this form of *control* to fulfil the purposes of our study.

Overall, Pavlou and Gefen [36] argue that market-driven mechanisms, like *control*, are powerful means to ensure desired

organizational outcomes. Indeed, empowering users with refined and easy-to-understand privacy settings is likely to support their beliefs that SNS provider is acting in their best interests thereby providing basis for *trust*. Indeed, investigating user behaviour on FB, Krasnova et al. [23] show that perceptions of *control* lead to enhanced *trusting beliefs* and reduced *privacy concerns* on the network. Moreover, users are even willing to pay significant sums of money to have more refined control over accessibility of their information on SNSs [21]. Overall, adjusting profile visibility constitutes a central strategy when it comes to resolving a conflicting pressure between the desire to self-present and the need to keep one's information private [48]. We assume that by actively defining the accessibility of their information to the outside world, SNS users are likely to feel themselves in the driver's seat when it comes to managing their privacy. These feelings are likely to lessen individual perceptions of riskiness of the SNS experience as well as contribute to the development of more positive attitudes towards the network, including platform *enjoyment*.

Taken together, studies from various contexts underscore the role of *control* in mitigating user *privacy concerns*, ensuring *trust* and enabling desired outcomes. However, even though the need to *control* one's outcomes is likely to be universal across cultures, we expect significant differences in the weight and consequences different cultures associate with it.

The role of *control* as an intercultural phenomenon has been intensively discussed in the context of organizational fairness. In these studies *control* is viewed as a major dimension of procedural justice and reflects the freedom to *voice* an opinion about how one's personal information will be used [31]. Konovsky [19] argues that cultural *individualism* may be influential in determining attitudes to *voice*. However, empirical evidence regarding the importance of *voice* perceptions is controversial. On the one hand, Leung and Lind [27] find that people in high-IDV cultures (like USA) show higher preference for *process control* as opposed to collectivistic cultures (like China). The reason for these differences may lie in the inherent competitiveness of high-IDV cultures. Collectivists, on the other hand, are more inclined towards harmony and hence attach less importance to *voice* in the decision-making process. On the other hand, some studies find that justice perceptions are equally important across *individualistic* and *collectivistic* cultures [33].

A number of studies support importance of voice in the trust-building process for countries with high IDV levels. For example, in an organizational context, Pillai and Williams [38] find that procedural justice is a more important predictor of trust in the US, than in Germany. Furthermore, Lim et al. ([28] p. 548) argue that as individualistic cultures build trust on a calculative basis, they are more likely to look for cues that indicate opportunism or trustworthiness of the provider. In this case, available privacy controls, "*vendor's recourse and refund policy and/ or the existence of third-party certifications, such as eTrust, BBB Online*" may provide such assurances. Furthermore, Dinev et al. [7] argue that people in individualistic and masculine societies will be more willing to depend on a trustee, if appropriate degree of control is provided. Taken together, if based solely on IDV dimension, the impact of control on trusting beliefs should be higher for the USA than for Germany. However, German users also exhibit high levels of UAI, which makes them particularly aware of possible vulnerability inherent in a trusting relationship. As a result, German users may be particularly sensitive to the availability of controls when deciding to intentionally expose themselves to privacy-related threats. Taken together, the impact of low IDV on the relationship between *control* and *trust* is likely to be *balanced out* by high level of UAI in Germany.

When it comes to the link between *control* and *privacy concerns*, UAI is likely to be a dominant factor in determining the strength of this relationship. Indeed, as people from high-UAI cultures experience stronger fear for the unknown, they actively seek for means to relieve this anxiety [15]. Dinev et al. ([7], p. 395) argue that people in risk-averse cultures would "*attempt to control almost everything in order to avoid the unexpected*". Therefore, it is conceivable that German SNS users would attach higher relevance to privacy controls when forming their judgements about privacy risks.

Whether or not the influence of *control* on *enjoyment* will be stronger in Germany as opposed to the USA is a complex issue. On the one hand, as collectivists prefer in-group communication, they are likely to find it more enjoyable when they are sure that no out-group members have knowledge of it Furthermore, collectivists are more likely to rely on procedures which retain inter-personal harmony [33]. In this case privacy settings offer excellent means to regulate the outgoing information without offending anyone. Individualists, on the other hand, may enjoy self-expression beyond their group of friends and hence be more relaxed in the absence of mechanisms controlling their information stream. Whether this argument would hold, however, is debatable. Even though Germany is much more collectivistic than the USA, its IDV index is much higher than the world average (67 vs. 43 [14]). Moreover, Lind and Early [29] argue that even the most individualist people care about group-related issues.

Taken together, we hypothesize that:

Hypothesis 3a: *The positive relationship between Perceived Control and Enjoyment* ***will not differ significantly*** *for US and German SNS users*. ***Hypothesis 3b:*** *The negative relationship between Perceived Control and Privacy Concerns* ***will be stronger*** *for German than for US SNS users*. ***Hypothesis 3c:*** *The positive relationship between Perceived Control and Trust in SNS provider* ***will not differ significantly*** *for US and German SNS users*.

Knowledge: Even when favourable information-handling practices and controls are in place, users may be unaware of their existence or content [45]. Many privacy policies are written in a complicated legalistic language illegible for an ordinary user. In 2006 Acquisti and Gross [1] found that around one third of their respondents were not aware about searchability control options, even though they were implemented on FB. Recognizing the ubiquitous nature of this problem for online companies, Malhotra et al. [31] acknowledge awareness about information-handling procedures as the key element of online privacy. In their view, increasing user awareness is likely to enhance *trust* and mitigate *privacy concerns*.

Even though people are expected to rely on their *knowledge* to categorize their experiences as threatening or safe, the role of *knowledge* in the context of privacy is ambiguous. For example,

Bonneau and Preibusch [3] develop a game-theoretical model for the privacy communication game on SNSs. In order to increase self-disclosure on SN platforms they recommend minimizing privacy priming for non-fundamentalists by hiding the privacy policy into the backrooms of SN websites. These privacy policies should, however, be fair enough to ensure the needs of privacy fundamentalists are addressed and their complaints prevented. Even though this approach partly reflects the current behaviour of many SNS providers, the long-term impact of such strategy is hard to predict. If a SNS provider avoids directly informing users about its information-handling practices, the media will do so, once the network becomes popular. This negative publicity may lead to undesirable 'halo' and 'sleeper' effects, under which a single negative piece of information (often from an untrustworthy source) spills over to damage the whole image of the provider. Lacking factual knowledge about privacy practices on their SNS, users may attribute unjustified level of risk to their self-disclosures – a highly undesirable development for any SNS provider. In support of our argument, Krasnova et al. [23] show a positive impact of *awareness* of enhancing *trust in SNS provider*.

Even though 'notice' constitutes a basic element of fair information practices important for both Germany and the USA, and, hence, is likely to mitigate *privacy concerns*, enhance *trust* and ensure *enjoyment* in both countries, its impact is likely to be contingent on culture. Indeed, Doney et al. [9] argue that culture plays a significant role in how individuals process information and integrate it into their decision-making process.

Overall, information-based cues may facilitate the *trust*-building process for both individualists and collectivists. However, whereas individualistic cultures may feel more confident in assessing the cons and pros of the provider's defection behaviour (calculative-based thinking), collectivists are likely to have a hard time assessing the predictability and benevolence of the provider on the basis of available facts [28]. Even though this argument speaks for a slightly higher importance of knowledge for US subjects, studies on the impact of *interactional justice* - reflective for the transparency and communication style of the trustee - on *trust* reveal no significant differences between individualistic and collectivistic countries [33]. Moreover, high UAI inherent in German culture may also intensify the value attached to *knowledge* when forming *trusting beliefs*. By and large, we expect no major differences in the link between *knowledge* and *trusting beliefs* for US and German subjects.

Similar to *perceived control*, we expect *knowledge* to play a more salient role in mitigating *privacy concerns* in such highly risk-averse society as Germany. Indeed, by getting informed about information-handling procedures, risk-averse SNS users may feel at least passively in control of their information [31] and hence perceive much less risk when communicating on the platform.

Finally, Dinev et al. [7] argue that while *collectivistic* cultures, like Germany, are more careful in forming their attitudes, people from *individualistic* societies, like USA, feel empowered with *knowledge* and, hence, form their perceptions (e.g. regarding *enjoyment*) more readily. On the other hand, high risk-averseness in Germany is likely to level up this effect. By and large, we expect no major differences in the link between *knowledge* and *enjoyment* perceptions of the US and German respondents. Taken together, we hypothesize that:

Hypothesis 4a: *The positive relationship between Knowledge and Enjoyment* ***will not differ significantly*** *for German than for American SNS users.* **Hypothesis 4b:** *The negative relationship between Knowledge and Privacy Concerns* ***will be stronger*** *for German than for American SNS users.* ***Hypothesis 4c:*** *The positive relationship between Knowledge and Trust in SNS provider* ***will not differ significantly*** *for American and German SNS users*.

5. EMPIRICAL STUDY

5.1 Survey development, design and sampling

Participants from Germany and the USA were recruited by posting announcements on university mailing lists, campus bulletin boards and on FB groups throughout Fall 2008 and Winter 2009. German and American respondents were offered a reward of EUR5 or $5 respectively. A total of 138 German subjects and 193 American subjects took part in the survey. 40.6% / 65.3% of the German / US sample were female and 57.2% / 34.2% were male. 85.5%/ 42% of the participants in Germany/ USA were between 20 and 29 years old. In the USA 51.3% of the participants were between 18 and 20 years of age. Overall, both samples were dominated by students – an important group of FB audience. Recognizing some demographic differences, we consider both samples to be comparable.

A questionnaire was initially developed in English and then carefully translated into German. English and German versions of the survey were offered to German residents. Validity of the translation was ensured as described in [25]. Each construct was modeled as reflective and measured on a 7-point Likert scale (unless specified otherwise). We relied on the pre-tested scales where possible. Scales for *Self-Disclosure (SD)*, *Legal Assurance (LA)* and *Enjoyment (EN)* are partly presented in [22] as well as [25] and included 6, 3 and 2 items respectively. 5 items for *Trust in SNS Provider (Tr)* were adapted from McKnight et al. [32]. 6 items for *Privacy Concerns (PC)* and 3 items for *Perceived Control (PCtrl)* are presented in Krasnova et al. [23]. Items for *Knowledge (KN)* were self-developed as shown in Table 1. Overall, the paper by Krasnova and Veltri [25] provides a good overview of the scales used in the study.

Table 1. Construct Operationalization

KN	1. I am well-informed about FB privacy policies; I know exactly: 2. ... how the information I provide on FB is allowed to be used by other users or companies; 3. ... how the information I provide on FB can and cannot be used by FB; 4. ...how existing laws regulate the use of my information on FB.

5.2 Evaluation of the Model

Our model has been tested using the Partial Least Squares methodology. The reasons for the choice of this approach was the non-normality of our data as well as a limited size of the German sample (less than 200 observations), as typically required by covariance-based methods [5]. Models for both countries were estimated separately using SmartPLS 2.0.M3 [42]. Measurement Model (MM) was evaluated in the first step. Parameters for

Indicator Reliability, Composite Reliability (CR) and Average Variance Extracted (AVE) were assessed to ensure Convergent Validity. Only 2 items in the German sample had loadings of 0.67 and 0.69, with all other items in both samples exceeding 0.7 threshold [16]. The CR values for all constructs in both models exceeded the required level of 0.7 [11]. The AVE values for all measured constructs by far surpassed the threshold level of 0.5 [41]. Finally, Cronbach's Alpha (CA) – a measure of Internal Consistency – was higher than a threshold of 0.7 for all constructs in both models [16]. Taken together, all criteria for Convergent were met. In the next step, Discriminant Validity was assessed, by ensuring that the square root of AVE for each construct was higher than the correlation between this construct and any other construct in a model [11]. This requirement was fulfilled for all constructs in both models (all results are available upon request). Summarizing, the MMs for both countries were well-specified. In the next step, the Structural Model (SM) was evaluated. We find that our 'PC' variables explain 24.0% and 15.8% of variance in *Self-disclosure* in the USA and Germany respectively. As we aimed to integrate only practice-relevant factors into our model - possibly omitting such influential variables as expected benefits of relationship maintenance or self-presentation - this level of explanatory power is adequate.

Table 2. Standardized path coefficients and p-values for MGA

Hyp.	A → B	Path Coefficient		p-value for MGA
		GER	USA	USA/GER
H 1a	**EN → SD**	**0.139*** **	**0.268*** **	**0.176**
H 1b	**PC → SD**	**-0.181** **	**0.032**	**0.065**
H 1c	**Tr → SD**	**0.070**	**0.329*** **	**0.026**
H 2a	LA → EN	0.101	-0.009	0.184
H 2b	**LA→ PC**	**-0.245*** **	**-0.175**	**0.353[1]**
H 2c	LA → Tr	0.216***	0.299***	0.244
H 3a	PCtrl → EN	0.022	0.348***	0.004
H 3b	**PCtrl→ PC**	**-0.143*** **	**0.056**	**0.063**
H 3c	**PCtrl → Tr**	**0.140*** **	**0.207*** **	**0.270**
H 4a	**KN → EN**	**-0.206**	**0.046**	**0.025[2]**
H 4b	KN → PC	-0.021	0.066	0.294
H 4c	KN → Tr	-0.079	0.113*	0.059

*Significance: * at 10%; ** at 5%; *** at 1% or lower*

Next, path coefficients were evaluated based on PLS algorithm. In line with the accepted practice, significance of path coefficients was determined via a bootstrapping procedure by setting the number of cases equal to sample size and the number of bootstrap repetitions to 200. Results are summarized in Table 2. Finally, Multi-Group Analysis (MGA) comparing path coefficients across two models was conducted. We note that MGA with non-normal data in PLS is still 'terra incognita' for most researchers in the field. As a temporary solution, some relied on a parametric approaches, disregarding distributional characteristics of their data (e.g. [18]). In a recent study, Henseler et al. [13] propose a PLS-MGA procedure, which is free of distributional assumptions. The accompanying spreadsheet implementation of their solution is, however, limited to only 100 bootstrap repetitions. As we intended to use 200 bootstrap repetitions, a testing procedure in GNU R was implemented, which builds on the spreadsheet formula but helps to overcome existing limitations [46]. P-values obtained via our PLS-MGA implementation of Henseler et al. [13] approach are presented in Table 2. The supported hypotheses are selected **in bold** in Table 2. Considering that our study is exploratory in nature, a significance level of 10% was considered acceptable. Finally, for the ease of comprehension, our results for both countries are also sketched in Figures 2 and 3: paths selected **in bold** are significant at least on 5%-level; a path selected **in dashed bold** is significant on 10%-level.

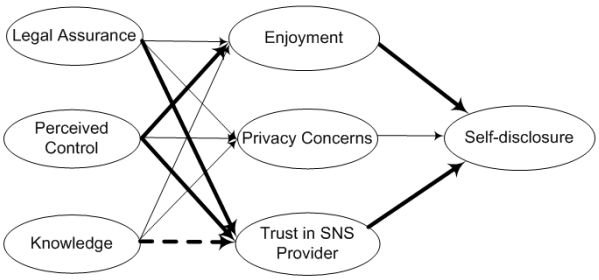


Figure 2. Results of the Structural Model for the USA.

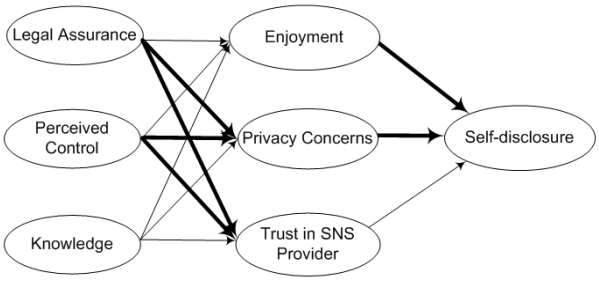


Figure 3. Results of the Structural Model for Germany.

6. THEORETICAL FINDINGS

Our results deliver important *theoretical* insights. First, we find that 'PC' of US and German subjects is characterized by distinct cognitive patterns. On the one hand, both cultures appear to be equally motivated by *enjoyment* in their decision to self-disclose (H1a supported). On the other hand, whereas Germans rather adjust their self-disclosure on the basis of *privacy concerns* (H1b supported), Americans base their decisions on their perceptions of the *trust in SNS Provider* (H1c supported). It appears that privacy-relevant mechanisms are helpless in ensuring *Enjoyment* in German culture. At the same time, and contrary to our expectations, a feeling of being in *control* emerges as a powerful booster for *enjoyment* for the individualistic American counterparts (H3a rejected). Even though users from both cultures report experiencing *enjoyment* from communicating with their

[1] Even though the MGA p-value is insignificant, Legal Assurance does not exert a significant impact on Privacy Concerns in the USA. We therefore consider H2b to be supported

[2] Even though the difference between path coefficients is significant, Knowledge does not exert a significant impact on Enjoyment in both USA and Germany. We therefore consider H4a to be supported.

peers, more *individualistic* American users may be increasingly apprehensive about the opportunistic behaviour of other users. As a result, they are likely to find their SNS experience as more pleasing once control means are in place. We find remarkable differences in the formation of *privacy concerns*. None of the factors we tested appears to exert an impact on the magnitude of *privacy concerns* in the USA. Insignificance of *legal assurances* may be explained by the autonomous character of individualistic US culture, which speaks against strong reliance on the government. Germans, on the other hand, are more collectivistic and risk-averse, which may explain their preference for legalistic remedies (H2b supported). Furthermore, as in the case with *enjoyment*, insignificance of *legal assurances* in mitigating *privacy concerns* may be explained by the relative unimportance of institutional privacy as opposed to social privacy for US subjects [39]. A closer look at the answers to "*privacy concerns"* items reveals that US subjects are particularly fearful of losing control over their data online, and less so when it comes to "behind-closed-doors" commercial processing. Whereas *legal assurances* may prevent abuse of personal data by corporations and other legal entities, they are ineffective in preventing the bullying, ridicule or secret sharing in a social environment. The same logic may explain the insignificance of *knowledge* about practices of SNS Provider in mitigating *concerns* over social privacy for US subjects. This, however, does not explain insignificance of *knowledge* in mitigating *privacy concerns* of German respondents, who express high anxiety over commercial use of their information (H4b rejected). One possible explanation may be that as Germans are more distrustful of companies, they may not internalize privacy-related claims of geographically distant FB. While Germans view active *control* as a salient determinant of their *privacy concerns*, we find it not relevant whatsoever for American subjects (H3b supported). On the one hand, this complete detachment may signal the irrational nature of privacy perceptions of US users. Furthermore, American users may pessimistically (or realistically?!) assess the level of protection offered by even most refined controls. They may assume that no matter how hard they protect their information, it can still be spread if "friends" choose to copy it into a malicious email. This logic is plausible, as *individualism* allows for self-serving behaviour of others [9]. With an average user having 130 FB friends ([10], this arguments may indeed be reflective of the true state of things. In both countries *control* perceptions and *legal assurance* were equally important in determining *trust in SNS provider* (H3c supported; H2c rejected). This result rejects our hypothesis (H2c) about the relative unimportance of legal means in ensuring *trust* for American vs. German subjects. It is possible that as the legal framework determines the negative side of the deviant behaviour, American users are likely to integrate it when forming *trust* on a calculative basis. Finally we find *knowledge* to be insignificant for the formation of *trusting beliefs* for German, and only weakly significant (at 10%-level) for American subjects (H4c rejected). As mentioned above, people from individualistic cultures are more proficient in aligning their behaviour with available cues. At the same time, socially and geographically-distant German users may find it difficult to collect evidence relevant for the development of prediction-based trust, which is typical for collectivistic cultures [9]. Taken together, results of our study provide a unique theoretical insight into the cross-cultural generalizability of the 'PC' theory. Distinct differences identified in our study signal that many privacy-mechanisms are culturally-determined and, hence, SNS providers cannot rely on the success of the "proven" means in ensuring positive outcomes when crossing boundaries, as discussed in the following section.

7. MANAGERIAL IMPLICATIONS

Our results provide a starting point for SNS providers who are looking for *practical* recommendations on their path to internationalization. We find that stimulating *enjoyable* experiences appears to be an important internationalization strategy. Recognizing that *socialization* on SNSs typically translates into *enjoyment*, one possibility is to open up as many channels of active communication and passive following as possible. Even though FB probably had exactly the same idea when introducing News Feed, may users find it useless and boring due to ineffective information filtering [20]. Hence, there is a pressing need to improve the criteria for the selection of the social content. Even though *trust in SNS provider* emerges as a relevant determinant of *self-disclosure* for US users, a quick look at the responses at the item level shows that US (as well as German) subjects are at best *slightly positive* about the trustworthiness of FB. Boyd and Hargittai [4] argue that a lot of this distrust comes from a heated media-driven discussion of inadequate approach of FB in managing user privacy. Amidst these debates, cross-cultural *legal assurances* may come as a much-needed help, as individualistic cultures are likely to rely on them when determining provider's losses in case of a broken trust – a backbone of calculative trust-building process [9]. Hence, by relating information misuse to financial repercussions (e.g. monetary fines in case of a litigation), policy-makers can make the process of calculative thinking more concrete. Finally, user *control* and involvement into privacy-relevant decisions appears to be an indispensible part of the trust-building effort in both Germany and the US. Indeed, in an individualistic culture like the US even rank-and-file members expect to be informed, asked and involved when relevant decisions are made. Our data shows that *self-disclosure* decisions of German users are contingent on the magnitude of *privacy concerns* they perceive. Again, providers may mitigate these negative perceptions by giving users more *control* over their information as well as, paradoxically, by supporting *legal enforcement* of fair information-handling practices. Other measures, outside of our model, may include feedback mechanisms when privacy-relevant decisions are being planned. As FB privacy record demonstrates: When user involvement in privacy-relevant decisions, even good ones, is missing, the consequences are overblown beliefs of privacy threats [44]. Altogether our results show that if financial means of SNS providers are limited, they should in the first place direct their investments in enhancing platform *enjoyment*, granting users with more *control* and, paradoxically, lobbying for more *legalistic safeguards* of user privacy. Although effects of *control* differ from country to country, our study shows that *control* perceptions influence "disclosure-relevant" constructs in both countries: *privacy concerns* in Germany as well as *trusting beliefs* and *enjoyment* in the USA. We note that even though increasing user *knowledge* regarding privacy issues maybe a good idea (significant at 10% level in the USA), it should not be the first priority of SNS providers when expanding internationally. On the policy side, our study reveals a paramount role of *global* institutional assurances in enhancing platform *trust* in both

countries, as well as in mitigating *privacy concerns* in Germany. So far limited, more regulation is likely to follow shortly as announced by the EU Justice Commissioner in January 2010 [37]. We argue that SNS providers should welcome these efforts as they are likely to motivate communication on their sites. Taken together, our results call for greater involvement of policy-makers in safeguarding user privacy. By taking a more active stance, politicians could help retain the unique potential of SNSs to create and maintain social capital.

8. CONCLUSION

The aim of this paper was to investigate intercultural dynamics behind 'PC' on SNSs. Using samples from Germany and the USA, major culturally-determined differences were discovered. Our results provide evidence that while some elements, such as *enjoyment*, are equally important across cultures, relevance of other constructs varies from country to country. For example, while Americans base their self-disclosure decisions on the basis of *trust*, Germans are driven by *privacy concerns*. Furthermore, even though *control* and *legal assurances* are important in both countries, the mechanisms of their integration into individual privacy decisions are distinct. These differences signal that SNS provider should adopt more flexible strategies when expanding internationally. From the policy perspective, our findings support legal intervention into privacy regulation on SNSs, as besides protecting users these measures are likely to help SNS providers in supporting sustainability of their networks. Our study suffers from several limitations, which, however, offer exciting venues for future research. First, both American and German samples were dominated by students. Even though student samples are acceptable when the research question is "universalistic" in nature and involves general psychological constructs [26], we encourage validation of our findings on the basis of more representative samples. Second, by adopting a 'PC' lens we assume that SNS users have a stable preference for privacy. John et al. [17], however, argue that privacy preferences are context-dependent and a behavioural perspective would be more appropriate when studying privacy-related decisions. Hence, future researchers may extend our study by accounting for these cognitive limitations. Finally, as our study was exploratory in nature, it was expected that while many of the culture-relevant hypotheses would get supported, some would not. In this respect, our analysis provides evidence for the complexity of cultural impact on privacy-relevant perceptions and behaviour.

9. REFERENCES

[1] Acquisti, A. and Gross, R. 2006. Imagined Communities: Awareness, Information Sharing and Privacy on The Facebook. In *Proceedings of 6th Workshop on PET*s (Cambridge, UK, June 28 - 30, 2006).

[2] Beretsky, S. 2009. Woman Loses Sick-Leave Benefits for Depression Thanks to Facebook Pics. http://psychcentral.com/blog/archives/2009/11/20/woman-loses-sick-leave-benefits-for-depression-thanks-to-facebook-pics, last accessed on August 10, 2010.

[3] Bonneau, J. and Preibusch, S. 2009. The privacy jungle: On the market for data protection in social networks. In *Proceedings of The Eighth WEIS* (London, UK, June 24-25, 2009).

[4] Boyd, d. and Hargittai, E. 2010. Facebook Privacy Settings: Who Cares? First Monday, 15, 8.

[5] Byrne, B. M. 2001. *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*. Lawrence Erlbaum Associates, USA.

[6] Dinev, T., Bellotto, M., Hart, P., Russo, V., Serra, I. and Colautti, C. 2006. Privacy Calculus Model in e-Commerce - a Study of Italy and the United States. *EUR J INFORM SYST*, 15 (1 August 2006), 389–402. DOI=10.1057/palgrave.ejis.3000590, 389-402.

[7] Dinev, T., Goo, J., Hu, Q. and Nam, K. 2009. User behaviour towards protective information technologies: the role of national cultural differences. *Information Systems Journal*, 19, 4 (July 2009), 391-412. DOI= 10.1111/j.1365-2575.2007.00289.x.

[8] Dinev, T. and Hart, P. 2006. An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research*, 17, 1 (March 2006), 61-80. DOI= 10.1287/isre.1060.0080

[9] Doney, P. M., Cannon, J. P. and Mullen, M. R. 1998. Understanding the Influence of National Culture on the Development of Trust. *Academy of Management Review*, 23, 3, 601-620.

[10] Facebook. 2010. Press Center: Facebook Statistics. http://www.facebook.com/press/info.php?statistics, last accessed on August 15, 2010.

[11] Fornell, C. and Larcker, D. F. 1981. Structural Equation Models with Unobservable Variables and Measurement Errors. *Journal of Marketing Research*, 18, 1, 39-50. DOI= 10.2307/3151312.

[12] Gefen, D., Rao, V. S. and Tractinsky, N. 2003. The Conceptualization of Trust, Risk, and their Relationship, Electronic Commerce: The Need for Clarifications. In t*Proceedings of HICSS* (Hawaii, USA, January 6-9, 2003).

[13] Henseler, J., Ringle, C. M. and Sinkovics, R. R. 2009. The use of partial least squares path modeling in international marketing. In *Advances in International Marketing (AIM)*, R. R. Sinkovics and P. N. Ghauri, Eds. Bingley, 277-320.

[14] Hofstede, G. 1991. *Cultures and Organizations: Software of the Mind*. McGraw-Hill, New York.

[15] Hofstede, G. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. Sage Publications, Thousand Oaks.

[16] Hulland, J. 1999. Use of partial least-squares (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal*, 20, 2 (February 1999), 195-204. DOI= 10.1002/(SICI)1097-0266(199902)20:2<195::AID-SMJ13>3.0.CO;2-7.

[17] John, L. K., Acquisti, A. and Loewenstein, G. F. 2009. The Best of Strangers: Context Dependent Willingness to Divulge Personal Information. (July 6, 2009). Available at SSRN: http://ssrn.com/abstract=1430482.

[18] Keil, M., Tan, B. C. Y., Wei, K.-K., Saarinen, T., Tuunainen, V. and Wassenaar, A. 2000. A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. *MIS Quarterly*, 24, 2 (June 2000), 299-325 .

[19] Konovsky, M. 2000. Understanding Procedural Justice and Its Impact on Business Organizations. *Journal of Management*, 26, 3, 489-511. DOI= 10.1177/014920630002600306.

[20] Koroleva, K., Krasnova, H. and Guenther, O. 2010. Stop Spamming Me!' - Exploring Information Overload on Facebook. In *Proceedings of AMCIS* (Lima, Peru, August 13-15, 2010).

[21] Krasnova, H., Hildebrand, T. and Günther, O. 2009. Investigating the Value of Privacy on Online Social Networks: Conjoint Analysis. In *Proceedings of ICIS* (Phoenix, Arizona, USA, December 15-18, 2009).

[22] Krasnova, H., Kolesnikova, E. and Günther, O. 2009. It Won't Happen To Me!": Self-Disclosure in Online Social Networks. In *Proceedings of 15th AMCIS* (San Francisco, USA, August 6-9, 2009).

[23] Krasnova, H., Kolesnikova, E. and Günther, O. 2010. Leveraging Trust and Privacy Concerns in Online Social Networks: an Empirical Study. In *Proceedings of 18th ECIS* (Pretoria, South Africa June 7-9, 2010).

[24] Krasnova, H., Spiekermann, S., Koroleva, K. and Hildebrand, T. 2010. Online social networks: why we disclose. *J INFORM TECHNOL*, 25, 2 (June 2010), 109-125.

[25] Krasnova, H. and Veltri, N. F. 2010. Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA. In *Proceedings of HICSS* (Koloa, HI, January 5-8, 2010).

[26] Kruglanski, A. W. 1975. The human subject in the psychology experiment: Fact and artifact. In *Advances in experimental social psychology*, L. Berkowitz, Ed. Academic Press, New York, 101-147.

[27] Leung, K. and Lind, E. A. 1986. Procedure and culture: Effects of culture, gender, and investigator status on procedural preferences. *Journal of Personality and Social Psychology*, 50, 1134-1140.

[28] Lim, K. H., Leung, K., Sia, C. L. and Lee, M. K. O. 2004. Is eCommerce Boundary-Less? Effects of Individualism-Collectivism and Uncertainty Avoidance on Internet Shopping. *Journal of International Business Studies*, 35, 6, 545-559. DOI:10.1057/palgrave.jibs.8400104.

[29] Lind, E. A. and Earley, P. C. 1992. Procedural Justice and Culture. International Journal of Psychology, 27, 2, 227-242.

[30] Luo, X. 2002. Trust production and privacy concerns on the Internet A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management*, 31, 2, 111--118. DOI:10.1016/S0019-8501(01)00182-1.

[31] Malhotra, N. K., Kim, S. S. and Agarwal, J. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15, 4, 336-355. DOI=10.1287/isre.1040.0032.

[32] McKnight, D. H., Choudhury, V. and Kacmar, C. 2002. The Impact of Initial Consumer Trust on Intentions to Transact with a Web site: a Trust Building Model. *J STRATEGIC INF SYST*, 11 (December 2002), 297 - 323. DOI=10.1016/S0963-8687(02)00020-3.

[33] Morris, M. W. and Leung, K. 2000. Justice for All? Progress in Research on Cultural Variation in the Psychology of Distributive and Procedural Justice. *Applied Psychology-An International Review*, 49, 1, 100-132.

[34] Muniz, A. and O'Guinn, T. 2001. *Brand Community*. Journal of Consumer Research 27, 412-432.

[35] Park, H. 1993. Cultural impact on life insurance penetration: A cross-national analysis. *International Journal of Management,* 10, 3, 342-350.

[36] Pavlou, P. A. and Gefen, D. 2004. Building Effective Online Marketplaces with Institution-Based Trust. *Information Systems Research*, 15, 1, 37-59.

[37] Phillips, L. 2010. New EU Privacy Laws Could Hit Facebook. Bloomberg Businessweek, January 29, 2010. http://www.businessweek.com/globalbiz/content/jan2010/gb20100129_437053.htm, last accessed on August 22, 2010.

[38] Pillai, R. and Williams, E. S. 1999. Are the Scales Tipped in Favor of Procedural or Distributive Justice? An Investigation of the U.S., India and Germany. In *Proceedings of National Academy of Management Meeting* (Chicago, IL).

[39] Raynes-Goldie, K. 2010. Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday*, 15, 1-4 (January 2010).

[40] Reiss, S. 2004. Multifaceted Nature of Intrinsic Motivation: The theory of 16 basic desires. Review of General Psychology, 8, 3, 179-193.

[41] Ringle, C. M. 2004. *Gütemaße für den Partial Least Squares-Ansatz zur Bestimmung von Kausalmodellen*. Working paper 16, University of Hamburg.

[42] Ringle, C. M., Wende, S. and Will, A. 2005. SmartPLS, Release 2.0.M3. University of Hamburg, Hamburg, Germany. http://www.smartpls.de, last accessed on September 22 2009.

[43] Rosen, P. and Sherman, P. 2006. Hedonic Information Systems: Acceptance of social networking websites. In *Proceedings of AMCIS* (Acapulco, August 4-6. 2006).

[44] Sinclair, C. 2010. Students: panic over online privacy, identity is overblown. http://dmlcentral.net/blog/chris-sinclair/students-panic-over-online-privacy-identity-overblown, last accessed on August 14, 2010 .

[45] Son, J. Y. and Kim, S. S. 2008. Internet Users' Information Privacy-Protective Responses: A Taxonomy and a Nomological Model. *MIS Quarterly*, 32, 3, 503-529.

[46] Theel, C. 2010. *Analysis of Social Network Site Usage Behavior The Role of Gender: An Application of Partial Least Squares Path Modeling Multi Group Analysis*. Diploma Thesis, Humboldt-Universität zu Berlin.

[47] Triandis, H. C. and Suh, E. M. 2002. Cultural influences on personality. *Annual Review of Psychology*, 53, 1, 133-160. DOI=10.1146/annurev.psych.53.100901.135200.

[48] Tufekci, Z. 2008. Can you see me now? Audience and disclosure regulation in Online Social Network Sites. Bulletin of Science, *Technology & Society*, 28, 1, 20-36.

Technology Adoption by Elderly People – An Empirical Analysis of Adopters and Non-Adopters of Social Networking Sites

Christian Maier
Centre of Human Resources Information Systems
Otto-Friedrich University Bamberg
christian.maier@uni-bamberg.de

Sven Laumer
Centre of Human Resources Information Systems
Otto-Friedrich University Bamberg
sven.laumer@uni-bamberg.de

Andreas Eckhardt
Centre of Human Resources Information Systems
Goethe-University Frankfurt am Main
eckhardt@wiwi.uni-frankfurt.de

ABSTRACT

This research paper analyzes the impact of attitudinal, control and normative beliefs on the intention to use social network sites (SNS) by people older than 50. Using the *Model of Adoption of Technology in Households* (MATH) and the data of 115 social network site adopters and 53 non-adopters it can be shown that the intention of adopters and non-adopters has been influenced by different reasons. Perceived Ease of Use and Normative Beliefs have only a significant impact for adopters. Moreover, this research paper unfolds Fear of Technology as a strong influence factor for non-adopters in regard not to use SNS in their daily routine. The paper concludes with a discussion of an age-sensitive design of SNS in order to address the digital divide.

Keywords

Adoption, Non-Adoption, MATH, Elderly People, Social Network Sites

1. INTRODUCTION

Due to new information and communication technologies, organizations can simplify the work of their employees, which is the largely overlooked perspective in IS research [21]. In addition households could integrate these technical innovations within their daily routine to handle ordinary or uncommon tasks within short periods of time. One essential renewal in the last years was the introduction of Social Network Sites (SNS), which can be defined as "*online shared interactive spaces, in which a group of people use a repertoire of technological features (forums, newsgroups, messaging) to carry out a wide range of social interaction*" ([42]; [44]). Actual, a lot of different SNS compete to be the market leader, however, at the moment Facebook [30], with more than 400 million active users, is the most used SNS around the world. On the other side, certain countries as Germany [66] have other online communities with a similar high number of users. In Germany, over 30 Million people are members of social communities on the internet [10].

These users can inform all their friends and acquaintances with just one message, communicate or chat to maintain social relationships. Apart from that, many people use SNS to share private information like photos or videos or try to enlarge their circle of friends. Others just pursue the aim to collaborate or to have fun while playing online games and compete with friends ([24]; [51]; [62]; [65]; [67]; [72]). Additionally further SNS (such as Xing or LinkedIn) support the application process of job seekers by providing the possibility to upload CVs, connect with their job network or communicate with recruiters and headhunters for job offers [74]. In Germany the three most important reasons to participate in a SNS are to stay in contact with family and friends, to exchange information about common interests and to search for new friends [10]. Nonetheless, these potentials of SNS can only be realized if people participate within the same social network.

Although modern information technology offers various advantages and is used by many people – often daily ([43], [65], [72]) – the amount of people that are not willing to use and adopt to SNS is surprising ([41]; [57]; [58]; [75]). In Germany there are around 50 Million people who do not have a profile in a SNS. This accounts for almost two thirds of the people living in Germany[1]. Such a non-adoption behavior of IT in general has been recognized within IS research and potential reasons were raised and identified concerning different applications. Different reasons have been identified in previous research such as fear and threats as concern for privacy ([7]; [8]), psychological issues like resistance [49]; or simply social issues as age, education or income ([4]; [25]). In addition Peter Mertens analyzed why IT implementation projects fail [56].

Nonetheless, if people reject using new technologies or applications as social network platforms non-adoption will entail various problems. From a societal point of view the most important one is the advancing spread of society in a group of people adopting new technologies and one rejecting it. This phenomenon is actually discussed and known as Digital Divide or

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Digital Inequality [25]. It describes the amount of people, who have limited access to the internet or do not have the ability to use computers effectively and efficiently. Major reasons for the Digital Divide in Germany are the lacking availability of broadband internet access points [37] and especially demographic factors such as level of education, gender and age [46].

Concerning the factor age for use and acceptance of the internet there are distinct differences in the German population. More than 90 per cent of the young people between 14-and 29 years are internet users. In contrast, only 48.5 per cent of people between 60-69 years and just 19 per cent of the people 70+ years of age are internet users [1]. A large proportion of these people do not adopt the internet and its applications. Reasons for this non-adoption lie in the rapid development of the internet in the past 20 years and the related dissemination of information and communication technology. For example, people who retired around the millennium did mostly not come in contact with new media during their working career [73].

In order to counter the phenomenon of Digital Divide, the German Federal Government introduced several initiatives to reduce resistance and foster internet use of elderly people. SNS exist that target specifically the elderly population. In Germany feierabend.de is one example of a platform designed to support the social interaction of people aged 50 and older. This specific platform was awarded in 2008 as the "Best Community" in Germany by the German Federal Department of Economics and Technology as the platform supports especially the generation 50+ to find their way into and through the World Wide Web. The platform is designed to support the exchange of information and experiences as well as interactions of people with similar interests. For example feierabend.de established over 100 regional groups to enable meetings of their members within their city or region. These regional groups enabled feierabend.de to connect the online and offline lifes of elderly people. However, with only 600,000 visitors each month this particular SNS as well as other similar platforms only reach a small part of the potential user group of people at the age of 50 and older.

Based on the previous analysis, this paper focuses on people with at least 50 years of age and leaves out the "*wired from birth"* [12] generation. By using the *Model of Adoption of Technology in Households* (MATH; [70]; [13]) this paper will analyze factors leading to adoption or non-adoption behavior of SNS by elderly persons. This research is in line with Brown who argued that adoption research in the household context should focus on SNS [12] and with Pak et al. (2009) who identified age-sensitive design of online services as an important aspect of IS research [60].

Therefore this paper analyzes which factors of MATH have an influence on the decision to adopt a modern technology as SNS. Apart from that it is investigated, which MATH construct has the strongest predictive value and if there are differences for adopters and non-adopters in relation to the observed antecedents of the intention to use SNS.

For this purpose, this paper provides an overview of the research background and relevant literature in Section 2 dealing with the Digital Divide in general, SNS as well as IT adoption and non-adoption in the household context. Based on this, Section 3 contains the central hypotheses and explains the used research design. Section 4 comprises the research results which are then discussed in Section 5.

2. Research Background

Within in this section the Digital Divide in general, SNS and IT adoption and non-adoption in the household context are discussed in order to provide the relevant background information for the developed research model.

2.1 IT Adoption and Non-Adoption in Households

IT adoption in general is a highly studied research area within the IS discipline. According to Williams et al. [76], since 1985 345 paper on technology adoption were published in the top 19 peer-reviewed journals of the IS community. Nonetheless, most of these articles analyzed IT adoption in organizations. In principle, IT-adoption and non-adoption can be investigated within organizational [71] and private contexts ([13]; [12]). In order to analyze the private domain, Venkatesh and Brown processed the *Model of Adoption of Technology in Households* (MATH; [70]; [13]), which is based on the *Theory of Planned Behavior* [2] and explains the Behavioral Intention with the help of Attitudinal Beliefs, Normative Beliefs and Control Beliefs. Attitudinal Beliefs subsume Utilitarian Outcomes (degree of effectiveness and utility of using PC within households), Hedonic Outcomes (degree of pleasure or fun) and Social Outcomes (degree of status, power or knowledge resulting from PC household adoption). Normative Beliefs consider the impact of friends, family members and acquaintances and Control Beliefs regard possible inhibitors as cost, difficulty of use or Fear of Technology, which can end in rejecting a new technology. The resulting model was enlarged in 2005 as Brown and Venkatesh [13] identified Age, Income and Marital Status as moderator effects.

In terms of age Brown and Venkatesh showed that in general age is a moderator for Utilitarian, Hedonic and Social Outcomes as well as for Normative and Control Beliefs. The relationship between Utilitarian Outcomes and Behavioral Intention is moderated in such way that it is increasingly significant with age and even more for those who are married. The relationship between Hedonic Outcomes and Intention is moderated by age such that with increasing age Hedonic Outcomes are less important. In terms of Social Outcomes the impact of status gains on Intention to Use increase with age. Also Normative Beliefs are moderated by age such that friends and family as well as secondary sources are more important for elderly people. In terms of Control Beliefs (Fear of Technology and Perceived Behavioral Control or Perceived Ease of Use) a moderation effect by age were identified. Consequently, these antecedents are more important for older people.

Another distinguishing criterion within IT adoption research is the motivation why people use IT. Generally, people can use it because of a voluntary incentive or due to mandatory settings. Social network sites, which are the underlying technology within this paper, are a good research domain to analyze adoption behavior in households [12]. Within such a setting many people – especially elderly people – reject using new technologies because they are not in a position to handle technologies and are not willing to ask for help if something did not work as planned [11]. This could be one reason, why the diffusion of broadband in

households moves slower than expected ([22]; [26]). Based on this observation, Choudrie and Dwivedi [21] investigated the adoption of broadband in households with the help of MATH. According to Venkatesh and Brown [70] they identified several barriers as high costs, ease or difficulty of PC and internet use, lack of skill and lack of needs, which could result in a rejection of new technologies. For non-users only the lack of knowledge played a subordinated role in order to understand non-adoption behavior. On the other side, it was possible to show that each attitudinal factor was important to predict the usage behavior.

The complex theme "non-adoption" has not yet been researched as extensive as the actual adoption decision [47]. Nonetheless, several IS researchers started to investigate this behavior ([36]; [54]) and tried to motivate for further research endeavor. A recent publication within MISQ identified perceived values, switching costs or support as factors which can tip the balance and lead to non-adoption [45]. Such factors differ depending on the underlying context, so that other authors identified loss of status or power, uncertainty [40], pressure, exchange [29] or perceived threat ([7]; [9]) as significant influence factors which increase the probability to reject technologies. Eckhardt et al. [27] focused on the other side of social influence and investigated what groups exert an influence on the decision of people to refuse adopting a technology. A research model which explicitly should explain why people do not adopt social network platforms with the help of the *Theory of Reasoned Action* (TRA; [31]; [3]) was conducted by Laumer et al. [49]. In doing this, the authors disclosed negative significant correlations between an individual's personality trait resistance and each TRA construct.

2.2 Social Network Sites

Internet usage and cognition changed due to new opportunities within information and communication technologies. One of the most influential alteration emerged through Social Network Sites (SNS) as Facebook or the VZ-network (meinVZ, StudiVZ, SchülerVZ), which are famous SNS for German students and pupils**.**

Nowadays, about 11.44 per cent of the total population of the world is registered within Facebook [30]. Focusing more sophisticated countries as USA, Sweden, Canada or UK, this percentage rate rises up to 40 per cent. Such a high number of users could be explained by the variety of SNS possibilities. Each SNS user can communicate with friends or strangers, maintain relationships, enlarge their circle of friends, share private information, collaborate or just have fun ([24]; [51]; [62]; [65]; [67]; [72]). Due to this, many people integrated social network sites in their daily routine ([43], [65], [72]) and spend there between 10 minutes and 3 hours every day ([28]; [68]; [62]).

Contrary to internet flirtation pages, in which people search new friends and try to meet them afterwards in reality, social network sites are used in most instances to keep in touch with friends and acquaintances, which are known from the real offline world. Only afterwards, these known people will be added in the online friends list. This behavior is called Offline-to-Online phenomenon and is a distinctive characteristic of Facebook and comparable platforms ([28]; [51]; [55]; [44]; [65]).

Regarding different platforms Facebook with around 13 million users in January 2010 is the number one in Germany in terms of total users considering the VZ platforms as different ones. The VZ community has 14.4 million users in total. Also important are wer-kennt-wen.de, stayfriends.de and myspace.com. Feierabend.de has around 600.000 regular visitors as illustrated by Figure 1.

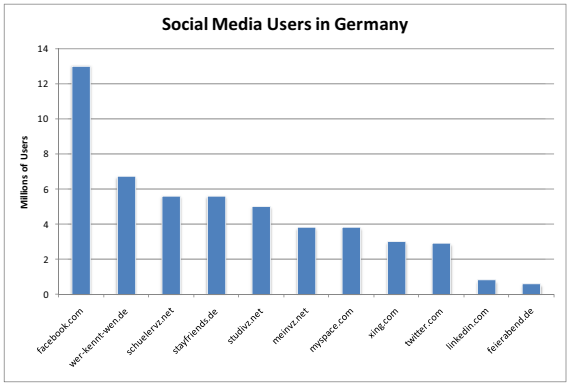


Figure 1: General SNS users in Germany [23]

In Germany, Facebook users are mainly students or young professionals between the age of16 and 28. On the other side, Figure 2 shows that with an increasing age, the user percentage decreases continuously. Consequently, only 5.12 per cent of all German Facebook users are at least 50 years old. Considering the whole German age distribution, which illustrates that the majority of people are older than 40, it is obvious that the percentage rate for elderly Facebook users is very small.

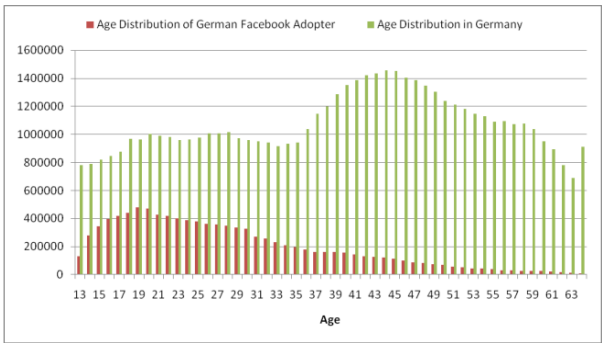


Figure 2: Age distribution of Facebook users

These figures indicates that SNS in general are mostly used by people younger than 30 and that platforms aimed at elderly are used only by a few people in relation to the posible number of users.

2.3 Digital Divide

Contrariwise to persons using SNS, people refusing such technologies, can get social problems through losing social contacts. This is one problematic consequence of the often discussed issue named Digital Divide.

The underlying question of the phenomenon Digital Divide is, whether people have access to internet or not. Afterwards, the scientific focus changed and age, income, rural residence, education, gender or race were studied together with their influence on non-usage of people. Along with it, researchers

investigated not only non-adoption reasons but also differences in people's online skills and thus the ability to find effectively and efficiently information on the web [34]. The latter is often called Second-Level Digital Divide or Digital Inequality and distinguishes self from Digital Divide by focusing not only on the question whether people have access to internet or not. Moreover it focuses on skills and knowledge of people using several technologies such as computers, internet or SNS [69].

Lots of problems, which were discussed through the rise of ICT, as privacy issues, interface issues, a lack of incentives or too complex technologies for most of the households (e.g. [69]) bias elderly in a more serious manner than younger persons. It is not self-evident that each person had contact with modern ICT within their workplace or has friends, acquaintances or family members who can explain them how to handle each new application. Another important facet for elderly persons is their preference to sustain their habitual daily routine and their reluctance to change their way of life. If people had no contact with ICT like computers or social network sites, such a technology or application can change one's life in dramatically way. Because of this, especially the elderly people try to maintain their status quo [45] and burke new innovations. In this context, the extent of an inherit attitude towards changing the status quo has to be regarded as well ([8]; [7]; [49]; [59]).

To overcome this problem in Germany, the Federal Government identified this issue and started initiatives to introduce elderly or inexperienced people to internet possibilities. Next to this, the program of the Federal Government also focuses on population groups with different backgrounds, women in rural areas or internet-interested people and thus takes account of the phenomenon Digital Inequality. Apart from the Federal Government, many other initiatives try to give elderly or unprivileged people an understanding of new and modern ICT. For example, the social network site Feierabend.de tries to address exactly this group of elderly people (50 years or older) and provides them a platform to stay in contact with friends, to enlarge their circle of friends or just to discuss topics which are important and interesting for elderly persons as acoustic hearing apparatus. Apart from that, this SNS throws light on privacy problems and alerts for tricksters and other potential traps.

Based on this general research background of digital inequality and SNS as well as the theoretical background of IT adoption in households the following sections describes the used research model and design to investigate adoption of SNS by elderly people.

3. Research Model and Design

Within this section, our research model will be developed. Based on the *Model of Adoption of Technology in Households* (MATH) ([13]; [70]), the influence of different constructs will be analyzed for adopters and non-adopters of SNS. Finally, the used data sample is provided and the research design will be explained.

3.1 Research Model

The general theoretical foundation for the presented research model is the MATH, which investigates the influence of Attitudinal, Control and Normative Beliefs on Behavioral Intention. With the help of this model, both adopters and non-adopters behavioral intention will be analyzed separately.

For both groups, the six hypotheses as arranged by Brown and Venkatesh ([13]; [70]) will be adapted and analyzed for elderly people. Thereby, the hypotheses are:

H1: Utilitarian Outcomes (Attitudinal Beliefs) has a direct positive influence on Intention of elderly people.

H2: Hedonic Outcomes (Attitudinal Beliefs) has a direct positive influence on Intention of elderly people.

H3: Social Outcomes (Attitudinal Beliefs) has a direct positive influence on Intention of elderly people.

H4: Subjective Norm (Normative Belief) has a direct positive influence on Intention of elderly people.

H5: Perceived Ease of Use (Control Beliefs) has a direct positive influence on Intention of elderly people.

H6: Fear of Technology (Control Beliefs) has a direct negative influence on Intention of elderly people.

The research model is illustrated by Figure 3.

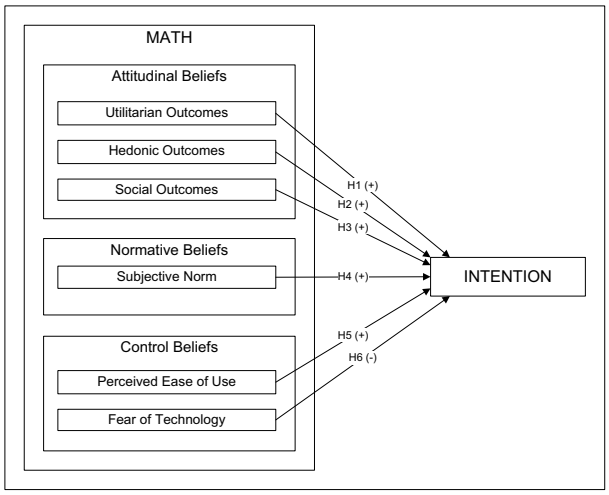


Figure 3: Research Model

3.2 Research Design and Participants

For the evaluation of the research model data of SNS (non)-adoption was collected within a general study of IT usage. The aim of this study was to explain why people do (not) use particular online services even if they have access to the internet. Therefore an online survey was conducted to collect empirical data. In order to reach people who are used to the internet in general and with different social background, demographics and knowledge background, this method seemed to be the most appropriate. Using this methodology and focusing on SNS as well as on elderly people within the study who are used to the internet and have a profile in an SNS or not could be researched.

Based on this data, SPSS Statistics 17.0 and Smart PLS ([63]) were utilized to analyze the influence of the six MATH constructs on intention to use SNS. The evaluation did not include incomplete data samples. As the focus within this paper is to analyze the adoption and non-adoption behavior of elderly people data of 53 SNS non-adopters and 115 SNS adopters, older than 50 years, is the underlying for this research endeavor.

The demographic information, separated by the actual adoption behavior could be seen in Table 1. In both groups are more men than women additionally more participants are between 50 and 54 years old as people older than 55. Nonetheless, the annual income and the whole demographics of both groups are comparable

Table 1: Research participants

Demographics of …		115 Adopters	53 Non-Adopters
Gender	Men	72.2%	75.5%
	Women	27.8%	24.5%
Age	50 - 54	55.5%	54.5%
	55 - 59	32.3%	34.0%
	60 - 64	12.2%	11.5%
Annual Income	< 20 K	25.2%	34.0%
	20 - 25 K	7.8%	11.3%
	25 - 35 K	16.5%	11.3%
	35 - 45 K	16.5%	11.3%
	45 - 55 K	7.8%	9.4%
	55 - 65 K	7.0%	3.8%
	65 - 80 K	4.4%	5.7%
	>= 80 K	14.8%	13.2%

4. Research Results

This section validates the research model for SNS adopters and non-adopters. Therefore, a measurement model and a structural model will be provided within the following two sections as we transferred our research model into a structural equation model and used Partial Least Squares for data analysis.

4.1 Measurement Model

Each of the seven constructs used – Social Outcomes, Hedonic Outcomes, Utilitarian Outcomes, Subjective Norm, Perceived Ease of Use, Fear of Technology and Intention – are measured with reflective indicators as in previous publications. Consequently, content validity, indicator reliability, construct reliability and discriminant validity have to be validated for each construct [5].

4.1.1 Content validity

While setting up the questionnaire, the aim was to refer to questions within the questionnaire, which were already used in empirical research by other researchers. However, following a recent discussion in IS research about the use of Fast Forms for empirical data collection [19] the item identified in prior research were converted to fast forms using semantic differentials and some of these questions had to be modified in order to fit the SNS context. The items used are illustrated at Table 2. These items have been pretested within the general study of IT usage.

Table 2: Measurement items

Item		Question	Semantic Differantials		
Intention	INT-1	I plan … use social network sites in the future.	not to use	…	to use
	INT-2	I intend … social network sites this year.	not to use	…	to use
	INT-3	I will … social network sites anymore.	not use	…	still use
	INT-4	I intend … social network sites w for application processes	not to use	…	to use
SO	SO-1	The usage of social network sites … my image.	decline	…	increase
	SO-2	People, who use social network sites, seem to be … intelligent.	less	…	more
	SO-3	People of my social envirnment, who use social network sites, have a … standing	bad	…	good
HO	HO-1	The usage of a social network site is …	objectionable	…	entertaining
	HO-2	The usage of a social network site is …	waste	…	exciting
	HO-3	The usage of a social network site is …	unpleasant	…	enjoyable
	HO-4	The usage of a social network site is …	boring	…	interesting
UO	UO-1	The usage of social network sites … to achieve my objectives	complicates	…	faciliates
	UO-2	The usage of social network sites makes it … ro achieve my objectives	more difficult	…	easier
	UO-3	All in all, the usage of a social network site is … .	useless	…	usefull
SN	SN-1	People, who have an influence on my behavior think that I should use social network sites.	totally disagree	…	totally agree
	SN-2	People, who are important for me think, that I should use social network sites.	totally disagree	…	totally agree
PEOU	PEOU-1	I t seems to be … to use social network sites.	very difficult	…	very easy
	PEOU-2	For me, it is … to learn how to handle social network sites.	very difficult	…	very easy
	PEOU-3	All in all, it is … to use social network sites.	very difficult	…	very easy
FOT	FOT-1	In my opinion, technologies change … .	very slow	…	very fast
	FOT-2	For me, it is … to acclimaatize to new technologies and standards.	very difficult	…	very easy

4.1.2 Indicator reliability

The proportion of the variance of an indicator, which derives from the relevant latent variables, will be shown by the indicator reliability. Each item should have at least a greater loading than 0.4, so that item SO-2 has to be removed within the non-adopter case [38]. For the rest, each value is greater than 0.7, whereby 50 per cent of the variance of a latent variable is explained by the used indicators [15]. All loadings have a significance level of $p < 0.001$ and are highly significant. This was calculated by using a bootstrap method with 5000 samples [35].

Table 3: Indicator reliability, construct reliability and discriminant validity for adopter

Item		Loading	Mean	AVE	CR	Latent Variable Correlation						
Intention	INT-1	0.971										
	INT-2	0.934										
	INT-3	0.970										
	INT-4	0.843	3.316	0,8666	0.9628	0,93091						
SO	SO-1	0.878										
	SO-2	0.901										
	SO-3	0.844	3.152	0,7645	0.9068	0,87436	0.5225					
HO	HO-1	0.828										
	HO-2	0.810										
	HO-3	0.981										
	HO-4	0.908	4.308	0,7587	0.9261	0,87103	0.5549	0.5711				
UO	UO-1	0.896										
	UO-2	0.893										
	UO-3	0.939	3.243	0,8268	0.9347	0,90929	0.6437	0.5892	0.6637			
SN	SN-1	0.975										
	SN-2	0.973	3.720	0,9485	0.9736	0,97391	0.5847	0.4234	0.5257	0.5753		
PEOU	PEOU-1	0.956										
	PEOU-2	0.945										
	PEOU-3	0.955	3.488	0,9061	0.9666	0,95189	0.4627	0.4953	0.5531	0.4455	0.4142	
FOT	FOT-1	0.887										
	FOT-2	0.845	3.605	0,7499	0.8570	0,86597	-0,0605	-0,1048	0.1563	0.078	-0,0046	0.2040
Note: All loadings are significant at p<0.001; Square Root of AVE is listed on diagonal by LVC												
Adopter												

4.1.3 Construct reliability

Composite Reliability (CR) and Average Variance Extracted (AVE) were used to assess the quality at the construct level [32]. Therefore, each CR value should be over 0.7 and AVE should be higher than 0.5 [6]. Both conditions are fulfilled for users and non-users as illustrated in Table 3 and Table 4

4.1.4 Discriminant validity

Discriminant validity describes the extent, to which measurement items differ from others which theoretically should not be equal [14]. In order to show this, the construct correlations should be

smaller than the root of the corresponding AVE ([38]; [32]). As Table 3 and Table 4 show, this criterion is fulfilled by the data collected for this study.

Table 4: Indicator reliability, construct reliability and discriminant validity for non-adopter

	Item	Loading	Mean	AVE	CR	Latent Variable Correlation						
Intention	INT-1	0.956										
	INT-2	0.958										
	INT-3	0.944										
	INT-4	0.895	2.176	0,8809	0,9673	0,93856						
SO	SO-1	0.808										
	SO-2											
	SO-3	0.874	2.984	0,7079	0,8288	0,84137	0.4551					
HO	HO-1	0.802										
	HO-2	0.952										
	HO-3	0.823										
	HO-4	0.934	3.651	0,7748	0,9319	0,88023	0.2223	0.5848				
UO	UO-1	0.907										
	UO-2	0.937										
	UO-3	0.942	2.591	0,8626	0,9496	0,92876	0.5380	0.7681	0.6659			
SN	SN-1	0.975										
	SN-2	0.957	2.818	0,9327	0,9652	0,96576	0.2977	0.4949	0.6786	0.5871		
PEOU	PEOU-1	0.946										
	PEOU-2	0.952										
	PEOU-3	0.977	2.737	0,9182	0.9712	0,95823	0.1994	0.6105	0.3312	0.5257	0.1884	
FOT	FOT-1	0.752										
	FOT-2	0.927	3.433	0,7123	0.0500	0,84398	-0,425	-0,2267	-0,217	-0,347	-0,4069	0.0759
Note: All loadings are significant at p<0.001; Square Root of AVE is listed on diagonal by LVC												
Non-Adopter												

As a consequence, it is possible to conclude that the measurement model has a high validity.

4.2 Structural model

After validating the measurement model, the structural model will be evaluated. In order to do this, the coefficient of determination (R^2) and the significance levels of the path coefficients [17] need to be observed.

Figure 4 shows that in the adopter case, 52.9 per cent of the variance of Intention can be explained by the six used constructs. Within the non-adopter case, 40.3 per cent of the variance is clarified. According to Chin [17] both models provide an acceptable goodness of fit. For non-adopters, two significant relationships can be confirmed. The first one is the negative influence of Fear of Technology on Intention and a positive impact of Utilitarian Outcomes on the dependent variable.

On the other side, these two relationships were also significant for the group of adopters. Apart from this, two more impacts were identified. These are the influence of Subjective Norm and Perceived Ease of Use on Intention. Only Social and Hedonic Outcomes seem to have no effect on Intention for both elderly adopters and non-adopters of SNS.

4.3 Group Comparison

Next to the previous results like the correlation between intention and each construct,

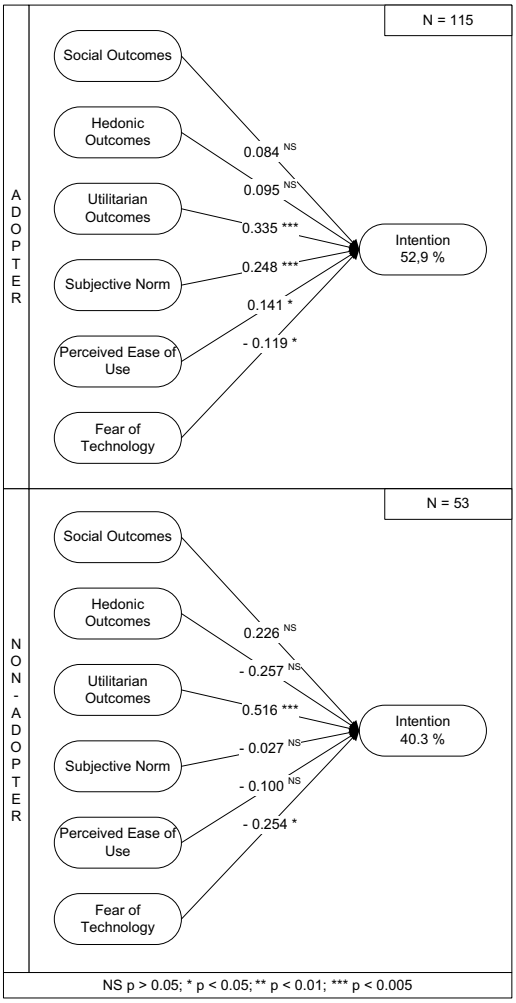


Figure 4: Structural Model

Table 5 represents the strength of effect for each construct and both groups. For adopters and non-adopters, the construct Utilitarian Outcomes has the highest impact on intention. Apart from that Fear of Technology plays an important role for non-adopters, whereas for adopters, this aspect is only of little importance.

Whether comparisons of means are investigated, only the aspect Social Outcomes is not significantly different for both adopter groups. The responsiveness of all other constructs is significantly different.

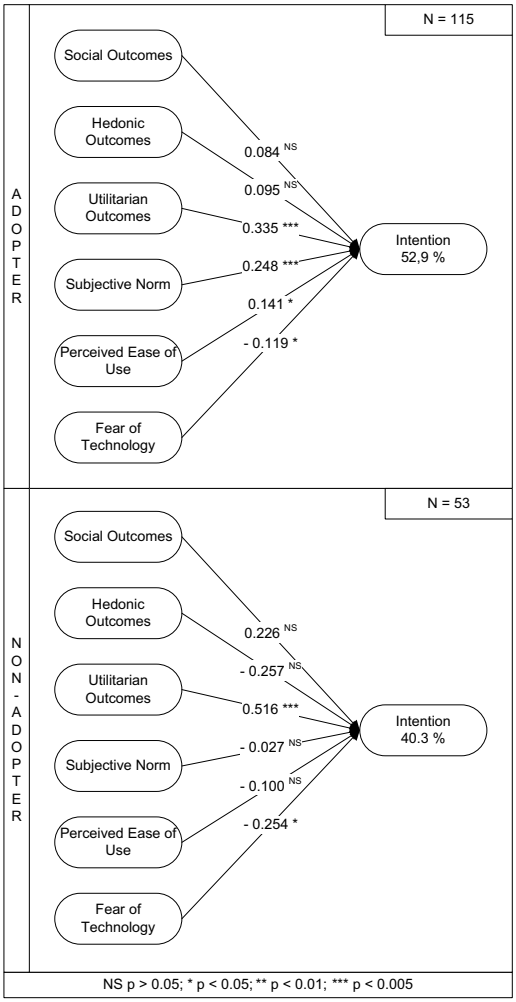


Figure 4: Structural Model

Table 5: Strength of Effect & Comparison of means

Inputfactor	Group	Correlation with Intention	Strength of Effect	Mean	Comparison of means		
					Levene-Test & Homogeneity of Variance	T	Sig.
Intention	Adopter	-	-	3.316	Yes (0.520)	5.635	0.000
	Non-Adopter	-	-	2.176			
Social Outcomes	Adopter	Not significant	1.6%	3.152	Yes (0.124)	1.529	0.133
	Non-Adopter	Not significant	2.9%	2.984			
Hedonic Outcomes	Adopter	Not significant	1.6%	4.308	Yes (0.322)	2.561	0.013
	Non-Adopter	Not significant	2.7%	3.651			
Utilitarian Outcomes	Adopter	Significant	14.1%	3.243	Yes (0.532)	3.622	0.001
	Non-Adopter	Significant	13.1%	2.591			
Subjective Norm	Adopter	Significant	7.9%	3.720	Yes (0.302)	2.683	0.010
	Non-Adopter	Not significant	0.0%	2.818			
Perceived Ease of Use	Adopter	Significant	0.4%	3.488	Yes (0.420)	3.596	0.001
	Non-Adopter	Not significant	3.1%	2.737			
Fear of Technology	Adopter	Significant	2.3%	3.605	Yes (0.150)	2.064	0.042
	Non-Adopter	Significant	11.4%	3.433			

Finally, the significance of the path coefficients was compared for adopters and non-adopters using the proposed procedure by Chin and Dibbern [18]. While doing so, it can be identified, that all paths are highly different for both adopters and non-adopters. In particular for non-adopters Fear of Technology has an higher mean value for non-adopters and the mean of the generated bootstrapped samples is significant different from adopters. Moreover for Utilitarian Outcomes the mean value is higher for non-adopters and significant different from adopters. For adopters, Subjective Norm and Perceived Ease of Use has a higher mean for adopters and is significant different from non-adopters. In terms of the two insignificant paths in both samples Social Outcomes has a higher mean for non-adopters and Hedonic Outcomes for adopters. Both are significant different between the two groups tested. The comparison of path coefficient is illustrated by Table 6.

Table 6: Comparison of Path Coefficient

Path	Group			Comparison of Path Coefficient		
		Mean Value	Standard Deviation	Levene-Test	T	Sig.
SO --> INT	Adopter	0.103	0.068	No (0.000)	17.088	0.000
	Non-Adopter	0.223	0.142			
HO --> INT	Adopter	0.112	0.078	No (0.000)	-43.454	0.000
	Non-Adopter	-0.281	0.182			
UO --> INT	Adopter	0.349	0.098	No (0.000)	17.866	0.000
	Non-Adopter	0.534	0.210			
SN --> INT	Adopter	0.244	0.090	No (0.000)	-57.252	0.000
	Non-Adopter	-0.164	0.131			
PEOU --> INT	Adopter	0.148	0.079	No (0.000)	-48.119	0.000
	Non-Adopter	-0.130	0.102			
FOT --> INT	Adopter	-0.118	0.065	No (0.000)	-27.812	0.000
	Non-Adopter	-0.280	0.079			

4.4 Limitations

This paper is a first try to analyze the intention of adopters and – separately – non-adopters within the MATH for elderly persons. Consequently, the results cannot be generalized limitless. First of all, the presented results derive from one online survey, so that only persons with internet access could participate. It is conceivable that the elderly people without internet access cannot participate and might show other reasons and therefore other correlations which lead to a non-adoption of SNS. On the other hand, it was important to ensure that each SNS non-adopter knows SNS and the involved advantages and disadvantages. This issue was addressed with different questions such as "I know the possibility to engage in social network sites like Facebook", "Advantages of social network sites are (participants could chose between different items or 'I don't know')" or "The first time I heard about social network sites was in (date)". Consequently, each non-adopter analyzed within this paper knows about SNS but does not use it. By collecting data in another way, it could be more problematic to separate between adopters, non-adopters and non-adopters which do not know about the existence of SNS.

Another crucial aspect within this publication is that only one technology was analyzed. Although Brown [12] advised using SNS when household adaption should be investigated, the model has to be confirmed by other researchers observing other technologies. The same will be true for the underlying culture. Dependent on the cultural background, the model could offer different correlations [33].

The last restriction is the relative small number of SNS non-adopters (N = 53) which participated. Considering, that the impact of six constructs on Intention was investigated, 60 data samples should have been the underlying basis [20]. By increasing the data sample up to the crucial threshold (by adding non-users which do not know SNS or by not eliminating incomplete samples), the results were still the same.

5. Discussion and Future Research

In general the findings of this paper show, that elderly people's adoption of SNS is determined by other perceptions and constructs than their non-adoption of SNS. For adopters a significant impact of Normative Beliefs measured as Subjective Norm, Perceived Ease of Use as a part of Control Beliefs on Intention to use SNS could be observed. Contrary, for non-adopters these perceptions have no significant influence on their intention. Nonetheless, these results do not implicate that non-adopters regard new IS as simple to use. By regarding the comparison of means, non-adopters consider the handling of SNS more cumbersome to use than people adopting SNS. The same could be monitored for Subjective Norm. Elderly non-adopters sense that the perceived pressure to adopt social network sites applied by their circle of friends and acquaintances is less than by the social environment of adopters. Future research could consider whether elderly SNS non-adopters have a smaller circle of friends or analyze whether their friends are also SNS non-adopters. Taking into account the Utilitarian Outcomes, which is the most important factor influencing the Intention, reveal that elderly non-adopters face SNS and their possibilities as less useful than elderly adopters.

Most of the people using such modern ways to communicate just use it to stay in contact with friends, they know from the real offline world, and do not try to enlarge their circle of friends by finding new contacts with equal interests. This phenomenon is called Offline-to-Online paradigm [65] and reveals that SNS users first know people from school, job, leisure activities or other activities based in the offline world and afterwards add these people in online communities to their friend lists. As the example of feierabend.de illustrates, SNS for elderly people are designed to enable social exchange between their users in both cases, known form the offline world or not. Therefore another opportunity for future research is to analyze whether the discussed offline-to-online paradigm also holds for elderly people.

Moreover other research activities identified enjoyment as the major predictor for using social networks whereas Usefulness is just less relevant [64]. The findings of this research cannot confirm this totally for elderly people. Our research indicates that people, which are older than 50 years, emphasize utilitarian facets of social networks and regard hedonic outcomes as less relevant. Nonetheless, elderly SNS adopters report having significantly more fun in using SNS than non-adopters, but in both cases, no significant influence on the usage decision could be observed. This could be explained by Phang et al. [61] who investigated the individual behavior in online communities depending on the distinct usage causes. They differentiated between knowledge seeking and knowledge contribution and support that usefulness is essential for people searching knowledge. Since, it is likelier that younger persons use the internet and social network sites to play games or funny quizzes, elderly people will utilize platforms as Facebook or feierabend.de as a source of information (for example about friends or acquaintances) whereby the usefulness is a more crucial issue.

Another important difference between adopters and non-adopters of SNS is the perceived Fear of Technology. In both cases, an influence on Intention could be monitored as well as significant unequal means. Thereby especially non-adopters have problems to engage with new technologies as computers, internet or social networks. For this reason, Fear of Technology has a strong impact on the decision for non-adopters. For these anxiously elderly people, new initiatives should be initiated to make SNS and other technologies accessible to them. The Digital Divide can be addressed, if these persons could be prepared for using new technologies. Nonetheless, this is not only a challenge for the Federal Government, to ensure that people can use modern ICT at home, it is also essential to handle these technologies within organizational contexts. If this group of people, who reject to adopt new technologies, is not able to deal with them or if they fear them, they will hinder the operating schedule. This is clearly reflected by Luftman and Kempaiah [53], which ranked the management of change as the sixth most critical issue for Chief Information Officers. This also comprises IS modifications, but these can only be successful if employees are able to engage in new situations [49] and are not afraid of IS. The latter could diminished by offering regular IS courses and IT trainings by the organization and thus, organizations could create competitive advantages. Feierabend.de has implemented an extended support for elderly people (e.g. offering services to scan pictures, hotline activities, extended explanation of services, etc.) Summing up, it is essential to give people an understanding of IT and IS because of both, to maintain a balanced economy – or sustainability at the corporate level – and from a sociological point of view to reduce the Digital Divide and Digital Inequality.

In general regarding the MATH the results of this research have some implications for the understanding of technology adoption in a private environment. By focusing on people older than 50 and by distinguishing between adopters and non-adopters the results indicate that within the group of people older than 50 the importance of each factor within the MATH is different for adopters and non-adopters. Elderly adopters are mainly driven by Utilitarian Outcomes, Subjective Norm, Perceived Ease of Use and Fear of Technology. In contrast elderly non-adopters are mainly influenced by Utilitarian Outcomes, and Fear of Technology. Moreover regarding significant difference in the mean of each construct the results indicate that Social Outcomes, Hedonic Outcomes, Utilitarian Outcomes, Subjective Norm and Perceived Ease of Use have a higher acceptance by adopters than by non-adopters. In contrast non-adopters are more afraid in terms of Fear of Technology. In addition, regarding the strength of effect the results show that for adopters and non-adopters the strongest effect can be observed for Utilitarian Outcomes as well as for Fear of Technology. In general these results point out that the relevance of each antecedent is different for adopters and non-adopters.

Apart from that by considering the R^2 of both adopters and non-adopters it is obviously that the explanation power is higher for adopters than for non-adopters. Consequently it would be quite conceivable to develop a model which explains – in particular – the non-adoption behavior. Therefore, first attempts which discussed such an approach could be found in the literature [48]. Nonetheless, it would be necessary to identify different reasons leading to non-adoption, as a status quo bias [42], resistance ([49]; [7]; [8]) or other inhibitors [16] as fear or threats.

Nonetheless this research shows that the MATH model is also valid for elderly people. Although, lots of future research will be necessary to understand the behavior of the elderly people altogether. This research was just a first attempt to enlighten

factors influencing adoption and non-adoption of the elderly people.

In terms of the methodology used it could be evaluated that the proposed Fast Form approach [19] is applicable to the MATH and empirical studies using MATH as all items are statistically useable for the model evaluation.

By analyzing the domain social network sites it is interesting that Hedonic Outcomes as fun provided by the platform have no impact on the intention. Given that fun is no significant impact factor for elderly people it is probable that these people for example do not play online social games in SNS. Thereby researchers could analyze which SNS services and applications (e.g. enlarge circle of friends, stay in contact with friends, communicate, exchange pictures or videos, have fun by playing online games) are utilized by which SNS users. In doing so it could be investigated which reasons or perceptions are responsible for such a SNS behavior. Furthermore, the usefulness of social network sites is essential for people to register in such communities. Consequently it would be interesting to know which services are explicitly most important for elderly people. Apart from that, the correlation between Perceived Ease of Use and Intention discloses a certain level of PC and internet skills.

In terms of practical implications this research shows that adopters and non-adopters are different regarding their motivation to adopt or not to adopt SNS. Given that it is important to motivate elderly people to participate in those networks one should focus on describing the Utilitarian Outcomes and by allaying the fear of technologies. As these two aspects are the most important perceptions for those who do not want to use SNS in the future. In contrast those who have experience with SNS point out that they perceive positive Utilitarian Outcomes, that the platforms are easy to use and do not perceive as much as non-adopters a fear due to the technology. In addition their normative beliefs encourage them to continue using SNS. Therefore for SNS provider for elderly people it is important to point out the usefulness of their platforms as feierabend.de does by establishing regional groups and supporting offline activates of the users. Moreover the Fear of Technology can be addressed as feierabend.de does by explicitly focusing on explaining and supporting the usage of the platform. This is a first step towards an age-sensitive design of online services as demanded by Pak et al. 2009 [60], which reveals that online services for elderly people should be connected with offline activities in order to support the use and usefulness of these platforms.

6. Conclusion

Why do elderly people adopt or not adopt SNS, is the main research question of this research. By using MATH the results indicate that adopters are mainly driven by Utilitarian Outcomes, Normative Beliefs, Perceived Ease of Use and Fear of Technology and non-adopters by Utilitarian Outcomes and Fear of Technology. As a consequence the research provides evidence that beside the moderator age within the math it is important to consider the difference between adopters and non-adopters within the group of elderly persons as well.

7. REFERENCES

[1] (N)Onliner Atlas: Eine Topographie des digitalen Grabens durch Deutschland: Nutzung und Nichtnutzung des Internets, Strukturen und regionale Verteilung. In: TNS Infratest Juni, München (2009).

[2] Ajzen, I.: The theory of planned behavior. In: Organizational Behavior and Human Decision Processes 50 (1991) 2, S. 179–211.

[3] Ajzen, I.; Fishbein, M.: Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ 1980.

[4] Azari, R.; Pick, J. B.: Technology and society: socioeconomic influences on technological sectors for United States counties. In: International Journal of Information Management 25 (2005) 1, S. 21–37.

[5] Bagozzi, R. P.: The Role of Measurement in Theory Construction and Hypothesis Testing: Toward a Holistic Model. In: Ferrell, O. C.; Brown, S. W.; Lamb, C. W. (Hrsg.): Conceptual and theoretical developments in marketing. Chicago, Ill. 1979.

[6] Bagozzi, R. P.; Yi, Y.: On the Evaluation of Structural Equation Models. In: Journal of the Academy of Marketing Science 16 (1998) 1, S. 74–94.

[7] Bhattacherjee, A.; Hikmet, N.: Physicians' resistance toward healthcare information technology: a theoretical model an empirical test. In: European Journal of Information Systems 16 (2007) 6, S. 725–37.

[8] Bhattacherjee, A.; Hikmet, N.: Physicians' Resistance toward Healthcare Information Technologies: A Dual-Factor Model. In: Hawaii International Conference on System Sciences (2007).

[9] Bhattacherjee, A.; Sanford, C.: Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model. In: MIS Quarterly 30 (2006) 4, S. 805–25.

[10] BitKom: 30 Millionen Deutsche sind Mitglieder in Communitys.
URL: http://www.bitkom.org/de/themen/36444_62772.aspx. Abrufdatum 21.08.2010.

[11] Blackwell, A. F.: End-user Developers at Home. In: Communications of the ACM 47 (2004) 9, S. 65–66.

[12] Brown, S. A.: Household Technology Adoption, Use, and Impacts: Past, Present, and Future. In: Information Systems Frontiers 10 (2008) 4, S. 397–402.

[13] Brown, S. A.; Venkatesh, V.: Model of Adoption of Technology in Households: A Baseline Model Test and Extension Incorporating Household Life Cycle. In: MIS Quarterly 29 (2005) 3, S. 399–426.

[14] Campell, D. T.; Fiske, D. W.: Convergent and discriminant validation by the multitrait-multimethod matrix. In: Psychological Bulletin 56 (1959) 2, S. 81–105.

[15] Carmines, E. G.; Zeller, R. A.: Reliability and validity assessment, [Nachdr.]. Newbury Park, Calif. 2008.

[16] Cenfetelli, R. T.: Inhibitors and Enablers as Dual Factor Concepts in Technology Usage. In: Journal of the Association for Information Systems 5 (2004) 11-12, S. 472–92.

[17] Chin, W. W.: The Partial Least Squares Approach to Structural Equation Modeling. In: Marcoulides, G. A. (Hrsg.): Modern methods for business research. Mahwah N.J. u.a. 1998.

[18] Chin, W. W.; Dibbern, J.: An Introduction to a Permutation Based Procedure for Multi-Group PLS Analysis: Results of Tests of Differences on Simulated Data and a Cross Cultural Analysis of the Sourcing of Information System Services Between Germany and the USA. In: Esposito Vinzi, V.; Chin, W. W.; Henseler, J.; Wang, H. (Hrsg.): Handbook of Partial Least Squares. Concepts, Methods and Applications. Berlin, Heidelberg 2010.

[19] Chin, W. W.; Johnson, N.; Schwarz, A.: A fast form approach to measuring technology acceptance and other constructs. In: MIS Quarterly 32 (2008) 4, S. 687–703.

[20] Chin, W. W.; Newsted, P. R.: Structural Equation Modeling analysis with Small Samples Using Partial Least Squares. In: Hoyle, R. H. (Hrsg.): Statistical strategies for small sample research. Thousand Oaks, Calif. 2000.

[21] Choudrie, J.; Dwivedi, Y. K.: Investigating Factors Influencing Adoption of Broadband in the household. In: Journal of Computer Information Systems 46 (2006) 4, S. 25–34.

[22] Choudrie, J.; Lee, H. J.: Broadband development in South Korea: institutional and cultural factors. In: European Journal of Information Systems 13 (2004) 2, S. 103–14.

[23] COMPASS HEADING: Nutzerzahlen Sozialer Netzwerke Januar 2010. URL: http://www.compass-heading.de/cms/nutzerzahlen-sozialer-netzwerke-januar-2010/. Abrufdatum 21.08.2010.

[24] Correa, T.; Hinsley, A. W.; de Zúniga, H. G.: Who interacts on the Web?: The intersection of users' personality and social media use. In: Computers in Human Behavior 26 (2010), S. 247–53.

[25] Dewan, S.; Riggins, F. J.: The Digital Divide: Current and Future Research Directions. In: Journal of the Association for Information Systems 6 (2005) 13.

[26] Dwivedi, Y. K.; Choudrie, J.; Brinkman, W.-P.: Development of a survey instrument to examine consumer adoption of broadband. In: Industrial Management & Data Systems 106 (2006) 5, S. 700–18.

[27] Eckhardt, A.; Laumer Sven; Weitzel, T.: Who influences whom Analyzing workplace referents' social influence on IT adoption and non-adoption. In: Journal of Information Technology 24 (2009) 1, S. 11–24.

[28] Ellison, N.; Steinfield, C.; Lampe, C. A. .: The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. In: Journal of Computer-Mediated Communication 12 (2007) 4, S. 1143–68.

[29] Enns, H. g.; Huff, S. L.; Higgins, C. A.: CIO Lateral Influence Behaviors: Gaining Peers' Commitment to Strategic Information Systems. In: MIS Quarterly 27 (2003) 1, S. 155–76.

[30] Facebook: Statistics. URL: http://www.facebook.com/press/info.php?statistics. Abrufdatum 01.07.2010.

[31] Fishbein, M.; Ajzen, I.: Belief, attitude, intention and behavior. An introduction to theory and research. Reading, Mass. 1975.

[32] Fornell, C.; Larcker, D. F.: Evaluating Structural equation models with unobservable variables and measurement error. In: Journal of Marketing Research 18 (1981) 1, S. 39–50.

[33] Gallivan M.J.; Srite, M.: Information Technology and Culture: Identifying Fragmented and Holistic Perspectives of Culture. In: Information & Organization 15 (2005) 2, S. 295–338.

[34] Hargittai, E.: Second-Level Digital Divide: Differences in People's Online Skills. URL: http://www.firstmonday.dk/issues/issue7_4/hargittai.

[35] Henseler, J.; Ringle, C. M.; Sinkovics Rudolf R.: The use of partial least squares path modeling in international marketing. In: Advances in International Marketing 20 (2009), S. 277–319.

[36] Hirschheim, R.; Newman, M.: Information Systems and User Resistance: Theory and Practice. In: The Computer Journal 31 (1988) 5, S. 398–408.

[37] Holznagel, B.: Frequenzeffizienz und Rundfunkspektrum. In: MMR (2008) 4, S. 207–15.

[38] Hulland, J. S.: Use of partial least squares (PLS) in strategic management research: A review of four recent studies. In: Strategic Management Journal 20 (1999) 2, S. 195–204.

[39] Hutter, T.: Facebook Demographische Zahlen der Länder > 1 Mio. Benutzer per 31.05.2010 (2010). URL: www.thomashutter.cin.

[40] Jiang, J. J.; Muhanna, W. A.; Klein, G.: User resistance and strategies for promoting acceptance across system types. In: Information and Management 37 (2000) 3, S. 25–36.

[41] Johansen, R.; Swigart, R.: Upsizing the individual in the downsized organization. Managing in the wake of reengineering, globalization, and overwhelming technological change. Reading, Mass 1996.

[42] Jones, Q.; Ravid, G.; Rafaeli, S.: Information Overload and the Message Dynamics of Online Interaction Spaces: A theoretical model and empirical exploration. In: Information Systems Research 15 (2004) 2, S. 194–210.

[43] Joos, J. G.: Social media: New frontiers in hiring and recruiting. In: Employment Relations Today 35 (2008) 1, S. 51–59.

[44] Khan, Z.; Jarvenpaa, S. L.: Exploring temporal coordination of events with Facebook.com. In: Journal of Information Technology 25 (2010) 2, S. 137–51.

[45] Kim, H.-W.; Kankanhalli, A.: Investigating User Resistance to Information Systems Implementation: A Status Quo Bias Perspective. In: MIS Quarterly 33 (2009) 3, S. 567–82.

[46] Kubicek, H.; Welling, S.: Vor einer digitalen Spaltung in Deutschland? Annäherung an ein verdecktes Problem von wirtschafts- und gesellschaftspolitischer Brisanz. In: Medien und Kommunikationswissenschaft 48 (2000) 4, S. 497–517.

[47] Lapointe, L.; Rivard, S.: A Multilevel Model of Resistance to Information Technology Implementation. In: MIS Quarterly 29 (2005) 3, S. 461–91.

[48] Laumer, S.; Maier, C.; Eckhardt, A.: Towards an Understanding of an Individual's Resistance to Use an IS - Empirical Examinations and Directions for Future Research. Phoenix (AZ) 2009.

[49] Laumer, S.; Maier, C.; Eckhardt, A.: Why do they resist? - An empirical analysis of an individual's personality trait resistance regarding the adoption of new information systems. European Conference on Information Systems, South Africa 2010.

[50] Lewin, K.: Frontiers in Group Dynamics: Concept, Method and Reality in Social Science; Social Equilibria and Social Change. In: Human Relations 1 (1947) 1, S. 5–41.

[51] Lewis, K.; Kaufman, J.; Christakis, N.: The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network. In: Journal of Computer-Mediated Communication 14 (2008) 1, S. 79–100.

[52] Lewis, K.; Kaufman, J.; Gonzales, M.; Wimmer, A.; Christakis, N.: Taste, Ties, and Time: A new social network dataset using Facebook.com. In: Social Networks 30 (2008) 4, S. 330–42.

[53] Luftman, J.; Kempaiah, R.; Rigoni E. H.: Key Issues for IT Executives 2008. In: MIS Quarterly Executive 8 (2009) 3, S. 151–59.

[54] Markus, M. L.: Power, Politics, and MIS Implementation. In: Communications of the ACM 26 (1983) 6, S. 430–44.

[55] Mayer, A.; Puller, S.: The Old Boy (and Girl) Network: Social network formation on university campuses. In: Journal of Public Economics 92 (2008) 1-2, S. 329–47.

[56] Mertens, P.: Fehlschläge bei IT-Großprojekten der öffentlichen Verwaltung - ein Beitrag zur Misserfolgsforschung in der Wirtschaftsinformatik. München 2008.

[57] Moore, G. A.: Crossing the chasm. Marketing and selling high-tech products to mainstream customers, Rev. ed. New York 1999.

[58] Norman, D. A.: Things that make us smart. Defending human attributes in the age of the machine, [10. print.]. Reading, Mass. 1999.

[59] Oreg, S.: Resistance to Change: Developing an Individual Differences Measure. In: Journal of Applied Psychology 88 (2003) 4, S. 680–93.

[60] Pak, R.; Price, M.; Thatcher, J. B.: Age-Sensitive Design of Online Health Information: Comparative Usability Study. In: Journal of Medical Internet Research 11 (2009) 4.

[61] Phang, C. W.; Kankanhalli, A.; Sabherwal, R.: Usability and Sociability in Online Communities. In: Journal of the Association for Information Systems 10 (2009) 10, S. 721–47.

[62] Raacke, J.; Bonds-Raacke, J.: MySpace and Facebook: Applying the Uses and Gratifications Theory to Exploring Friend-Networking Sites. In: CyperPsychology & Behavior 11 (2008) 2, S. 169–74.

[63] Ringle, C. M.; Wende, S.; Will, A.: SmartPLS, 2.0 (beta) 2005.

[64] Rosen, P.; Sherman, P.: Hedonic Information Systems: Acceptance of Social Networking Websites. Mexiko 2006.

[65] Ross, C.; Orr, E. S.; Sisic, M.; Arseneault, J. M.; Simmering, M. G.; Orr, R. R.: Personality and motivations associated with Facebook use. In: Computers in Human Behavior 25 (2009) 2, S. 578–86.

[66] studiVZ: Data and Facts. URL: http://www.studivz.net/l/about_us/1/. Abrufdatum 14.07.2010.

[67] Subrahmanyam, K.; Reich, S. M.; Waechter, N.; Espinoza, G.: Online and offline social networks: Use of social networking sites by emerging adults. In: Journal of Applied Developmental Psychology 29 (2008) 6, S. 420–33.

[68] Valenzuela, S.; Park, N.; Kee. K. F.: Is there social capital in a social network site? Facebook use, and college students' life satisfaction, trust, and participation. In: Journal of Computer-Mediated Communication 14 (2009) 4, S. 875–901.

[69] Venkatesh, A.: Digital home technologies and transformation of households. In: Information Systems Frontiers 10 (2008) 4, S. 391–95.

[70] Venkatesh, V.; Brown, S. A.: A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges. In: MIS Quarterly 25 (2001) 1, S. 71–102.

[71] Venkatesh, V.; Morris, M. G.; Davis, G. B.; Davis, F. D.: User Acceptance of Information Technology: Toward a unified View. In: MIS Quarterly 27 (2003) 3, S. 425–78.

[72] Walther, J. B.; Van der Heide, B.; Kim, S.-Y.; Westerman, D.; Tong, S. T.: The Role of Friends' Appearance and Behavior on Evaluations of Individuals on Facebook: Are We Known by the Company We Keep? In: Human Communication Research 34 (2008) 1, S. 28–49.

[73] Wege ins Netz: Der Wettbewerb für mehr digitale Kompetenz (2010). URL: http://www.bmwi.de/BMWi/Navigation/Technologie-und-Innovation/Informationsgesellschaft/internet-erfahren,did=339642.html. Abrufdatum 21.08.2010.

[74] Weitzel, T.; Eckhardt, A.; Laumer, S.: A Framework for Recruiting IT Talent: Lessons from Siemens. In: MIS Quarterly Executive 8 (2009) 4, S. 175–89.

[75] Wiener, L. R.: Digital woes. Why we should not depend on software. Reading, Mass 1993.

[76] Williams, M. D.; Dwivedi, Y. K.; Lal, B.; Schwarz, A.: Contemporary trends and issues in IT adoption and diffusion research. In: Journal of Information Technology 24 (2009) 1, S. 1–10.

A Theory of User Acceptance of IS Project Management Methodologies: Understanding the Influence of Psychological Determinism and Experience

Kunal Mohan
EBS Business School
Söhnleinstraße 8D, 65201 Wiesbaden
kunal.mohan@ebs.edu

Frederik Ahlemann
EBS Business School
Söhnleinstraße 8D, 65201 Wiesbaden
frederik.ahlemann@ebs.edu

ABSTRACT
Despite the overwhelming advantages of using an IS project management methodology, organisations are rarely able to motivate their staff to use them: Consequently, this lack of methodology usage by individuals fails to deliver the expected advantages of better quality, control, less time and effort. We analyse the determinants of an individual's intention to use IS project management methodology in order to enable organisations to engineer those that meet the needs of actual users and are really used by them. Results from an exploratory field study conducted in a service organisation, are used to construct a conceptual model. Based upon this research model, we posit that: a) value of a methodology, b) workgroup influence, c) self-beliefs, d) organisational characteristics, and e) previous habits influence intention to use a methodology. Additionally, we find that the strength of these relationships depends upon the needs of an individual and the degree of prior experience they have in using similar methodology.

General Terms
Management, Human factors, Theory

Keywords
Methodology adoption, Usage, IS project management

1. INTRODUCTION
In the search for ways to arrive at replicable, pragmatic, cost-effective, and timely solutions to real-world problems in systematic and predictable ways [1], organisations either adopt or customise and adaptively apply IS project management (ISPM) methodologies, which consist of tested bodies of methods, rules and procedures. Some of the most fundamental concepts that justify the use of structured methodologies, as identified by Fitzgerald, [2] are: i) they reduce complexity by subdividing the projects development and management process into plausible and coherent steps, ii) they increase transparency and therefore control of the activities, thus reducing risk and uncertainty of projects, and iii) they provide a goal-oriented framework that helps to direct the application of techniques and resources at appropriate times during the project [2]. Despite the advantages of using an ISPM methodology, only 50% of organisations are actually able to make their staff use such methodologies [3]. In the context of software development, a project survey conducted by Russo et al. [4] showed that only 6% of organisations claim that their methodologies are always used as specified. Cicmil et al. [1] also found that resistance towards the acceptance of project management methodologies is high because the users do not have faith in the concept, fear power loss, or lack adequate training and support from upper management. Organisational theorists have long recognised that behavioural resistance of individuals against the use of new methodologies is because they might not share the goals of the organisations in which they work, and that exerts pressure on them to use the new methodologies [5]. As such, the roots of lacking methodology acceptance, lies – among other factors – in the failure to understand the attitudes of individuals towards using a methodology. This lack of understanding ultimately leads to the development and implementation of ISPM methodologies that might be considered unsuitable, and are as a result rejected by individuals [6].

In the past, research projects attempted to analyse only a few of the above-mentioned problems. These projects focused on cognitive user decision-making in narrow and specific organisational contexts (mostly in the field of software engineering). However, these research projects have not provided any concrete answers [7]. Some studies have also attempted to examine usage behaviour of individuals regarding IS methodologies from a technology adoption perspective. They view software development methodologies as technology innovations, and make use of Diffusion of Innovations Theory (DOI) and the Technology Acceptance Model (TAM) (for e.g. [8,9]). Strikingly, our literature review revealed that none of the studies conducted in the past attempted to examine the effect of individual's automatic behaviour (e.g. habits and emotions) in the context of methodology usage. Research has also not attempted to understand the effect of individual users' deep-rooted personal characteristics and traits, such as their needs, expectancies, age and gender. The expenditure of time and effort in developing and implementing ISPM methodologies makes this a critical area of IS concern [8]. This leads us to some fundamental questions regarding the use of methodologies, which this study attempts to answer:

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

a) What are the determinants of an individual's decision to accept and use an ISPM methodology?
b) How do basic needs of individuals and other contextual factors such as methodology experience influence the predictive power of the different determinants of ISPM methodology acceptance?

Our study is a step toward filling the gap in the ISPM evaluation, development and adoption literature, which till now has not developed a theoretically and practically complete and relevant understanding of the determinants that influence the acceptance of ISPM methodologies, and has also not studied the effect of personal traits such as needs. We heed the call of Pfleeger [10], who appeals that the field of MIS needs to better understand the role of people in the adoption process, drawing upon social science models as appropriate to further this understanding [9]. We have identified needs theories – e.g. Maslow's hierarchy of needs theory [11] and Murray's theory of psychogenic needs [12] – that help us understand how, when and which specific needs are more important to people, and the social science model of Triandis' Theory of Interpersonal Behaviour (TTIB) [13], to provide a comprehensive theoretical basis for analysing the aforementioned research questions.

The remainder of this *research-in-progress* paper is organised as follows: In section 2 we discuss the foundations, which aids the reader's understanding of the context of our research. We discuss prior research on the topic in order to clarify what has been done and what needs to be done. In section 3 we provide an overview of the methods we use, and why we use them. Section 4 explains the basic theoretical concepts that provide the framework for our conceptual model. We present our research model and hypotheses, pointing out validated survey instruments that might be used to operationalise the underlying constructs in the next phase of our research (which involves testing the proposed model). In Section 5, we discuss limitations of the study and outline the next steps in the course of our research. To conclude, in section 6, we discuss the implications and contributions of our study.

2. FOUNDATIONS

In our research, we focus on examining the behaviour of *individual users* of a methodology instead of an organisation because, although a particular ISPM methodology is developed and implemented by an organisation, the extent of its use is usually determined by the actual users of the methodology [9,14]. Additionally, we also focus only on the use of ISPM *methodologies* instead of ISPM methods/techniques (e.g. stakeholder analysis, earned value analysis, network planning, risk analysis etc.) and tools, since tools (e.g. project management information system, excel/ word based Templates, ARIS etc.), techniques and methods can be used in the absence of a formal methodology, and the use of a methodology represents a radical change compared to the use of methods/tools [8]. Reasons why the adoption and use of new ISPM methodologies might be so different and so much more challenging than the adoption of specific methods and tools lie partly in the tacit organisational and individual problems that are caused by the introduction of new methodologies. For example, the stress associated with the learning of a new methodology, the fear and impact on self-esteem and identity associated with the organisational restructuring or re-engineering, and the emotional costs of role conflict and ambiguity or workplace transformation might be serious inhibitors of ISPM methodology acceptance and usage [15].

In the context of methodology adoption, Khalifa and Verner [14] found that better process and product quality has a substantial effect on a software developer's decision to use waterfall and prototyping methodologies. Application of both technological and behavirol models such as TAM and TPB come to similar conclusion and state that *usefulness*, a characteristic of a methodology is the single most important determinant of methodology acceptance and use by its actual users [16,8,9]. Subsequent research has therefore focused on this particular variable but neglected other potential crucial attributes of a methodology. For example Riemenschneider et al. [9] apply five theoretical models and conclude that "…if a methodology is not regarded as useful by developers, its prospects of successful deployment may be seriously undermined". Hardgrave and Johnson [16] also conclude that "…software developers do not view their personal benefits separately from organisational benefits" [16]. Therefore, the personal usefulness (PU) of a methodology might not affect their decision to use it. Hardgrave and Johnson [16] come to this conclusion because they could not psychometrically separate their PU construct from their organisational usefulness (OU) construct. We suggest differentiating between OU and PU based on other grounds and seek to provide a solution in our conceptual model.

However, critics have suggested that TAM and TPB are too parsimonious and need to be expanded by integrating variables specific to the methodology under investigation [17]. Nevertheless, even when a handful researchers attempt to examine other methodology attributes, the attributes are found to be either not significant or their effect negligible – e.g. [8,9] partly because these studies neglect to integrate other nontechnical and noneconomic variables from related theoretical perspectives [17]. As Warner [18] observers, the concept of adoption is a complex social phenomenon which involves both technical and nontechnical factors and sociologists would undoubtedly agree with this view. Unfortunately, the several different disciplines, generally concentrating on their individual variables, have neglected to incorporate the personality attributes in understanding the methodology acceptance problem. As such, little is known about the interactive effects of the attributes of methodologies and the nontechnical personality characteristics and it seems reasonable that variables from both sets are important in explaining the problem at hand [18].

3. METHODOLOGY

An exploratory investigation was conducted to examine practitioner perceptions towards methodology acceptance and usage. We accompanied a large multinational professional service firm (140,000+ employees) in its endeavour to develop three IS management methodologies: a) IT project and portfolio management, b) IT benefits management, and c) Enterprise architecture management (Table 1 provides an overview of the methodologies studied).

Table 1. Overview of Methodologies

	Methodology A	**Methodology B**	**Methodology C**
Name	IT Benefits Management	IT Project- /Portfolio Management	Enterprise Architecture

			Management
Descrip-tion	Development of a methodology to manage, so that potential benefits arising from IS projects are realised	Development of a comprehensive PPM methodology to ensure the efficient and effective execution of IT projects	Development and implementation of a methodology to improve the alignment of business and IT in an enterprise
Data Sources	Interviews, document review, field notes	Interviews, field notes, questionnaires, protocols	Interviews, field notes, document review
Individ-uals/ dep-artments involved	Project managers, PMO, Corporate Controlling (CC), benefits managers, consultants	Project managers, Project management office (PMO), CC, IS managers, consultants	Enterprise architects, CIO, business analysts and functional managers

The ability to observe the development process of various methodologies deemed the organisation as a fruitful ground for our investigation. Multiple data collection methods are applied, based on a) archival sources, b) unstructured and semi-structured interviews (lasting 30 to 60 minutes each) with individuals involved in the management of the organisations' IS/IT, c) protocols, document review and field notes of multiple workshops (each lasting five to eight hours) involving representatives of upper management, corporate controlling (CC), IT project managers (PM), and the Project Management Office (PMO). An exploratory investigation, involving such a diverse segment of users, developers and supporters provided us with a holistic understanding of the development of individuals' beliefs, attitudes and usage behaviours. In-depth interviews allowed us to better understand the process by which people reach decisions about using a particular methodology since "…it records more fully how subjects arrive at their opinions. The way subjects ramble, hesitate, stumble, and meander, as they formulate their answers, tips us off to how they are thinking and reasoning." [19]. The interactive workshops in particular allowed us to gain a deep understanding of the interplay between different organisational members/departments. Field notes and protocols that were gathered in the workshops and team meetings, in which individuals shared their thoughts and emerging ideas, provided clues about relationships, anecdotes and informal observations [20]. Interviews and workshop protocols were generally conducted and written by two investigators, face-to-face. In order to strengthen the internal validity and generalisability of our research, existing literature and theories, the TTIB framework was used to form a priori concepts/codes, to develop the interview guide and to structure field notes/protocols [16]. Prior to a workshop or an interview, we created a text document based on the TTIB concepts/codes. For every dimension, we left a blank page in which we noted our observations and interpretations. Such a prestructured document helped us to swiftly note our observations and thoughts, and to allocate them to the right concept/code without having to interrupt the participant. This also helped minimise data loss as a result of the investigator not being able to keep up with the fast pace of the workshops and interviews. In short, the a) investigator, b) theory, and c) method triangulation technique that is applied in our study provides stronger substantiation of constructs and propositions. Table 2 provides an example of the qualitative data we collected in 8 semi-structured interviews with project managers and in 15 methodology development workshops.

Table 2. Example of Qualitative Data on Methodology Acceptance and Use

Dimension	Expressed by	Interview/ Workshop Participant Comments, Observations and General Findings
Task-oriented usefulness	Project manager (PM), Project team member (PTM)	Majority of the interviewees mentioned the usefulness of a particular methodology in achieving project goals to be a key determinant of their decision to use the methodology. A project manager gave an example of a Software Development methodology that was developed by the organisation over a period of 2 years and with considerable resources. He mentioned that the methodology was never used the way it was supposed to be used because it was so complex, comprehensive and "over-engineered" that most managers felt that it was counterproductive.
Pleasure/ Enjoyment	PM, PTM	Interviewees occasionally mentioned experiencing 'pride' when using a methodology because they had mastered its use. One person felt 'loyal' towards the organisation by strictly using the methodology as requested of him. Some IT managers hinted at experiencing 'excitement' at the thought of being able to experiment with various methods and techniques, or felt a sense of 'accomplishment' or 'self-actualisation' by using a methodology.
Materialism	PM	An interviewee mentioned that he used a methodology as insurance in case projects fail. By adhering strictly to the methodology, he can deny responsibility for the failed project and simply "blame the methodology". In such a scenario, a methodology is used because through its use the user can avoid negative career or monetary consequences because of failed projects.
Awareness	PL, CC, PMO, PTM	Some of the workshop participants were influenced by the opinions of external consultants who were experts on methodology engineering. Workshop participants reacted positively to solutions and explanations provided by these experts and actively sought their advice by asking questions.
Capabilities and Experience	PM, PTM	We observed that inexperienced users often doubted their skills and knowledge regarding the correct use of a new methodology. We also noticed that project managers with more than 5 years' experience were more actively involved in the interactive workshops and provided suggestions on how to improve the methodology. Project managers with less than 2 years' experience with methodologies repeatedly mentioned in interviews and workshops that they needed better training in the use of complicated methodologies. Project managers raised a number of questions regarding the effort involved in learning

		the new methodology and the support provided from the organisation.
Organisa-tional support	PM	Demands for more support and political backing were reported in the interviews. One of the managers mentioned that "we don't get help from the project management office when we run into conflicting situations regarding using a methodology. The only way we resolve the problem is through using our social networks and getting help from experienced colleagues. A person who doesn't have a good social network because he is new in the organisation finds it extremely hard to use the methodology correctly".
Voluntari-ness of use	PM, PMO, CC	During a workshop, when PMO and CC representatives tried to force project managers to adopt a certain way of executing an IT project management methodology, one of the managers replied aggressively, saying "I refuse to do this. I will not use the methodology like this. It will never work". A PMO representative stated in an interview that "we cannot force them (project managers) to do something they don't find right. There are always some loopholes in procedures and they will use these loopholes to use the methodology the way they want to". We therefore infer that even though organisations can deploy obligatory methodologies, their actual use cannot be forced and thus correct usage is ultimately a voluntary user act.

4. CONCEPTUAL FRAMEWORK AND HYPOTHESES

The decision to adopt a methodology requires time, energy, and careful consideration on the potential user's behalf [21]. Since intention to use a methodology is a measure of the strength of a person's intention to actually use it, literature suggests that a person's intentions indicates how hard he or she is willing to try and how much effort he or she is planning to exert to actually use the methodology. Research on behavioural decision-making also posits, based on a number of empirical studies, that there is a strong relationship between behavioural intention and the actual behaviour, i.e. intention to use a methodology and its adoption and actual use [22], [23], [24], [17]. We construct our theoretical framework (see Figure 1) based on a subset of the TTIB model according to which, an individual's intention to use a methodology depends on cognitive as well as automatic behavioural influences, namely:

a) the person's attitude towards the methodology (his or her belief that using the methodology will lead to certain favourable or unfavourable outcomes) examined in section 4.1,

b) subjective norms (which refer to perceived pressure and influence exerted from a person's social environment, forcing him or her to either use the methodology or to not use it) examined in section 4.2,

c) perceived behavioural control (the person's belief that he or she has adequate external and internal control over the use of a specific methodology) examined in section 4.3 and 4.4, and

d) habits (the persons subconscious use of a specific methodology) examined in section 4.5.

Results of our exploratory investigation (see table 2) provide further empirical evidence for the existence of the above mentioned antecedents of a person's intention to use a methodology.

4.1 Value

The usefulness of an ISPM methodology is reflected in the value that would be generated through its use, originating in an individual's mind through cognitive mechanisms that relate to goal attainment [23]. Although past research has focused primarily on task-related *utilitarian value*, which seeks to provide instrumental value to the user – such as increasing task performance, efficiency and productivity [25], research in the field of consumer behaviour suggests that there are other sources of value related to one's personal goals– *hedonic value* [26] – that are more subjective and personal than its utilitarian counterpart [25] and *materialistic value* [27] – that focuses on the acquisition of worldly possessions. Hedonic value is generated as a result of pleasurable experiences that a person might encounter through sensations generated on multiple sensory channels by using a methodology [25]. Hedonic value may therefore be defined as the extent to which the activity of using a methodology is perceived as being enjoyable in its own right, apart from any performance consequences that may be anticipated [28]. Even though individuals may not expect using methodologies in organisational settings to prompt high levels of fun, we argue – similar to Venkatesh [24] – that *methodology enjoyment* is still expected to be a relevant factor in influencing an individual's perceptions of a methodology, as "...enjoyment not only includes the desire for fun but also involves, among others, exploration, discovery, challenge, loyalty and curiosity".

Materialistic value, on the other hand, is based on an orientation that describes material goods and money as important for personal happiness and social progress. According to Belk [27] "…at the highest levels of materialism, such possessions assume a central place in a person's life and are believed to provide the greatest sources of satisfaction and dissatisfaction". In the context of our study, materialism refers not only to monetary advantages but also involves intrinsic rewards such as respect, status and acceptance [29]. Our proposed study of usefulness of a methodology from three distinct perspectives captures more details about an individual's attitude towards the use of a methodology, and might be the solution to the psychometric problem faced by Hardgrave and Johnson [16] (see research methodology section). While utilitarian value can be considered to be primarily a manifestation of organisational usefulness, hedonic and materialistic value typically reflect personal usefulness.

4.2 Workgroup Influence

Extensive research on human behaviour shows that a methodology's use is influenced by an individual's perception that people who are important to him think he should or should not use it[30]. According to Venkatesh and Davis [17], the reason why workgroup influences directly impact a person's intention to use a methodology is because people may choose to use the methodology, even if they don't have a favourable attitude towards its use or the consequences of its use, if they believe that "...one or more referents they think would, and they are sufficiently motivated to comply with the referent's opinion". In

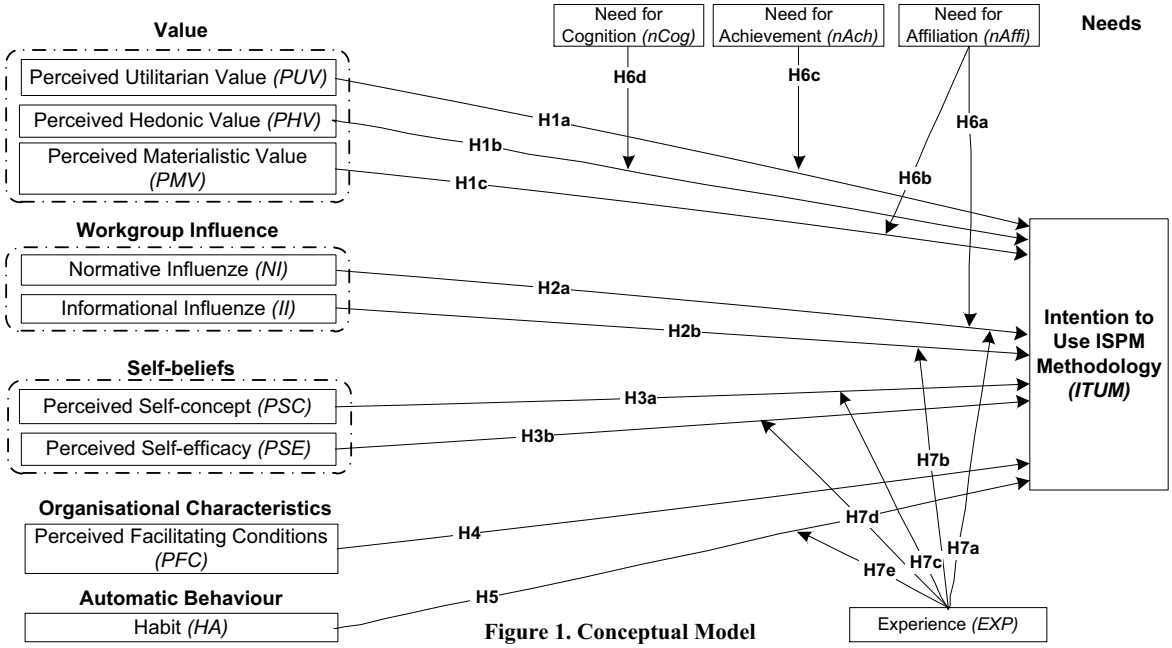


Figure 1. Conceptual Model

order to fully understand the effect of social influences on a user's behaviour, Deutsch and Harold [31] suggest two dimensions of workgroup influence – a) normative and b) informational influence. *Normative influence* (NI) refers to an individual's tendency to conform to group members' expectations. It implies that a person's decision to use a methodology is influenced either a) by the user's motivation to conform to the opinions of his work environment (e.g., colleagues, supervisor) in order to realise a reward or avoid a punishment mediated by them, or b) by the user's motivation to satisfy his notion of self-definition by doing what his or her peers (whom he or she wants to be like) do [32]. On the other hand, *informational influences* (II) refer to the tendency to perceive information gained from others as indicative of reality [31], and implies that a person's decision to use a methodology is influenced by the information provided by "mediums of knowledge" such as experts or publications on the topic [32]. Informational influence is indicative of uncertainty on the part of the influenced. In other words, an individual relies on information from others to make informed choices and to reduce uncertainty regarding the "soundness" of his intention to use an ISPM methodology about which he himself has little knowledge. A number of studies have explored these theoretical mechanisms and have found significant support for the ability of workgroup influences to affect a person's intention to act in a particular manner [30,33].

Self-beliefs

In the 1950s, coinciding with the zenith of behaviouristic influence, the "humanistic revolt" in psychology called for renewed attention to inner experience, to internal processes and to humans' self-relevant perceptions. Since then, there has been a resurgence of interest in this field, confirmed by numerous studies, to promote an emphasis on the importance of healthy and positive self-perceptions. Bandura [34] suggests that such self-perceptions involve processes of self-reflection, through which individuals are able to evaluate their experiences and thoughts, and determine what they will do with their knowledge and skills, i.e. their competence. Judgment of one's personal competence reflected in one's self-beliefs therefore not only determine what a person decides to do but also "…how much effort people will expend on an activity, how long they will persevere when confronting obstacles, and how resilient they will prove in the face of adverse situations" [35]. Consequently, in the context of our study, the more positive the self-beliefs, the stronger the intention to use an ISPM methodology, the greater the effort invested to use it, and the stronger the persistence and resilience. Two types of such self-beliefs have been especially dominant in motivation research — self-efficacy and self-concept beliefs.

Based on Bandura's [34] research, *self-efficacy* — a core construct in his social cognitive theory — refers to the belief that one has the capability to perform necessary actions in order to be able to use an ISPM methodology. In the context of this study, *perceived self-efficacy* (PSE) refers to the degree to which a person believes that using a particular methodology would be a) free of physical and mental effort and b) easy to learn [22]. It is important to note that self-efficacy judgments are very task and situation specific. Individuals make use of these judgments in reference to some very specific goals and characteristics of ISPM usage that cannot be generalised to other domains [34].

While self-efficacy beliefs are sensitive to contextual factors, *perceived self-concept* (PSC) beliefs are general or domain specific feelings of self-image. Self-efficacy and self-concept represent different views of oneself. The difference between the two constructs lies primarily in the notion that self-efficacy is *task-specific*, whereas self-concept is *domain-specific* i.e. "self-concept is measured at a more general level of specificity and includes the evaluation of such competence and the feelings of self-worth associated with the behaviours in question" [35]. For example, in the domain of mathematical academic performance, a typical self-efficacy task-specific question is, "How confident are

you that you can successfully solve … equation", which differs from the general domain specific item, "I am good at working with numbers".

Although a number of studies in the past two decades have examined self-efficacy and self-concept individually, few researchers have explored the relationships between them [35]. As is the case with behavioural decisions, it is likely that different situations call forth different self-beliefs. In the context of our study, we suggest that when individuals are familiar with task demands (i.e. what is required to successfully use an ISPM methodology), they may call on the task-specific *self-efficacy* beliefs to help them decide whether to use the methodology or not [35]. But when task demands are unfamiliar (for e.g. because the user has never used ISPM methodologies before and therefore cannot judge the skills required to master the methodology), they generalise from prior attainments that are perceived as similar to the required task [34]. So, they call upon the domain-specific *self-concept* and gauge their perceived competence with their self-concept beliefs, which they consider to more closely correspond to the novel requirements of using a ISPM methodology. (Consider, for example, the person's self-concept that he is good at strictly follow procedures. If a person concludes that he is generally good at following rules, he might also conclude that he will be able to use the specific ISPM methodology. This might be the case because, although not familiar with the specific requirements of using an ISPM, the individual does know that like other methodologies an ISPM requires the user to rigorously follow structured procedures, an act that he considers himself to be good at). The reason why we attempt to study the mutual effect of self-efficacy and self-concept is not to maximise the explanatory power of our model but rather to fill this important gap in existing literature. The *empirical focus* of this argument (self-efficacy vs self-concept) centres on the question of which self-belief provides the greater explanation and prediction of an individual's intention to use a methodology; the *conceptual focus* centres on which beliefs individuals attend to as they go about the business of day-to-day living [35].

4.3 Organisational Characteristics

Whereas self-belief is understood as the user's confidence in his ability to independently use a methodology (reflecting an internal locus of control), social psychology literature suggests that there is another dimension to the effective use of a methodology. This dimension is termed *facilitating conditions* – the user's perceived control over external resources that are needed to use a methodology (reflecting an external locus of control) [36]. As such, facilitating conditions can be understood as the degree to which a user believes that organisational resources are available that will help him use a methodology. In the context of our research, these desirable organisational resources can be considered as support offered by organisational units (such as the PMO or top management) in the form of guidance in the correct usage of methodologies, or political backing. The more a user believes that he or she can get such external resources when he or she needs them, the more confidence he or she will have in successful usage, and the more inclined he or she will be to use that methodology.

4.4 Automatic Behaviour

Plato theorised that "the human mind possessed three distinct faculties: *cognition* or knowing, *emotion* or feeling, and *conation* or willing" [37]. Two of these constructs, cognition and conation, represent people's conscious (intentional) behaviour when trying to explain and predict the use of methodologies. While past research in the field of human usage behaviour has focused primarily on understanding an individual's *planned* (i.e. intentional) decision-making, we also need to consider a person's subconscious (automatic) behaviour, also known as habits, which refer to the non-intentional, automatically inculcated reactions [38]. Habit is portrayed "…as a well-learned action sequence, originally intentional, that may be repeated as it was learned without conscious intention when triggered by environmental cues in a stable context" [38]. In the context of using a new ISPM methodology, we suggest that individuals in organisations might be reluctant to change their habits, which they have learned unconsciously through past repetitions, and might therefore be unwilling to adopt new methodologies. As such, including the habit construct in our behavioural model adds further explanatory power for methodology usage.

4.5 Moderating Influence of Personal Characteristics

Researchers, attempting to understand and predict behaviour with the help of causal models, as is the case with positivistic confirmatory research, base it (usually unknowingly) on the philosophic idea of *Determinism* – a view that every event, including human cognition, behaviour, decision and action is causally determined by an unbroken chain of prior occurrences. Based upon this determinism of human behaviour, the theory of orectic psychological determinism states that people must always act upon their greatest desire and needs [39]. To do otherwise would be irrational. As such, based on needs theory (e.g. [11], [12], [40]) in humanistic psychology, individuals are expected to use a ISPM methodology based on their perceptions that it will enable them to fulfil their specific needs. Many definitions of basic needs have been proposed, of which the one provided by Ryan and Deci [41] is most consistent with the scope of this study. They indicate that "a basic need, whether it be a physiological need or a psychological need, is an energizing state that, if satisfied, conduces toward health and well-being but, if not satisfied, contributes to pathology and ill-being" [41]. This implies that the factors that will be most influential in helping an individual decide for or against the adoption and use of an ISPM methodology are those that satisfy his basic needs. The inability to do so might result in serious discomfort, and this dissatisfaction might be visible in the individual's rejection of the particular methodology. Needs of an individual are thus expected to play a *moderating* role (as depicted in our research model in Figure 1) and influence the explanatory power of the determinants of intention to use ISPM methodology.

In our research, we specifically focus on moderating effects because – besides the examination of direct effects – scholars are increasingly seeking to understand complex relationships [42]. While the literature emphasises the need to take moderation effects [43], its neglect has led to a lack of relevance as "…relationships that hold true independently of context factors are often trivial" [42]. For the purpose of our study, we employ Murray's theory of psychogenic needs [12], and Reiss's theory of

16 basic desires [44] as these are considered the most fundamental and comprehensive list of underlying psychological human needs and motivational processes:

4.5.1 *Need for Affiliation (nAffi*) is the desire to achieve acceptance from one's social surroundings [12]. Individuals with a high need for affiliation tend to enjoy being with other people, making friends, and maintaining personal relationships. Affiliation-oriented employees tend to gravitate towards behaviors that allow them to develop warm and caring relationships with other employees. In a work environment, materialistic endowments such as rewards, promotion, gifts and praise from peers have been found, in a number of studies, to be conveyors of, and adequate substitute for, positive interpersonal relationships and feelings of acceptance [27]. Additionally, since individuals with high nAffi seek to develop strong social relationships with individuals in their work environment, they will more likely comply with requests, appeals and influences of their seniors, peers and colleagues in order to gain their acceptance. Based on this reasoning, nAffi is expected to have a moderating effect on the strength of the effect of PMV → ITUM, and NI → ITUM.

4.5.2 *Need for Achievement (nAch)* refers to an individual's desire to do things better, accomplish difficult tasks, overcome obstacles, become an expert and achieve high performance standards, or a need for significant task related accomplishment [12]. People high in nAch aspire to accomplish difficult tasks where success depends primarily on their efforts. The more complex a task is, the more gratification/satisfaction people with high nAch are expected to feel, since being successful at tasks in which others have failed symbolises and communicates personal competence. Individuals with high nAch are expected to expend more effort, persevere longer when confronted with obstacles and show resilience in the face of adverse situations [35]. Such individuals are more focused on internal motivation and personal achievement rather than external rewards and recognition. As such, employees with a high nAch will only use a methodology if they can be convinced that the methodology will enable them to achieve high performance, productivity and become good at their job. We therefore propose that nAch will have a moderating effect on the strength of the effect of PUV → ITUM.

4.5.3 *Need for Cognition (nCog)* is the desire for knowledge and reasoning [12,44], as well as the need to explore and discover. They tend to be information seekers, engage in and enjoy effortful cognitive activity. Individuals high in need for cognition naturally tend to seek, acquire, think about and reflect back on information and experiences by experimenting and exploring, to make sense of a newly implemented methodology [45]. Therefore, people high in the nCog are more likely to want to try out a new ISPM methodology because they enjoy and find this process of exploring and understanding new concepts for themselves highly satisfying. Consequently, we expect nCog to will have a moderating effect on the strength of the effect of PHV → ITUM. Neglecting to integrate nCog in past research might explain why many empirical studies in the field of MIS report that hedonic value has only weak or no effect on human motivation.

Empirical research has shown that the above-mentioned needs are largely unconnected to one another [44,46]. Although the list of needs in the literature is extensive, we consider these three needs to be representative of the most fundamental high-level primary needs in the context of influence tactics, in the sense of being innate or "hard-wired" [46]. Other secondary needs can be derived from these high-level primary needs. For example, Murray's *need for play, need for curiosity*, and *need for understanding* may be attributed to *nCog*, the *need for contrarience*, and the *need for acquisition* may be derived from the *nAch*. The *need for family* – as proposed by Reiss [44] – and the *need for social recognition* may be attributed to the *nAffi*, and the *need to compete or win* can also be derived from the *nAch* [46]. Another reason to study fewer needs (rather than more) is related to the value of a parsimonious approach: as the list of needs increases, the utility of the approach diminishes. A long, unwieldy list of needs is precisely the reason why earlier needs-related theories fell out of favour [47].

4.5.4 *Experience (EXP)* is defined as the degree of knowledge or skill in the use of ISPM methodologies that is acquired over time through practical use, and has a significant impact on a person's behaviour [23]. Experience has been reported in a number of conflicting studies to affect the degree to which individuals use methodologies. For example, Fitzgerald [2] found that experienced software developers were less likely to follow a methodology rigorously, whereas less experienced developers were more likely to do so. On the other hand, Leonard-Barton [48] suggests that experienced developers are more likely to use a methodology. However, Kozar [49], and Lee and Kim [50] report that more experienced developers are less likely to follow methodologies and procedural formalisation, partly because of the accumulation of systems development know-how among them. Consequently, conflicting results reported in previous studies warrant a deeper investigation in how individuals' experience might affect their intention to use an ISPM methodology.

Bandura [34] suggested that one's experiences are the most influential source of self-beliefs. According to him, positive experiences increase one's self-confidence in one's abilities and, as such, lead to positive self-beliefs. Research based on experiential learning theory and social cognitive theory, especially in the context of the development of managerial skills, also states that work experience leads to increased organisation-based self-beliefs [51]. This implies that for individuals with *high* experience in the use of ISPM methodologies, perceived self-efficacy will have high explanatory power (since self-efficacy is task-specific, individuals who have had prior experience with the use of ISPM methodologies are in a better position to judge if they have the necessary skills to use the methodology). On the other hand, for individuals with *low* or *no* prior experience, perceived self-concept will have higher explanatory power, since task demands are unfamiliar to them. Furthermore, as the intensity of a habit is a function of past activities and is developed over time through repeated use, we expect that habit is stronger in the case of experienced users. Experience is also expected to influence the effect of workgroup pressure on methodology usage, since inexperienced users are more likely to consult information sources and adopt the views of experienced seniors and colleagues whom they consider to be experts in the correct usage of the methodology [31,17]. We therefore include the experience construct in the conceptual research model as a moderator that affects the relationship between intention to use ISPM

methodologies and workgroup influences, self-beliefs and automatic behaviour. The related research hypotheses are summarised in Table 1. The table also provides an overview of some studies that have used highly validated instruments to operationalise the constructs of our conceptual model.

Table 3. Research hypotheses and prior operationalisation of respective constructs

H1a, H1b, H1c: PUV[b], PHV[c], PMV[d] are positively associated with ITUM[a]
H2a, H2b: NI[e], II[f] are positively associated with ITUM[a]
H3a, H3b: PSC[g], PSE[h] are positively associated with ITUM[a]
H4: PFC[i] is positively associated with ITUM[a]
H5: HA[j] is negatively associated with ITUM[a]
H6a, H6b: The influence of NI and PMV on ITUM will be moderated by *nAffi*[k] so that the effect will be stronger for individuals with the specific need.
H6c: The influence of PUV on ITUM will be moderated by *nAch*[k] so that the effect will be stronger for individuals with the specific need.
H6d: The influence of PT on ITUM will be moderated by *nCog*[l] so that the effect will be stronger for individuals with the specific need.
H7a, H7b, H7d, H7e: The influence of NI, II, PSC, PSE, HA on ITUM will be moderated by *EXP*[m] so that the effect will be stronger for individuals with more experience.
H7c: The influence of PSC on ITUM will be moderated by EXP[m] so that the effect will be stronger for individuals with less or no experience.

[a][17,23]; [b][22,25,24,23]; [c][26,28,25,52]; [d][53,27]; [e][17,23]; [f][54,55]; [g][56]; [h][57,58]; [i][36,24,23]; [j][33]; [k][59,60]; [l][45]; [m][17,23]

5. LIMITATIONS AND NEXT STEPS

Although, our proposed model examines the methodology acceptance issue in a holistic manner, the present research has some limitations that should be noted. Firstly, the exploratory field study was conducted in a single organisation and country. Although the company is a multi-national organisation with operations all over the world, there might be structural and cultural influences that vary in different market sectors and countries, and need to be taken into consideration when evaluating the consistency of our findings. Secondly, our study of intention to use might be a particular limitation of this research because intentions (even though they play a major role in determining actual use) do not always lead to actual use. As such, future research might be able to build upon our findings and study actual document use. Regarding operationalization of the proposed constructs there might be a possibility that prior instruments might not be suitable to establish appropriate levels of discriminant validity in the context of our study and therefore new scales might need to be developed.

In developing the initial set of items, we will follow the advice of Straub [61] and employ a rigorous step-by-step iterative process, as well as utilise the existing literature (see Table 3 for an overview of the prior operationalization of constructs). After obtaining the initial battery of items, two researchers will conduct expert interviews with six subject matter experts (three academics and three practitioners) to obtain specific information as to whether the initial items are comprehensible, valid and complete [61]. In order to further improve content and construct validity, we will subsequently conduct a Q-sorting and item ranking in two rounds. In the final step, the questionnaire will be subject to a pre-test, based on a convenience sample with individuals who represent the target population. The final survey instrument will be web-based, administered to a diverse population of ISPM methodology users, to collect quantitative data, needed for testing the model and hypotheses. In order to understand cultural influences, data will be collected from the USA, Germany, Austria, Switzerland and India. We will attempt to include more countries, especially developing and Asian nations such as Japan, China, as well as African nations, as research based on Hofstede's cultural dimensions [62] has shown that individuals from these nations, when compared to Western nations, are governed by different attitudes, preferences and norms.

6. CONCLUSION

Our work seeks to further the research on individual acceptance and adoption of ISPM methodologies by unifying the theoretical perspectives on *cognitive* and *automatic behaviour* and the *needs of individuals* within a single model. Based on validated theories - particularly from the fields of sociology and psychology - and an exploratory field study, we propose a conceptual model. This research model holds that personal traits of individual – especially their needs and experience – determine that determinants of ISPM methodology acceptance has a larger effect on the individual's intention to actually use the methodology. The proposed multidimensionality of a methodology's value represents a departure from traditional operationalisation (which is based solely on task-oriented advantages) and might reveal more complex and until now unknown interaction effects on human behaviour, especially in regard to the use of new methodologies. Furthermore, the mutual study of the influence of self-efficacy and self-concept beliefs is an attempt to provide much needed conceptual clarification on which self-belief is a stronger predictor of methodology adoption, and under which circumstances. While we propose that the predictive power of the self-beliefs varies with individuals' experience with ISPM methodology use, future research should attempt to dig deeper and find further factors that might help to understand the theoretical functioning of the two self-beliefs.

The proposed study of the interaction effects of needs and experience from a temporal point of view is a new approach. While needs are long-lived traits, experience changes gradually. Our findings might have major implications not only for the MIS research community but also for related fields in that it might be able to explain a) how needs change over time with experience for men and women, and b) how these changes determine which determinants of intention to use a methodology becomes more important over time with experience. Human needs have always played a key role in organisational development, and the proposed study is an attempt towards "humanising" organisational ISPM methodologies [63], that is, to enable organisations to be more responsive to human concerns when developing and implementing new methodologies. However, our study of intention to use might be a particular limitation of this research because intentions (even though they play a major role in determining use) might not always lead to actual use. As such, future research might be able to build upon our findings and study actual documented use.

Our research also has significant implications for practitioners. Each of the proposed constructs reveals a different aspect of human behaviour and personality, and each can serve as a point of attack for organisations in their attempts to steer them in the desired direction [30]. Our findings could help organisations to manage the selection, development, introduction, adoption and use of new methodologies. We propose that future research should study the determinants of the constructs identified in this study, as well as the interrelationships between them. Another very promising field of focus is cultural influences on human behaviour. Although the understanding of cultural influences has been repeatedly emphasized by top journal editors – e.g., Straub [64] – it is seldom incorporated in research, generally because of the difficulty of data collection. If successful in collecting sufficient data for statistical analysis from a wide range of *different types* of cultures – categorised by Hofstede [62] – our study, as proposed, will further improve the generalisability of our findings, as well as seek to reveal new avenues and "blue ocean ideas" [64] for future research. A better understanding of these determinants would enable us to design organisational interventions that would increase new ISPM methodology usage in order to improve productivity and quality, as well as to reduce effort.

In conclusion, user acceptance of ISPM methodology remains a complex and elusive, yet extremely important, phenomenon. Past research has made progress in unravelling some of its mysteries. The development and testing of our model seeks to advance theory and research on this crucial matter.

7. REFERENCES

[1] Cicmil S., Đorđević Z., and Zivanovic S., 2009, "Understanding the adoption of project management in Serbian organizations: Insights from an exploratory study.," Project Management Journal, **40**(1), pp. 88-98.

[2] Fitzgerald B., 1998, "An empirical investigation into the adoption of systems development methodologies," Inf. Manage., **34**(6), pp. 317-328.

[3] Glass R. L., 1999, "A Snapshot of Systems Development Practice," IEEE Softw., **16**(3), pp. 112-111.

[4] Russo N. L., Hightower R., and Pearson J. M., 1996, "The Failure of Methodologies to Meet the Needs of Current Development Environments," Proceedings of the British Computer Society's Annual Conference on Information System Methodologies, pp. 387-393.

[5] Teodoro M. P., 2009, "Bureaucratic Job Mobility and The Diffusion of Innovations.," American Journal of Political Science, **53**(1), pp. 175-189.

[6] Winter M., Smith C., Morris P., and Cicmil S., 2006, "Directions for future research in project management: The main findings of a UK government-funded research network," International Journal of Project Management, **24**(8), pp. 638-649.

[7] Wynekoop J. L., and Russo N. L., 1995, "Systems development methodologies: unanswered questions," Journal of Information Technology (Routledge, Ltd.), **10**(2), p. 65.

[8] Hardgrave B. C., Davis F. D., and Riemenschneider C. K., 2003, "Investigating Determinants of Software Developers' Intentions to Follow Methodologies.," Journal of Management Information Systems, **20**(1), pp. 123-151.

[9] Riemenschneider C. K., Hardgrave B. C., and Davis F. D., 2002, "Explaining Software Developer Acceptance of Methodologies: A Comparison of Five Theoretical Models," IEEE Transactions on Software Engineering, **28**(12), pp. 1135-1145.

[10] Pfleeger S. L., 1999, "Understanding and improving technology transfer in software engineering.," Journal of Systems & Software, **47**(2/3), p. 111.

[11] Maslow A. H., 1954, Motivation and Personality, Harper & Brothers.

[12] Murray H. A., 1938, Explorations in Personality, John Wiley & Sons Inc.

[13] Triandis H. C., 1980, "Values, attitudes, and interpersonal behavior," Nebr Symp Motiv, **27**, pp. 195-259.

[14] Khalifa M., and Verner J. M., 2000, "Drivers for Software Development Method Usage.," IEEE Transactions on Engineering Management, **47**(3), p. 360.

[15] Vickers M. H., 1999, "Information technology development methodologies.," Journal of Management Development, **18**(3), p. 255.

[16] Hardgrave B., and Johnson R., 2003, "Toward an information systems development acceptance model: the case of object-oriented systems development," Engineering Management, IEEE Transactions on, **50**(3), pp. 322-336.

[17] Venkatesh V., and Davis F. D., 2000, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies.," Management Science, **46**(2), p. 186.

[18] Warner K. E., 1974, "The Need for Some Innovative Concepts of Innovation: An Examination of Research on the Diffusion of Innovations.," Policy Sciences, **5**(4), pp. 433-451.

[19] Gerring J., 2006, Case Study Research: Principles and Practices, Cambridge University Press.

[20] Eisenhardt K. M., 1989, "Building Theories from Case Study Research.," Academy of Management Review, **14**(4), pp. 532-550.

[21] Rogers E. M., 2003, Diffusion of Innovations, 5th Edition, Free Press.

[22] Davis F. D., 1989, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.," MIS Quarterly, **13**(3), pp. 319-340.

[23] Venkatesh V., Morris M. G., Davis G. B., and Davis F. D., 2003, "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly, **27**(3), pp. 425-478.

[24] Venkatesh V., 2000, "Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model.," Information Systems Research, **11**(4), p. 342.

[25] van der Heijden H., 2004, "User Acceptance of Hedonic Information Systems," MIS Quarterly, **28**(4), pp. 695-704.

[26] Babin B. J., Darden W. R., and Griffin M., 1994, "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value.," Journal of Consumer Research, **20**(4), pp. 644-656.

[27] Belk R. W., 1985, "Materialism: Trait Aspects of Living in the Material World.," Journal of Consumer Research, **12**(3), pp. 265-280.

[28] Davis F. D., Bagozzi R. P., and Warshaw P. R., 1992, "Extrinsic and Intrinsic Motivation to Use Computers in

the Workplace.," Journal of Applied Social Psychology, **22**(14), pp. 1111-1132.
[29] Arnolds C. A., and Boshoff C., 2000, "Does higher remuneration equal higher job performance?: an empirical assessment of the need-progression proposition in selected need theories.," South African Journal of Business Management, **31**(2), p. 53.
[30] Ajzen I., 1991, "The theory of planned behavior," Organizational Behavior and Human Decision Processes, **50**(2), pp. 179-211.
[31] Deutsch M., and Harold G. B., 1955, "A Study of Normative and Informational Social Influences on Individual Judgment," Journal of Abnormal and Social Psychology, **51**(3), pp. 629–636.
[32] Burnkrant R. E., and Cousineau A., 1975, "Informational and Normative Social Influence on Buyer Behavior.," Journal of Consumer Research, **2**(3), pp. 206-215.
[33] Lumayem M., and Hirt S. G., 2003, "Force of Habit and Information Systems Usage: Theory and Initial Validation.," Journal of the Association for Information Systems, **4**, pp. 65-95.
[34] Bandura A., 1986, "Social foundations of thought and action: a social cognitive theory.," Social foundations of thought and action: a social cognitive theory.
[35] Pajares F., 1997, "Current directions in self-efficacy research," Advances in motivation and achievement, **10**, pp. 1–49, 1–49.
[36] Bhattacherjee A., Perols J., and Sanford C., 2008, "Information Technology Continuance: A Theoretical Extension and Emperical Test.," Journal of Computer Information Systems, **49**(1), pp. 17-26.
[37] Hirschman E. C., and Holbrook M. B., 1982, "Hedonic Consumption: Emerging Concepts, Methods and Propositions.," Journal of Marketing, **46**(3), pp. 92-101.
[38] Ortiz de Guinea A., and Markus M. L., 2009, "Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use.," MIS Quarterly, **33**(3), pp. 433-444.
[39] Bolles R. C., 1963, "Psychological Determinism and the Problem of Morality," Journal for the Scientific Study of Religion, **2**(2), pp. 182-189.
[40] Alderfer C. P., 1972, Existence, Relatedness, and Growth; Human Needs in Organizational Settings, Free Press.
[41] Ryan R. M., and Deci E. L., 2000, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.," American Psychologist, **55**(1), p. 68.
[42] Henseler J., and Fassott G., 2010, "Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures," Handbook of Partial Least Squares, pp. 713-735.
[43] Chin W. W., Marcolin B. L., and Newsted P. R., 2003, "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study.," Information Systems Research, **14**(2), pp. 189-217.
[44] Reiss S., 2004, "Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires.," Review of General Psychology, **8**(3), pp. 179-193.
[45] Cacioppo J. T., Petty R. E., Feinstein J. A., Blair W., and Jarvis G., 1996, "Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition," Psychological Bulletin, **119**, pp. 197–253.
[46] Sun R., 2009, "Motivational Representations within a Computational Cognitive Architecture," Cognitive Computation, **1**(1), pp. 91-103.
[47] Deci E. L., 2000, "The Darker and Brighter Sides of Human Existence: Basic Psychological Needs as a Unifying Concept.," Psychological Inquiry, **11**(4), pp. 319-338.
[48] Leonard-Barton D., 1987, "Implementing Structured Software Methodologies: A Case of Innovation in Process Technology.," Interfaces, **17**(3), pp. 6-17.
[49] Kozar K. A., 1989, "Adopting Systems Development Methods: An Exploratory Study.," Journal of Management Information Systems, **5**(4), pp. 73-86.
[50] Lee J., and Kim S., 1992, "The relationship between procedural formalization in MIS development and MIS success: a contingent analysis," Inf. Manage., **22**(2), pp. 89-111.
[51] Brutus S., Ruderman M. N., McCauley C. D., and Ohlott P. J., 2000, "Developing from Job Experiences: The Role of Organization-based Self-Esteem.," Human Resource Development Quarterly, **11**(4), pp. 367-80.
[52] Lesser J., and Madabhushi L., 2001, "Measurement of Consumer Intrinsic Motivation: Exploratory Assessment of Its Two Primary Dimensions and Theoretical Bases," Marketing Management Journal, **11**(2), pp. 81-96.
[53] Belk R. W., 1984, "Three scales to measure constructs related to materialism: reliability, validity, and relationships to measures of happiness," Advances in Consumer Research, **11**(1), pp. 291-297.
[54] Bearden W. O., Netemeyer R. G., and Teel J. E., 1989, "Measurement of Consumer Susceptibility to Interpersonal Influence.," Journal of Consumer Research, **15**(4), pp. 473-481.
[55] Bearden W. O., Netemeyer R. G., and Teel J. E., 1990, "Further Validation of the Consumer Susceptibility to Interpersonal Influence Scale.," Advances in Consumer Research, **17**(1), pp. 770-776.
[56] Rosenberg M., 1965, Society and the adolescent self-image, Princeton University Press.
[57] Compeau D. R., and Higgins C. A., 1995, "Computer Self-Efficacy: Development of a Measure and Initial Test.," MIS Quarterly, **19**(2), pp. 189-211.
[58] Lee C., and Bobko P., 1994, "Self-Efficacy Beliefs: Comparison of Five Measures.," Journal of Applied Psychology, **79**(3), pp. 364-369.
[59] Edwards A. L., 1959, Edwards personal preference schedule, Psychological Corporation, New York.
[60] Frs R. H., and Knox A. B., 1972, "A Validity Study of Scales to Measure Need Achievement, Need Affiliation, Impulsiveness, and Intellectuality," Educational and Psychological Measurement, **32**(1), pp. 147-154.
[61] Straub D. W., 1989, "Validating Instruments in MIS Research.," MIS Quarterly, **13**(2), pp. 147-169.
[62] Hofstede D. G., 2003, Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations, Sage Publications, Inc.
[63] Alderfer C. P., 1977, "Organizational Development,"

Annual Review of Psychology, **28**, p. 197.

[64] Straub D. W., 2009, "Creating Blue Oceans of Thought Via Highly Citable Articles.," MIS Quarterly, **33**(4), pp. iii-vii.

The MATH of Internet Adoption: Comparing Different Age-Groups

Bjoern Niehaves
European Research Center for Information Systems
University of Muenster
Leonardo-Campus 3
48149 Muenster
+492518338087
bjoern.niehaves@ercis.uni-muenster.de

Ralf Plattfaut
European Research Center for Information Systems
University of Muenster
Leonardo-Campus 3
48149 Muenster
+492518338081
ralf.plattfaut@ercis.uni-muenster.de

ABSTRACT

Modern societies share two common trends: First, elderly people form a strongly growing group in the population (societal aging) and, second, the importance of information and communication technology is growing rapidly. However, the elderly are often excluded from benefiting from IT-enabled service delivery: An age-related digital divide exists. Current research lacks understanding what reasons prevent elderly to use the internet. Therefore, this paper examines the intention to use the internet in a private manner among the elderly. For higher explanatory power we also included two other age-groups (G1: <40; G2: 40-59; G3: >59). Here, we build a survey instrument based on the Model of Acceptance of Technology in Households (MATH) and test the model against comprehensive survey data (n=501). We find out that MATH is able to explain between 42% and 81% of the variance in private internet usage intention. Moreover, several differences in driver for usage intention exist, e.g. was the importance of applications for fun much higher in the first age group than among the other. Potentially fruitful avenues for future research are discussed.

Keywords

Model of Acceptance of Technology in Households, MATH, digital divide, demographic change, elderly

1. INTRODUCTION

Western Societies share two common trends. First, the share of elderly people is constantly rising [1,2]. Today, about 17% of Europe's population is older than 65 years. This share will rise up to 28% by 2050 [3]. Using this measure Germany has been among the oldest countries in the world and is the oldest country in the European Union with a share of 20.4% [4]. This trend is mainly due to better health care and food supply which results in a rising life expectancy. Moreover, the fertility rate defined as children per woman is only about 1.5 in the European Union [4].

Second, the importance of information, information processing, and communication is growing in industrialized countries. This phenomenon has been named "information society" [5-7]. In today's information societies electronic communication and commerce, the exchange of information, and the usage of information technology becomes more and more important. This development does not only occur in the workplace but also in private life.

However, even in high developed countries information and communication technologies as the internet are not used by everyone. A digital divide between those that use and those that do not use the internet exists. In the literature several reasons for the digital divide can be found, e.g. social status, ethnicity, education, income, or age [8-11].

Moreover, information and communication technologies are used as a measure to ease the life of elderly people. Here, concepts as ambient assisted living (AAL) aim at increasing the autonomy, self-confidence, and mobility of the elderly. However, AAL projects always require the ability to use modern technologies. Hence, an understanding of drivers for internet usage can help to close the digital divide and, thus, to prepare today's and tomorrow's elderly for AAL [12]. Subsuming, our research question is:

RQ: Which differences in the factors and their importance for private internet usage intention exist in different age groups?

This paper is structured as follows. First, we will present our theoretical background. We will shortly present our theoretical background consisting of digital divide and technology acceptance research. Afterwards, we will present the Model of Acceptance of Technology in Households (MATH) which will be used as our research model. In the next section we will show our research methodology including data collection and analysis. Then, we will present our results which will be discussed in the following section. The paper closes with conclusions, limitations, and future research.

2. THEORETICAL BACKGROUND

2.1 Digital Divide Research

Since more than a decade ago, digital divide has been established as a major research theme. The field was opened by first contributions around the year 2000 (A short introduction into digital divide and its history can be found in [13]). Generally, digital divide refers to the gap between, first, those who do have effective access to ICT ad use the potentials of these technologies and, second, those who do not have this access or those who do

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

not use corresponding technologies. In digital divide research two different streams are observable with a difference in the unit of analysis. On the one hand, the unit of analysis can be the difference in technology adoption between different countries. Typically, developed countries have a high rate of (both early and late) adopters while developing countries have lower rates. Here research is about the influence of different variables, as culture, income, education, on the adoption [13,14]. On the other hand, the unit of analysis can be single users or groups of users in a specific region or country. Here, several groups of people are excluded from benefiting from the merits of ICT due to certain factors [15,16]. Examples for such groups are people with migration background, elderly people, or less educated citizens. In this specific study we understand digital divide as an emerging polarization phenomenon in a specific society (here: a German municipality, see below), which creates a gap between those people who do have access to and use the potentials of ICT (on-liners) and those who do not (off-liners) [17].

Digital divide research often focuses on the access to and usage of the internet. Several theoretical contributions and models provide explanation for internet usage behavior (often with concentration on the usage in a private manner; here: private internet usage). Early research on this theme (first generation digital divide research) has focused on issues of physical access. Here, researchers and practitioners focused on the provision of computers and internet connection to off-liners. Projects to bridge the digital divide were established and encompassed free internet access at local libraries or comparable centers or the free provision of computers to elderly people [18]. This research was somehow limited in terms of explanatory power. Hence, second generation digital divide research extended this narrow focus on physical access and included factors as motivational or skill access [9,19,20].

2.2 Technology Acceptance Research

The field of technology acceptance research originates in psychology. Here, several theories exist to explain reasoned action [21,22] or planned behavior [23,24]. The idea of individual technology acceptance was prominently introduced into IS by Davis with his Technology Acceptance Model (TAM) [25,26]. TAM consists of two independent and two dependent (lateral) variables. Perceived usefulness and perceived ease of use of a certain technology both influence the behavioral intention to use this technology. The behavioral intention then translates into actual system usage. TAM has been criticized for its lack of falsifiability, its limited explanatory and predictive power, and even its triviality [27,28]. Other models as extend TAM – a prominent example is the Unified Theory of Acceptance and Use of Technology [29]. While UTAUT and TAM are theories to explain technology adoption in workplaces and in private environments, the Model of Adoption of Technology in Households (MATH) focuses on personal technology (in early studies: personal computer) adoption [30-32].

3. RESEARCH MODEL

MATH was created to explain the adoption of technology in households. Its key constructs were derived in a qualitative longitudinal study of personal computer usage in households (Venkatesh and Brown 2001). Later, Brown and Venkatesh (2005) used these constructs and created a comprehensive multi-item measurement model. This model was tested in a quantitative study to predict the adoption of personal computers in households. One of its construct is Utility for work-related use. It is defined as the extent to which using a PC enhances the effectiveness of work-related activities (Venkatesh and Brown 2001, Venkatesh et al. 2003, Aijzen 1991, Davis et al. 1989). As our study focuses only on private internet usage this latent variable was not considered in our study. Moreover, we changed items to measure the workplace referents' influence to also measure the influence of the extended social network (e.g. acquaintances from political or sport activities). According to MATH attitudinal, normative, and control believes influence behavioural intention. We argue that the influences of the independent variables on behavioural intention are moderated by the age of the respondents (see Figure 1).

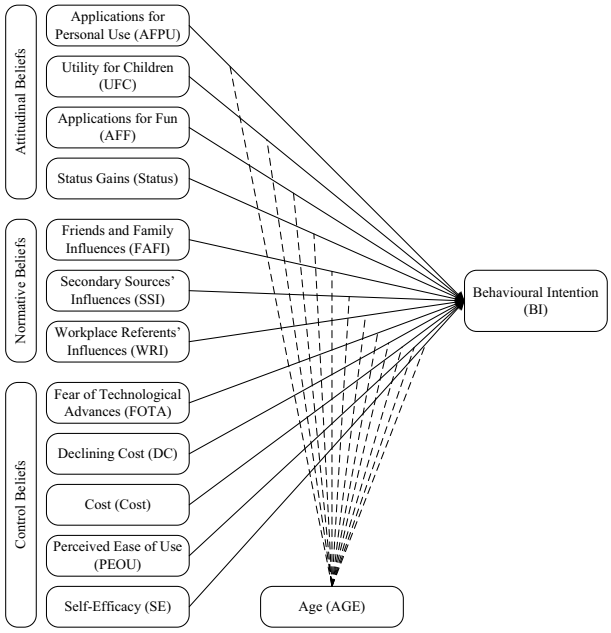


Figure 1: Model of Acceptance of Technology in Households

These moderating effects were modeled using three different groups for analysis (see below). Moreover, it led to the following hypothesis:

1. AFPU: Internet adoption for personal use requires a serious intent in internet usage. We assume that this seriousness is higher in older age groups. Hence, we hypothesize that the influence of AFPU is higher for people aged 40 or higher.
2. AFF: Young people tend to use technology for its own sake. Gaming is usually associated with the younger generation. Hence, we assume that age will moderate the influence of applications for fun so that the influence is higher for young people.
3. FAFI: Young and medium-aged people have bigger social networks. Hence, we hypothesize that the influence of friends and family is more important for them.
4. SSI: The influence of traditional media is suspected to be of greater importance for old people.

5. PEOU: Elderly people are more likely to need the right skills for internet usage. Here, the perceived ease of use plays an important role. Hence, the influence of PEOU will be greater for people age 60 or older.
6. Self-efficacy: In the same way, the influence of self-efficacy will be higher for old people.

For UFC, Status, WRI, FOTA, DC, and Cost we do not hypothesize any influence of Age as these variables seem to be relevant for all age groups.

4. RESEARCH METHODOLOY

4.1 Data Collection

We constructed a questionnaire following our presented research model. The constructs and items are well established in the literature.[1] In a pilot study (n=7, random selection) the respondents gave positive feedback to our questionnaire. Hence, the study did not result in any changes. We used the validated questionnaire to gather data within a medium-sized city located in Western Europe between September and October 2009. We used three unique data-gathering strategies simultaneously: First, we extracted contact data of 1,500 randomly chosen citizens out of the cities resident registration. Each addressee received a personal letter from the mayor announcing the aim of the questionnaire, the questionnaire itself, and a stamped return envelope. Second, we placed additional 1,500 questionnaires at the cities' town-hall and local libraries. Third, we called slightly more than 100 randomly chosen people and interviewed them via phone. Hence, we avoided problems as mentioned by [33]. To lever the response rate we raffled three material prizes among all respondents. Additionally, we held a press conference with the mayor to announce the start of the survey and issued another press release in the middle of the data collection phase. This led to good coverage of our survey in the local media. All in all, we received 501 questionnaires (see section 4.3 for sample demographics). An additional non-response analysis did not reveal any biases.

4.2 Data Analysis

As stated above we employed a paper-based questionnaire to gather our data. Hence, for analysis we entered our data into an online tool. Here we used SPSS 17.0.0 for first analysis. To be able to answer our research question we split the data in three disjoint sets with respect to the respondents' age. G1 (younger) includes the respondents aged 39 or younger, G2 (middle-aged) includes the respondents between age 40 and 59, and G3 (older) covers the older adults (60 years or older). We chose this grouping following [34] as it results in adequate and comparably high numbers in all groups. To further analyze our datasets with regards to the presented research model we employed the partial least squares (PLS) path modeling algorithm [35-37]. To run the corresponding algorithm we used the SmartPLS 2.0 (M3) Software Package [38]. In correspondence to other MATH studies, all constructs were modeled using reflective indicators [32]. While running the PLS algorithm we employed the centroid weighting scheme. The centroid weighting scheme does not tend to slightly overestimate effects as the factor weighting scheme [39]. Our datasets include some missing values (for more details see sample demographics). These missing values were treated using the mean replacement algorithm [40].

4.3 Sample Demographics

Our sample consists of data of 501 respondents. As described above we split the data in three datasets. G1 consists of all respondents aged 39 or younger (n=186, approx. 37% of all subjects). In this group about 2% of all items were missing. The mean age was around 28 years with a standard deviation of 8. The youngest respondent was 13. About 37% of all respondents were male. G2 consists of 199 middle-aged respondents (age between 40 and 59, approx. 40% of all subjects). Here, about 3% of all items were missing. In this age group every possible year of birth occurred. About 36% of all respondents were male. In G3 we subsume 116 respondents aged 60 or older (approx. 23% of all). The oldest respondent was 83. Here, 53% of all respondents were male (see Table 1).

Table 1: Sample demographics

	G1	G2	G3
N	186	199	116
Missing Values	2.22%	2.91%	8.45%
Age - Mean	28.18	48.77	68.03
Age - Std. Dev.	8.09	5.3	5.5
Sex - Male	68	71	62
Sex - Female	118	127	54

5. RESULTS

The constructs used in our study are well known and have been proven to be valid. However, using standardized measures we can show that some minor problems exist with regards to construct validity. The measurement model estimations of the different age groups are presented in Tables 2 to 4. Here, ICR stands for the internal consistency reliability (Cronbach's Alpha). Generally, an ICR above .9 is considered as excellent, one between .7 and .9 as high, and one between .5 and .7 as moderately high [41]. A lower ICR is a signal for problematic construct validity. In our study, only the ICR of workplace referent's influences (WRI) is low for the young age group. This could be due to a different understanding of the studied social group. However, as the corresponding reliability is above .5 for all other age groups, we did not change the items at this stage of research. All other reliabilities are over .5, sometimes even higher than .9. Moreover, all correlations between the constructs (off-diagonal elements in the tables) are lower than the square roots of the shared variance between constructs and their respective measures. This is a good indicator for convergent and discriminant validity [42].

[1] An overview of the items used can be requested from the authors.

	ICR	Mean	S-Dev	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Applications for personal use	.778	4.856	1.363	.824												
2 Utility for children	.892	5.228	1.239	.325	.888											
3 Applications for fun	.802	5.123	1.202	.494	.530	.797										
4 Status gains	.746	4.005	1.272	.285	.178	.374	.785									
5 Friends and family influences	.895	4.595	1.480	.320	.134	.337	.403	.870								
6 Secondary sources' influences	.890	4.483	1.377	.301	.158	.356	.403	.443	.904							
7 Workplace referents' influences	.427	4.272	1.392	.286	.234	.269	.383	.582	.432	.795						
8 Fear of technology advances	.667	3.912	1.318	.216	.087	.049	.159	.197	.169	.167	.668					
9 Declining cost	.666	4.845	1.011	.293	.087	.079	-.081	.062	.202	.045	.227	.776				
10 Cost	.776	2.778	1.300	-.025	-.027	-.001	.081	.016	-.026	.045	-.002	-.238	.789			
11 Perceived ease of use	.652	5.709	.892	.566	.306	.499	.201	.329	.281	.317	.093	.202	-.083	.708		
12 Self-efficacy	.663	6.214	.868	.438	.161	.443	.193	.336	.239	.212	.043	.121	-.176	.695	.774	
13 Behavioral intention	.866	6.598	.948	.444	.160	.485	.196	.243	.230	.236	.171	.169	-.107	.551	.571	.889

Table 2: Measurement Model Estimation for Group 1 (G1)

	ICR	Mean	S-Dev	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Applications for personal use	.736	4.767	1.278	.789												
2 Utility for children	.907	5.327	1.127	.027	.906											
3 Applications for fun	.804	4.408	1.227	.317	.245	.791										
4 Status gains	.677	3.945	1.090	.194	.054	.216	.770									
5 Friends and family influences	.888	4.098	1.462	.197	.098	.255	.306	.864								
6 Secondary sources' influences	.839	4.556	1.208	.300	.061	.193	.332	.461	.869							
7 Workplace referents' influences	.503	4.021	1.469	.269	-.031	.187	.186	.540	.476	.817						
8 Fear of technology advances	.562	3.202	2.425	-.225	-.085	-.210	-.074	-.155	-.166	-.159	.688					
9 Declining cost	.652	4.950	1.021	.309	.042	.039	.132	.114	.035	.057	-.092	.759				
10 Cost	.838	2.937	1.437	.042	-.131	.125	.091	-.057	-.064	-.039	.176	-.213	.867			
11 Perceived ease of use	.712	5.235	.962	.357	.146	.483	.027	.115	.212	.156	-.383	.194	-.116	.733		
12 Self-efficacy	.667	5.926	.991	.417	-.028	.395	.099	.084	.252	.265	-.380	.166	-.005	.642	.781	
13 Behavioral intention	.809	6.358	1.208	.390	-.111	.180	.145	.205	.191	.239	-.320	.249	-.133	.339	.561	.850

Table 3: Measurement Model Estimation for Group 2 (G2)

	ICR	Mean	S-Dev	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Applications for personal use	.791	4.505	1.375	.836												
2 Utility for children	.819	4.961	1.108	.122	.857											
3 Applications for fun	.851	4.330	1.351	.632	.348	.830										
4 Status gains	.742	3.975	1.234	.162	.169	.295	.809									
5 Friends and family influences	.863	4.141	1.445	.293	.129	.293	.284	.839								
6 Secondary sources' influences	.853	4.362	1.421	.182	.353	.376	.449	.324	.879							
7 Workplace referents' influences	.636	3.772	1.442	.269	.172	.338	.247	.371	.318	.853						
8 Fear of technology advances	.713	3.610	3.129	-.211	-.009	-.224	.034	.038	.065	-.016	.565					
9 Declining cost	.674	4.529	.946	.306	.330	.347	.206	.039	.213	.124	-.107	.781				
10 Cost	.763	3.385	1.189	-.018	-.129	-.105	.068	.114	.041	-.056	.052	-.146	.820			
11 Perceived ease of use	.844	4.743	1.329	.553	.271	.737	.276	.112	.276	.353	-.246	.431	-.122	.827		
12 Self-efficacy	.824	4.875	1.642	.524	.245	.648	.295	.235	.436	.439	-.220	.316	-.088	.766	.860	
13 Behavioral intention	.893	5.153	2.031	.487	.196	.550	.253	.326	.449	.514	-.195	.302	-.093	.650	.878	.908

Table 4: Measurement Model Estimation for Group 3 (G3)

We employed bootstrapping (with 500 iterations) using randomly selected sub-samples for testing the significance of the PLS model.[2] In general, the item loadings show that the latent variables are measured by the corresponding items, as almost all items have comparably high loadings (Table 5, please note that AGE is a single-item variable). However, the latent variable FOTA appears to be measured incorrect: In G1 there are low item loadings for FOTA1 and in G2 and G3 for FOTA2. Here, we also analyzed the average variance extracted. Here, we observe high values so that all variables can still be considered valid [41].

[2] Significance is depicted using the asterisk symbol (* means significant on a .95, ** on a .995, and *** on a .999 level).

		G1	G2	G3
AFPU	AFPU1	.875	.891	.901
	AFPU2	.741	.589	.674
	AFPU3	.850	.854	.912
UFC	UFC1	.965	.824	.832
	UFC2	.858	.930	.867
	UFC3	.837	.960	.871
AFF	AFF1	.828	.767	.889
	AFF2	.487	.604	.656
	AFF3	.900	.902	.865
	AFF4	.898	.860	.887
Status	SS1	.750	.713	.898
	SS2	.643	.799	.707
	SS3	.933	.794	.810
FAFI	FAFI1	.906	.884	.897
	FAFI2	.873	.876	.872
	FAFI3	.861	.877	.826
	FAFI4	.841	.816	.756
SSI	SSI1	.923	.813	.868
	SSI2	.910	.905	.902
	SSI3	.878	.888	.867

		G1	G2	G3
WRI	WRI1	.864	.847	.799
	WRI2	.718	.786	.904
FOTA	FOTA1	.303	.812	.642
	FOTA2	.918	-.327	-.444
	FOTA3	.635	.809	.590
DC	DC1	0.89	0.63	0.77
	DC2	0.9	0.84	0.88
	DC3	0.45	0.78	0.67
Cost	COST1	0.92	0.8	0.83
	COST2	0.67	0.9	0.89
	COST3	0.76	0.9	0.74
PEOU	EE1	.836	.849	.907
	EE2	.633	.657	.854
	EE3	.505	.559	.660
	EE4	.807	.827	.865
SE	SE1	.584	.572	.798
	SE2	.825	.822	.863
	SE3	.880	.908	.916
BI	BI1	.955	.921	.929
	BI2	.780	.710	.857
	BI3	.924	.904	.936

The paths of MATH have been proven to be significant to explain behavioral intention to adopt technology in households in previous studies [32]. However, in this application of the model bootstrapping suggests that many relationships are not significant among the different age groups (see Table 6).

Table 5: Path Coefficients (Dependent Variable: BI)

	G1	G2	G3
R^2	.456	.418	.812
AFF	.318 **	-.052	-.091
AFPU	.086	.164 *	.037
Cost	-.035	-.121	-.032
DC	.030	.080	.053
FAFI	-.092	.149 *	.101
FOTA	.124	-.095	-.031
PEOU	.170	-.072	-.008
SE	.305 **	.505 ***	.807 ***
SSI	-.027	-.062	.091
Status	-.016	.052	-.064
UFC	-.154	-.119	-.052
WRI	.095	-.007	.132 **

The coefficient of determination (R^2) is here defined as the proportion of variance explained by the model (and not by random error or non-included constructs). Considering that applications of MATH in other contexts resulted in R^2-values between .50 and .74 [32] our study results shows one considerably high coefficients of determination. In G1 we can explain about 46%, in G2 about 42%, and in G3 about 81% of the corresponding variance in behavioral intention to use the internet. The value for G3 is astonishingly high although we did not model any product terms for moderating effects. Usually moderating effects modeled using product terms result in higher coefficients of determination.

Table 6: PLS-MGA-Results (Dependent Variable: BI)

	G1 vs. G2	G1 vs. G3	G2 vs. G3
AFF	> *	> **	>
AFPU	<	>	>
Cost	>	<	<
DC	<	<	>
FAFI	< *	< *	>
FOTA	> *	>	<
PEOU	>	>	<
SE	<	< ***	< *
SSI	>	<	< *
Status	<	>	>
UFC	<	<	<
WRI	>	<	<

To evaluate whether the path coefficients differ significantly we employed PLS-Multigroup-Analysis (PLS-MGA) as suggested by [43]. PLS-MGA does not require any distributional assumptions and is used with the help of bootstrapping results. Table 7 gives an overview whether path coefficients in a specific group are larger (>) or smaller (<) then the corresponding other. Moreover, significance of this comparison is given as calculated using PLS-MGA.

6. DISCUSSION

Our study includes several findings that are important for theory. First, all but one constructs of our measurement model seem to be reliable and consistent. Only the influence of workplace referents shows a low Cronbach's Alpha in the first age group. Reasons for this could lie in the different understanding of the underlying questions between age groups. Moreover, we changed the questions of this variable to better fit the extended social network.

Second, the further analysis of the measurement model highlights differences in the mean values of the latent variables between the age groups. Apparently, the young age group (G1) believes strongly that the internet offers applications for fun (AFF in G1 is 5.123). The other age groups are lower (AFF in G2 is 4.408 and in G3 4.330). Here, as suspected, younger people use the internet more for fun-related purposes. Moreover, the average perceived influence of friends and family declines with the age (FAFI in G1: 4.595; in G2: 4.098; in G3: 4.141). Apparently, as usage is generally declining with the age pressure from friends is declining as well. The same holds true for the perceived costs of internet usage. Here, the young generation thinks of the internet as cheap while older users see the costs more dominantly (Cost in G1: 2.778; in G2: 2.937; in G3: 3.384). This is in line with the different usage behavior. Studies as well as media coverage report an always-on mentality among young internet users while old users see the internet more as a tool you explicitly have to "turn on" [44]. Two other latent variables support this perception very well. Perceived ease of use (PEOU in G1: 5.709; in G2: 5.235; in G3: 4.743) as well as self-efficacy (SE in G1: 6.214; in G2: 5.926; in G3: 4.875) are both high among the young generation and lower in the older ones. Here, we argue that the big experience of the young generation and availability through multiple devices especially to them are factors for this phenomenon.

Third, analyzing the path models we can see that only a minority of all paths are significant. However, this is in line with previous MATH studies [31,32]. Interestingly, the influence of SE is very high among all age-groups. Apparently, the perceived amount of existing knowledge is a good predictor of the intention to use the internet. Analyzing the differences in the path coefficients using our hypotheses yields the following results: First, AFPU is a significant path in the medium age group. However, the relationship is not significant in all other groups. Moreover, there are no significant differences in importance among the groups. Hence, hypothesis 1 is falsified. Second, AFF has a high and significant influence in the young generation and an even negative influence in both other groups. This difference could be shown to be significant. Hence, our second hypothesis is supported by this study. Third, we hypothesized that the influence of FAFI is higher for young and medium-aged people. However, the path coefficient is only positive in G2 and G3 (significant only for G2). The influence in G1 is significantly smaller than in both other groups. Hence, our third hypothesis is falsified. The influence is highest in the medium age group. Fourth, traditional media (SSI) has only a positive influence on G3. Although this influence is not significant, it is significantly higher than the influence in G2. Hence, hypothesis 4 is partially verified. Fifth, the influence of PEOU is not significant in all age groups. As there are no differences between the groups, H5 is falsified. Sixth, SE has a high and significant impact on BI in all age groups. However, as hypothesized this influence is significantly highest in G3. Hence, H6 is supported by this study. For the impact of UFC, Status, WRI, FOTA, DC, and Cost on BI we did not hypothesize any influence of age. Here, it has to be mentioned that WRI has a significant influence on BI in G3. Moreover, the influence of FOTA is significantly higher for the young group.

Fourth, the coefficient of determination (R^2) is generally in range of prior studies using MATH [31,32]. Although several relations have been shown to be of limited significance the model is able to predict a good share of the variance in usage intention. This holds especially true for the age group of the elderly.

Moreover, our results have several implications for practice. Many public and private organizations start projects and initiatives to bridge the age-related digital divide. These projects follow different ideas and have varying successes. However, especially with regards to future requirements (e.g. for AAL), successful e-Inclusion strategies are needed. Here, organizations should construct their initiatives recognizing the presented results. Decision makers should, for example, think about addressing the social environment of citizens through strong disseminators enrooted in the corresponding milieus.

7. CONCLUSION, LIMITATIONS, AND FUTURE RESEARCH

This paper analyzes influencing factors for the intention to use the internet in a private manner. Here, we presented a research model based on MATH and established six hypotheses. To elaborate on the moderating effects of age on the internet adoption we use three different age groups. With data collected using an extensive survey in 2009 we could analyze the responses of more than 500 different subjects. Here, we used the PLS path modeling method (SmartPLS was the software package used). Our results suggest that MATH is of great use in predicting usage intention among all age groups, especially among the elderly (60 years and older). Our study highlights the importance of self-efficacy for the intention to use the internet: Among all age groups SE had one of the highest influences. Moreover, we showed that this influence is highest in the old age-group. Furthermore, we could show that in the young generation the influence of applications for fun was significantly higher than in all other groups. Hence, we contribute the following findings: First, in the young generation fun is the single most important driver for internet usage. Second, in the old generation self-efficacy plays the most important role. Third, in the medium age-group self-efficacy, friends and families opinion, and applications for personal use form a mixture of influence on behavioural intention.

However, our study is limited to a certain extent due to several issues. First, the representativeness of samples is always open to discussion. Here, it could be questioned whether a sample of 501 respondents is big enough. We argue that our sample was randomly chosen and that a non-response analysis yielded no bias. Second, we gathered our results in one city. While we have a good chance that our sample represents the inhabitants of this municipality, the generalizability to the region or bigger geographical units has yet to be proven. However, we believe that our results will, to a great extent, hold true in other settings in Western European countries as well.

Our paper shows several potentially fruitful avenues for future research. Future studies could aim at testing the generalizability by replication in other social or cultural settings. Up to now it is questionable whether results out of data gathered in one city can be generalized. Here, comparative studies could be valuable as well. Furthermore, new moderating variables could be introduced. These could either be more classical, as gender, ethnicity, or education or be completely different as psychological variables. Moreover, our study had a slight problem in the reliability of the WRI construct. Here, future studies should reassess the usage of the construct and aim at improving it with the help of other items. In the end, this could help to increase explanatory power.

8. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Volkswagen Foundation (Schumpeter-Fellowship of the first author).

9. REFERENCES

[1] G.F. Anderson and P.S. Hussey, “Population aging: a comparison among industrialized countries,” *Health Affairs*, vol. 19, 2000, pp. 191-203.

[2] M. Fougère and M. Mérette, “Population ageing and economic growth in seven OECD countries,” *Economic Modelling*, vol. 16, Aug. 1999, pp. 411-427.

[3] S. Kröhnert, I. Hoßmann, and R. Klingholz, *Die demografische Zukunft von Europa: Wie sich die Regionen verändern*, Deutscher Taschenbuch Verlag, 2008.

[4] Eurostat, “Population, Demography: Main tables,” 2010.

[5] F. Machlup, *The production and distribution of knowledge in the United States*, Princeton, USA: Princeton University Press, 1962.

[6] A.S. Duff, D. Craig, and D.A. McNeill, "A note on the origins of the'information society'," *Journal of Information Science*, vol. 22, 1996, pp. 117-122.

[7] S. Nora and A. Minc, *Die Informatisierung der Gesellschaft [Broschiert]*, Campus Verlag GmbH, 1997.

[8] R. Agarwal, A. Animesh, and K. Prasad, "Social Interactions and the 'Digital Divide': Explaining Variations in Internet Use," *Information Systems Research*, vol. 20, 2009, pp. 277-294.

[9] J.A.G.M. van Dijk, "Digital divide research, achievements and shortcomings," *Poetics*, vol. 34, Aug. 2006, pp. 221-235.

[10] F. Bélanger and L. Carter, "The impact of the digital divide on e-government use," *Communications of the ACM*, vol. 52, 2009, pp. 132-135.

[11] L. Carter and F. Bélanger, "The utilization of e-government services: citizen trust, innovation and acceptance factors," *Information Systems Journal*, vol. 15, Jan. 2005, pp. 5-25.

[12] Ambient Assisted Living Association, "Objectives - Ambient Assisted Living," 2010.

[13] N. Corrocher and A. Ordanini, "Measuring the digital divide: a framework for the analysis of cross-country differences," *Journal of Information Technology*, vol. 17, Mar. 2002, pp. 9-19.

[14] J. James, "Reconstruing the digital divide from the perspective of a large, poor, developing country," *Journal of Information Technology*, vol. 19, Aug. 2004, pp. 172-177.

[15] L. Kvasny and M. Keil, "The challenges of redressing the divide : a tale of two US cities," *Information Systems Journal*, vol. 16, 2006, pp. 23-53.

[16] J. Hsieh, A. Rai, and M. Keil, "Understanding digital inequality: Comparing continued use behavioral models of the socio-economically advantaged and disadvantaged," *Management Information Systems*, vol. 32, 2008, pp. 97-126.

[17] European Commission, *eInclusion@EU: Strengthening eInclusion & eAccessibility across Europe. Analytic framework - eInclusion and eAccessibility priority issues*, 2004.

[18] J.K. Eastman and R. Iyer, "The elderly's uses and attitudes towards the Internet," *Journal of Consumer Marketing*, vol. 21, Jan. 2004, pp. 208-220.

[19] P.J. Dimaggio and E. Hargittai, "From the 'Digital Divide' to `Digital Inequality': Studying Internet Use As Penetration Increases," 2001.

[20] M. Warschauer, *Technology and social inclusion: Rethinking the digital divide*, Cambridge, MA: MIT Press, 2004.

[21] M. Fishbein, "Attitude and the prediction of behaviour," *Readings in attitude theory and measurement*, M. Fishbein, ed., New York: Wiley, 1967, pp. 447-492.

[22] M. Fishbein and I. Ajzen, *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, 1975.

[23] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, Dec. 1991, pp. 179-211.

[24] S. Taylor and P.A. Todd, "Understanding Information Technology Usage: A Test of Competing Models," *Information Systems Research*, vol. 6, Jun. 1995, pp. 144-176.

[25] F.D. Davis, "A technology acceptance model for empirically testing new end-user information systems: Theory and results," 1986.

[26] F.D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, vol. 13, 1989, pp. 319 - 340.

[27] M. Chuttur, "Overview of the Technology Acceptance Model: Origins, Developments and Future Directions," *Sprouts: Working Papers on Information Systems*, vol. 9, 2009.

[28] Y. Lee, K.A. Kozar, and K.R.T. Larsen, "The technology acceptance model: Past, present, and future," *Communications of the Association for Information Systems*, vol. 12, 2003, pp. 752-780.

[29] V. Venkatesh, M.G. Morris, G.B. Davis, and F.D. Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly*, vol. 27, 2003, pp. 425 - 478.

[30] V. Venkatesh and S. a Brown, "A Longitudinal Investigation of Personal Computers in Homes: Adoption Determinants and Emerging Challenges," *MIS Quarterly*, vol. 25, Mar. 2001, p. 71.

[31] S. Brown, V. Venkatesh, and H. Bala, "Household Technology Use: Integrating Household Life Cycle and the Model of Adoption of Technology in Households," *The Information Society*, vol. 22, Sep. 2006, pp. 205-218.

[32] S.A. Brown and V. Venkatesh, "Model of adoption of technology in households: A baseline model test and extension incorporating household life cycle," *MIS quarterly*, vol. 29, 2005, pp. 399-426.

[33] M. Warschauer and T. Matuchniak, "New Technology and Digital Worlds: Analyzing Evidence of Equity in Access, Use, and Outcomes," *Review of Research in Education*, vol. 34, Mar. 2010, pp. 179-225.

[34] S. Czaja and J. Sharit, "Age differences in attitudes toward computers," *Journals of Gerontology:Psychological Sciences*, vol. 53B, 1998, pp. 329-340.

[35] G. Marcoulides, W. Chin, and C. Saunders, *Foreword: A Critical Look at Partial Least Squares Modeling*, 2009.

[36] J. Henseler and G. Fassott, *Handbook of Partial Least Squares*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[37] W.W. Chin and J. Dibbern, *Handbook of Partial Least Squares*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[38] C.M. Ringle, S. Wende, and S. Will, "SmartPLS 2.0 (M3) Beta," 2005.

[39] B. Wilson and J. Henseler, "Modeling Reflective Higher-Order Constructs using Three Approaches with PLS Path Modeling: A Monte Carlo Comparison," *Australian and New Zealand Marketing Academy Conference*, Otago, Australia: 2007.

[40] A.A. Afifi and R.M. Elashoff, “Missing Observations in Multivariate Statistics: I. Review of the Literature,” *Journal of the American Statistical Association*, vol. 61, 1966, pp. 595-604.

[41] P. Hinton, C. Brownlow, and I. McMurray, *SPSS Explained 2nd Edition*, Routledge Chapman & Hall, 2005.

[42] C. Fornell and D.F. Larcker, “Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics,” *Journal of Marketing Research*, vol. 18, 1981, pp. 382-388.

[43] J. Henseler, C.M. Ringle, and R.R. Sinkovics, “The use of partial least squares path modeling in international marketing,” *Advances in International Marketing*, vol. 20, 2009, p. 277–319.

[44] M. Dworschak, “Internet: Null Blog,” *Spiegel Online*, 2010.

Micro-Blogging Adoption in the Enterprise: An Empirical Analysis

Valentin Schöndienst
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Germany
+49/302093-5742
valentin@schoendienst.net

Hanna Krasnova
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Germany
+49/30/2093-5742
krasnovh@wiwi.hu-berlin.de

Oliver Günther
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Germany
+49/30/2093-5742
guenther@wiwi.hu-berlin.de

Dirk Riehle
Friedrich-Alexander-Universität Erlangen-Nürnberg
Schlossplatz 4
91054 Erlangen, Germany
+49/913185-0
dirk@riehle.org

ABSTRACT
Given the increasing interest in using social software for company-internal communication and collaboration, this paper examines drivers and inhibitors of micro-blogging adoption at the workplace. While nearly one in two companies is currently planning to introduce social software, there is no empirically validated research on employees' adoption. In this paper, we build on previous focus group results and test our research model in an empirical study using Structural Equation Modeling. Based on our findings, we derive recommendations on how to foster adoption. We suggest that micro-blogging should be presented to employees as an efficient means of communication, personal brand building, and knowledge management. In order to particularly promote content contribution, privacy concerns should be eased by setting clear rules on who has access to postings and for how long they will be archived.

Keywords
Keywords: Micro-Blogging, Social Software, Adoption, Privacy, Enterprise 2.0, Twitter, Technology Acceptance

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

1. INTRODUCTION
More and more companies recognize the power of social software to improve communications and knowledge exchange within their organizations [6, 7]. Almost 50% of businesses are about to introduce social software [42]. Outside corporate infrastructures, user-generated activity streams have proven to be a highly efficient means for filtering, spreading, and allocating information [29]. In this context, micro-blogging is about to become "a standard feature in enterprise social software platforms" and "is earning its place alongside other channels, enabling new kinds of fast, witty, easy-to-assimilate exchanges" [16].

To date, there are no studies investigating return on investment in enterprise micro-blogging. However, discussions in the blogosphere [10, 13, 35, 36] as well as a few documented cases [13, 5, 24] indicate substantial business value. Indeed, micro-blogging is considered to be a light-weight means of enhancing communication and collaboration. By allowing people to share short messages and tag them with keywords, enterprise micro-blogging lets employees determine which information they receive based on who publishes it and what it is about. That way, users keep in touch with specific persons across geographical and functional dispersion and find colleagues who share interests or work on similar issues.

As opposed to email, phone and instant messaging, micro-blogging is non-intrusive as people voluntarily subscribe (and unsubscribe) to other people's feeds. This makes for the evolutionary nature of micro-blogging – which lets employees efficiently spread news or questions and at the same time discharge email and instant messaging traffic. Management can use micro-blogging to keep track of ongoing issues, decisions, solutions, and ideas at the grassroots level.

Recognizing the considerable potential of micro-blogging, more than a dozen vendors are attempting to tap into this market by offering systems that enable employees to micro-blog [19]. At the same time, large IT companies are exploring ways to add micro-blogging functionality to their product portfolio (e.g., SAP's ESME project, Oracle's OraTweet, IBM's BlueTwit).

Analogously to social software in the consumer space, the success of micro-blogging in the enterprise is expected to be user-driven. Indeed, micro-blogging implies social interaction and self-disclosure, as opposed to the use of conventional office software. As a result, bringing social software into the workspace taps into dimensions that go beyond traditional technology acceptance models. Whereas employee participation is essential for the success of micro-blogging in corporate environments, attempts to understand people's motivation to micro-blog has concentrated on the use of Twitter [e.g., 3, 22, 25, 31]. Micro-blogging adoption in the enterprise, on the other hand, has received little academic attention. This paper seeks to fill this gap. It provides an empirical analysis of factors determining an employee's decision to adopt micro-blogging at the workplace.

To this end, we proceed as follows: In the following section, we review related work. Then we derive the research model and test it using a dataset from an online survey with 100 participants. Here we use Structural Equation Modeling with Partial Least Squares (PLS). Finally, we conclude by discussing managerial implications for the introduction of an enterprise micro-blogging system.

2. RELATED WORK

Despite a growing body of research concerning the use of micro-blogging services like Twitter, there have only been a few attempts to investigate micro-blogging in a work environment. Ehrlich and Shami [11] conducted a content analysis and compare the messages of 34 employees posted publicly on Twitter to the ones they posted on an intra-firm micro-blogging platform. They found significant differences in individuals' use of the two systems. Employees used the company-internal system mostly to engage in "Q&A" and personal, directed interaction, whereas Twitter was mostly used to share information with a larger community.

Semi-structured interviews shed more light on the underlying motivation of employees to use micro-blogging. In particular, communication benefits such as the ability to share information in real-time and to be aware of what colleagues are working on were mentioned by the interviewees. At the same time, employees saw micro-blogging as a way to enhance their reputation.

Zhao and Rosson [44] interviewed 11 active Twitter users and built on theories on interpersonal communication and social psychology to explore potential benefits micro-blogging can bring to informal communication at work. Their results hint at relational benefits such as building common ground, managing people perceptions, and creating a sense of connectedness as well as personal benefits such as the acquisition of valuable information.

Similarly, Barnes et al. [3] used an interpretive approach to investigate the case of a firm's use of an in-house social software development that shows close similarities to micro-blogging. They found that employees regarded micro-blogging as a good information source and efficient tool to stay aware of activities in the company. On the flipside, they found hints that employees have concerns regarding privacy aspects such as the monitoring of their work.

While these studies hint at use cases and benefits the subject of adoption has received no particular attention. Meanwhile, the results suggest a number of enabling factors as well as potential inhibitors, which are not considered in traditional technology acceptance theories such as the Technology Acceptance Model [e.g., 37], or the Unified Theory of Acceptance and Use of Technology (UTAUT) [38]. In this context, Günther et al. [18] present the results of four focus group sessions on the adoption of enterprise micro-blogging and argue for modifications and extensions of traditional technology acceptance theory. In the absence of empirically validated research on enterprise micro-blogging, we build on these recommendations and evaluate their respective relevance empirically.

3. RESEARCH MODEL

Venkatesh et al. [18] identify similarities among construct operationalizations in various technology acceptance theories and present an integrated model: the UTAUT. Günther et al.'s [18] model builds around the core constructs of UTAUT. In the following, we present refinements to the model which aim at providing a better understanding of the specific nature of micro-blogging adoption.

3.1 Dependent Variables

Following UTAUT, Günther et al. [18] use *Behavioural Intention* as a major dependent variable in their model. Investigating the case of micro-blogging adoption, Böhringer [5], however, finds that approximately one third of all users contribute less than one message per week. Meanwhile, these users do not necessarily reject the system: "On the contrary many say that they use the micro-blogging system quite often. However, they are mostly passive users and do not actively contribute to a great extent" [5, p. 7]. Yet, successful implementations of enterprise micro-blogging systems require a balance of active and passive users. Indeed, without continuous influx of fresh information the contents of the system will become obsolete – a dangerous development for platform sustainability. At the same time, consumption of the content is a primary goal of micro-blogging introduction in the first place. Both usage patterns – reading and contributing - imply adoption and they are not mutually exclusive. With this discussion in mind, we split the traditional *Behavioural Intention* construct into two independent constructs: *Intention to Contribute* and *Intention to Follow*.

Intention to Contribute (CTB) refers to an individual's intention to actively micro-blog. This implies posting messages to share information, contribute content, and respond to others.

Intention to Follow (FLW) refers to an individual's intention to use the system to retrieve information and read through colleagues' postings. In the following sections we seek to explain the dynamics behind individual micro-blogging adoption with respect to these two dependent variables.

3.2 Independent Variables

Venkatesh et al. [38, p. 447] define ***Performance Expectancy*** as "the degree to which an individual believes that using the system

would help him or her to attain gains in job performance." The construct integrates similar concepts, such as perceived usefulness and relative advantage, from other technology acceptance theories. Within these theories the respective constructs related to *Performance Expectancy* are the strongest predictors of intention [38]. As for the use of micro-blogging at the workplace, some people derive utility from being able to conveniently reach people and share information, while others are motivated to use the system because they can access useful content. In the absence of a single *Behavioral Intention* construct, we therefore hypothesize that *Performance Expectancy* will have a significant positive influence on both *Intention to Contribute* (H1a) and *Intention to Follow* (H1b).

Reputation refers to "the degree to which use of an innovation is perceived to enhance one's [..] status in one's social system" [30, p. 195]. For an individual employee micro-blogging provides a means to make accomplishments and competences visible. For example, posting news related to some project can give an indication of one's contributions as well as capabilities. From an individual viewpoint elevated status may contribute to self-assertion in interactions with colleagues and clients. In this context, *Reputation* "provides a general basis for greater productivity" [37, p. 189] and is therefore theorized to have significant positive influence on *Performance Expectancy* (H5a).

Expected Relationships refer to the degree to which one believes one can build or improve mutual relationships through the use of a micro-blogging system. By sharing information on both work-related issues as well as personal interests, micro-blogging allows users to learn about colleagues and lets them get in touch in an unobtrusive manner. Given findings that the scope of one's network within a company correlates with one's productivity [41] *Expected Relationships* is hypothesized to have a significant positive influence on *Performance Expectancy* (H5b).

Communication Benefits refer to improved information flow, i.e. easier and quicker information exchange throughout the organization. Related categories were mentioned 62 times throughout Günther et al.'s [18] focus group sessions and turned out to be the most relevant construct. Being able to keep track on issues within the organization as well as being able to quickly spread information may support task accomplishment and increase one's productivity. Consequently, *Communication Benefits* is hypothesized to have a significant positive influence on *Performance Expectancy* (H5c).

On the other hand, micro-blogging adds another channel to a multitude of communication channels already in place. Put into perspective to other means of communication, character-limited micro-blogging messages may replace interruptive phone calls and extensive emails. That way, looking at all communication means at an employee's disposal, micro-blogging could reduce the overall amount of information one has to cope with. However, perhaps with trivial Twitter-postings in mind [28], Günther et al.'s [18] focus group participants attributed micro-blogging a low ***Signal-to-Noise Ratio*** and expressed fear that useful content would be rare. Therefore, for the purpose of our study *Signal-to-Noise Ratio* refers to the possible negatively shaded perception that using micro-blogging would lead to more incoming information one has to cope with. Negative effects of interruptions and the role of attention economics on overall performance have been shown by Gonzalez and Mark [17]. As a result, it is hypothesized that *Signal-to-Noise Ratio* will have a significant negative influence on *Performance Expectancy* (H5d).

Günther et al.'s [18] focus group results further suggest individuals' ***Privacy Concerns*** to be an important inhibitor for usage intention. The introduction of this construct into technology acceptance research is innovative. Indeed, as each posting is linked to its author colleagues could draw conclusions on someone's work or private life. Further, superiors could track postings in order to monitor someone's work. In any event, *Privacy Concerns* derive from self-disclosure and can therefore be theorized to have a significant negative influence on *Intention to Contribute* (H2a) but no significant influence on *Intention to Follow* (H2b).

Günther et al. [18] integrate *Codification Effort* as another antecedent of the Performance Expectancy. This construct is, however, omitted from our study for the sake of brevity and due to its close resemblance to *Effort Expectancy*.

Effort Expectancy "is the degree of ease associated with the use of the system" [38, p. 450] as well as related constructs within other models were found to be significant only at the earliest point of time of measurement and became nonsignificant as experience increased [38]. The functionality of micro-blogging systems is very limited as there are just two things one can do: write very short messages and access others' messages in various ways. Due to the simplicity of micro-blogging systems we expect efforts associated with learning and operating the system to not be a significant factor in an employee's adoption decision. Consequently, we theorize *Effort Expectancy* to significantly influence neither *Intention to Contribute* (H3a), nor *Intention to Follow* (H3b).

According to Günther et al.'s [18] focus group results, *Organizational Culture* is an important determinant of an employee's decision to use micro-blogging. If there is a certain degree of consensus to collaborate in an organization, i.e. ***Collaborative Norms*** [4] one is more likely to use a micro-blogging system to seek information as well as to provide content. Consequently, we theorize *Collaborative Norms* to have significant positive influence on both *Intention to Contribute* (H4a) and *Intention to Follow* (H4b).

Venkatesh et al. [38] show that the influence of ***Facilitating Conditions*** on *Behavioral Intention* becomes nonsignificant when both *Performance Expectancy* and *Effort Expectancy* are included in the model. Therefore, we exclude the construct from our considerations on determinants of *Intention to Contribute* and *Intention to Follow*.

To provide a better overview, all hypotheses formulated above are summarized in Table 1.

Table 1: Overview Hypotheses

No.	Hypothesis
H1a	Performance Expectancy will have a significant positive effect on Intention to Contribute.
H1b	Performance Expectancy will have a significant positive effect on Intention to Follow.
H2a	Privacy Concerns will have a significant negative effect on Intention to Contribute.

H2b	Privacy Concerns will not have a significant effect on Intention to Follow..
H3a	Effort Expectancy will **not** have a significant effect on Intention to Contribute.
H3b	Effort Expectancy will **not** have a significant effect on Intention to Follow
H4a	Collaborative Norms will have a significant positive effect on Intention to Contribute.
H4b	Collaborative Norms will have a significant positive effect on Intention to Follow.
H5a	Reputation will have a significant positive effect on Performance Expectancy.
H5b	Expected Relationships will have a significant positive effect on Performance Expectancy.
H5c	Communication Benefits will have a significant positive effect on Performance Expectancy.
H5d	Signal-to-Noise Ratio will have a significant negative effect on Performance Expectancy.

4. EMPIRICAL STUDY

4.1 Survey Design and Sampling

As of today, established micro-blogging implementations at enterprises are still rare. Therefore, we followed the approach of Zhao & Rosson [44] and conducted a survey among users of Twitter. Before proceeding to the survey questions, participants were asked to imagine that a Twitter-like system was introduced in their place of employment. All subsequent questions related to this hypothetical situation.

Invitations to participate in the online survey were spread through various News Groups and Twitter accounts. Also, people were asked to "re-tweet" the invite so that the survey link could spread across social networks. Each participant had a 50% chance to get a 5 Euro Amazon.com gift certificate.

The responses were collected in August 2009. The overall gross sample consisted of 100 participants. After deleting unusable data sets, a final net sample of 82 observations was obtained. The respondents were between 19 and 65 years old with an average of 32 years. Men were overrepresented with 72%. 46% of the respondents worked for companies with less than 500 employees 31% were employed by companies with more than 500 employees, 23% did not provide information about their company size.

As familiarity with Twitter was a precondition for participation in the study, we expect our respondents to have above average open-mindedness toward social software, at least in the private domain – a possible limitation of our study.

4.2 Development of Measurement Scales

All constructs in the study involved multiple items as reflected in Table 2. Content validity of the constructs was ensured by relying on pre-tested scales where possible. In order to maintain content validity of the adopted and self-developed scales, experts in the field of social software as well as ordinary users were asked to verify them. Additionally, during the pre-test phase, unclear items with low inter-item correlation coefficients were removed. All items were anchored on a seven-point Likert scale.

Table 2: Overview Items

Construct	Items (source)
Intention to Follow, FLW	1. I would keep myself updated through the system. 2. I would read through the messages of others. 3. I would check out the postings of others. (self-developed)
Intention to Contribute, CTB	1. I would use the system to share snippets of information. 2. I would use the system to contribute. 3. I would take an active part in discussions on the system. 4. I would provide content on the system. (self-developed)
Performance Expectancy, PE	1. I would find the system useful in my job. 2. Using the system would enable me to accomplish tasks more quickly. 3. Using the system would increase my productivity. (Based on: [38])
Signal-to-Noise Ratio, SN	Looking at all communication means at my disposal (e.g. email, phone, instant messaging), if I used the system… 1. I would have to cope with more useless information. 2. I would receive more irrelevant messages. 3. I would spend more time on handling incoming information. 4. I would be confronted with higher amounts of unnecessary information. (self-developed)
Communication Benefits, CBE	If I used the system… 1. I could keep myself better informed about issues within my organization. 2. I would be more aware about what is going on within my organization. 3. I could communicate more quickly with my colleagues. 4. I could exchange knowledge more quickly. (self-developed)
Expected Relationships, ER	Using the system… 1. would strengthen the ties between other employees and me. 2. would establish new contacts with other employees in my organization. 3. would help to create stronger relationships with other employees in the organization. 4. would expand the scope of my association with other employees in the organization. (Based on: [23])
Effort Expectancy, EE	My interaction with the system would be clear and understandable. 1. It would be easy for me to skillfully use the system. 2. I would find the system easy to use. 3. Learning to operate the system would be easy for me. (Based on: [38])
Reputation, RE	If I used the system…

	1. I would make my accomplishments more visible to others.a 2. I would improve my status among my colleagues. 3. my specific competences would be more visible to others. 4. I would gain more recognition in my organization. (Based on: [30, 39])
Privacy Concerns, PV	If I used the system, someone in my organization... 1. may be able to learn more about my private life than I am comfortable with. 2. may be able to deduce some information that I consider private. 3. may be able to learn something about me which will invade my privacy. (Based on: [43])
Collaborative Norms, CN	1. Cooperation is encouraged in my organization. 2. Collaboration is considered important in my organization. 3. Knowledge sharing is considered important in my organization. 4. Knowledge sharing is encouraged and frequent in my organization. (Based on: [12, 27])

4.3 Research Methodology

Structural Equation Models (SEMs) may be estimated based on the analysis of either the covariance structure [26] or the variance structure [40]. Both approaches are suitable to empirically evaluate the relations among latent variables. However, the conceptual and methodological differences make the latter approach, i.e. PLS, most suitable for the purposes of our study. First of all, the sample size of 82 meets PLS's requirements of having at least 10 times the number of observations as there are exogenous constructs influencing the most complex endogenous construct [2]. Consequently, we expect the results not to be affected by sample size limitations. Secondly, in the absence of theory on the adoption of social software at the workplace our research is of rather exploratory nature than of a confirmatory one. This makes the case for the variance-based PLS approach which is generally preferred for theory building and prediction and requires fewer statistical assumptions [14].

All calculations were carried out using SmartPLS 2.0 [34], a statistical package developed for the estimation of SEMs using the PLS approach.

4.4 Model Evaluation

We follow Chin [8] and Ringle [33] and analyze the measurement model and the structural model separately in order to evaluate the overall quality of the model.

4.4.1 Evaluation of the Measurement Model

In order to examine the measurement model, Convergent Validity and Discriminant Validity were assessed. In line with Bagozzi and Philips [1, p. 468] "convergent validity is the degree to which two or more attempts to measure the same concept are in agreement". Convergent validity was ensured via three criteria: (1) Indicator Reliability, (2) Composite Reliability (CR), and (3) Average Variance Extracted (AVE).

In our model, Indicator Reliability is assured as all indicators have loadings higher than 0.8 [21]. For Composite Reliability, all values by far exceed the minimum required threshold of 0.6 [20, 33]. Further, all latent variables had AVEs of well over a recommended threshold of 0.5 [15]. Finally, Internal Consistency was evaluated by calculating Cronbach's Alpha (CA). For all latent variables, Cronbach's Alpha is above 0.7 confirming Internal Consistency [32]. Taken together, Convergent Validity can be assumed as summarized in Table 3.

To ensure Discriminant Validity square root AVE values for each latent variable must exceed the correlation between the respective and all other latent variables of the model [15]. This requirement is met for all latent variables as shown in Table 4. Hence, Discriminant Validity can be assumed.

Table 3: Quality Criteria of the Constructs

Construct	Indicators	AVE	CR	CA
CBE	4	0,685	0,897	0,847
CN	4	0,762	0,927	0,921
CTB	4	0,824	0,949	0,929
EE	4	0,877	0,955	0,932
ER	4	0,847	0,957	0,940
FLW	3	0,768	0,908	0,847
PE	3	0,828	0,935	0,896
PV	3	0,961	0,987	0,980
RE	4	0,853	0,959	0,943
SN	4	0,820	0,948	0,926

Table 4: Latent Variable Correlation, Respective Square Roots of AVE in *bold italic*

	CBE	PE	CN	CTB	EE
CBE	***0,827***				
PE	0,765	***0,910***			
CN	0,079	0,137	***0,873***		
CTB	0,593	0,515	0,133	***0,908***	
EE	-0,010	0,029	0,305	0,132	***0,936***
ER	0,612	0,571	-0,183	0,267	-0,155
FLW	0,634	0,697	0,132	0,488	-0,018
PV	-0,110	-0,035	0,097	-0,306	-0,090
RE	0,600	0,595	0,102	0,561	0,186
SN	-0,286	-0,337	0,061	-0,131	0,063
	ER	**FLW**	**PV**	**RE**	**SN**
ER	***0,920***				
FLW	0,534	***0,876***			
PV	-0,173	-0,100	***0,980***		
RE	0,528	0,522	0,006	***0,923***	
SN	-0,290	-0,267	0,347	-0,041	***0,906***

4.4.2 Evaluation of the Structural Model

The R^2 measure describes how much of the respective endogenous variable's variance is explained by the model's exogenous

variables. A good R^2 value should be above 40% [33]. Given the explorative nature of our research 50% for *Intention to Follow* is good, a value of 36% for *Intention to Contribute* is acceptable. The proposed decomposition of the *Performance Expectancy* construct explains 64% of its variance.

In the next step, path coefficients and their significance levels were evaluated on the basis of bootstrapping results as summarized in Table 5 and Figure 1.

We find that *Communication Benefits*, *Reputation*, and *Signal-to-Noise Ratio* are significant determinants of *Performance Expectancy*. Furthermore, *Communication Benefits* have a stronger influence on *Performance Expectancy* than *Reputation* and *Signal-to-Noise Ratio* (the comparison of the relative strength of the path coefficients has rendered t-values exceeding the benchmark of 1.96). the *Expected Relationships* construct, however, has no significant impact on *Performance Expectancy*.

Both *Privacy* and *Performance Expectancy* are significant determinants of *Intention to Contribute*. Comparison of the absolute values of path coefficients (disregarding the sign of the effect) reveals no significant difference in the strength of their impact on *Intention to Contribute* (t-value=1.54). Further, *Performance Expectancy* has a significant effect on *Intention to Follow*.

Interestingly, *Collaborative Norms* and *Effort Expectancy* have no significant effect at all, neither on *Intention to Contribute*, nor on *Intention to Follow*.

Table 5: Path Coefficients. Significance Levels, Hypothesis Evaluation

Dep. Var.	Ind. Var.	Path Coeff.	Hypothesis
CTB	CN	0,072	H4a: rej.
CTB	EE	0,070	H3a: supp.
CTB	PE	0,493**	H1a: supp.
CTB	PV	-0,290**	H2a: supp.
FLW	CN	0,067	H4b: rej.
FLW	EE	-0,066	H3b: supp.
FLW	PE	0,686**	H1b: supp.
FLW	PV	-0,089	H2b: supp.
PE	CBE	0,540**	H5c: supp.
PE	ER	0,078	H5b: rej.
PE	RE	0,224*	H5a: supp.
PE	SN	-0,151*	H5d: supp.
*significant at 5% level, **significant at 1% level			

4.4.3 Ad hoc Analysis of Direct Effects

Even though our model is strongly based on theory, the presence of direct effects from antecedents of *Performance Expectancy* to *Intention to Contribute* and *Follow* cannot be excluded. In order to check for this possibility, ad hoc analysis of direct effects has been conducted. To limit other influences, only one path was checked at a time. Results of our evaluations are presented in Table 6. We find that desire to gain *Reputation* as well as expectations regarding *Communication Benefits* will also *directly* motivate users to *Follow* and *Contribute*. Furthermore, *Expected Relationships* will positively influence users' *Intention to Follow*.

Table 6: Path Coefficients for Direct Effects

Dep. Var.	CTB	FLW
RE	0,408**	0,201*
CBE	0,444**	0,244*
SN	0,171	0,059
ER	-0,071	0,224*
*sign. at 5% level, **sign.at 1% level		

Implications of our findings are discussed in the following sections.

5. DISCUSSION AND IMPLICATIONS

5.1 Theory Implications

Much more than conventional office software, social software, and in particular micro-blogging, support social interaction and imply self-disclosure. As a result, bringing applications like micro-blogging into the workspace goes beyond traditional technology acceptance theory. Against this background, the aim of this study was to extend the classical technology acceptance research in order to help companies in integrating social software products in employees' daily routines. Results of our study have significant implications for theory.

First, a number of new constructs were introduced and empirically tested. The results show that *Communication Benefits*, *Reputation*, and *Signal-to-Noise Ratio* are reliable and valid constructs that are significant in predicting the formation of employees' *Performance Expectancy* (R^2 amounts to 64%). This deepens the understanding of the rather abstract concept of *Performance Expectancy* in the domain of social software adoption. Additionally, ad hoc analysis of direct effects has shown that *Reputation* and *Communication Benefits* have a direct impact on employee's intention to adopt micro-blogging, expressed in both *Intention to Follow* and *Intention to Contribute*. Further, even though *Expected Relationships* do not influence *Performance Expectancy*, we find them to have a direct impact on the *Intention to Follow*.

Further, we introduce *Privacy Concerns* construct into technology acceptance theory and show that the construct is a significant determinant of an employee's decision to contribute content. It turns out that there is no significant difference in the strength of its impact compared to the *Performance Expectancy* construct which is at the core of traditional technology acceptance models [e.g., 38]. This indicates the necessity to include *Privacy Concerns* into technology acceptance theory when social applications are explored in the work context.

Finally, our results show that when investigating acceptance of a system that provides a platform for both content contribution and content consumption, adoption becomes a two-dimensional variable, with distinct dynamics behind its formation. For example, *Privacy Concerns* determine employees' intention to contribute content, but do not exert a significant effect on their intention to follow others' contributions.

Overall, considering that our model was built by closely following existing theory, the presence of significant direct effects identified in our ad hoc analysis is interesting. On a more general level, this

result calls for more theory-building efforts in the area of social software adoption.

5.2 Managerial Implications

In our study, we distinguish between adoption in terms of following posts from colleagues and adoption in terms of writing messages. We find that that an employee's perception whether micro-blogging would "help him or her to attain gains in job performance" [38, p. 447] has a strong positive influence on adoption in terms of both contributing and following. Given the key role of *Performance Expectancy* construct, also in other technology acceptance models, it is important to understand which factors are crucial for its formation. Our results allow us to derive practical recommendations on how to foster adoption:

5.2.1 Communication Benefits

We find that expected *Communication Benefits* of micro-blogging constitutes the most crucial component in employees' formation of *Performance Expectancy.* In addition, desire to contribute and follow may be motivated as shown in our ad hoc analysis. Therefore, micro-blogging's usefulness should be promoted by presenting its advantages to employees. Examples of these communication benefits are the fast spreading of relevant messages, the social filtering which increases the signal-to-noise ratio, the ease of using micro-blogging when compared with other social media channels, and the implicit collective assignment of relevance to messages by the mechanism of "re-tweeting" them.

5.2.2 Signal-to-Noise Ratio

Concerns that micro-blogging would lead to information overload (i.e. worsening of *Signal-to-Noise Ratio*) have a significant negative impact on *Performance Expectancy*. It is imperative to ease these objections in order to foster adoption. Management should promote the use of filtering tools as well as help clarify the potential misunderstanding that employees must follow every posting—this is not true and consequently, utilizing a micro-blogging service promises to require less attention than using email, in comparison. Given that employees can attune their use in reaction to how much they get out of the service, much of the fear of a low *Signal-to-Noise Ratio* is unfounded. Therefore, the "natural selection" mechanism of following and un-following other users should be illustrated carefully upon the introduction of the system: users choose freely whose messages they follow. That way, users can fine-tune information inflow while users who are excessively posting irrelevant information will shortly have no followers.

At the same time, public recommendations on how to use the micro-blogging system could reduce concerns that colleagues could excessively post irrelevant messages.

5.2.3 Reputation

Our results show that an individual employee's possibility to increase his/her status in an organization by posting messages has a significant positive influence on his/her *Performance Expectancy.* In addition, direct effects on the Intention to Contribute and Follow have been confirmed. However, while this is a driver for individuals to adopt micro-blogging, it may be an inhibitor on an organizational level as executives may fear increased transparency. The introduction of micro-blogging in the enterprise thus becomes a political issue. This is also true for other "Enterprise 2.0" technologies that affect hierarchies and power structures.

The introduction of internal micro-blogging should be viewed as a move towards increased transparency of communication, not only top-down, but across the organization. Top management needs to lead the way, as micro-blogging is no panacea to turn a secretive organization into a transparent Enterprise 2.0 firm.

Management needs to understand (and act accordingly) that micro-blogging can lay open the role of important communication hubs in organizations—people who spend time on connecting other people but are usually not recognized (nor rewarded) for such activity. Being able to make such contributions visible provides another incentive for employees to adopt micro-blogging.

5.2.4 Expected Relationships

Ad hoc analysis reveals that expectations that one can build or improve mutual relationships through the use of a micro-blogging system will motivate users to follow each other's contributions. Research shows that whom people know is highly correlated with what they come to know [9]. Hence, opportunities to maintain and extend one's network should be communicated as a key benefit of the system.

5.2.5 Privacy Concerns

Concerns that colleagues and superiors could deduce private information from one's postings turned out to be a significant inhibitor for people to actively use micro-blogging on the job. Giving employees control on who can access their postings and for how long they will be archived could mitigate these concerns. More over, helping users better understand how to use a tool like micro-blogging should help them overcome fears of unintentionally misbehaving and harming themselves.

Yet, *Intention to Follow* colleagues' messages turned out to not be significantly affected by *Privacy Concerns*.

5.2.6 Collaborative Norms

A common understanding to share knowledge and collaborate in an organization (i.e. *Collaborative Norms*) does neither significantly facilitate *Intention to Contribute* nor *Intention to Follow*. While this result seems surprising, it shows that enterprises with a rather competitive organizational climate are not excluded from implementing micro-blogging successfully.

5.2.7 Effort Expectancy

The nonsignificance of *Effort Expectancy* confirms our hypothesis that the use of an application as simple as micro-blogging is not inhibited by efforts related to operating or learning the system.

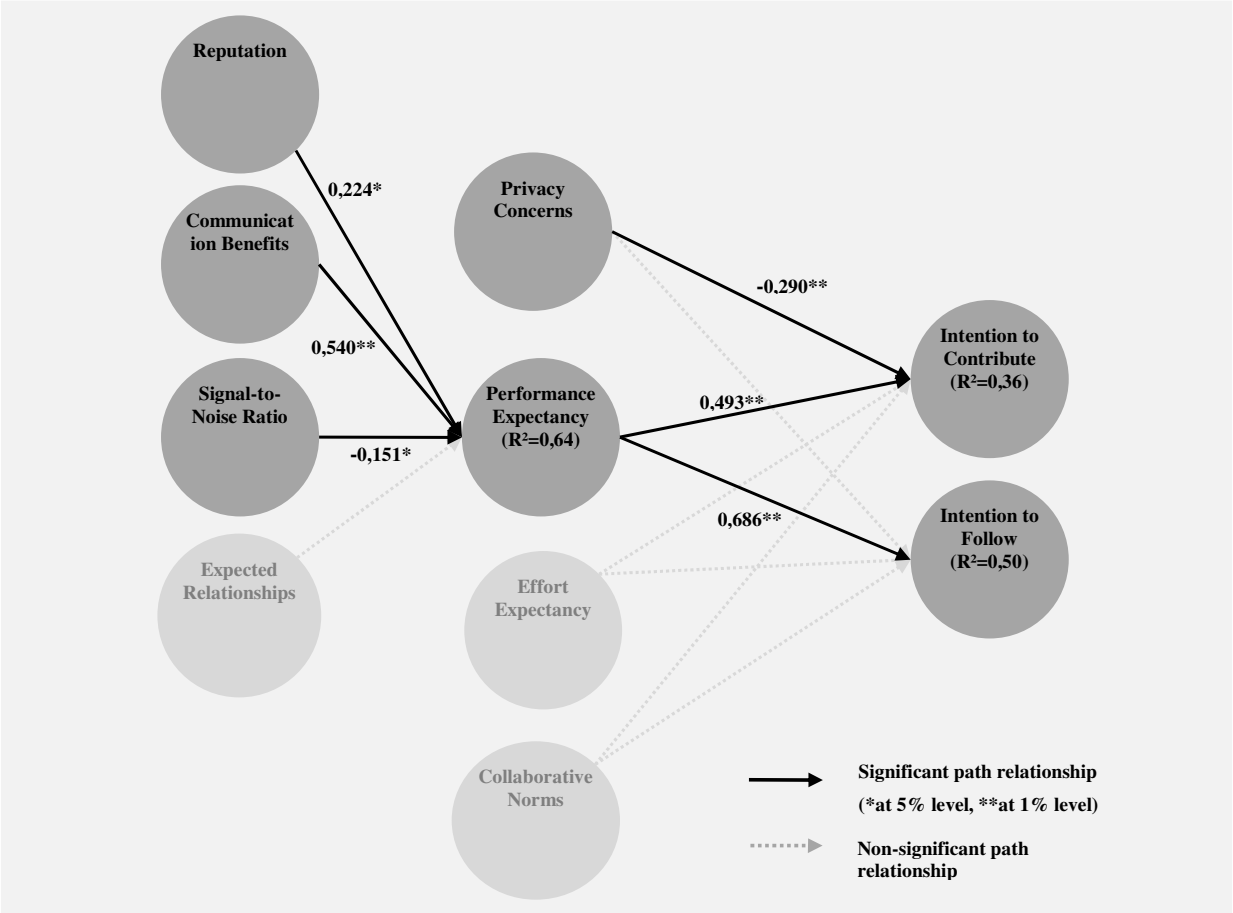


Figure 1: Results, Path Coefficients

6. CONCLUSION

A successful implementation strategy needs to point out the utility of micro-blogging. We suggest to position the system as an efficient means for communication, personal brand building, and knowledge management. While users can be won over by advertising improved communications and personal brand building, management may be more interested in the knowledge management aspects of the application.

In order to particularly facilitate contributions, and thereby content generation, privacy concerns should be eased by setting up clear rules on who has access to one's postings and for how long they will be archived.

In contrast to conventional office software, micro-blogging implies social interaction and self-disclosure. This applies to social software in general. As a result, bringing applications like micro-blogging into the workspace goes beyond traditional technology acceptance theory. Our findings show that technology acceptance models need to be adjusted in order to account for this new paradigm.

Meanwhile, the results underline the need to limit employees' exposure to information input and indicate the growing importance of personal brand building and privacy issues in today's enterprises.

While recruiting participants from Twitter allowed us to investigate employees from a variety of company backgrounds, our future research will further test the model presented here in real enterprise settings. It will also take other micro-blogging paradigms into account.

7. REFERENCES

[1] Bagozzi, R. P., and Phillips, L. W. 1982. Representing and testing organizational theories: A holistic construal. *Administrative Science Quarterly*, 459–489. DOI= http://dx.doi.org/10.2307/2392322

[2] Barclay, D., Higgins, C., and Thompson, R. 1995. The partial least squares (PLS) approach to causal modeling: Personal computer adoption and use as an illustration. *Technology studies*, 2(2), 285–309. DOI= http://dx.doi.org/10.1109/HICSS.2010.406

[3] Barnes, S.J., and Böhringer, M. 2009. Continuance Usage Intention in Microblogging Services: The Case of Twitter. In *Proceedings of the 17th European Conference on Information Systems (ECIS)*, Verona, Italy, June 08-10, 2009.

[4] Bock, G. W., Kankanhalli, A., and Sharma, S. 2006. Are norms enough? The role of collaborative norms in promoting organizational knowledge seeking. *European Journal of Information Systems*, 15(4), 357. DOI= http://dx.doi.org/10.1057/palgrave.ejis.3000630

[5] Böhringer, M. 2009. Information und Kontext im Enterprise 2.0; In *Workshop-Proceedings of the 9th Mensch & Computer conference*, Workshop on "Enterprise 2.0 - Web 2.0 im Unternehmen", Berlin, September 6th-9th, 2009, 130-134.

[6] Bughin, J., Chui, M., and Miller, A. 2009. How companies are benefiting from Web 2.0: McKinsey Global Survey Results, *McKinsey Quarterly*.

[7] Caya, P., and Nielsen, J. 2009. Enterprise 2.0: Social Software on Intranets. *Nielsen Norman Group Report*.

[8] Chin, W. W. 1998. Issues and opinion on structural equation modeling. *Management Information Systems Quarterly*, 22(1), 7–16.

[9] Cross, R. L., Parker, A., and Sasson, L. 2003. Networks in the knowledge economy. *Oxford University Press*, USA.

[10] Dawson, R. 2008. Micro-blogging in the enterprise: an idea whose time has come? Retrieved on August 31, 2009, from http://rossdawsonblog.com/weblog/archives/2008/10/microblogging_i.html.

[11] Ehrlich, K. and Shami, S. 2010. Microblogging inside and outside the workplace. In *ICWSM '10: Proc. of the Int. Conf. on Weblogs and Social Media*, AAAI. DOI= http://dx.doi.org/10.1145/1460563.1460615

[12] Fisher, R. J., Maltz, E., and Jaworski, B. J. 1997. Enhancing communication between marketing and engineering: the moderating role of relative functional identification. *The Journal of Marketing*, 54–70.

[13] Fitton, L. 2008. Best Buy's "Enterprise Twitter". Retrieved on August 31, 2009, from http://www.readwriteweb.com/archives/best_buy_enterprise_twitter.php.

[14] Fornell, C., and Bookstein, F. L. 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, 440–452.

[15] Fornell, C., and Larcker, D. F. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research,*. 39–50.

[16] Gartner (2009). Gartner's 2009 Hype Cycle Special Report Evaluates Maturity of 1,650 Technologies. Retrieved on August 31, 2009, from http://www.gartner.com/it/page.jsp?id=1124212.

[17] González, V. M., and Mark, G. 2004. "Constant, constant, multi-tasking craziness": managing multiple working spheres. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 113–120. ACM New York, NY, USA. DOI= http://dx.doi.org/10.1145/985692.985707

[18] Günther, Krasnova, H., Riehle, D., and Schöndienst, V. 2009. Modeling microblogging adoption in the enterprise. In *Proceedings of the Fifteenth Americas Conference on Information Systems*, San Francisco, California August 6th-9th 2009

[19] Hinchcliffe, D. 2009. Twitter on your intranet: 17 microblogging tools for business. Retrieved on August 31, 2009, from http://blogs.zdnet.com/Hinchcliffe/?p=414&page=3.

[20] Homburg, C., and Baumgartner, H. 1995. Beurteilung von Kausalmodellen: Bestandsaufnahme und Anwendungsempfehlungen. *Marketing ZfP*, 17(3), 162–176.

[21] Homburg, C., and Giering, A. 1996. Konzeptualisierung und Operationalisierung komplexer Konstrukte: Ein Leitfaden für die Marketingforschung. *Marketing ZfP*, 18(1), 5–24.

[22] Huberman, B., Romero, D., and Wu, F. 2008. Social networks that matter: Twitter under the microscope. *First Monday [Online],* 14, 1.

[23] Hsu, C. L., and Lin, J. C. C. 2008. Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, 45(1), 65–74. DOI= http://dx.doi.org/10.1016/j.im.2007.11.001

[24] Ives, B. 2009. Implementing Enterprise Micro-messaging with Yammer at Océ. Retrieved on August 31, 2009, from http://blog.contentmanagementconnection.com/Home/20966.

[25] Java, A., Song, X., Finin, T., and Tseng, B. 2007. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 56–65. ACM New York, NY, USA. DOI= http://dx.doi.org/10.1145/1348549.1348556

[26] Jöreskog, K. G. 1977. Structural Equation Models in Social Sciences: Specification, Estimation and Testing. In *Krishnaiah, P. R. (ed.): Applications of statistics.* Amsterdam, pp. 265-287.

[27] Kankanhalli, A., Tan, B. C. Y., and Wei, K. K. 2005. Contributing knowledge to electronic repositories: an empirical investigation. *Mis Quarterly*, 29(1), 113–43.

[28] Kelly, R. (2009). Twitter Study Reveals Interesting Results About Usage – 40% is "Pointless Babble". Retrieved on August 31, 2009, from http://www.pearanalytics.com/2009/twitter-study-reveals-interesting-results-40-percent-pointless-babble/.

[29] Kwak, H., Lee, C., Park, H., and Moon, S. 2010. In *Proceedings of the 19th International World Wide Web (WWW) Conference*, April 26-30, 2010, Raleigh NC (USA)

[30] Moore, G. C., and Benbasat, I. 1991. Development of an instrument to measure the perceptions of adopting an

information technology innovation. *Information Ssystems Research*, 2(3), 192–222.

[31] Naaman, M., Boase, J., and Lai, C.-H. 2010. Is it really about me? Message content in social awareness streams. In *Proc. CSCW* 2010. DOI= http://dx.doi.org/10.1145/1718918.1718953

[32] Nunnally, J. C. (1978). *Psychometric Theory*. 2nd Edition. McGraw-Hill, New York, NY.

[33] Ringle, C. M. (2004). Gütemaße für den Partial Least Squares-Ansatz zur Bestimmung von Kausalmodellen. *Industrielles Management*; Universität Hamburg, Institut für Industriebetriebslehre und Organisation, Arbeitspapier, 16, 1–32.

[34] Ringle, C.M., Wende, S. and Will, A. (2005). *SmartPLS 2.0 (beta)*. Retrieved August 4, 2009, from http://www.smartpls.de.

[35] Röhrborn 2008. Enterprise Microblogging: ein neuer Hype? Ja und Nein. Retrieved on August 31, 2009, from http://www.humannetworkcompetence.de/2008/09/12/enterprise-microblogging-ein-neuer-hype-ja-und-nein/.

[36] Strout, A., Cascio, J. 2008. Is the Enterprise Ready for Microblogging Tools like Twitter? Retrieved on August 31, 2009, from http://mashable.com/2008/09/30/enterprise-microblogging/.

[37] Venkatesh, V., and Davis, F. D. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204.

[38] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.

[39] Wasko, M. M., and Faraj, S. 2000. "It is what one does": why people participate and help others in electronic communities of practice. *Journal of strategic information systems*, 9(2-3), 155–173. DOI= http://dx.doi.org/10.1016/S0963-8687(00)00045-7

[40] Wold, H. 1982. Soft modeling: the basic design and some extensions. *Systems under indirect observation*, 2, 589–591.

[41] Wu, W., and Aral, B. 2008. Mining Face-to-Face Interaction Networks Using Sociometric Badges: Predicting Productivity in an IT Configuration Task. In *International Conference on Information Systems*, Paris, France.

[42] Young, O. G., Burris, P., and Davis, Z. R. 2009. The Enterprise 2.0 Buyer Profile: 2009. In *Forrester Research*.

[43] Xu, H., Dinev, T., Smith, H. J., and Hart, P. 2008. Examining the Formation of Individual's Privacy Concerns: Toward an Integrative View. In *Proceedings of ICIS 2008*

[44] Zhao, D., and Rosson, M. B. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. In Proc. *GROUP 2009*. ACM Press. DOI=http://dx.doi.org/10.1145/1531674.1531710

Managing the Implementation and Use of Real Time Collaboration: When Vision Meets Reality

Nadine Vehring
University of Muenster
Leonardo Campus 3
48149 Muenster
+49 (0)251 83-38125
nadine.vehring@ercis.de

Stefan Klein
University of Muenster
Leonardo Campus 3
48149 Muenster
+49 (0)251 83-38110
stefan.klein@ercis.de

Malte Kramer
University of Muenster
Leonardo Campus 3
48149 Muenster
+49 (0)251 83-38113
malte.kramer@ercis.de

ABSTRACT

The paper studies the specifics of implementing a platform technology in an organization which regards IT as core competence and prides itself as particularly employee focused and friendly. As result of strategic alignment, management has developed a vision of how Real Time Collaboration can contribute to the future concept of the company and the enactment of the organizational leitidee of an integrated service unit. The paper reports on the challenges of managing the transition from a voluntary use in an experimental setting towards mandatory use in operational processes. In line with the corporate culture, management is relying on self-organizing forces on the team level while it is at the same time carefully monitoring the adoption process and cautiously sets rules and guidelines to facilitate team-based modes of use.

Keywords

Real Time Collaboration, Lotus® Sametime®, RTC adoption, RTC use, RTC implementation, strategic alignment, infrastructure management

1. INTRODUCTION

Real Time Collaboration (RTC) is increasingly seen as part of social networking services or social media. Companies are struggling to find an appropriate managerial response to the opportunities and threats of these media in a corporate environment [1].

This paper reports about the early stages of the implementation and adoption of RTC in a large financial services company. In order to capture multiple perspectives on RTC and to understand processes of alignment or misalignment between management's visions for RTC and the perceptions of employees, we have conducted interviews at different levels of the organization and confronted management with preliminary results of our study.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

The in depth case study elaborates the specifics of implementing – or seeding – a platform technology in an organization which regards IT as core competence and prides itself as particularly employee focused and friendly. Management has positioned RTC as a building block of the future concept of the organization (strategic alignment) and has articulated ideas of how the current organizational vision of an integrated service unit can be enacted by using RTC.

The aim of this paper is to enable a better understanding of the implementation and adoption of RTC in a specific organizational environment. In particular we want to elaborate possible challenges that may arise within and between the different stages of the implementation process and to discuss possible management reactions and interventions. Furthermore, we want to clarify the roles and requirements of the involved parties at the different organizational levels (company, group, individual). Obviously this does not provide the basis for statistical generalizability. Instead we are looking for theoretical generalizability, i.e. more general messages or insights provided in the case.

Our paper proceeds as follows. We begin by presenting the concept of RTC and some aspects of the prior research on RTC. We describe our research method and the case company in section three and the implementation and adoption process of Sametime® in section four. In section five we discuss possible conflicts and the reaction of the management. Moreover we reflect on the relevance of the different organizational levels. We end with a short conclusion.

2. REAL TIME COLLABORATION

In the following, we will describe the concept of RTC. Furthermore, we will present relevant aspects of prior research on RTC. Before describing the research design in the following section, we will present our research questions at the end of this section.

2.1 Conceptualization of RTC

Real Time collaboration is an emerging genre of communication and collaboration systems [2]. Resulting from a market convergence of the telecommunication and groupware market, RTC systems are a combination of communication technologies, like Voice-over-IP (VoIP) telephony and instant messaging, and various collaborative applications [3].

According to [3], RTC usually comprises four building blocks (see Table 1).

Table 1: Building Blocks of Real Time Collaboration Systems

Building Blocks	Description
Unified Communication	RTC is based on the concept of Unified Communications which refers to the integration of various information and communication channels.
Presence information	The status information can give information about the availability of the user and his media and communication devices.
eCollaboration portfolio	RTC systems can comprise features of groupware applications, e.g. team calendars or document folders.
Contextualization	RTC systems can be integrated within the context of the user, e.g. with organizational processes and business applications.

Although it is possible to describe the specific components or features of RTC technologies, such as text chat, presence information or application sharing, the technology itself is flexible and open to diverse modes of use [4]. Therefore we are looking at RTC in this paper as a platform technology or infrastructure, which provides a rich set of affordances (e.g. [5], [6]). This view differs from the usual understanding of technology as an application with a predefined purpose and a clearly defined task environment. As such it comprises specific management challenges [7].

The market for RTC products can be divided into two segments: the mass market for private customers and the market for business customers [2]. The market for private customers mainly covers systems that integrate VoIP with instant messaging and presence signaling and can be downloaded for free, like Skype and MSN Web Messenger. On the opposite side, telecommunication system providers (e.g. Alcatel, Nortel and Siemens) and traditional software companies such as Microsoft and IBM offer complex and large-scale integrated RTC systems for the business segment [8].

IBM's Lotus® Sametime® is a large-scale integrated RTC system and comprises the aforementioned building blocks:

1. Sametime® users can communicate by using various communication channels, e.g. chat, VoIP telephony and video telephony.
2. Presence information is available for all users who are signed-in on the system.
3. Sametime® includes multiple collaboration features, such as group chat, application sharing or document sharing.
4. There are multiple options to integrate Sametime® into organizational processes.

2.2 Prior work and research questions

Research on RTC can be divided into two main categories. The first category of research focuses on the sense-making of RTC and tries to answer questions such as: what is it and in which ways is it different from other communication technologies? (e.g. [9])

The second category of research concentrates on the adoption and use of RTC at the individual level (e.g. [8]) and at the group level (e.g. [10], [4]) and mainly emphasizes the (transformation of) routines of daily communication.

However, the design and management of the implementation of RTC has been rarely addressed in the literature so far. Thus, we have addressed this void in earlier work by discussing the managerial and organizational design of Real Time Collaboration in a services company [7]. We have identified an interdependent set of management interventions, which provided a supportive environment for the strategic orientation, the organizational design and the early stages of an RTC implementation (see Table 2).

Table 2: Managerial tasks and responsibilities

(Adapted from [7], p. 6)

	Managerial tasks	Specifically …
Strategic orientation	Framing	Vision of the communication environment and the strategic role of communication routines. Application or infrastructure: scope and modes of use.
Organizational design	Context setting	Management approach: corporate policies vs. hands-off, decentralized approach. Related organizational approach: operational integration and control vs. self organization.
	Embedding	… into the organizational culture. … into the organizational structure (responsibilities, mandates etc.) and relating to organizational levels (corporate, business unit, group, individual).
	Rule setting	Defining the scope and level of policies and rule setting. Developing, negotiating, setting and - over time - adjusting rules.
	Creating support infrastructures	.. for routine and emerging forms of use. Responding to user requests and needs.
Implementation process	Managing the implementation	Procedural and developmental view: planned vs. emergent development, tactics of scoping and roll-out.

In this paper we are focusing on a later stage of the implementation, at which we are able to observe the first responses to the introduction of Sametime® and related managerial interventions. We use the combined evidence of voices from management and employees to reflect management challenges across the different stages of the implementation process and on the different organizational levels. Our reflections are led by two research questions:

1. Which are the specific management challenges related to the implementation of a versatile platform technology (RTC)?
2. Are there particular modes of management intervention that facilitate vision and culture compatible adoption of RTC?

3. RESEARCH DESIGN

In line with earlier work on CSCW implementation, which emphasizes the relevance of the organizational context (e.g. [11]) as well as the processes of appropriation by the users [12], we have done a thorough multi-level analysis, juxtaposing managerial actions and conceptualizations with voices from the ground, i.e. feedback from teams. We believe that detailed, multi level – specifically micro level – analyses contribute to a better understanding of implementation and use of RTC.

In comparison to the implementation of software with a particular focus, the point in the case company is to roll-out Sametime® to anybody, without requiring the users to ask or justify. In that sense we regard it as a platform technology or infrastructure.

3.1 Method

We are drawing on a case study about a medium-sized financial services company, identified by the pseudonym MUFIN.

We have conducted interviews concerning the implementation process of Sametime® at different levels of the company (management, employees) and confronted management with preliminary results of our interviews (see Figure 1).

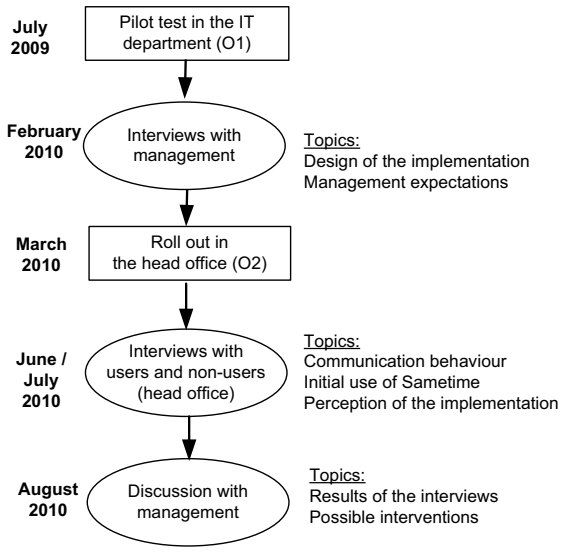


Figure 1: Overview of the data collection steps.

In February 2010, we conducted an extensive interview with the responsible managers for the implementation of Sametime®. Driving questions of this interview were the main rationale for implementing Sametime®, the organizational design of Sametime®, and management expectations concerning the use of Sametime®. To get a deeper understanding of the implementation process and possible implications for the employees, we subsequently have interviewed representatives of the workers' council, the HR department, the IT compliance and data protection office, and the line management.

In June and July 2010, subsequent to the implementation of Sametime® in the head office of MUFIN, we conducted semi-structured interviews with 13 employees. These employees were all members of one operating department of MUFIN, but they belonged to nine different teams. Key questions of these interviews were general communicational behavior, and actual use and perception of the implementation process of Sametime®.

Subsequent to a first analysis of these interviews, we presented preliminary results to the responsible IT managers in August 2010 in order to discuss possible implications for future stages in the adoption of Sametime®. Moreover the presentation provided us with a unique opportunity to double-check and verify our interpretation of the interview results and to extend our understanding of the organizational setting.

3.2 The case company

MUFIN is a financial services company operating in a tightly regulated, yet highly competitive market. The services can be characterized as information products and services. Because of this, the IT department plays a prominent role for the company. Besides the development, implementation and maintenance of the IT, the IT department is also responsible for the organizational development.

Information systems are ubiquitous in the organization and are provided as a working environment for knowledge workers in order to support their daily work, foster their productivity as well as contributing to employee empowerment, work enrichment and flexibility.

MUFIN operates a total of 10.000 workstations, 6.000 of which are run by the 2.100 sales organizations that are spread around the country. Another 4.000 are located in the head office. The IT department has about 500 staff members.

Besides the IT department, the head office comprises several operating departments. These operating departments are again subdivided into several divisions which consist of small teams of 8 to 12 employees. These teams function as a back office and support the sales organizations in their daily work.

MUFIN has a strong and explicit organizational culture and a long tradition as an employee-focused company. MUFIN is regarded as a family friendly employer and has supported telework for years. Although there are structural frictions and conflicts between the employees in the head (back) office and the sales agents, the management of MUFIN emphasizes and pursues the vision of an integrated services unit.

Corresponding to the organizational culture, MUFIN's management practices a participatory management style and recognizes its responsibility towards the workforce. Management does not only regularly involve the workers' council in decision making, but tries to achieve consensus with the council prior to organizational changes.

Concerning its organizational strategy, MUFIN is positioned as a service and customer oriented organization. Furthermore, MUFIN presents itself as an innovative organization in which IT is regarded as a core competence.

4. THE SAMETIME® IMPLEMENTATION PROCESS

The implementation process of Sametime® has been designed as a phased process across different organizational levels (see Figure

2), which reflect the organizational and strategic vision (next section) as well as a preliminary view on its implementation.

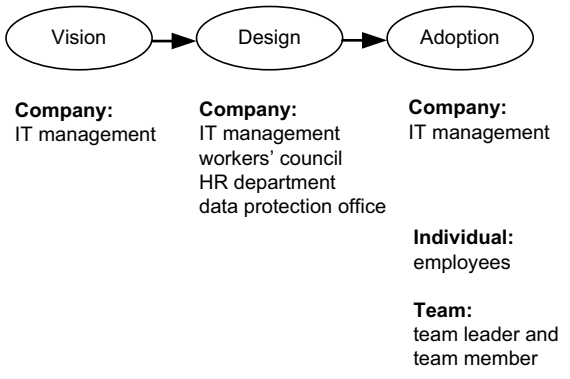


Figure 2: Stages of the Sametime® implementation process.

4.1 The vision

MUFIN's IT management views Sametime® as an enabler for organizational development and in line with the strategy of becoming an innovative service organization. Therefore it aims at a companywide implementation.

The vision for Sametime® (and subsequent Unified Communication and Collaboration (UCC) and social media technologies) comprises a set of related elements (V1-V4):

V1. Sametime® is seen as a productivity tool (operational strategy), which helps to simplify daily routines, some of which are such mundane practices as coordinating the lunch break. Moreover, facilitating and accelerating communication and collaboration across the company is seen as a key asset for a service company.

V2. Sametime® is seen as part of the HR strategy to maintain the reputation of an attractive workplace, in particular for younger employees, who are used to Internet, mobile and social media in their private and professional life.

V3. Sametime® is seen as building block to implement the vision of an integrated services unit, which will increasingly rely on knowledge sharing between back office and sales organizations. As the complexity of the services is increasing, the sales and customer service units will have to rely on experts in order to provide competent and timely responses. Application sharing and text chat are seen as supportive of the strategy.

V4. Sametime® is seen as an instrument to create a customer-focused service company. Therefore it is envisioned to link service agents into the customer portal in order to increase interactivity and customer value.

4.2 Organizational design

The implementation and roll-out of Sametime® is positioned in line with the overall strategy, vision and organizational culture of MUFIN. Although the management has the idea of implementing Sametime® throughout the entire company, they have decided to pursue a stepwise approach (O1-O4). This approach reflects MUFIN's policy to integrate all relevant parties into processes of organizational change. It allowed MUFIN to firstly concentrate on the design of the implementation in the head office and to decide on the design of the roll-out across the sales organizations at a later stage. While this process may be slower in the beginning, it is not only expected to yield more sustainable results and avoid conflicts with the workforce, but also to provide additional learning opportunities.

O1. Pilot test in the IT department (July 2009) in order to trial the technology and build organizational support for the roll-out.

The pilot test allowed the IT department to test Sametime®, to observe the uptake and employees' responses and to prepare the organization for a roll-out. In contrast, the design of the second stage of the roll-out required the involvement of the workers' council. Amongst other things, this was crucial because of the presence feature of Sametime® which could be perceived as an instrument for surveillance and control [13] and could lead to a breach of the employees' privacy [8]. In order to cope with this issue, management and workers' council reached an agreement on the design of the second stage comprising two principles:

1. Sametime® will be implemented as an open infrastructure.

   Throughout the head office, access to Sametime® will be provided for everybody without specific request. It will be provided as an open infrastructure without specific usage scenarios. Employees are free to use it as it suits them.

2. The use of Sametime® is voluntary.

   Sametime® will be provided for voluntary use. Furthermore, management has committed itself not to use Sametime® for monitoring their employees. The IT compliance and data protection manager has been involved throughout the project to ensure compliance with corporate guidelines and data protection regulation.

This agreement regulates the implementation and use of Sametime® in the head office for one year. By implementing Sametime® as an open infrastructure and on a voluntary basis, management recognizes the platform character of RTC and provides space for experimentation, adaptation and appropriation in order to encourage use and adoption [12].

To support the employees, the IT department has provided a list of company specific rules (chat etiquette), a functionality and user's guide and a discussion forum. Furthermore, there are contact persons in every operating department that are trained on Sametime®.

O2. Roll-out across the head office as a platform technology (March 2010). This phase reflects V1 & V2: the infrastructure view combined with an organizational learning approach.

While Sametime® has been conveniently integrated into the workplace infrastructure, e.g. single-sign on, it has not yet been formally integrated into the operational applications and services processes.

After the first year of use, management will conduct a review of the results. Moreover, the workers' council and management will have to decide on the future use of Sametime® by negotiating an employment agreement (e.g. [14]).

While there is a clear vision of how Sametime® could be positioned in the future of MUFIN, there is no specific plan yet for the design of phases 3 and 4:

O3. Roll-out to the sales organizations in order to forge the link between head office and sales organization (planning status). This phase is seen as the implementation of V3. However, management is aware that the notion of voluntary use will not be sufficient to support the collaboration between head office and sales organizations.

O4. Integrating chat functions into the customer portal (planning status) is seen as part of the implementation of V4.

4.3 Management's expectations

Prior to the second stage of the roll-out (O2), we asked the IT management about the expected use of Sametime® in the operating departments of the head office. Their expectations of the adoption and use of Sametime® were shaped by past experience and in harmony with the organizational culture:

1. Earlier cases of adoption of new processes or new IT

   Adoption of new processes or IT in the operating departments have shown that "peer pressure" and mutual help at the team level are often more effective than command and control structures. Therefore, management has anticipated some sort of team-based self-organization and adjustment in the adoption of Sametime®. Moreover, management has been monitoring the IT helpline in order to identify issues that might require particular attention and support.

2. The actual use of Sametime® within the IT department

   The IT department has been using Sametime® for over six months. They have established shared practices, e.g. they use Sametime® to support informal or ad-hoc communication or to negotiate availability (outeraction) [15].

4.4 Adoption in the head office

The adoption and initial use of Sametime® in the head office is affected by the individual employees' communication patterns. In the following, we will present communication patterns and initial use of Sametime®. Afterwards we will discuss employee's perceptions and management's monitoring of the implementation process.

4.4.1 General communication behavior

To understand the adoption and initial use of Sametime® in the selected operating department, it is important to be aware of employees' communication patterns.

Daily communication can be classified according to three different main communication partners:

1. Communication with the sales units

   During the day, there is a lot of communication between the sales agents and the employees. They communicate by using different communication media, e.g. telephone, e-mail, fax, and letters. While telephone is seen as the preferred medium for direct, personal and urgent communication, e-mail is mostly used to document requests and agreements.

2. Team communication

   The communication within the teams also plays a prominent role. For the team members it is very important to know about the availability of their colleagues in order to be able to provide the right information when someone calls for their colleagues. To inform others about their absence from their desk or work station, team members usually write e-mails to all team members (one-to-many communication). In addition, there is also direct one-to-one communication between the team members, e.g. to discuss problems.

3. Interdepartmental communication

   There is also interdepartmental one-to-one communication with members of other teams, divisions or operating departments, but this communication does not play a dominant role.

Besides business communication, there is also informal communication, e.g. communication to coordinate lunch breaks. This communication does not necessarily take place within the teams but normally crosses teams, divisions and departmental boundaries.

4.4.2 Initial use of Sametime®

Our interviews reveal that the frequency and intensity of use varies among the 13 interviewees (see Table 3).

Table 3: User types after four month of initial use

Frequency	Description
Extensive use	Some interviewees use Sametime® on a regular basis for different communication purposes.
Casual use	Some interviewees use Sametime® only occasionally.
Passive use	Few interviewees only connect to Sametime® to be able to receive group messages.
No use	Few interviewees do not use Sametime® at all because of concerns about privacy or disruptions.

After four month of use in the selected operating department, three features of Sametime® are mainly used (see Table 4).

Table 4: Dominant features

Feature	Examples for the use
One-to-one chat	- Coordination, e.g. to ask for a call back - Information gathering, e.g. about a customer - Informal communication
Group message	- Information about absence (team level) - Informal communication, e.g. to coordinate lunch
Presence information	- General sense of who is around - Coordination of availability

In the following, we concentrate on the role of group messages that inform about presence and absence, as this is one of the dominant communication practices on the team level (see section

4.4.1). Group messages in general allow informing a group of people about something which is only relevant at a special moment, e.g. "I will be in a meeting for the next 30 minutes". In contrast to email and chat, group messages can not be answered. They are non-persistent and only appear in a pop-up window on the screen of the addressed persons until the pop-up window is closed.

In order to use Sametime® for this purpose, it is necessary that all team members are connected to Sametime® and able to receive group messages so that all team members are able to reach all other team members through this medium [16]. However, as the use of Sametime® is voluntary, there are some employees who currently do not use Sametime® (see Table 3).

We had access to nine different teams. In five of them, the team leader and the team members had discussed and jointly agreed on the use of Sametime®. One of the interviewees reported[1]:

> „… in our team, all team members directly said 'ok, let's use it'. […] Someone from the IT department came to one of our team meetings and presented Sametime to us. […] we all said 'Ok, that is a good solution' because we always had the problem: how to inform the other colleagues."

In the remaining four teams, some employees did not use Sametime® at all. Moreover, in some of these teams there had been no discussion about the use of Sametime®. One interviewee explained:

> "No, we have not talked about how we want to deal with Sametime. We don't want to define rules … we cannot say: 'You have to connect to Sametime' … because of the aspect of voluntariness … Because of this we cannot say ‚Please, you all should use Sametime'."

The interviews have provided background information about these four teams. The first is that the employees who use Sametime® expect rules and commitment for a common use of Sametime®, at least on the team level. They want to be able to be effectively informed about presence and absence of their team colleagues. The second aspect is that some of the team leaders are cautious to discuss about a possible use of Sametime® at the team level as this might contradict the notion of voluntary use.

4.4.3 Perception of the implementation process

Most of the interviewees stated that the initial use of Sametime® was fairly easy, straightforward and self-explanatory. Because of this there was very little need for the information documents and discussion forum provided by the IT department.

In contrast to the expectations of the workers' council, there was little concern about privacy issues. In the majority of cases, the interviewees explained that trust is a very important aspect of the organizational culture. This trust is reciprocal: the company trusts its employees and employees trust the company. Therefore, surveillance and control has not been a major issue so far. One team reported that their team leader insisted that all team members would log into Sametime® as soon as they start to work. However, the team leader was reprimanded by management because her behavior was against the established rules.

[1] All quotes have been translated into English.

For most of the interviewees, the broader vision guiding the implementation of Sametime® was not noticeable. They were not able to imagine future possible fields of application. In fact, they perceived Sametime® as just another communication medium. One interviewee stated:

> "I think that it is just another possibility to communicate with each other."

As we confronted them with a possible connection between the head office and the sales organizations via Sametime®, they responded quite defensively, e.g.

> "Oh, no. No, because it will be too much. […] I just try to imagine, there is a lot to do for us. [..] There are a lot of telephone calls and mails from the sales organizations. If they would be connected to Sametime, I think I would sign off from it."

> "No, this will be too much … because, as soon as the sales organizations are connected to Sametime, they will probably say that the head office has to use it. "

Having little concerns about the use within the head office, there were profound concerns about including the sales organizations. These concerns were mainly dominated by the relationship between the head office and the sales organizations and by the amount of communication taking place between them. Amongst other things, the interviewees feared that the amount of communication between sales organizations and head office – which is already very high and sometimes intrusive and interrupting – would increase significantly. Becoming visible on Sametime® to the sales organizations was seen as creating expectations on the sales side for immediate response. Therefore they wished for clearly defined rules that could manage the communication via Sametime® between the sales organizations and the head office.

4.4.4 Monitoring of the implementation process

As management has set-up the first phase as an experimental phase, they have been monitoring the early stages of adoption. In particular they realized that the helpline was rarely used. Instead users asked each other for help, tricks or shortcuts for the use of Sametime®. While IT management had decided early on to use the base version and not to invest in customizing, suggestions for functional improvements were taken onboard. E.g. the group message function, which normally only pops-up on the screen for a few seconds, was made persistent until deleted by the user in order to ensure that the message was noted.

Management responded to the issues described in section 4.4.3 by planning a small information campaign in order to raise the awareness about the benefits of a mutual commitment – and not a formal order – to use Sametime® at the team level. They are considering to change the setting at the system level from opt-in, i.e. actively start Sametime®, (as it is right now) to opt-out.

5. DISCUSSION

The implementation of Sametime® at MUFIN is a story about introducing a platform technology for voluntary use and appropriation by teams and individuals. At the same time the implementation is perceived by management as part of a strategic vision for an innovative, modern and employee friendly company, which excels in a competitive market because of its customer focus and

integrated service offered jointly by sales units and back office. This gives rise to a number of interrelated questions:

1. Alignment (design): What is the scope of alignment and how is it achieved and maintained throughout the implementation process?
2. Alignment (leadership): Which instruments does management use to encourage and facilitate voluntary use across the company? What are the mechanisms of learning and adjustment during the experimental phase (monitoring and response)?
3. Transformation: New rules for new modes of use.

The concept of strategic alignment on corporate level was developed during the early 1990ies (e.g. [17], [18]). Over the years, the concept was a) applied in more specific settings, such as channel strategies (e.g. [19]) and b) broadened to cover social and cultural aspects. [20] for example identify several dimensions of alignment in the MIS literature: strategic/intellectual, structural, social and cultural in their state-of-the-art paper.

In conformity with these extensions, we are applying the concept of alignment to make sense of the implementation of RTC against a broader strategic, organizational, and cultural background.

5.1 Alignment and design challenges

MUFIN presents itself as an innovative and IT-oriented services company with a clear focus on its customers and employees. The vision for Sametime® has been portrayed as in line with and supportive of the overall business strategy (section 4.1). The notion of RTC as infrastructure is reflected in V1 and V2, which are quite broad. Yet they are in line with the espoused organizational culture and values of respect, empowerment and individual responsibility (organizational alignment, e.g. [21]). This is reflected in the participatory approach of an early and active involvement of the workers' council, the HR, the IT compliance and data protection office and line management.

The outcome of the negotiations is an agreement for a one year experimental and voluntary use throughout the head office without a specific functional scope or operational integration. Management is expecting that this one year will create a momentum of use and appropriation throughout the organization and will yield a level of familiarity which typically mitigates fears. Moreover, this approach suggests that management expects – based on prior experience and the organizational culture of decentralized responsibility – a dynamic of self-organization at the group level.

5.2 Initial phase of the adoption process

Although the introduction of Sametime® had been positioned as emergent and open-ended, the different parties who took part in the negotiations had different expectations regarding the likely outcome. The workers' council was expecting employees to be concerned about control and surveillance (individual level). Management expected some sort of self-organizing within the teams in the process of adopting Sametime® (team level).

In contrast to the expectations of the workers' council, very few concerns about surveillance and control have been articulated. On the contrary, the interviewees articulated a high level of trust into the integrity of management and acknowledged that management had the resources to control, if they decided to do so, anyway. We see this as clear evidence for an open and trustworthy organizational culture.

With respect to self-organization within the teams, evidence was more limited than expected (see section 4.4.2). In order to be able to use Sametime® as a reasonable replacement for e-mail to communicate absence and presence, it is inevitable that all team members use Sametime®. While in some teams all members had agreed on the use of Sametime®, in other teams some members chose not to use Sametime®. As a result, the users and team leaders who were aware of Sametime®'s potential for the team communication, wished for more specific rules and commitment to Sametime® at least at the team level (see section 4.4).

However, this view appears to be at odds with the agreement on voluntary use. The team leaders and team members have to find ways to agree on the use of Sametime® within their team. Finding an agreement is obviously not trivial because even addressing the issue might be seen as an infringement on the idea of voluntary use. This is supported by the incident about the team leader who insisted that all team members would log into Sametime® (see section 4.4.3).

What we see is a fine line between the idea of voluntary use, which has been chosen for good reasons, and subtle management guidance and interventions in order to facilitate the appropriation of Sametime® at the group level. At the same time, past investments into a cooperative culture and results-based management of the teams seem to pay off as the teams have shown willingness to take responsibility, to coordinate their work and to voluntarily make mutual commitments in order to create a more professional working environment.

5.3 Alignment and leadership

The agreement of an experimental phase of voluntary use limits the scope for management intervention. Yet, management has used subtle measures such as training members of the various operational teams as liaisons to the IT department and briefing the team managers in order to motivate them. The technical access was simplified (single-sign on) and simple ground rules have been defined (chat etiquette etc.) in order to facilitate the early stages of use.

In order to mitigate fears about surveillance and control, management sent a clear message that use was voluntary and team managers were not supposed to enforce the use of Sametime®.

Yet management has been monitoring the uptake and the use of the helpline in order to see whether further interventions are needed. At this time they are considering a) to share stories of Sametime® adoption and appropriation and b) to encourage teams to create a mutual commitment for use in order to benefit as a team. While maintaining the principle of voluntary use, management plans to discuss with the workers' council and the line managers about how to build commitments on the team level.

According to the agreement on voluntary use, there is no automatic login for Sametime®. Every employee actively has to decide if he wants to use Sametime® day-to-day (opt-in model). This turns out to be very cumbersome for the regular users. Because of this the management suggests to change this to an opt-out model. This would imply that Sametime® is started automatically, but every user is free to disconnect from Sametime®. Although the aspect

of voluntariness still remains, it will be possible to get those team members who normally just forget to connect with Sametime®.

5.4 Alignment and the implementation strategy

For the medium and long term, management has tangible ideas about specific and operational usage scenarios for Sametime® (see section 4.1). However, the very same principles of experimentation and voluntary use which have been productive during the early phase of implementation will no longer work if Sametime® is to be used in operational, customer-facing processes.

Our interviewees (team level) reported profound reservations about the idea of creating a Sametime® link between the sales organizations and the back office. Their response seemed to reflect well known conflicts between customer facing and back office units.

Still, management quite explicitly upheld the idea of an integrated service unit covering sales organizations and back office (V3 see section 4.1). They see this as a key asset for the (future) success of the company and have invested in the past to bridge the gaps between the two types of units. The use of Sametime® fits this strategy if management succeeds to prepare the organization for such a move and to successfully mitigate concerns about constant interruptions and undue delegation of work from the sales units to the back office units.

The general approach could be similar to the design of the experimentation phase, where general principles have been established and enforced in order to mitigate concerns. I.e. management would have to pay close attention to the emerging practices of communication and indeed collaboration between sales organizations and back office units. Management reported about a similar approach when the helpline was introduced in the IT department in order to protect software developers from constant interruptions by help seeking users. The communication line (helpline first) was established and enforced.

On another level, V3 incorporates a different mode of use: creating a back office team of experts with defined rosters implies an organizational setting with distinct roles and commitments. In such a scenario the use of Sametime® shifts from a person focus (presence signal and communication channel) to a role focus (a person with a particular expertise will be on duty).

While a specific plan of how to manage the tradition from an experimental use towards new modes of use and a differentiated set of application scenarios which require clear commitments and rules has not yet been articulated, management seems to rely on established organizational practices and a mature organization, willing to trust and follow management.

In sum, we see that the process of implementation and adoption implies changes and interrelated adjustments on all organizational levels, the entire company, the teams as well as the individuals, in order to reap the potential benefits of Sametime® (see Table 5).

Table 5: Organizational levels

Level	Challenges
Company	**Vision of an integrated service unit** Challenges: - strategic and cultural alignment - management of the implementation - ongoing monitoring and interventions - adjustment of the design - organizational development and rule setting
Team	**Adoption on the team level** Challenges: - appropriating Sametime® and adjusting practices at the team level - accepting responsibility and showing commitment at the team level - coordination of team rules
Individual	**Individual adoption of Sametime®** Challenges: - appropriating Sametime® and adjusting communication practices - managing the portfolio of communication media - taking responsibility for the adjustment of work practices (balancing availability and interruptions).

6. CONCLUSION

One of the general messages of the case is that context matters when we study platform technologies. Therefore we have taken pains in understanding and reconstructing the relevant context and link it to the dynamics of adoption that we have observed.

The paper describes the introduction of Sametime® as part of the communication infrastructure of a services company. It illustrates management's efforts to align the rules and guidelines of Sametime® use with a clearly articulated and enacted organizational culture of participation, empowerment and respect. Management recognizes different perspectives on RTC across different organizational levels and units.

Even in the context of such an open and employee-focused culture, RTC technologies that make the users' presence status visible across the organization raise concerns about surveillance and control. In order to mitigate these fears, management suggested a one year period of experimental, voluntary use, expecting that the self-organizing mechanisms at team level would create an organizational momentum of adoption and appropriation. Moreover, they expect that increasing familiarity with the technology will also help to dispel fears.

Still the overall vision for the use of Sametime® at MUFIN is much more ambitious and is seen as a building block to ensure the company's future competitiveness, productivity and responsiveness, as well as attractiveness for employees. More sophisticated modes of use for RTC with communication features embedded

into operational processes require an explicit commitment by the employees and clear organizational rules.

Management is patiently waiting for the results of the experimental phase before engaging in detailed planning of the next phase. This seems to reflect on the one side a learning attitude towards new technologies ("let's see how the organization will respond") but also a high level of confidence in the self-organizing abilities and dynamics of the teams.

Management walks a fine line between providing space for experimentation and setting guidelines. It is monitoring the ongoing processes of adoption and appropriation and intervening with discretion. The likely outcome is a broad use of RTC as part of the organization's communication media and channels combined with a set of specific modes of use within particular teams and user communities. The dynamics of the later might facilitate an extension of use for the former.

7. REFERENCES

[1] Berlecon Research. 2009. Perspektive Unified Communications – Wie weit sind deutsche Unternehmen. Berlecon Report, May 2009. Accessed 2010-02-25, DOI=http://www.damovo.de/DE/forms/feedback_studie_perspectiveuc_09.asp.

[2] Frößler, F. 2008. *A Practice Theoretical Analysis of Real Time Collaboration Technology – Skype and Sametime® in Software Development Projects*. Cuvillier, Göttingen.

[3] Riemer, K. and Frößler, F. 2007. Introducing Real-Time Collaboration Systems: Development of a Conceptual Scheme and Research Directions. *Communications of the Association for Information Systems*. 20, 204-225.

[4] Riemer, K., Frößler, F and Klein, S. 2007. Real Time Communication – Modes of Use in Distributed Teams. In Proceedings of the 15th European Conference on Information Systems (St. Gallen, Switzerland, June 7-9, 2007.

[5] Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

[6] Norman, D. A. 1988. *The Psychology of Everyday Things*. Basic Books, New York.

[7] Klein, S., Vehring, N. and Kramer, M. 2010. Introducing Real Time Communication: Frames, Modes & Rules. In *Proceedings of the 23rd Bled eConference on eTrust: Implications for the Individual, Enterprise and Society* (Bled, Slovenia, June 20-23, 2010), 591-606.

[8] Cameron, F. C. and Webster, J. 2005. Unintended consequences of emerging communication technologies. Instant Messaging in the workplace. *Computers in Human Behaviour*. 21, 85-103.

[9] Lazar, I. 2006. Integrating Telephony, IM, Video and Mobility with Presence. *Business Communications Review*. June 2006, 28-31.

[10] Herbsleb, J. S., Atkins, D., Boyer, D. G., Handel, M. and Finholt, T.A. 2002. Introducing Instant Messaging and Chat in the Workplace. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves* (Minneapolis, Minnesota, USA, April 22-25, 2002), 171-178.

[11] Sanderson, D. 1991. The CSCW Implementation Process: An Interpretative Model and Case Study of the Implementation of a Videoconference System. In *CSCW '92 Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, 370-377.

[12] Dourish, P. 2003. The Appropriation of Interactive Technologies: Some Lessons from Placeless Documents. *Computer Supported Cooperative Work (CSCW)*. 12, 4, 465-490.

[13] Sewell, G. 1998. The Discipline of Teams: The Control of Team-based Industrial Work through Electronic and Peer Surveillance. *Administrative Science Quarterly*, 43, 2, 397 – 428.

[14] Herholtz, P. and Frede, W. 2002. Vorschläge für eine effizientere IuK-Mitbestimmungspraxis. *Computer Fachwissen*. 1/2002, 12-18.

[15] Nardi, B. A., Whittaker, S. and Bradner, E. 2000. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of the 2000 ACM conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA, Dec 2-6, 2000), 79-88.

[16] Markus, M. L. 1990. Towards a "Critical Mass" Theory of Interactive Media. In *Organizations and communication technology*, J. Fulk and C. Steinfield, Eds. Sage Publications, Newburry Park, CA, 194-218.

[17] Baets, W. 1992. Aligning Information Systems with Business Strategy. *Journal of Strategic Information Systems*. 1, 4, 205-213.

[18] Henderson, J. C. and Venkatraman, N. 1993. Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, 32, 1, 4-16.

[19] Müller-Lankenau, C., Wehmeyer and K., Klein, S. (2006): *Strategic Channel Alignment: An Analysis of the Configuration of Physical and Virtual Marketing Channels,* in: Information Systems and e-Business Management (ISeB), 4 (2), 187 - 216.

[20] Chan Y. E. and Reich, B. H. 2007. State of the Art. IT Alignment: What Have we Learned? *Journal of Information Technology*. 22, 297-315.

[21] Semler, S. W. 1997. Systematic Agreement: A Theory of Organizational Alignment. *Human Resource Development Quarterly*. 8, 1, 23-40.

Track 6: Neue Technologien und Anwendungen – New Technologies and Applications

Bettina Berendt
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200, B-3001
Heverlee +32-16-327700
bettina.berendt@cs.kuleuven.be

Oliver Günther
Humboldt-Universität zu Berlin
Institute of Information System
Spandauer Str. 1, D-10178 Berlin
+49-30-2093-5742
guenther@wiwi.hu-berlin.de

Kerstin Schäfer
Humboldt-Universität zu Berlin
Institute of Information System
Spandauer Str. 1, D-10178 Berlin
+49-30-2093-5742
schaefke@wiwi.hu-berlin.de

VORWORT

Der Track „Neue Technologien und Anwendungen“ verzeichnete 69 Einreichungen. Hiervon wurden 8 nach der ersten Begutachtung akzeptiert, 13 weitere wurden unter Auflagen angenommen, 48 wurden abgelehnt. Von den 13 Arbeiten, die unter Auflagen akzeptiert wurden, haben letztlich 12 reüssiert, was auf eine Annahmequote von 29% hinausläuft.

Ungeachtet des weiten thematischen Fokus des Tracks haben sich anhand der Einreichungen klare Schwerpunkte herauskristallisiert. Arbeiten in diesem Track sollten sich neuen Hard- und Softwaretechnologien zuwenden und zeigen, wie diese betriebliche und organisatorische Abläufe verändern.

Ein typisches Beispiel aus dem Hardwarebereich wären RFID-Tags und Funksensoren, die es erlauben, Prozesse und Objekte zu verfolgen und im Detail in Informationssystemen abzubilden. Derartige Technologien werden neben operativen Anwendungen selbstverständlich auch Auswirkungen auf die internen Organisationsstrukturen oder auch die Kooperationsbeziehungen zwischen Unternehmen haben.

Im Softwarebereich wären z.B. agile Entwicklungsmethoden oder auch soziale Netzwerke als neue Technologien anzusehen, die betriebliche und organisatorische Abläufe nachhaltig verändern können.

Nun ist Veränderung zwar immer ein interessantes Forschungsthema, andererseits ist Veränderung zwar oft, aber nicht immer gut für Produktivität und Wettbewerbsfähigkeit. Eingereichte Arbeiten sollten daher ihren Schwerpunkt nicht auf die betrachtete Technologie an sich legen, sondern auf die ökonomischen und organisatorischen Implikationen derselben.

Die angenommenen Arbeiten haben genau diesen Spagat zwischen technischem Potenzial und gesellschaftlicher und wirtschaftlicher Relevanz vollzogen. Web 2.0, RFID sowie Mobilität waren wie erwartet Schwerpunkte der eingereichten Arbeiten, aber es wurden auch eine ganze Reihe anderer Informationstechnologien von unseren Autoren thematisiert. Daneben wurde in den meisten Arbeiten die wichtige Frage der Wirtschaftlichkeit – die ROI-Frage – angesprochen.

Wir wünschen den Lesern eine angenehme und informative Lektüre.

EDITORIAL

The track “New Technologies and Applications” attracted 69 submissions. After careful reviewing, 8 papers were accepted immediately, 13 were subject to a conditional accept, and 48 were rejected. 12 of the 13 papers that were accepted conditionally succeeded in the end, which amounts to an overall acceptance ratio of 29%.

Despite the wide focus of the track, papers clustered along some very clearly defined areas of interest. The main idea was to investigate the impact of recent hardware and software technologies on processes in business and public administration.

New information technologies, such as RFID or social networks, have a profound impact on the way we do things. They change the granularity of how we capture events and processes in our IT. They change the way companies or government agencies are organized and how they cooperate. They change the way people work together in organizations of any kind. This is true for hardware and software technologies alike – regarding software, just look at agile development methods or social online networks.

While change is always an interesting area of research, it is not always good from a productivity or competition point of view. That is why we asked potential authors from the start to not only look at a technology but also at its potential economic or societal impact.

Most successful submissions followed this lead. As we expected, Web 2.0, RFID, and mobility were the major areas of emphasis, even though a large range of other emerging information technologies were mentioned. Productivity and ROI were a major focus as well – just as we had intended.

We wish our readers a pleasant journey through the many and diverse contributions we decided to publish.

Auswirkungen der Verfügbarkeit von Kontextinformationen über mobile Nutzer auf die Vermarktung mobiler Werbeträgerleistung

Andreas Albers
Goethe Universität Frankfurt
Grüneburgplatz 1
60323 Frankfurt
+49 69 79834667
andreas.albers@m-chair.net

ZUSAMMENFASSUNG

Die Verfügbarkeit von Kontextinformationen (z.B. aktueller Aufenthaltsort) über Nutzer erlaubt es Werbetreibenden ihre mobilen Werbekampagnen zielgenau auf die Nutzungssituation ihrer Rezipienten auszurichten und dadurch eine hohe Aufmerksamkeit bei diesen zu generieren. Allerdings haben Kontextinformationen besondere Eigenschaften, die vor allem ökonomische Auswirkungen auf die traditionelle Vermarktung von Werbeträgerleistung durch Mobile Marketing Plattformanbieter (wie z.B. Google oder Yahoo) haben. Ziel der Arbeit ist es diese Auswirkungen zu identifizieren, durch Designempfehlungen für Mobile Marketing Plattformen zu adressieren und so Mobile Marketing Plattformanbietern die Nutzung der Vorteile von Kontextinformationen zu ermöglichen. Zur Evaluierung der Ergebnisse der Arbeit wurde eine Expertenumfrage durchgeführt.

Stichworte

Mobile Werbeträgerleistung, Pricing, Kontextinformationen, mobile Werbekampagnen, Mobile Marketing Plattformen.

1. EINLEITUNG

Die hohe Mobilfunkpenetration in der Bevölkerung in Kombination mit dem zunehmenden Erfolg der mobilen Datenkommunikation erlaubt es Werbetreibenden theoretisch mehr Rezipienten als über jedes andere Medium zu erreichen (Lispman 2007). Dabei machen eine hohe Interaktivität, 1:1-Kommunikation sowie die tiefe Verankerung in der Privatsphäre von Nutzern dieses Medium zum Persönlichsten aller derzeit existierenden Medien (Sultan und Rohm 2005). Nicht zuletzt deshalb lässt sich in den letzten Jahren eine kontinuierlich steigende Nachfrage nach der Aufmerksamkeit mobiler Nutzer von Seiten der Telekommunikationsunternehmen (z.B. Orange), Internet Provider (z.B. Google), Mobilgerätehersteller (z.B. Nokia) und Softwareproduzenten (z.B. Microsoft) beobachten. In der Folge ist mobile Werbung derzeit das am stärksten wachsende Marketingmittel (Strategy Analytics 2008 und Kishore 2009).

Bei dieser Entwicklung zeigt sich, dass es für Werbetreibende immer wichtiger wird, ihre werbliche Kommunikation zielgenau und bedürfnisgerecht auf ihre Rezipienten abzustimmen. Während derzeit immer noch Mobile Marketing auf Basis von SMS-Nachrichten den Markt dominiert, geht es mit dem Trend zum mobilen Web bzw. mobilen Apps in Richtung banner- oder text-basierter Werbung (Strategy Analytics 2008). In diesem Umfeld lassen sich zwar die klassischen inhaltebezogenen Methoden zur Ausrichtung von mobiler Werbung auf bestimmte Zielgruppen aus dem stationären Internet übertragen (Ehrlich 2008), allerdings stellt das mobile Internet zusätzlich explizite Kontextinformationen über einen Nutzer (z.B. aktueller Aufenthaltsort oder eindeutige Nutzerkennung) zur Verfügung.

Die Verwendung und weitere Anreicherung dieser Kontextinformationen erlaubt zusätzlich die Ausrichtung einer Werbeanzeige auf die aktuelle Nutzungssituation eines mobilen Nutzers – sein Einverständnis über die Erhebung der Kontextinformationen vorausgesetzt (Opt-In). So könnte beispielsweise ein Werbetreibender über einen mobilen Nutzer wissen, dass dieser männlich, 20-30 Jahre alt und Student ist, und sich gerade zur Mittagszeit in München am Hauptbahnhof aufhält. Daraufhin könnte für seine Nutzungssituation relevante Werbung (z.B. für ein Restaurant in der Nähe) übermittelt werden.

Während Kontext bzw. Kontextinformationen nicht für jede Form von mobilen Werbekampagnen nützlich sein müssen, können, bei angemessenem Einsatz, nichtdestotrotz Vorteile für die folgenden, involvierten Stakeholder festgestellt werden:

- **Mobile Marketing Plattform Anbieter (MMPA)**: MMPAs vermarkten den Kommunikationskanal zwischen mobilem Nutzer und Werbetreibenden zur Umsetzung von Werbekampagnen in Form von *mobiler Werbeträgerleistung* (z.B. ein virtueller Werbeplatz für Werbebanner auf der mobilen Portal-Website des MMPA). Die Bereitstellung des

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Zugangs zu Kontextinformationen für Werbetreibende erlaubt es MMPAs sich gegenüber Anbietern traditioneller Mobile Marketing Plattformen (MMPs) zu differenzieren. Im Weiteren stellt die situationsbezogene Beschreibung mobiler Nutzer für Werbetreibende einen höheren Wert dar – die zudem zu einer höheren Zahlungsbereitschaft bei Letzteren führen kann (Figge 2007).

- **Mobile Nutzer:** Durch Preisgabe eigener Kontextinformationen, versetzen mobile Nutzer Werbetreibende in die Lage, ihre derzeitigen Informationsbedürfnisse durch ein entsprechendes werbliches Angebot zu befriedigen (Figge und Theyson 2006).
- **Werbetreibende:** Zur Durchführung von mobilen Werbekampagnen muss ein Werbetreibender zunächst mobile Werbeträgerleistung erwerben. In diesem Zusammenhang erlauben es Kontextinformationen die Relevanz eines mobilen Nutzers vor Kauf zu prüfen und dann erst eine Kaufentscheidung zu treffen. Somit ist es möglich, dass mobile Werbung nur an relevante Rezipienten versendet wird (Figge und Theyson 2006; Figge 2007).

Trotz der aufgezeigten Vorteile von kontextsensitiven, mobilen Werbekampagnen zeigt Albers (2007, 2008), dass beim Einsatz von Kontextinformationen im Zusammenspiel mit mobilen Werbekampagnen potentiell Effekte auftreten können, die insbesondere ökonomische Auswirkungen für die traditionelle Vermarktung von mobiler Werbeträgerleistung durch MMPAs haben. Ziel des vorliegenden Papiers ist es deshalb diese potentiellen Auswirkungen zu diskutieren und entwickelte Empfehlungen für das Design von MMPs zu präsentieren, damit die dargestellten Vorteile von Kontextinformationen an die Stakeholder mobiler Werbekampagnen weitergegeben werden können.

Dazu beschreibt Kapitel 2 zunächst die analysierte Umsetzung kontextsensitiver, mobiler Werbekampagnen, die durch MMPAs ermöglicht wird, und grenzt somit den Forschungsbereich der Arbeit ein. Innerhalb des definierten Schwerpunkts der Arbeit dokumentiert Kapitel 3 den Stand der Forschung zu kontextsensitiven, mobilen Werbekampagnen. Kapitel 4 stellt dann das Forschungsziel sowie die der Arbeit zugrundeliegenden Forschungsmethodik vor. Im Anschluss daran präsentiert Kapitel 5 die Analyse potentieller, relevanter Auswirkungen von Kontextinformationen. Auf Basis dieser Erkenntnis werden dann entwickelte Designempfehlungen für MMPs vorgeschlagen. Die empirische Validierung dieser Designempfehlungen mittels Expertenbefragung findet in Kapitel 6 statt. Abschließend fasst Kapitel 7 die Arbeit zusammen und zeigt einen Ausblick auf weiteren Forschungsbedarf in diesem Bereich.

2. MOBILE MARKETING PLATT-FORMEN FÜR KONTEXTSENSITIVE, MOBILE WERBUNG

Bis zum heutigen Tag profitieren mobile Werbekampagnen überwiegend von der tiefen Verankerung des mobilen Mediums in der Privatsphäre ihrer Nutzer – und damit von deren persönlichen Natur (Birkel 2008). Neben dieser immanenten Eigenschaft erlauben es Kontextinformationen (z.B. Aufenthaltsort, Uhrzeit, Identifizierbarkeit, etc.), die individuelle, aktuelle Situation von mobilen Nutzern für Werbetreibende abzubilden und damit eine effektivere Ansprache durch mobile Werbekampagnen zu ermöglichen (Figge 2007).

In ihrem vielzitierten Beitrag definieren Dey und Abowd (1999) einen solchen Kontext eines Nutzers als „…Informationen, die verwendet werden können um die Situation einer Entität zu beschreiben. Eine Entität ist eine Person, ein Ort oder ein Objekt, die als relevant für die Interaktion zwischen Nutzer und Anwendung, einschließlich Nutzer und Anwendungen selbst, erachtet wird“. Dabei unterscheiden die Autoren zwischen primärem und sekundärem Kontext. Der *primäre Kontext* definiert sich dabei aus den *direkt* über eine Entität verfügbaren Informationen, repräsentiert durch *Ort, Person, Zeit, und Aktion*. Im Rahmen der vorliegenden Arbeit werden diese Informationstypen wie folgt auf Informationen, die über das Mobilfunknetz bzw. das mobile Internet erhoben werden können, abgebildet:

- **Aufenthaltsort:** Der Aufenthaltsort mobiler Nutzer kann entweder durch eine zellbasierte Ortung über das Mobilfunknetz erhoben (Schiller 2003) oder durch eine gerätebasierte Ortung auf Basis des Global Positioning Systems (GPS) bzw. drahtlosen Netzwerkzugriffs (WLAN) ermittelt werden (Albers et al. 2005).
- **Identität bzw. Identifizierbarkeit:** Die weltweit eindeutige Mobile Subscriber Integrated Services Digital Network Number (MSISDN), gespeichert auf der Subscriber Identity Modul (SIM) Karte des Nutzers, ist die Voraussetzung für den Zugriff eines jeden mobilen Endgeräts auf das Mobilfunknetz (Schiller 2003). Somit sind Nutzer im Mobilfunknetz jederzeit durch eine eindeutige Nummer über den Mobilfunkbetreiber identifizierbar und damit wiedererkennbar.
- **Uhrzeit:** Die Bestimmung des Zeitpunkts an dem ein mobiler Nutzer auf das Mobilfunknetz zugreift ist trivial und wird direkt vom Mobilfunkbetreiber zur Verfügung gestellt (Robra-Busssantz 2005).
- **Nutzeraktion:** Zur Bestimmung des Nutzerverhaltens können vom Internet Provider oder anderen Webanalyse Unternehmen eingegebene Suchbegriffe, abgesendete Web Formulare oder angeklickte Links des Nutzers erfasst und entsprechend als Nutzeraktionen interpretiert werden (Steimel et al. 2008).

Aus den aufgeführten primären Kontextinformationen kann im Weiteren der *sekundäre Kontext* ermittelt werden. Dieser definiert sich durch indirekte Informationen, die aus dem primären Kontext abgeleitet werden können. In Rahmen dieser Arbeit werden dazu die folgenden sekundären Kontextinformationen berücksichtigt:

- **POS Distanz:** Übermittelt ein Werbetreibender im Rahmen einer Werbekampagne seinen Point-of-Sales (POS) Standort (z.B. ein Restaurant), so kann zusammen mit dem aktuellen Aufenthaltsort des Nutzers die Distanz zwischen POS und Nutzer ermittelt werden (Huff 1964).
- **Verbleibende Besuchszeit:** Übermittelt ein Werbetreibender im Rahmen einer Werbekampagne die Öffnungszeiten seines POS, kann auf Basis der aktuellen Tageszeit die verbleibende Besuchszeit für einen Nutzer ermittelt werden.

- **Persönliche Präferenzen:** Durch die eindeutige Identifizierbarkeit eines mobilen Nutzers können für diesen vorab gespeicherte, persönliche Präferenzen aus einer entsprechenden Datenbank ermittelt werden (vgl. Database Marketing). Diese Informationen können dabei z.B. aus sozio-demographischen, geographischen oder transaktionsbasierten Daten bestehen und zu einem Nutzerprofil zusammengeführt werden (Nieschlag et al. 2001).
- **Aktuelles Bedürfnis:** Aus den Nutzeraktionen lässt sich im Weiteren auch ggf. das aktuelle Bedürfnis eines Nutzers in seiner aktuellen Situation ableiten (Ehrlich 2008). Beispielsweise dann, wenn ein Nutzer eine bestimmte Portalkategorie (Restaurants) oder Themengebiet (Mittagessen) auswählt und damit eine Handlungsabsicht verknüpft werden kann (vgl. Behavioral Targeting).

Auf Basis dieser Kontextinformationen ermöglicht dann eine MMP die Bereitstellung von mobiler Werbeträgerleistung zur Umsetzung kontextsensitiver, mobiler Werbekampagnen. Abbildung 1 beschreibt diesen Ablauf und damit das Zusammenspiel der beteiligten Akteure für dieses Nutzenangebot.

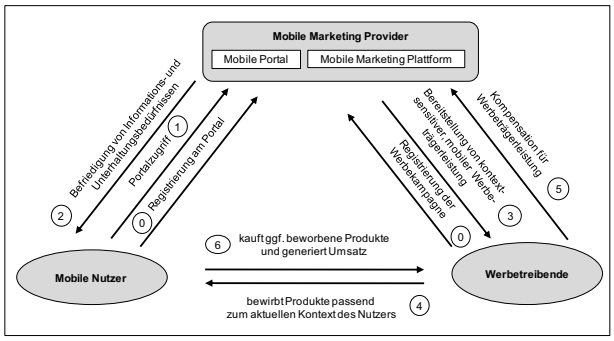


Abb. 1: Vermarktung kontextsensitiver, mobiler Werbeträgerleistung über eine MMP

Voraussetzung für die Vermarktung mobiler Werbeträgerleistung durch den MMPA ist die Registrierung des Nutzers am MMP Portal (0). Hierbei kann der Nutzer beispielsweise persönliche Präferenzen übermitteln und Privatsphäreneinstellungen hinsichtlich der Offenlegung seiner Kontextinformationen festlegen. Auf der anderen Seite müssen Werbetreibenden ihre geplanten mobilen Werbekampagnen beim MMP registrieren (0). Dabei übermitteln sie die Inhalte der Werbekampagnen und geben dabei ihre Zielgruppenprofile (z.B. männlich, 20-30 Jahre, max. 500m POS Distanz) sowie ihr Budget an.

Sind diese Vorbereitungen abgeschlossen, greift der Nutzer auf das mobile Portal des MMPA zu (1). Bei diesem Zugriff werden zunächst seine freigegebenen und technisch verfügbaren Kontextinformationen erhoben und zu einer Nutzungssituationsbeschreibung (NSB) zusammengeführt. Diese NSB dient zum einen der Personalisierung der für den mobilen Nutzer bereitgestellten Inhalte (2) und zum anderen zur Ausrichtung einer mobilen Werbekampagne auf die Bedürfnisse des mobilen Nutzers in seiner aktuellen Situation (Targeting). Zur Umsetzung ihrer Werbekampagnen wird den Werbetreibenden im Anschluss mobile Werbeträgerleistung durch den MMPA zur Verfügung gestellt (3). Diese beinhaltet, dass mobile Werbekampagnen für die passenden Nutzungssituationen auf dem mobilen Endgerät des Nutzers angezeigt werden (4). Für eine angezeigte Werbekampagne entrichtet der Werbetreibende schließlich ein Entgelt an den MMPA (5). Wird der Nutzer letztlich von der mobilen Werbekampagne angesprochen, kauft er ggf. das beworbene Produkt und generiert somit Umsatz für den Werbetreibenden (6).

Beim Vergleich von traditioneller und kontextsensitiver, mobiler Werbeträgerleistung lassen sich zunächst zwei offenkundige Unterschiede hinsichtlich des Targetings von Werbekampagnen feststellen (Albers 2008; Birkel 2008):

- **Eindeutigkeit von Kontextinformationen:** Im Gegensatz zu Nutzerprofilen bei traditioneller mobiler Werbeträgerleistung repräsentieren Kontextinformationen ein explizites (nicht statistisch ermitteltes) Wissen über individuelle, mobile Nutzer.
- **Kontrolle der Nutzer über Kontextinformationen:** Kontextinformationen müssen im Gegensatz zu traditionellen Nutzerinformationen i.d.R. explizit durch den Nutzer preisgegeben werden (z.B. durch Eingabe von persönlichen Präferenzen).

Diese zwei signifikanten Unterschiede stellen für die vorliegende Arbeit bereits eine ausreichende Motivation dar, um weitergehende Auswirkungen von Kontextinformationen auf die Vermarktung von traditioneller, mobiler Werbeträgerleistung für MMPAs zu untersuchen.

3. STAND DER FORSCHUNG

Das Konzept kontextsensitiver, mobiler Werbung ist nicht neu. Allerdings spielt es immer noch eine untergeordnete Rolle in der Praxis. Kommt diese Werbeform im mobilen Web oder für mobile Apps zum Einsatz, dann wird überwiegend der aktuelle Aufenthaltsort eines mobilen Nutzers verwendet um physische Angebote oder POS in der geographischen Nähe eines Nutzers zu bewerben (Location-Based Advertising) (Steimel et al. 2008).

Im Vergleich zur Praxis spielt kontextsensitive, mobile Werbung bisher ebenfalls nur eine untergeordnete Rolle (Leppäniemi et al. 2006). Jedoch werden bereits Eigenschaften von Kontextinformationen, die über deren Status als reine Zusatzinformationen für klassische Nutzerprofile hinausgehen, beleuchtet. In diesem Zusammenhang liefern Ranganathan (2002) und Varshney (2003) eine generelle Übersicht von Chancen und Herausforderungen von kontextsensitiver, mobiler Werbung. Die übrigen Beiträge zu diesem Thema lassen sich dann grob in die Bereiche Design, Nutzerakzeptanz und ökonomische Aspekte von kontextsensitiver, mobiler Werbung einteilen.

In Bezug auf Designaspekte deckt der größte Teil der Beiträge zu kontextsensitiver, mobiler Werbung die Entwicklung von Architekturrahmenwerken und prototypischen Implementierungen zur Demonstration der (technischen) Machbarkeit ab (E-LBA Project 2002; Ojala et al. 2003; Aalto et al. 2004; Bulander et al. 2005; Hristova und O'Hare 2004; Kurkovsky und Harihar 2006; Tang et al. 2007; Thawani et al. 2007).

Forschung zur Nutzerakzeptanz wird hauptsächlich im Zusammenhang mit Fragestellungen zur Privatsphäre mobiler Nutzer, dem Design von Benutzeroberflächen und dem tatsächlichen Nutzen kontextsensitiver, mobiler Werbung für mobile Nutzer betrieben (Aalto et al. 2004; Klafft und Spiekermann 2006; Bruner 2007).

Hinsichtlich ökonomischer Aspekte kontextsensitiver, mobiler Werbekampagnen existieren nur einige wenige Beiträge. Deren Schwerpunkt ist dabei die Erweiterung, Anpassung und Optimierung von MMPA Geschäftsmodellen (Figge 2007; Tripathi 2003; Gopal und Tripathi 2006).

Zusammengefasst zeigt sich, dass der aktuelle Stand der Forschung sich derzeit hauptsächlich auf Konzepte und Architekturen zur Einbindung von Kontextinformationen in mobile Werbekampagnen konzentriert. Die ökonomischen Auswirkungen von Kontextinformationen auf die traditionelle Vermarktung mobiler Werbeträgerleistung durch MMPAs werden dabei kaum oder nur am Rande adressiert.

4. FORSCHUNGSZIEL UND -METHODIK

Das Forschungsziel der vorliegenden Arbeit ist es, potentielle Auswirkungen der Bereitstellung von Kontextinformationen für mobile Werbekampagnen auf die traditionelle Vermarktung von mobiler Werbeträgerleistung zu untersuchen und ggf. entsprechend der Geschäftsziele von MMPAs adressierbar zu machen.

Aufgrund des kaum vorhandenen Einsatzes kontextsensitiver, mobiler Werbung in der Praxis lassen sich potentielle Auswirkungen von Kontextinformationen nur schwer empirisch erheben. Deshalb werden zunächst relevante Auswirkungen mittels Literaturrecherche explorativ identifiziert und diskutiert. Dazu folgt diese Analyse einer vereinfachten Wertschöpfungskette für mobile Werbeträgerleistung bei der Kontextinformationen erhoben, aufbereitet und schließlich für die Veräußerung von mobiler Werbeträgerleistung genutzt werden.

Die ermittelten Auswirkungen liefern dann im Anschluss die Grundlage zur Entwicklung von Designempfehlungen für MMPs. Dabei stehen die ökonomischen Implikationen für MMPs, die sich aus der Verwendung von Kontextinformationen ergeben, im Fokus der Arbeit.

Aufgrund der Tatsache, dass kontextsensitive, mobile Werbekampagnen bisher kaum in der Praxis vorkommen, war die Durchführung von Case Studies oder Nutzerumfragen zur Evaluierung der erarbeiteten Designempfehlungen wenig sinnvoll. Deshalb wurde eine *Expertenbefragung* in Form von vorstrukturierten, qualitativen Interviews als Methode gewählt. In diesem Zusammenhang wurden vier Hypothesen in Bezug auf die Nutzung von Kontextinformationen sowie deren Auswirkungen und Adressierung aufgestellt und damit die wichtigsten Eckpunkte des Nutzens der entwickelten Designempfehlungen geprüft.

Die Experten wurden aus einem Portfolio von Unternehmen ausgewählt, das möglichst viele relevante Aspekte des zugrundeliegenden Wertschöpfungsnetzwerks einer MMP abdeckt. Dabei wurde zunächst die Sicht eines MMPAs (*allesklar.com*) und einer Mobile Advertising Agentur (*InteractiveMedia*) vertreten. Im Weiteren hat das Unternehmen *Deloitte* schwerpunktmäßig Identity Management- und Privatsphärenaspekte adressiert. Hinsichtlich der Geschäftsmodelle im mobilen Internet konnte ein Experte der Unternehmensberatung *Detecon International GmbH* befragt werden. Zu IKT-Infrastrukturfragestellungen hat insbesondere das Systemhaus *T-Systems* und den Telekommunikationsausrüsters *Nokia-Siemens Networks* Stellung genommen[1].

Die Auswertung der erhobenen Daten erfolgte auf Basis der qualitativen Inhaltsanalyse (QIA). Sie stellt ein Werkzeug bereit, um jegliche Art von Material, das einer Kommunikation entstammt, zu analysieren und auszuwerten (Mayring 2008). Unter Anwendung der *Zusammenfassenden QIA* wurden daraufhin die Antworten der Experten für jede gestellte Interviewfrage zu generellen Aussagen zusammengefasst und diese abschließend interpretiert.

Der dadurch entstandene Beitrag der Arbeit kann dann vor dem Hintergrund der folgenden Einschränkungen abschließend begutachtet werden:

- Es wurde nur ein monopolistischer MMPA angenommen. Die Berücksichtigung von Wettbewerb ist wichtig und sinnvoll, wurde aber aus Komplexitätsgründen als zukünftige Forschung zurückgestellt.
- Ebenfalls aus Komplexitätsgründen wurde nur eine beschränkte Anzahl von Kontextinformationen untersucht. Dabei beschränkte sich die Arbeit auf die durch Dey und Abowd (1999) vorgegebenen Typen von primären Kontextinformationen (Ort, Zeit, Identität und Aktion) sowie den daraus abgeleiteten sekundären Kontextinformationen.

Es wurde dabei auch unterstellt, dass es einem MMPA technisch möglich ist, alle relevanten Kontextinformationen eines mobilen Nutzers für eine mobile Werbekampagne zu erheben. In der Realität könnten diese auch über verschiedene Service Provider verteilt sein, auf die ein MMPA nicht unbedingt Zugriff hat (FIDIS 2005).

5. DESIGNEMPFEHLUNGEN FÜR MOBILE MARKETING PROVIDER PLATTFORMEN

Motiviert durch die in Kapitel 2 identifizierten, besonderen Eigenschaften von Kontextinformationen, werden nun deren potentiellen Auswirkungen auf die Vermarktung traditioneller Werbeträgerleisung präsentiert und durch entwickelte Designempfehlungen adressiert.

5.1 Auswirkungen von Kontextinformationen auf mobile Werbeträgerleistung

Die wichtigsten Auswirkungen von Kontextinformationen werden anhand einer vereinfachten Wertschöpfungskette für mobile Werbeträgerleistung analysiert (vgl. Abb. 2).

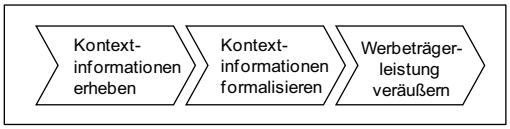


Abb. 2: Vereinfachte Wertschöpfungskette für kontextsensitive, mobile Werbeträgerleistung

Dadurch besteht die Wertschöpfungskette aus der Erhebung von Kontextinformationen über mobile Nutzer, deren Formalisierung

[1] Obwohl es für die verschiedenen Fragestellungen bevorzugte Unternehmen gab, haben alle Unternehmen zu allen Interviewfragen Stellung genommen.

zur Weiterverarbeitung sowie letztlich der Nutzung von Kontextinformationen zur Veräußerung mobiler Werbeträgerleistung. Entlang dieser drei Kettenglieder werden im Folgenden die Auswirkungen von Kontextinformationen dargestellt.

5.1.1 Erhebung von Kontextinformationen

Zunächst gilt es Kontextinformationen über den mobilen Nutzer zu erheben. Im Gegensatz zu Nutzerprofilen für traditionelle, mobile Werbeträgerleistung müssen Kontextinformationen (z.B. Aufenthaltsort) aktiv vom Nutzer per Opt-In preisgegeben werden (vgl. Kapitel 2). NSBs können somit für individuelle, mobile Nutzer unterschiedlich konkret oder detailreich ausfallen (z.B. könnte ein Nutzer seinen Aufenthaltsort privat verdecken während er im geschäftlichen Kontext sichtbar wäre). Dadurch hängt der Wert eines mobilen Nutzers für den Werbetreibenden neben dessen Relevanz für seine Werbekampagne zusätzlich von der Bereitschaft mobiler Nutzer ab, relevante Kontextinformationen über sich preiszugeben (Albers 2007, 2008). Da werbliche Angebote i.d.R. keinen direkten Anreiz für die Preisgabe von persönlichen Informationen bieten, gilt es für Werbetreibende bzw. MMPAs mobile Nutzer auf andere Weise dazu zu motivieren.

5.1.2 Formalisierung von Kontextinformationen

Ein weiterer Unterschied von Kontextinformationen gegenüber Nutzerprofilen aus dem klassischen Mobile Marketing besteht darin, dass mobile Nutzer diese Informationen nicht nur kontrollieren, sondern auch aktiv gegenüber einem MMPA bzw. Werbetreibenden kommunizieren müssen (z.B. persönliche Präferenzen). Somit gilt es über diese Informationen ein gemeinsames, semantisches Verständnis über alle Marktakteure (MMPA, mobile Nutzer und Werbetreibende) herzustellen (Schmid 1999, Figge 2007) und so eine effektive Ausrichtung von mobilen Werbekampagnen auf relevante mobile Nutzer zu ermöglichen.

5.1.3 Veräußerung von Werbeträgerleistung

Während Nutzerprofile im traditionellen Mobile Marketing ex post auf Basis von Nutzerstatistiken erstellt werden (Ehrlich 2008), stellen Kontextinformationen hingegen *explizite* Informationen über *individuelle* Nutzer dar, die vor dem Kauf von mobiler Werbeträgerleistung zur Verfügung stehen. Dies erlaubt es Werbetreibenden für ihre Werbekampagnen nur solche Nutzer zu selektieren, die für sie die größte Geschäftsrelevanz aufweisen (Albers 2007).

Vorhandene explizite Kontextinformationen führen zudem dazu, dass der potentielle Kontakt zu einem mobilen Nutzer durch nutzerindividuelle POS Distanz, Präferenzen, etc. einzigartig für jeden Werbetreibenden wird und somit einen individuellen Wert bekommt. Daraus resultiert letztlich wiederum eine nutzerspezifische Zahlungsbereitschaft für mobile Werbeträgerleistung (heterogenes Gut), deren Höhe nur Werbetreibende selbst genau kennen (Bohte et al. 2004).

Aufgrund des volatilen und heterogenen Charakters von NSBs müssen Werbetreibenden in die Lage versetzt werden, den Wert der kontextsensitiven, mobilen Werbeträgerleistung für einen mobile Nutzer zu bestimmen, da sie dieses Gut ansonsten nicht kaufen würden (Haase et al. 1998). Dadurch zeigt sich, dass die klassischen Preismodelle im Mobile Marketing auf Tausend-Kontakt-Preis Basis[2] nicht mehr in der Lage sind eine optimale Allokation mobiler Werbeträgerleistung für individuelle, mobile Nutzer herbeizuführen (Sharma et al. 2008).

Im Weiteren muss es Ziel von MMPAs sein, mobile Werbeträgerleistung zum optimalen Preis zu verkaufen ohne gleichzeitig dabei zu riskieren dieses Gut aufgrund falscher Preisgestaltung nicht zu verkaufen, da das Gut nicht gelagert werden kann (Bothe et al. 2004).

Schließlich dürfen MMPAs mobile Werbeträgerleistung nicht an beliebige Werbetreibende verkaufen. Sie müssen eine gewisse Relevanz zwischen beiden Parteien sicherstellen, um zu vermeiden, dass mobile Nutzer durch unpassende mobile Werbekampagnen gestört werden. Beispielsweise könnte ein mobiler Nutzer einen beworbenen POS auch physisch statt nur virtuell besuchen und somit einen größeren Schaden haben, falls dieser nicht für ihn relevant ist (Feldmann 2005).

5.2 Entwicklung von Designempfehlungen für MMPs

Auf Basis der in Abschnitt 5.1 diskutierten Auswirkungen von Kontextinformationen auf die Vermarktung traditioneller, mobiler Werbeträgerleistung werden nun die entwickelten Designempfehlungen für MMPs entlang der vorgestellten Wertschöpfungskette präsentiert. Zunächst werden dazu Aspekte der Erhebung von Kontextinformationen über mobile Nutzer diskutiert, dann Vorschläge zur Formalisierung letzterer angeführt und schließlich ein Preismechanismus zur Veräußerung kontextsensitiver, mobiler Werbeträgerleistung vorgeschlagen.

5.2.1 Empfehlungen für die Erhebung von Kontextinformationen

Die Aufmerksamkeit von Nutzern für mobile Werbekampagnen kann im Vergleich zu anderen physischen Gütern von MMPAs nicht direkt über einen Beschaffungsmarkt akquiriert werden. Stattdessen muss sie durch die Bereitstellung von mobilen Inhalten gewonnen werden. Dabei lassen sich redaktionelle, nutzer- und durch Werbetreibende generierte Inhalte (z.B. Gelbe Seiten) unterscheiden. Der Wert dieser Aufmerksamkeit wird signifikant durch die verfügbaren Kontextinformationen über mobile Nutzer bestimmt. Es sollte deshalb das Ziel von MMPAs sein, qualitativ hochwertige Kontextinformationen mittels folgender Maßnahmen zu akquirieren:

- **Schaffung von Anreizen:** Mobile Nutzer benötigen Anreize (z.B. finanzielle Anreize, soziales Ansehen oder Personalisierung), um persönliche Daten von sich preiszugeben (Kobsa 2007).
- **Bereitstellung von mobilem Identitätsmanagement:** Laut Friberg (2007) zeigen verschiedene Studien, dass durch die Möglichkeit mobiler Nutzer die Verwendung ihrer persönlichen Daten zu kontrollieren, Privatsphärenbedenken positiv beeinflusst werden können. Ein mobiles Identitätsmanagement System (mIdM) bietet in diesem Zusammenhang eine solche Kontrolle und stellt zudem weitere Mehrwertdienste (z.B. Single-Sign-On,

[2] Preis eines MMPA für 1000 Kundenkontakte mit einem bestimmten Profil im Rahmen der Umsetzung einer mobilen Werbekampagne

Nachverfolgung von getätigten Transaktionen, etc.) bereit (FIDIS 2005).

- **Umsetzung von Transparenz:** Die transparente Darstellung der Verarbeitung persönlicher Daten für mobile Nutzer fördert die Herstellung einer Vertrauensbeziehung zwischen mobilen Nutzern und dem MMPA. Auf der einen Seite können so die Angaben von Werbetreibenden überprüft werden, um die Auslieferung von irrelevanten mobilen Werbekampagnen aufgrund falscher Angaben (z.B. falsche POS Adresse oder Öffnungszeiten) zu reduzieren. Auf der anderen Seite kann für mobile Nutzer transparent gemacht werden, warum sie eine bestimmte mobile Werbekampagne erhalten haben (z.B. der werbende POS ist in der geographischen Nähe eines mobilen Nutzers) (Radmacher 2007).

5.2.2 Empfehlungen zur Formalisierung von Kontextinformationen

Um akquirierte Kontextinformationen zu nutzen, gilt es die Eigenschaften und die Qualität dieser Informationen für Werbetreibende bewertbar zu machen (Koppius 1999).

Zu diesem Zweck wird eine angemessene Darstellungsform für Kontextinformationen benötigt, die ein gemeinsames semantisches Verständnis aller Marktteilnehmer herstellen kann. Allgemein findet sich in der Literatur dazu entweder der Einsatz eines Produktkatalogs (Schmid 1999) oder die Verwendung von Ontologien (Daconta 2003).

Speziell für das Mobile Marketing entwickelte Figge (2007) ein semantisches Identitätskonzept (SIK), das aus drei Sub-Ontologien besteht. Diese erlauben es die Nutzungssituation eines mobilen Nutzers basierend auf dessen Kontextinformationen *Identität, Aufenthaltsort und Zeit* zu beschreiben.

Für die formale Darstellung von NSBs über mobile Nutzer kann das SIK verwendet werden, wenn es um drei Aspekte erweitert wird:

- Es fehlt für das SIK der Kontextinformationstyp *Aktivität*, der das aktuelle Bedürfnis eines mobilen Nutzers beschreiben kann.
- Es müssen die vorhandenen Kontextinformationen durch *Qualitätsinformationen* (z.B. Genauigkeit des Aufenthaltsorts eines mobilen Nutzers) ergänzt werden, um eine objektive Bewertung durch Werbetreibende zu ermöglichen.
- Es muss das SIK um *Regeln für die Verwendung von Kontextinformationen* (d.h. Privacy Policies) ergänzt werden, damit mobile Nutzer die Kontrolle über ihre persönlichen Daten behalten.

Nicht zuletzt müssen die verfügbaren Kontextinformationen in ein einheitliches Format (z.B. Uhrzeitformat) überführt und dann auf Basis des erweiterten SIK zu einer Nutzungssituationsbeschreibung zusammengeführt werden.

5.2.3 Empfehlungen für einen Preismechanismus zur Veräußerung von kontextsensitiver, mobiler Werbeträgerleistung

Aufgrund des Charakters kontextsensitiver, mobiler Werbeträgerleistung als volatiles, heterogenes Gut dessen tatsächlichen Wert nur Werbetreibende selbst kennen (vgl. Abschnitt 5.1.3), ist dessen klassische statische Bepreisung für MMPA nicht mehr optimal (es besteht nun eine individuelle Zahlungsbereitschaft seitens der Werbetreibenden, die von MMPA abgeschöpft werden kann). Daher wird als Preis- bzw. Allokationsmechanismus ein dynamisch-interaktiver Bepreisungsmechanismus (Skiera et al. 2005) für kontextsensitive, mobile Werbeträgerleistung vorgeschlagen.

Die Evaluierung existierender Bepreisungsmodelle in der Literatur ergab, dass als einziges Konzept die *Auktion* zu diesem Zweck geeignet scheint. Sie ermöglicht Werbetreibenden ihre individuelle Zahlungsbereitschaft aktiv gegenüber einem MMPA auszudrücken. Während die konkrete Gestaltung der Auktion von den Anforderungen bzw. Zielen eines MMPAs abhängt, zeigt die tiefe Verankerung des mobilen Mediums in der Privatsphäre der Nutzer, dass ein MMPAs zumindest ein gewisses Maß an Relevanz zwischen mobilem Nutzer und Werbetreibenden bzw. Werbekampagne sicherstellen sollte (vgl. Abschnitt 5.1.3) und diese Anforderung dann auch in den Allokationsprozess integriert werden muss.

Als geeigneter Auktionstyp zu diesem Zweck bietet sich daher das *Multi-attributive Auktionsformat* an. Diese Versteigerungsform hat ihren Ursprung in der Beschaffungsdomäne und erlaubt es Bietern, neben dem Preis, weitere Qualitätskriterien in das Gebot zu integrieren (z.B. Lieferzeiten, Produkt, etc.). Ihre zentrale Komponente ist die Nutzenfunktion des Gebotsempfängers. Als Input erwartet sie von jedem Bieter den *Preis* und die *Qualitätskriterien* für ein ökonomisches Gut. Der Bieter für den die Nutzenfunktion den höchsten Wert zurückliefert, gewinnt die Auktion (Bichler 2000).

Überträgt man nun das Konzept *Multi-attributive Auktion* auf die Versteigerung kontextsensitiver, mobiler Werbeträgerleistung, so kann zunächst die Relevanz zwischen mobilem Nutzer und Werbetreibenden durch den Abgleich zwischen NBS- und Zielgruppenprofil (z.B. POS Distanz eines mobilen Nutzers) als *Qualitätskriterium* bestimmt werden. Danach kann das Ergebnis dieses Abgleichs zusammen mit dem *Preis* in die Nutzenfunktion des MMPAs eingesetzt werden. Die Auswertung der Nutzenfunktion ergibt dann das endgültige Gebot für Auktion[3].

Abbildung 3 zeigt beispielhaft wie die Relevanz einer mobilen Werbekampagne für einen Nutzer in den Auktionsprozess integriert werden kann. Dazu dividiert der MMPA in diesem Szenario die abgegebenen monetären Gebote der Werbetreibenden bzw. der POS durch deren Distanz zum mobilen Nutzer und erhält damit ein virtuelles Gebot, das Relevanz und Gebotshöhe im Rahmen der Auktion ausbalanciert.

[3] Beispielweise wertet eine größere POS Distanz zu einem mobilen Nutzer das monetäre Gebot eines Werbetreibenden gegenüber dem mit einer kürzeren POS Distanz eines anderen Werbetreibenden, ab.

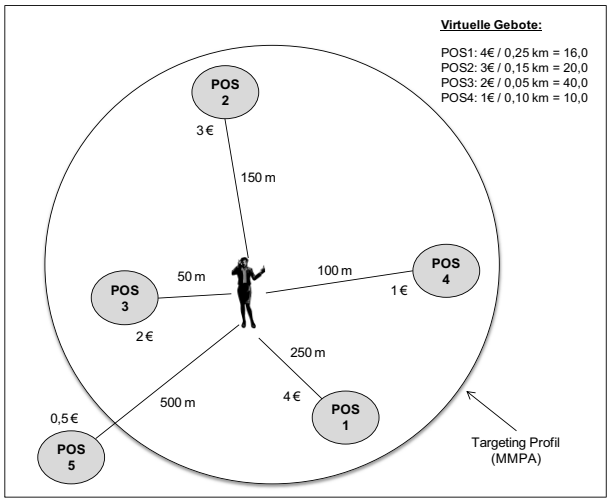


Abb. 3: Beispiel einer multi-attributive Auktion für mobile Werbeträgerleistung auf Basis der POS Distanz

In dem dargestellten Fall gewinnt POS3 die Auktion. Zwar hat dieser nur das dritthöchste Gebot abgegeben, aber er kann dieses Defizit durch eine kurze Distanz zum mobilen Nutzer gegenüber seinen Konkurrenten ausgleichen.

6. EMPIRISCHE EVALUATION

Um zu dokumentieren, dass Kontextinformationen insbesondere ökonomische Auswirkungen auf die traditionelle Vermarktung von mobiler Werbeträgerleistung haben und zu dessen Adressierung multi-attributive Auktionen einen Ansatz darstellen, der weiterverfolgt werden sollte, wurden Interviews mit Experten aus dem entsprechenden Praxisumfeld durchführt. Damit wurden die wichtigsten Eckpunkte der Designempfehlungen evaluiert und werden nun als zusammengefasste Ergebnisse im Folgenden vorgestellt.

6.1 Aufgestellte Hypothesen und Ergebnisse

Zur Evaluierung der wichtigsten Designempfehlungen wurden vier Hypothesen aufgestellt. Als Erstes sollte geprüft werden, ob MMPAs in Zukunft planen, Kontextinformationen in mobile Werbekampagnen zu integrieren (Hypothese *H1*).

Hypothese H1: *Mobile Marketinganbieter planen Kontextinformationen in mobilen Werbekampagnen zu integrieren.*

Als Ergebnis der Befragung konnte Hypothese *H1* aufgrund der Aussagen der Experten bestätigt werden. Kontextinformationen in Form von Aufenthaltsdaten über mobile Nutzer werden in Ansätzen bereits für mobile Werbekampagnen genutzt. Die verstärkte Integration, auch weiterer Arten von Kontextinformationen, ist durch MMPAs zukünftig fest geplant.

Darauf aufbauend adressierte Hypothese *H2* die Annahme das Kontextinformationen Auswirkungen auf die traditionelle Vermarktung mobiler Werbeträgerleistung durch MMPAs haben.

Hypothese H2: *Kontextinformationen haben Auswirkungen auf die traditionelle Vermarktung mobiler Werbeträgerleistung. Diese Auswirkungen werden alle relevanten Teile einer MMP (Geschäftsmodell, IKT-Infrastruktur und Allokations-mechanismus) betreffen.*

Auch diese Annahme wurde von den befragten Experten befürwortet. Als zentrale Begründung dafür wurden die besonderen Eigenheiten von Kontextinformationen genannt. Zum einen müssen Kontextinformationen im Gegensatz zu traditionellen mobilen Werbekampagnen aktiv von mobilen Nutzern preisgegeben werden. Deshalb muss auch für Kontextinformationen ein gemeinsames, semantisches Verständnis bei allen Marktteilnehmern (inkl. mobilen Nutzern) vorhanden sein. Im Weiteren wurde von den Experten festgestellt, dass die Existenz von Kontextinformationen bei mobilen Werbekampagnen zu einer individuellen Wertschätzung und damit Zahlungsbereitschaft für mobile Werbeträgerleistung durch Werbetreibende führen wird bzw. kann.

Im Anschluss wurde überprüft, ob die Designempfehlungen geeignet sind, die Auswirkungen von Kontextinformationen angemessen aus Sicht der Stakeholder (MMPAs, Werbetreibende und mobile Nutzer) zu adressieren.

Hypothese H3: *Die auftretenden Auswirkungen von Kontextinformationen in Bezug auf die Vermarktung kontextsensitiver, mobiler Werbeträgerleistung können durch die entwickelten Designempfehlungen zur Konstruktion von MMPs adressiert werden.*

Die durch Hypothese *H3* vertretene Annahme wurde von den Experten unterstützt. Dabei bestätigten diese den Bedarf nach einem dynamisch-interaktiven Preismechanismus um es Werbetreibenden zu ermöglichen ihre individuelle Zahlungsbereitschaft für kontextsensitive, mobile Werbeträgerleistung auszudrücken. Dies schafft die Möglichkeit für MMPAs ihre Einnahmen gegenüber einer statischen Bepreisung steigern. Allerdings sind, laut Experten, grundsätzlich nicht alle Werbetreibende bereit, sich auf diese Art der Bepreisung einzulassen. Im Weiteren wurde bestätigt, dass es für MMPAs erforderlich ist, ein gewisses Maß an Relevanz von mobilen Werbekampagnen für Nutzer sicherzustellen. Allerdings herrschte dabei Uneinigkeit bei den Experten über die Wahl der passenden Methode.

Abschließend wurden die Designempfehlungen für MMPs hinsichtlich ihrer technischen und rechtlichen Implementierbarkeit überprüft.

Hypothese H4: *Die vorgeschlagenen Designempfehlungen zur Konstruktion von MMPs sind technisch und rechtlich umsetzbar.*

Für Hypothese *H4* sahen die Experten in Bezug auf die technischen Umsetzbarkeit keine unüberwindbaren Hürden – lediglich kleinere oder größere Anpassungen vorhandener IdM-Systeme wurden vorausgesagt. Auch rechtliche Bedenken konnten ausgeschlossen werden – sofern sich MMPAs an die jeweiligen nationalen Datenschutzvorgaben und -richtlinien halten.

6.2 Schlussfolgerung

Aus den Ergebnissen der Evaluation lässt sich schlussfolgern, dass kontextsensitive, mobile Werbung in Zukunft an Relevanz gewinnen wird. Dabei werden durch die Integration von Kontextinformationen Auswirkungen auf die traditionelle Vermarktung mobiler Werbeträgerleistung auftreten. Diese können jedoch mit den entwickelten Designempfehlungen für MPPs grundsätzlich adressiert werden. Im Weiteren stehen der technischen und rechtlichen Umsetzung dieser Empfehlungen aus Sicht der Experten zunächst keine unüberwindbaren Hürden im Wege.

Jedoch sahen die Experten neben den Vorteilen von Kontextinformationen und deren Nutzung durch MMPAs auch insbesondere drei zentrale Hürden die für den Erfolg von kontextsensitiven, mobilen Werbekampagnen im Sinne dieser Arbeit, die beachtet werden müssen:

- **Dezentrale Speicherung von Kontextinformationen und Datenschutz:** Kontextinformationen über mobile Nutzer liegen in der Praxis nicht bei nur einem Identitätsanbieter oder nur ausschließlich beim mobilen Nutzer. Dadurch wird die Erstellung einer NSB für mobile Werbekampagnen in der Praxis erheblich schwieriger als in der Arbeit angenommen. Sollten sich hingegen alle Kontextinformationen über einen mobilen Nutzer zentral erfassen lassen, äußerten einige Experten wiederum Datenschutzbedenken. Folglich wird in diesem Zusammenhang nach einem Mittelweg gesucht werden müssen.
- **Einheitliche Beschreibung von Kontextinformationen:** Die Experten sahen die einheitliche Beschreibung von Kontextinformationen bzw. NSBs als eine wichtige Voraussetzung für ein gemeinsames Verständnis zwischen den Stakeholdern und für den Erfolg der entsprechenden Werbekampagnen. Allerdings sehen die Experten auch eine Hürde für eine MMP-übergreifende bzw. branchenweite Vereinheitlichung von NSBs – sofern sie nicht von Branchengrößen wie beispielsweise Google getrieben würde.
- **Individuelle Bepreisung von kontextsensitiver, mobiler Werbeträgerleistung:** Während sich die Experten weitgehend einig waren, dass eine statische Bepreisung für kontextsensitive, mobile Werbeträgerleistung nicht länger adäquat ist, so waren sich einige Personen uneinig ob die Bepreisung tatsächlich für individuelle Nutzer und deren NSBs stattfinden kann. Aufgrund des hohen Aufwands steht deshalb die Frage im Raum ob nicht eher nur kleine Nutzergruppen bepreist werden sollten und wenn ja, wie sich diese zusammensetzen.

7. ZUSAMMENFASSUNG UND DISKUSSION

Der große Erfolg mobiler Datenkommunikation in den letzten Jahren hat zu einem zunehmenden Wettbewerb um die Aufmerksamkeit mobiler Nutzer im Kontext von mobilen Werbekampagnen geführt. Folglich müssen MMPAs nach neuen Wegen suchen um sich von ihren Wettbewerbern zu differenzieren. Im Zusammenhang dieser Arbeit ermöglicht dies die Bereitstellung von Kontextinformationen für mobile Werbekampagnen durch MMPs, die es Werbetreibenden erlaubt mobilen Nutzern in ihrer aktuellen Nutzungssituation ein werbliches Angebot zu übermitteln.

Der Einsatz von Kontextinformationen für mobile Werbekampagnen generiert jedoch Auswirkungen auf die traditionelle Vermarktung mobiler Werbeträgerleistung. Diesen Auswirkungen müssen MMPAs begegnen, um letztlich von den Vorteilen von Kontextinformationen zu profitieren.

Die wichtigsten Ergebnisse der vorliegenden Arbeit in diesem Zusammenhang sind wie folgt:

- **Identifikation und Analyse der Auswirkungen von Kontextinformationen:** Die Identifikation und Analyse der Einflüsse von Kontextinformationen auf die traditionelle Vermarktung mobiler Werbeträgerleistung hat gezeigt, dass Kontextinformationen mehr als nur zusätzliche Informationen über mobile Nutzer darstellen. Ihre besonderen Eigenheiten verändern die Charakteristika mobiler Werbeträgerleistung als ökonomisches Gut. Somit liefert diese Analyse einen Beitrag zum Problemverständnis in der Mobile Marketing Domäne.
- **Entwicklung von Empfehlung zur Konstruktion MPPs:** Die entwickelten und evaluierten Designempfehlungen liefern Richtlinien für MMPAs zur Konstruktion von MPPs auf denen kontextsensitive, mobile Werbeträgerleistung gehandelt werden kann. Diese erlauben es die Auswirkungen von Kontextinformationen auf die traditionelle Vermarktung mobiler Werbeträgerleistung zu adressieren, damit MMPAs und weitere Stakeholder an den Vorteilen von Kontextinformationen partizipieren können.

Neben dem Beitrag der vorliegenden Arbeit lassen sich die folgenden Ansätze für zukünftige Forschung diskutieren:

- **Analyse der Vermarktung kontextsensitiver, mobiler Werbeträgerleistung in einem Markt, der Wettbewerb zwischen MMPAs aufweist:** Die vorliegenden Designempfehlungen wurden ausgehend von einer Monopolsituation entwickelt.
- **Entwicklung von Ansätzen zur Aggregation von verteilten, partiellen Identitäten mobiler Nutzer zu einer kontextsensitiven NSB:** In der vorliegenden Arbeit wird unterstellt, dass MMPAs Zugriff auf alle relevanten Kontextinformationen eines mobilen Nutzers haben. In der Praxis sind diese jedoch über mehrere partielle Identitäten und Identitätsanbieter eines mobilen Nutzers verteilt (FIDIS 2005).
- **Entwicklung eines Verhandlungsmechanismus für die Preisgabe von Kontextinformationen**: Mobile Nutzer verlangen typischerweise (monetäre) Anreize, um persönliche Daten von sich preiszugeben. Um zwischen Werbetreibenden und mobilen Nutzern zu einer Einigung bzgl. der preiszugebenen persönlichen Daten zu kommen, könnte ein Verhandlungsmechanismus diesen Vorgang unterstützen.

Die aufgeführten Ansätze für eine zukünftige Forschung zeigen u.a. das frühe Stadium der Arbeiten in Bezug auf die Vermarktung von kontextsensitiver, mobiler Werbeträgerleistung auf. Die durch die vorliegende Arbeit geschaffenen Erkenntnisse zu Auswirkungen von Kontextinformationen sowie die Entwicklung der Designempfehlungen bilden aber eine erste Grundlage zur Förderung der Verbreitung von kontextsensitiver, mobiler Werbung durch MMPAs in der Praxis und haben damit das Potential weitere Forschung in diesem Bereich zu stimulieren.

8. LITERATUR

[1] Aalto L., Göthlin N., Korhonen J., and Ojala T. 2004 "Bluetooth and WAP Push Based Location-Aware mobile advertising system", *Proceedings of the Second International Conference on Mobile Systems, Applications and Services*, Boston, MA, USA, 49-58.

[2] Albers A., Figge S., and Radmacher M. 2005 LOC3 - Architecture Proposal for Efficient Subscriber Localisation in Mobile Commerce Infrastructures, *Proceedings of the 2nd*

IEEE International Workshop on Mobile Commerce and Services (WMCS'05), Munich, Germany.

[3] Albers A. 2007 Allokation und Bepreisung mobiler Kundenkontakte im Mobile Marketing, *Proceedings of 7th Internationale Tagung Wirtschaftsinformatik*, Karlsruhe, Germany.

[4] Albers A. 2008 An Electronic Market Framework for context-sensitive Mobile Consumer Profiles in the Marketing Domain, *Proceedings of the 13th Americas Conference on Information Systems*, Keystone, CO, USA.

[5] Bichler M. 2000 *The Future of eMarkets - Multi-Dimensional Market Mechanisms*, Cambridge University Press, Cambridge, UK, 2000.

[6] Birkel, M. 2008 Mobile Lifestyle: Das Herz des Kundengewinnen, in: Thorsten Schwarz (Eds.): *Leitfaden Online Marketing,* marketing-börse, Waghäusel, Germany, pp. 481-485.

[7] Bohte, S.; Gerding, E. and Poutré, H. 2004 Market-Based Recommendation: Agents that Compete for Consumer Attention, *Proceeedings of the ACM Transactions on Internet Technology* (4:4), pp. 420-448.

[8] Bulander R., Schiefer G., and Decker M. 2005, Anonymity by Design - Eine Architektur zur Gewährleistung von Kundenschutz im mobilen Marketing, *Proceedings of the Conference on Mobile Business – Processes, Platforms, Payments. Proceedings zur 5. Konferenz Mobile Commerce Technologien und Anwendungen (MCTA 2005)*, Bamberg, Germany, 87-100.

[9] Bruner G., Attitude toward Location-Based 2007 *Advertising, Journal of Interactive Advertising* (7:2), 154-177.

[10] Daconta, M.; Obrst, L. and Smith, K. 2003 The semantic Web: A guide to the future of XML, Web Services, and Knowledge Management, Wiley, Indianapolis, USA.

[11] Dey, A. K. and Abowd G. D. 1999 Towards a Better Understanding of Context and Context-Awareness, *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, Karlsruhe, Germany, 304-307.

[12] Ehrlich M. 2008 *Die Entwicklung der Onlinewerbung*, Leitfaden Online Marketing, Waghäusel, Germany, 265-270.

[13] E-LBA Project 2002 *European Location Based Advertising*, http://www.e-lba.com, 2002, accessed 30-03-2010.

[14] Feldmann V. 2005 *Leveraging mobile media - Cross-media strategy and innovation policy for mobile media communication*, Physica-Verlag, Heidelberg, Germany.

[15] FIDIS 2005 Interoperability of Identities and Identity Management Systems: Structured account of approaches on interoperability, *http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp4-del4.1.account_interoperability.pdf,* accessed 2010-02-04.

[16] Figge S. and Theysohn S. 2006 Quantifizierung IKS-basierter Marketleistungen - Analyse eines werbefinanzierten Geschäftsmodells für den M-Commerce, *Wirtschaftsinformatik* (48:2) 96-106.

[17] Figge, S. 2007 *Innovatives Mobile Marketing - kontextabhängige Kundenansprache mit Hilfe mobiler Portale*, Kovac, Hamburg, Germany.

[18] Friberg Å. 2007, An Empirical Evaluation of Online Privacy Concerns with a Special Focus on the Importance of Information Transparency and Personality Traits, Luleå University of Technology, LULEÅ, Sweden.

[19] Goldhaber, M. 1997 The Attention Economy and the Net, http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/519/440, accessed 2009-12-13.

[20] Gopal R. and Tripathi A., Advertising via wireless networks, *Journal of International Communications* (4:1), 2006, 1-15.

[21] Haase, K.; Salewski, F. und Skiera, B. 1998 Preisdifferenzierung bei Dienstleistungen am Beispiel von 'Call-by-Call'-Tarifen, *Zeitschrift für Betriebswirtschaft (68:1)*, 1053-1072.

[22] Huff, D. 1964 Defining and Estimating a Trading Area, *Journal of Marketing* (28:1), 34-38.

[23] Hristova N., and O'Hare G. 2004 Ad-me: Wireless Advertising Adapted to the User Location, Device and Emotions, *Proceedings of the 37th Hawaii International Conference on System Sciences*, Hawaii, USA.

[24] Klafft, M. and Spiekermann, S. 2006 Lohnt sich kontextsensitives elektronisches Couponing? Eine Fallstudie aus der Fast-Food-Industrie, *Absatzwirtschaft* (9:1), 66-75.

[25] Kishore, A. 2008 *Mobile Advertising: The Service Provider Revenue Opportunty*, Heavy Reading, New York, USA.

[26] Kobsa, A. 2007 Privacy-Enhanced Web Personalization, in: W. Nejdl (Eds.): *The Adaptive Web: Methods and Strategies of Web Personalization,* Springer Verlag, Berlin, Heidelberg, Germany; New York, USA.

[27] Koppius, O. 1999 Dimensions of Intangible Goods, *Proceedings of the 32nd Hawaii Conference on System Sciences,* Hawaii, USA.

[28] Kurkovsky S. and Harihar K. 2006 Using ubiquitous computing in interactive mobile marketing, *Personal Ubiquitous Computing* (10:1), 227–240.

[29] Leppäniemi M., Sinisalo J., Karjaluoto I., and Karjaluoto H. 2006 Mobile Marketing Research 2000 -2005 - Emergence, Current Status, Future Directions, *International Journal of Mobile Marketing* (1:1) 2-12.

[30] Lispman, A. 2007 A Number of U.S. Computers Accessing the Internet Via Mobile Broadband Soars 154 Percent in 2007, http://www.comscore.com/press/release.asp?press=2099, 2008, accessed 24-03-2010.

[31] Mayring, P. 2008 *Qualitative Inhaltsanalyse*, Dr. nach Typoskr, Beltz, Weinheim, Germany.

[32] Nieschlag, R.; Dichtl E., and Hörschgen H. 2001, *Marketing*, 19., überarb. und erg. Aufl, Duncker und Humblot, Berlin.

[33] Ojala, T.; Korhonen, J.; Aittola, M.; Ollila, M.; Koivumäki, T.; Tähtinen J.; Karjaluoto, and Karjaluoto, H. 2003 SmartRotuaari – Context-aware Mobile Multimedia Services,Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia, Norrköping, Sweden.

[34] Radmacher, M. 2007 Elicitation of Customer Profile Elements by Transparent communicated Extensions, *Proceedings of the Recommender Systems 2007,* Minneapolis, Minnesota, USA

[35] Ranganathan A. and Campbell, R. H. 2002 Advertising in a Pervasive Computing Environment,: September 28, 2002, Atlanta, Georgia, USA ; *in conjunction with ACM Sigmobile Eighth International Conference on Mobile Computing and Networking*, ACM International Workshop on Mobile Commerce, New York, NY, ACM Press.

[36] Robra-Bussantz, S. 2005 Ubiquitous Customer Interface - Situationsbasierte Informationsversorgung des mobilen Kunden, *Wirtschaftsinformatik* (47:1), 25-35.

[37] Sharma, C.; Herzog, J. and Melfi, V. 2008 *Mobile advertising*, Wiley, Hoboken, NJ, USA.

[38] Schiller, J. H. 2003 *Mobile communications*, 2nd Edition, Addison-Wesley, London, UK.

[39] Schmid, B. 1999 Elektronische Märkte – Merkmale, Organisation und Potentiale, in: A. Hermanns and M. Sauter (Eds.): *Management-Handbuch Electronic Commerce,* Vahlen, Munich, Germany.

[40] Steimel, B.; Paulke, S. and Klemann, J. 2008 *Praxisleitfaden Mobile Marketing*, mind Business Consultants, Meerbusch, Germany.

[41] Strategy Analytics 2008 *Mobile Advertising: Forecast Methodology*, Strategy Analytics, USA.

[42] Sultan F. and Rohm, A. 2005 The Coming Era of "Brand in the Hand" Marketing, *MIT Sloan Management Review* (41:1), 83-90.

[43] Tang H., Shaoyi L., Shaoyi L., JingJun Xu D. 2007 An Open Platform for Context-aware Short Message Service, *Proceedings of the America's Conference of Information Systems,* Keystone, CO, USA, 2007.

[44] Thawani, A.; Gopalan, S. and Ramamritham, K. 2007 Context-aware Timely Information Delivery in Mobile Environments, *The Computer Journal (50:4)*, 460-472.

[45] Tripathi, A. 2003 *Decision Models for Wireless Advertising*, University of Connecticut, Connecticut, USA.

[46] Varshney U. and Vetter, R. 2003 Mobile Commerce: Framework, Applications and Networking Support, Mobile Networks and Applications (7:1), 185-198.

Towards Cycle-Oriented Requirements Engineering

Marina Berkovich
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr.3
85748 Garching
berkovic@in.tum.de

Christian Mauro
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr.3
85748 Garching
mauro@in.tum.de

Jan Marco Leimeister
Fachgebiet Wirtschaftsinformatik
Universität Kassel
Nora-Platiel-Str. 4
34127 Kassel
leimeister@uni-kassel.de

Felix Weyde
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr.3
85748 Garching
felix.weyde@in.tum.de

Helmut Krcmar
Lehrstuhl für Wirtschaftsinformatik
Technische Universität München
Boltzmannstr.3
85748 Garching
krcmar@in.tum.de

ABSTRACT

Product Service Systems (PSS) – a combination of software, hardware and service elements – are an emerging trend on the market. Their development encounters different difficulties, amongst them the existence of dynamic cyclic interdependencies. These cycles make the challenging requirements engineering (RE) for PSS even more complicated. Up to now there is no integrated RE approach for PSS. An important step in developing such an approach is to understand the effects of cycles on RE. In this paper the cycles and their causes are analyzed by an industrial case study, a literature review, and by expert interviews. The contribution of this paper is twofold. First, a detailed explanation of cycles' causes emerging in the development and RE of PSS is given and the effects of the resulting cycles on RE are elaborated. Second, a list of requirements to RE for PSS is derived, which describes how the cycles can be managed adequately.

Keywords

requirements engineering, cyclic interdependencies, cycle, complex solutions, product service systems.

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

1. INTRODUCTION

Differentiation from competitors is one of the most important aspects for the success of a company. Customers want a solution that solves their problems and provides advantages for them [2]. Thus, differentiation by physical technological products (tangible products) or services alone is not sufficient to assure the fulfillment of customers' and market's needs, nor is it adequate for long-term success [13, 37]. Therefore, companies offer "a combination of goods and services that are integrated and customized to meet the idiosyncratic requirements of a customer" [39]. That means companies offer complex solutions, referred to as PSS – product service systems – or hybrid products, consisting of integrated bundles of products and services [23, 38]. The product can be either hardware or software, or a combination of both. Especially software plays a major role in the realization of PSS, because it is essential for innovative multi-functional systems [25]. Accordingly, software has often the task to realize the integration of hardware and services. Integration of different domains as product, software and service engineering in the development process of PSS, different lifecycles of PSS' components, a high degree of technological integration, and a high level of customer integration characterize product service systems and influence their development, making it rather complex [10].

In the development of PSS requirements engineering *(RE)* plays a vital role. RE is one of the most important and also critical activities within a development process and is defined as "gathering, documenting and managing requirements" [3]. A successful RE is a precondition for the success of the development process [11]. Products need to be adapted to the customer requirements, while they are becoming more complex and innovative [28]. Changes of requirements in late phases of the development are up to 100 times more costly than during the initial development steps, during the RE phases [9]. Because of the challenges of PSS described above the RE for PSS is especially intricate and complex [5, 7]. Thus, the overall research question of our research is *how RE for PSS should look like*. Since the research alongside this research question is very extensive, this paper will focus on one special topic. The topic is

the influence of cyclic interdependencies on the requirements engineering of PSS.

An important aspect in the development of products and services is cyclic interdependencies (further called *cycles*). They lead to iterations, bad coordination between the areas involved in the development process, permanent adaption and changes of PSS' components and of the activities in the development process [20, 21]. Thus, the cycles cause increasing development time and costs. The costs are increasing because the additional iterations in the development cause additional effort and resource spending. For example, it is even possible that the product, which has been already built, has to be modified several times if the customers' requirements have changed. These additional iterations cause time delays as well, which may have a negative impact on the customer satisfaction. Furthermore, supplemental consultation with the customer may be necessary and can cause additional efforts on customer-side.

Due to the complexity and interdisciplinarity of PSS, the cycles are especially challenging for PSS. One of the challenges is different lifecycles of the PSS' components. For example, the hardware may be replaced by a new one within a timeframe in which the provided services keep the same. Notwithstanding, the modification of the hardware can lead to changes of the services which have to be considered. Thus, such differences in lifecycles cause the need for coordination and iterative adjustments of the PSS between different domains.

Many cycles arising during the development of PSS have their roots in requirements engineering. Thus, the first step is to understand the cycles during the RE of PSS profoundly, which is the focus of this paper. For this purpose, the cycles' causes are identified and then examined. For each cycle's cause it is elaborated what cycle it is inducing in the RE. The cycles are inevitable in the RE, but they should be minimized and controlled as far as possible. Based on the insights into cycles' causes and their effects on the RE, we derived a list of recommendations how the cycles in the RE for PSS can be managed adequately.

The research question for this paper is *how RE for PSS should look like in order to be able to deal with cycles.* The contribution of this paper is twofold: (1) a list of cycles and their causes emerging in the RE of PSS (2) a list of requirements for the RE for PSS in order to make it cycle-oriented.

To analyze the cycles, an industrial case study is presented. In this case study, the development process of a PSS in industry is analyzed. Furthermore, the analysis relies on an extensive literature review with the goal to find cycles that have already been described in research literature. We also conducted qualitative expert-interviews to capture current practices in RE in companies producing PSS, with special attention on emerging cycles and their management.

2. RESEARCH METHODOLOGY

Our research was conducted in two phases. The first phase can be seen as a pre-study for the second phase. Hence, the first phase is handled in brief in this paper, and more room is given to the second phase. In figure 1 an overview of the phases and the single steps within the phases is given.

2.1 Phase 1

In phase 1 the development process of PSS must be understood and the role of RE must be clarified. In **step 1a**, a literature review about PSS was conducted in order to understand their characteristics and their development process. A special focus was laid on the RE of PSS and its integration into the development process. The selection of the literature relied on high-quality journals and conferences of information systems ranked as A-Journals and A-Conferences by the WI-list [1]. All in all 88 publications were included in the review.

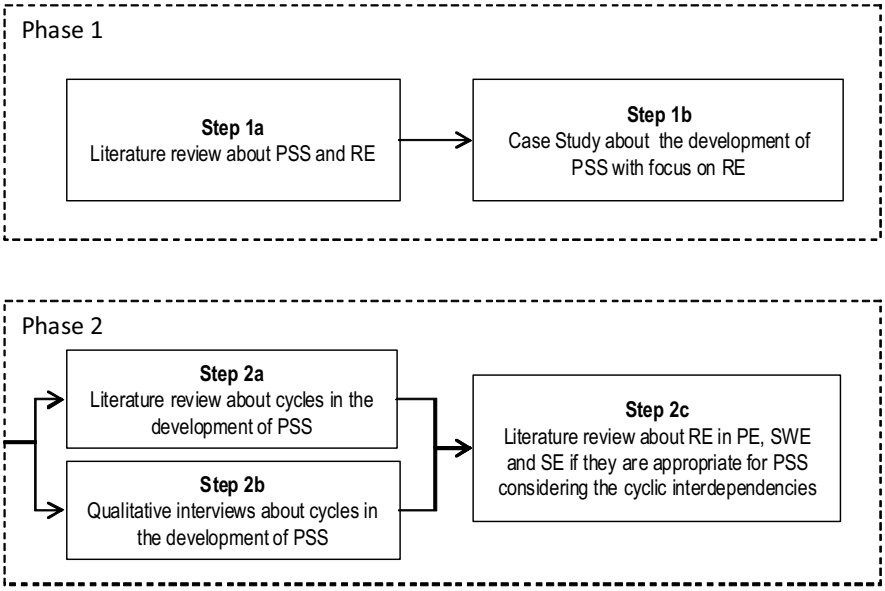


Figure 1: Research Design

Then, in **step 1b**, a case study about the development of PSS was conducted in order to understand their development in practice. The design of this phase is aligned according to Yin [42]. In this case study a large company of the chemical industry producing PSS was analyzed between July and November 2006 by interviewing employees of three departments. The result of phase 1 is a list of characteristics of PSS and of their development process. The insight that cyclical interdependencies are a challenge in the development and RE was made here.

2.2 Phase 2

Phase 2 relies on literature reviews and a series of expert interviews. They had a broader scope and were analyzed in multiple steps. The results of the first analysis did not take the cycles into account and have already been published. These precedent publications are summarized here in short:

- Within an empirical exploration of requirements engineering for hybrid products 15 expert interviews in industry were conducted and analyzed regarding the handling of RE for PSS in practice [7].
- A literature review had the goal to examine the current practices of RE in the three domains involved in the development of PSS. It summarized these practices and evaluated them based on defined criteria [5].
- A similar literature review as described in [5] was also conducted in the product development and published in Berkovich et al. [8]. The criteria for the evaluation of the practices were further elaborated and the techniques are described in more detail.

The second phase has the goal to identify the cycles' causes, to further analyze the role of cycles in the development process of the RE and to analyze especially the role of the RE in handling the cycles.

In **step 2a** a further literature review on cycles arising during the development of PSS (including also the RE) was conducted. The goal of the literature review was to identify cyclic interdependencies and their causes in the development, to understand how they were handled and which implications they had on the development and especially on RE. This literature review included (in addition to the literature on PSS) the top-10 textbooks on product, service and software engineering and design according to the sales ranking of amazon.com (accessed on 07.02.09), as well as publications according to the WI-list [1].

The results of **step 2b** present cycles' causes based on the qualitative expert interviews. These interviews were also used to categorize the cycles using expert opinion. A series of semi-standardized expert interviews was conducted, with the goal to capture current practices in RE in companies producing PSS, with a special attention on the handling of cycles during the development. As described earlier, an analysis of a part of these interviews was already published in Berkovich et al. [7]. The study design is summarized here in brief. In the course of the study 15 experts in companies involved in the development of PSS were interviewed. Some companies were selected because there were contacts to them; others were selected according to the list of Luenendonk [26]. The interviews were carried out between June and October 2008 and were guided according to the process steps of RE: requirements elicitation, analyses and negotiation, documentation and validation, change management and traceability [30, 41].

The result of **steps 2a and 2b** is a list of cycles' causes arising during the development process of PSS and having an impact on RE. During the interviews the cycles of the literature review were presented to the experts. Together with the newly found cycles the experts categorized the cycles according to their causes. The criterion for choosing the categories was that all cycles of one category should have the same impact on RE. The categories developed in each interview, were consolidated by the researchers.

In the **step 2c** we analyzed the existing approaches for RE in the product engineering, software engineering and service engineering about their handling of cycles with the goal to understand if they were appropriate for PSS and to state the requirements on a cycle-oriented RE for PSS. For our literature review we included the domains that are involved in the development of PSS.

As described earlier, the literature review presented in this paper relies on the same data as the reviews published in Berkovich et al. [5] and Berkovich et al. [8]. The research design is summarized here in short. The selection of textbooks relied partially on sales ranks on amazon.com; partly specialized books on RE known to us were also included. In product and service engineering RE is not seen as a special discipline. Thus, for these domains general textbooks on development methods were chosen. Additionally, conference papers and journal publications were selected according to the WI list [1] and a list of high-quality journals available at our institute. The result of this step is an analysis whether the cycles identified before are handled by the RE approaches adequately, if the methods for the handling of cycles used by these approaches are appropriate for PSS and which requirements must be fulfilled by a new RE approach for PSS to handle the cycles.

3. STEP 1A – LITERATURE REVIEW ABOUT PSS AND THE REQUIREMENTS ENGINEERING

3.1 Characteristics of PSS

In contrast to classical PSS consisting of a physical product and a service part, we also regard a combination of software and service or of hardware, software and service as a PSS [4, 10, 24]. The increasing importance of PSS has been shown by several studies. For example Sturm and Bading [35] asked 1641 medium-sized companies about the importance of integrated bundles of products and services and concluded that 59.8% assessed them as very important for the success on the market. A main driver in the development towards PSS is the awareness that a customer does not have an interest in a product or service per se, but he wants his problems to be solved [23].

PSS consist of both material and immaterial components developed by different domains, leading to interdisciplinary work and to the need for technological integration [7, 23]. Interdisciplinary work means that different domains have to

cooperate in the development process in order to achieve a comprehensive understanding of the problem to be solved. The technological integration means that the different parts of PSS have to be integrated technically and organizationally, both between each other and into the value-creation process of the customer [10]. PSS are regarded highly individualized [29], and this individualization is especially important in fulfilling customer wishes.

3.2 Requirements Engineering in the context of PSS

RE is generally understood as a cooperative, incremental and iterative process with the goal of identifying all requirements that are relevant for the system under development, creating a common understanding of the problem to be solved between all stakeholders, and then documenting these requirements in a form that can be used in later development steps [31].

Both in practice and in literature the RE approaches for PSS manage the requirements to a PSS as a whole and in integrated way just insufficiently. Also the domains involved in the development of PSS meet a challenge to gain a comprehensive understanding of the solution to be developed [5, 7]. However, it has been realized that an integrated RE approach would be especially important for PSS, because of their complexity and interdisciplinarity [33]. Berkovich et al. [6] identified the most important activities of RE for PSS (see Figure 2) based on the role of RE in the lifecycle of PSS and on the definition of RE in product engineering, software engineering and service engineering (e.g. [3, 12, 17, 40]): requirements elicitation, analysis and documentation, negotiation, validation, change management and tracing.

The requirements elicitation has the task of gathering all goals, wishes and requirements from the stakeholders. During the requirements analysis the initially elicited requirements are refined and concretized. In this step the initial requirements are "translated" into the language of the developers so that they can be used by the single domains. The requirements analysis includes also the search for conflicting requirements both between initial and concretized requirements and between concretized requirements belonging to the different domains. The purpose of requirements negotiation is to resolve conflicts regarding requirements between stakeholders. The requirements documentation has the task of documenting all requirements according to documentation standards [41]. The requirements validation has the task of assuring that the requirements actually represent the vision of the stakeholders. The last two phases – change management and tracing – are done during the development process. The traceability refers to the ability to follow the life-cycle of a requirement and to assess the impact of changes of requirements.

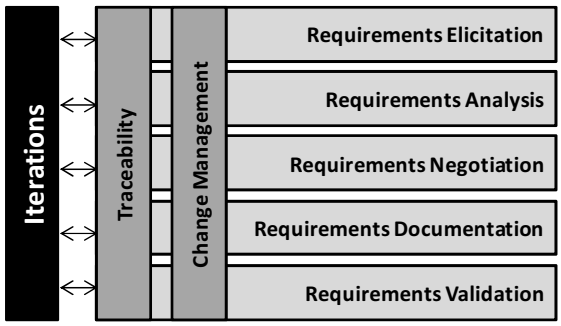


Figure 2: RE Activities and Iterations for PSS according to Berkovich et al. [6]

4. STEP 1B – CYCLES FOUND IN AN INDUSTRIAL CASE STUDY

In order to understand the development of PSS in practice we analyzed a company of the chemical industry. Further, we describe the parts of the case study that stay in relation to the cycles. The PSS of this company consisted of mainly two parts, according to which the company was organized: (1) Engineering division: construction of gas processing plants (2) Gas division: supplying gas to customers.

In greater detail, the first division develops both customer-individual facilities and standardized facilities based on serial production. The reason for offering standardized facilities is the cost-advantage. These facilities are cheaper, but may not fully meet customer requirements. The second division offers customers a supply of gas. Depending on the circumstances and requirements of the customer, a solution is provided that may consist of pipelines, transport of gas by trucks, or construction of plants at the customer's side. This second division orders the required engineering work from the first one. Additionally, services for already existing facilities are offered, like maintenance, installation, training and initial operation. The customers of the company simply acquire the service "gas supply" and are unconcerned about how the gas is brought to them. The contractor then constructs the necessary facilities and operates them.

In the development of the gas facilities, a concurrent development of hardware and software takes place. During this development process cycles can be observed. The engineering division is exclusively concerned with chemical processes and their implementation by hardware. All software development for control software is outsourced. The first step when starting a software development is to define the concretized requirements which can be realized by the domains. During the contracting phase a solution is iteratively developed. Due to the cost-pressure, the processes and also the software are adjusted during construction and operation of the facility (see Figure 3).

In this way, experiences can be transferred to the development. However, during the development some development steps have

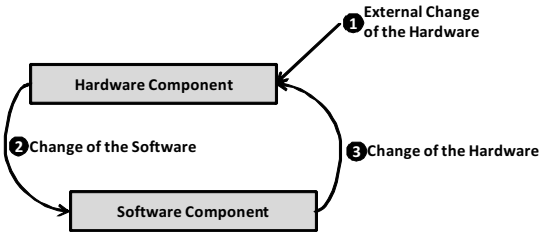


Figure 3: Cycle in the in the development of gas facilities caused by a concurrent development of hardware and software

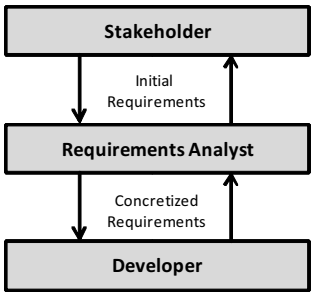


Figure 4: Cycle in the development of gas facilities caused by misunderstandings between the RE and the development

to be repeated because of changing requirements. Hence, the RE plays a major role in handling the cycles. For a most effective handling of cycles, the RE should be able to predict cycles and give the possibility to avoid them.

During the development also another type of cycle was observed. When the development is starting, the initial requirements are captured in a specification document. Then these requirements are concretized and a more precise specification is created. This detailed specification is given to the developers as the basis of their work. During the requirements validation of the concretized requirements as well as during the realization of requirements by the developers, variations and inconsistencies were found. They lead to a rework of the requirements specification, which created multiple cycles between initial stakeholders, requirements analysts, and developers (see Figure 4).

5. STEP 2A AND 2B – CYCLES' CAUSES INFLUENCING REQUIREMENTS ENGINEERING

In this section the results of the expert interviews and literature reviews are described. In our description of the cycles, we describe causes for cycles and group them in categories of similar causes (see Table 1). For each group, it is described which RE activities are iteratively repeated and form a cycle consequently, if the cycle's cause emerges.

Cycles in Literature –The cyclic dependencies within and outside of the development process are a challenge for companies to overcome. According to Langer et al. [21], a cycle is characterized by "a repeated succession of similar occurrences and of results initiated by them, such as sub-processes, artifacts, developments, etc., and the succession of different occurrences within one sequence, e.g., the innovation process". The development of PSS marked as very complex is affected by dynamic cyclic behaviors [21].

Comprehensive List of Cycles Causes – This section categorizes all causes of cycles based on the literature review and interviews into seven major categories. Each category describes the causes for the cycle, gives an example, and explains the implications for the requirements engineering.

Category 1*: Availability and maturity of technologies:* This category contains the causes for cycles that have to do with the availability of new technologies. They were identified in the literature review, for example, Dodgson et al. [14]. If a new technology emerges on the market, it is possible to use it to realize components such as hardware, software or service of the PSS.

An example of such a cycle is the introduction of VoIP, which offers the possibility of restructuring the communication infrastructure (hardware and software) of solutions under development (http://www.msnbc.msn.com/id/6354872/ accessed on 27.04.09). If a component or a part of it is realized by new technology, the concretized requirements can change. These changes can have an impact on the customer requirements of the overall product and on other concretized requirements, and thus have to be analyzed. It must also be checked whether the new technology realizes all requirements that were realized by the old one. Further, new conflicts due to changed concretized requirements may emerge and need to be considered.

If, for example, a new technology is available and should be

Table 1. Categories of Cycles' Causes

Nr.	Category of Cycles' Cause	Causes of Cycles
1	Availability and maturity of technologies	• new software technology / product technology / service technology emerges on the markets
2	Competitive trends	• product is no longer required by the customer • rivalry product is more attractive for the customer • service is no longer required • new market trend is observed
3	Different lifecycles of hardware / software	• hardware / software is out-dated
4	Change of customer-requirements	• change of customer-requirements
5	Financial cycles on the capital market	• product has to be released immediately to the customer • price of the product has to be reduced
6	Development cycles	• hardware / software / service cannot be realized • integration of domains' results failed due to derivations • internal stakeholders change the requirements
7	Legislative cycles	• laws / standards change

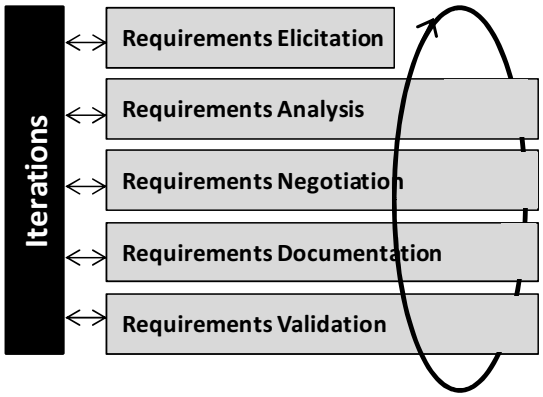


Figure 5: Cycle in the RE caused by "*Availability and maturity of technologies*"

used in the PSS, then the following steps are executed within the RE. Therefore, it is necessary, to check if the requirements emerging from the new technology are conform to the old requirements. First, the requirements emerging from the new technology are collected. Then, they are analyzed for conflicts with the existing requirements. If conflicts are present, they have to be solved by discussion between all stakeholders. Furthermore, the usual activities of RE are executed, such as documentation, and validation. Figure 5 shows this cycle in the RE exemplarily.

***Category 2**: Competitive trends:* The causes for this category of cycles were found in the interviews and the literature review, for example, Fuchs and Apfelthaler [15], Spath et al. [34] and Hauschildt and Salomo [16]. These cycles have to do with product management and market trends.

If a PSS or a service is no longer needed by the customer, it means that the customer-requirements have changed.

Given that the customer is interested in the overall product, the new customer-requirements have to be elicited and the RE process has to be repeated, but still reusing the old requirements as much as possible.

If new market trends are observed, the customer needs to be consulted as to whether he wants to integrate them into the PSS. Integrating them would mean adding new customer requirements and refining them to corresponding concretized requirements.

If a rivalry product seems more attractive for the customer since it has better characteristics than the considered PSS, the rivalry product has to be analyzed (for example Benchmarking techniques [40]) and new requirements need to be derived. Particularly quality requirements and attractiveness requirements can have an influence of the competitive advantage [27].

***Category 3**: Different lifecycles of hardware and software:* These causes of cycles were found in the literature review, for example, Pahl et al. [27], Lavagno and Passerone [22]. The components of PSS have different life-cycles [5], and thus if one component of a PSS (hardware, software and service) is out-of-date, it has to be replaced. Thereby, new detailed requirements

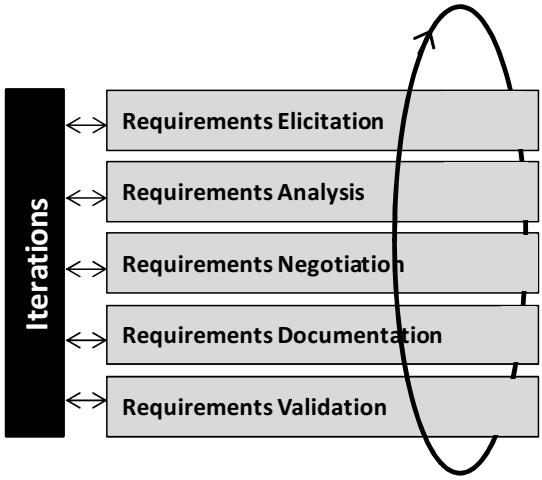


Figure 6: Cycle in the RE caused by "*Change of customer-requirements*"

can emerge within the development, which have an effect on customer requirements.

In general, if it is discovered that a hardware or software component is out-dated, all requirements affected by those components have to be determined. This can be done by using classical traceability techniques. Then, it has to be analyzed whether these requirements change when the old technology is replaced, and the potential effects on other requirements have to be analyzed. All this information is collected and given to the developers in order to give them the possibility to decide which next steps have to be taken.

An example for such cycles is ATMs, where the card-reading hardware had to be replaced due to new standards for credit cards (http://www.stern.de/wirtschaft/geld/87195.html accessed on 20.04.09). Also, the software had to be modified, which was not easily possible because the operating system was no longer supported by the supplier.

***Category 4**: Change of customer-requirements:* The customer-requirements can change during development and use of a product [19, 31]. In such cases it is important to clarify the impacts on concretized requirements, on the already developed system, and on the environment. The change of one customer-requirement can also have an impact on other customer-requirements.

This cycle can occur during development time as well as during the use of the PSS. If the customer requirements change, the elicitation of requirements has to be done, in order to elicit all changes requirements. Then, they are analyzed and checked for contradictions to existing requirements. In the requirements negotiation, the contradictions have to be resolved. Then, the new requirements are documented and validated. As we can see, all phases of the requirements engineering are repeated in this cycle. It is shown exemplarily in Figure 6.

***Category 5**: Financial cycles on the capital market:* These cycles were identified in the literature review, for example, Fuchs and Apfelthaler [15]. The cycles in this category can have the consequence that the product must be immediately released

to the customer or that the price of the product has to be reduced. In both cases it is necessary to decide which functionality of the product can be abandoned to achieve the necessary savings or the necessary delivery-time, for example the imitations of the iPod. After the release of the iPod, several competitors forced their products under development to move to an immediate market release.

In the case that the price of a product has to be reduced, the RE has to reprioritize all requirements. In order to do so, the requirements analysis has to be redone, because when prioritizing the requirements, the interdependencies between them have to be regarded. The reprioritization of the requirements and the omitting of low prioritized ones necessitate the repetition of the requirements negotiation and validation.

Category 6*: Development cycles:* These cycles were identified by the interviews. These cycles occur when it is impossible to realize a component (hardware, software, service) of the PSS.

If a component cannot be realized, it has to be decided whether the affected customer-requirement can be changed or whether the functionality can be realized by another component.

If the integration of the partial solutions experiences major problems, the implications for the customer-requirements have to be determined, and the requirements possibly have to be changed.

The change of requirements by internal stakeholders can also cause cycles. It has to be analyzed whether the customer-requirements are affected by these changes and whether coordination with the customer is necessary.

Category 7*: Legislative cycles:* These cycles emerge when a law or standard that is relevant for the PSS changes. These changes are typically known in advance and it can be planned for them. If such a change emerges, first all new requirements have to be elicited. Then, the new requirements are incorporated into the existing specification. That makes it necessary to newly execute the requirements analysis; in order to check which requirements are affected and to find and resolve newly originated conflicts. Then, the requirements negotiation and validation have to be repeated. As a last step, the effects of the changes of the requirements on the already build parts of the PSS have to be analyzed.

An example of such a cycle is the change of the laws regarding the system for managing the unemployment payments in Germany. After a major amendment of the law, the system had to be changed by 2005-01-01 (http://www.mid.de/fileadmin/documents/pdf/Anwenderberichte/Sonderdruck_eGov_0307_MID.pdf accessed on 18.04.09).

6. STEP 2C – CYCLE-ORIENTED REQUIREMENTS ENGINEERING

The cycles described in the previous section have an impact on the RE of PSS. In order to effectively handle these cycles, the RE has to support certain activities and exhibit certain characteristics. The objective of this section is to analyze which characteristics and activities are needed and to assess whether existent RE approaches are capable of implementing them.

Results of the Literature Review – RE is widely accepted in product engineering and is integrated in the engineering processes. The handling of requirements is, however, often limited to the first phases of the development [7]. The requirements are mostly documented in lists, and changes are then documented within those lists. In software engineering, RE is applied throughout the development process [31]. In contrast to product engineering, the management of requirements during the development is more advanced: changes are documented more carefully and methods for carrying out impact analyses of changes are available. Because service engineering is still a relatively young discipline, there is a general lack of systematic approaches [18]. Some process models of service engineering mention RE but without offering techniques.

Requirements for a Requirements Engineering Approach

As described in the research design, requirements to the RE of PSS – without cycles perspective – can be found in [7]. Based on the same data basis as in [7], we developed additional requirements to the RE of PSS, in order to make it cycle-oriented.

The main requirements to RE for PSS without cycle-orientation are summarized here: The first aspect is that a coordinated RE process for all components of PSS has to be established. The requirements for the different domains cannot be elicited, analyzed, etc., separately. Second, because of involving different domains in the development process, the interdisciplinary work needs to be improved. Third, the customer-integration is essential for successful RE and needs to be improved.

The requirements to the RE process that were derived from the needs of handling the cycles are described in detail here. Table 2 gives an overview of the requirements and the cycles that they were derived from. Each requirement is described here in detail:

Requirement 1: Examination of the influence of a cycle on the RE: If cycles arise, it is necessary to provide a method for determining whether the cycle has an influence on the RE. Scenario techniques [32], which are used to model interactions, are suited for this purpose. By modeling the interaction of the cycle and the development process, the influence can be detected.

Requirement 2: Iterative RE: All identified categories of cycles, except category 5, cause the requirements' changes. A proven concept of handling changes of requirements is an iterative approach. If a requirement is changed, the entire RE process or some activities of it can be repeated [34].

Requirement 3: Tracing from customer- to concretized requirements: The cycles of categories 1, 3, 5, 6, and 7 cause at first the change of concretized requirements, while cycles 2, 4 and 7 cause at first the change of customer-requirements and then of concretized requirements. It is necessary to analyze the impact on concretized requirements if customer requirements are changed, and vice versa. A possibility for tracking such interdependencies is the implementation of traceability. In software engineering, Sommerville and Sawyer [32] introduce traceability to enable tracing of requirements. Traceability matrices are a further means for implementing traceability [19]. For PSS, it must be considered that the concretized requirements concern different domains [7]. The concretized requirements of

Table 2. Requirements for RE to be cycle-oriented

Nr.	Requirements for RE	Cause for this requirement	Cycles involved
1	Examination of the influence of a cycle on the RE	Cycles exist in the development process	all
2	Iterative RE	Requirements can change during the development process due to reasons	1, 2, 3, 4, 6, 7
3	Tracing from customer- to concretized requirements and vice versa	Customer requirements can change Concretized requirements can change	1, 3, 5, 6, 7 2, 4, 7
4	Interdependencies between concretized requirements	On changes, interdependent requirements can also change.	1, 2, 3, 4, 6, 7
5	Identifying requirement conflicts	On changes, new conflicts may emerge.	1, 2, 3, 4, 6, 7
6	Reuse of requirements	After changes, already documented requirement can be reused.	1, 2, 3, 4, 6, 7
7	Defining a validity period of requirements	Previously known changes must be regarded	7
8	Prioritization of requirements	Selection of important requirements.	2, 5

each domain are documented using different notations and are not easily understandable by different people.

Requirement 4: Interdependencies between concretized requirements: If a concretized requirement is changed, further changes of concretized requirements can be triggered. It is important to redo the requirements analysis for the affected requirements, which can be determined by means of traceability. In the case of PSS, these traceability and analysis methods have to support the distribution of the concretized requirements over the domains. Because the requirements in these domains are documented differently, special methods have to be developed. For example, there is no formal way of defining requirements to services [7, 18]. Thus, a method has to be developed for establishing the traces between concretized requirements of different domains. If customer requirements are changed, an analog process and corresponding methods have to be provided.

Requirement 5: Identifying requirement conflicts: When concretized requirements are changed, new conflicts between changed and unchanged requirements can emerge, which are very important to analyze and resolve. Requirements 4 and 5 support the identification of conflicts by providing traceability between requirements; however, the main problem is the identification of conflicts. A common method is to formalize requirements by modeling them, e.g., modeling with use cases [19, 36]. Such mechanisms for identifying conflicts have to be adapted for the three involved domains, and the identification of conflicts between the two different domains must be possible. The identification and resolution of conflicts between customer-requirements can be done analog. Conflicts between customer requirements have to be resolved by established methods, e.g., workshops [3].

Requirement 6: Reuse of requirements: When the customer-requirements and the concretized requirements change (see requirement 2) it is necessary to reuse already defined requirements. By reusing requirements, time and costs can be saved [32]. The reuse can be supported by documenting interdependencies between requirements and by traceability.

Requirement 7: Defining a validity period of requirements: The legislative cycles (category 7) lead to previously known changes in the system on a previously known date. Hence, it is possible to take into consideration these changes during the development by specifying not only how each requirement has to be defined, but also in which period of time it is valid.

Requirement 8: *Prioritization of requirements:* Cycles of categories 2 and 5 lead to a reduction of the offered solution. Therefore, it is important to define priorities for requirements which specify how important a certain requirement is for the customer [28]. Based on these priorities, a decision on the reduction of the functionality can be taken. In the development of PSS it is important to find a prioritization which takes into account the different domains. Thus, the initial prioritization of the customer-requirements has to be transferred to the concretized requirements.

7. CONCLUSIONS AND THE OUTLOOK FOR FUTURE RESEARCH

In this paper a variety of research steps have been presented that were performed in order to understand the role of cycles in the RE of PSS and to derive requirements for the RE. A cycle in the development process covers a sequence of events and the results like sub processes, artifacts, etc. Such cycles complicate the planning of development processes and are problematic in conjunction with high time- and cost-pressure.

A main contribution of this paper is a comprehensive list of causes of cycles derived by literature reviews and expert interviews in industry. This way, the causes for the cycles and the implications of them on the RE were identified. The list of cycles comprises seven categories of different types of cycles and explains the causes of the cycles. It is interesting to see that the cycles' causes appear on different stages of the development and have both company-internal and external causes.

Although cycles appear also in classic requirements engineering processes of the single domains, in the context of PSS, they have even graver implications. Therefore, a special handling of the cycles in the RE of PSS is necessary. It was especially

surprising for us that most of the identified cycles have a direct effect on the requirements for the PSS. Thus, for the effective management of these cycles, the prerequisites need to be laid in RE.

The second part of the study addresses what a requirements engineering needs to provide in order to enable an effective handling of cycles. Eight requirements to a requirements engineering for PSS were derived and are described in detail. Because the cycles emerge during the development, a great challenge is to deal with changing requirements during the development process. This finding is reflected in the requirements to the RE which are largely concerned with impact analyses of changing requirements and traceability issues. The list of requirements to the RE does only comprise specific requirements staying in relation with cycles and PSS. To achieve this, all initially identified issues were compared to known RE approaches in the three domains relevant to PSS. By this means only unsolved issues have been recorded in the list of requirements. Surprisingly, these issues – such as traceability and change management – are already known to requirements engineering research, but not addresses adequately. To achieve a cycle-oriented RE it is therefore necessary to improve these topics of RE.

In our future research the results of this paper will be used to develop a cycle-oriented requirements engineering for PSS. First, the requirements to the RE approach will be incorporated into a process model for RE that describes the activities of RE. The process model has to regard the requirements presented in this paper in order to be cycle-oriented. By the execution of the activities of RE, a specification is created or an existing specification is modified and adapted. Because of the cycles, it is challenging to maintain a specification consistent and avoid that its complexity is growing over time. Therefore, a structure for the specification will be developed that addresses these problems.

8. ACKNOWLEDGEMENT

We thank the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG) for funding this project as part of the collaborative research center „Sonderforschungsbereich 768 – Managing cycles in innovation processes – Integrated development of product-service-systems based on technical products.

9. REFERENCES

[1] 2008. WI-Orientierungslisten *Wirtschaftsinformatik*.

[2] Abramovici, M. and Schulte, S. 2007. Optimising customer satisfaction by integrating the customer's voice into product development. In *Proceedings of the International Conference on Engineering Design, ICED'07* (Paris, France, 2007).

[3] Aurum, A. and Wohlin, C. 2005. *Engineering and Managing Software Requirements.* Springer, Berlin.

[4] Becker, J. and Krcmar, H. 2008. Integration von Produktion und Dienstleistung – Hybride Wertschöpfung. *Wirtschaftsinformatik 50*, 3 (2008).

[5] Berkovich, M., Esch, S., Leimeister, J.M. and Krcmar, H., Requirements engineering for hybrid products as bundles of hardware, software and service elements – a literature review. In *Proceedings of the 9. Internationale Tagung Wirtschaftsinformatik, WI 2009* (Wien, Österreich, 2009).

[6] Berkovich, M., Leimeister, J.M. and Krcmar, H., Ein Bezugsrahmen für Requirements Engineering hybrider Produkte. In *Proceedings of the Multikonferenz Wirtschaftsinformatik, MKWI 2010* (Göttingen, 2010).

[7] Berkovich, M., Leimeister, J.M. and Krcmar, H., An empirical exploration of requirements engineering for hybrid products. In *Proceedings of the XVIIth European Conference on Information Systems, ECIS 2009* (Verona, Italy, 2009).

[8] Berkovich, M., Leimeister, J.M. and Krcmar, H., Suitability of Product Development Methods for Hybrid Products as Bundles of Classic Products, Software and Serivce Elements. In *Proceedings of the ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE*, (San Diego, 2009).

[9] Boehm, B. and Basili, V.R. 2001. Top 10 list [software development]. *Computer 34,* 1 (2001), 135 - 137.

[10] Böhmann, T. and Krcmar, H. 2007. Hybride Produkte: Merkmale und Herausforderungen. In *Wertschöpfungsprozesse bei Dienstleistungen: Forum Dienstleistungsmanagement*. Gabler, 240-255.

[11] Browne, G.J. and Rogich, M.B. 2001. An empirical investigation of user requirements elicitation: Comparing the effectiveness of prompting techniques. *Journal of Management Information Systems 17,* 4 (2001), 223-249.

[12] Byrd, T.A., Cossick, K.L. and Zmud, R.W. 1992. A synthesis of research on requirements analysis and knowledge acquisition techniques. *MIS Quarterly 16,* 1 (1992), 117-138.

[13] Davies, A. 2004. Moving base into high-value integrated solutions: a value stream approach. *Industrial & Corporate Change 13,* 5 (2004), 727-756.

[14] Dodgson, M., Gann, D. and Salter, A. 2008. *The Management of Technological Innovation: Strategy and Practice: The Strategy and Practice*. Oxford University Press.

[15] Fuchs, M. and Apfelthaler, G. 2008. *Management internationaler Geschäftstätigkeit* Springer, Wien, 2008.

[16] Hauschildt, J. and Salomo, S. 2007. *Innovationsmanagement*. Vahlen.

[17] Hefley, B. and Murphy, W. 2008. *Service Science, Management and Engineering: Education for the 21st Century*. Springer, Berlin.

[18] Husen, C.v. 2007. *Anforderungsanalyse für produktbegleitende Dienstleistungen.* Fakultät Maschinenbau, Universität Stuttgart.

[19] Lamsweerde, A.v. 2009. *Requirements Engineering: From System Goals to UML Models to Software Specifications*. Wiley & Sons.

[20] Langer, S., Kreimeyer, M., Müller, P., Lindemann, U. and Blessing, L. 2008. Prozessmodellierung für das Zyklenmanagement der Entwicklung hybrider Leistungsbündel. In *Dienstleistungsmodellierung: Methoden, Werkzeuge und Branchenlösungen*, O. Thomas, and M. Nüttgens, Ed. Physica-Verlag, 71-89.

[21] Langer, S. and Lindemann, U. 2009. Managing cycles in development processes - analysis and classification of external context factors. In *Proceedings of the International conference on engineering design, ICED'09* (Stanford CA, USA, 2009).

[22] Lavagno, L. and Passerone, C. 2005. Design of Embedded Systems. In *Embedded Systems Handbook*, R. Zurawski, Ed. CRC Press Inc.

[23] Leimeister, J.M. and Glauner, C. 2008. Hybride Produkte – Einordnung und Herausforderungen für die Wirtschaftsinformatik. *Wirtschaftsinformatik*, *50,* 3 (2008).

[24] Leimeister, J.M., Knebel, U. and Krcmar, H. 2009. Hybrid Value Creation in the Sports Industry - the Case of the Mobile Sports Companion. *International Journal of Information Systems in the Service Sector (IJISSS).*

[25] Liggesmeyer, P. and Rombach, D. 2005. *Software-Engineering eingebetteter Systeme: Grundlagen-Methodik-Anwendungen*. Spektrum Akademischer Verlag.

[26] Lünendonk GmbH 2008. Lünendonk®-Listen 2008.

[27] Pahl, G., Beitz, W., Feldhusen, J. and Grote, K.-H. 2006. *Engineering Design: A Systematic Approach*. Springer, Berlin.

[28] Pohl, K. 2007. *Requirements Engineering. Grundlagen, Prinzipien, Techniken.* Dpunkt Verlag.

[29] Sawhney, M. 2006. Going beyond the Product: Defining, Designing and Devlivering Customer Solutions. In *The Service-dominant Logic of Marketing,* R.F. Lusch, and S.L. Vargo, Ed. M. E. Sharpe, New York, 365-380.

[30] Sommerville, I. 2004. *Software Engineering*. Pearson, Boston.

[31] Sommerville, I. and Kotonya, G. 1998. *Requirements Engineering: Processes and Techniques.* Wiley & Sons.

[32] Sommerville, I. and Sawyer, P. 1997. *Requirements Engineering: A Good Practice Guide*. Wiley & Sons.

[33] Spath, D. and Demuß, L. 2003. Entwicklung hybrider Produkte – Gestaltung materieller und immaterieller Leistungsbündel. In *Service Engineering - Entwicklung und Gestaltung innovativer Dienstleistungen*, Bullinger, H.-J. and Scheer, A.-W. eds., Springer, Berlin, Heidelberg, New York.

[34] Spath, D., Dill, C. and Scharer, M. 2001. *Vom Markt zum Markt*. Log_x.

[35] Sturm, F. and Bading, A. 2008. Investitionsgüterhersteller als Anbieter industrieller Lösungen – Bestandsaufnahme des Wandels anhand einer Umfrage. *Wirtschaftsinformatik 50,* 3 (2008).

[36] Sutcliffe, A. and Maiden, N. 1998. The domain theory for requirements engineering. *Software Engineering, IEEE Transactions on 24*, 3 (1998), 174-196.

[37] Tan, A.R., McAloone, T.C. and Gall, C., Product/Service-System Development - An explorative Case Study in a manufacturing Company. In *Proceedings of the International Conference on Engineering Design, ICED* (Paris, France, 2007).

[38] Tukker, A. 2004. Eight types of product-service system: eight ways to sustainability? Experiences from suspronet. *Business Strategy and the Environment*, *13*. 246-260.

[39] Tuli, K.R., Kohli, A.K. and Bharadwaj, S.G. 2007. Rethinking Customer Solutions: From Product Bundles to Relational Processes. *Journal of Marketing 71,* 3 (2007), 1-17.

[40] Ulrich, K. and Eppinger, S. 2003. *Product Design and Development* Mcgraw-Hill Professional.

[41] Watson, H.J. and Frolick, M.N. 1993. Determining information requirements for an EIS. *MIS Quarterly 17,* 3 (1993), 255-269.

[42] Yin, R.K. 2009. *Case Study Research. Design and Methods*. B&T.

SUPPORT OF MANAGERIAL DECISION MAKING BY TRANSDUCTIVE LEARNING

Hubertus Brandner
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Stefan Lessmann
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

Stefan Voß
Institute of Information Systems
University of Hamburg
Von-Melle-Park 5
D-20146 Hamburg, Germany

ABSTRACT

Transductive inference has been introduced as a novel paradigm towards building predictive classification models from empirical data. Such models are routinely employed to support decision making in, e.g., marketing, risk management and manufacturing. To that end, the characteristics of the new philosophy are reviewed and their implications for typical decision problems are examined. The paper's objective is to explore the potential of transductive learning for corporate planning. The analysis reveals two main factors that govern the applicability of transduction in business settings, decision scope and urgency. In a similar fashion, two major drivers for its effectiveness are identified and empirical experiments are undertaken to confirm their influence. The results evidence that transductive classifiers are well superior to their inductive counterparts if their specific application requirements are fulfilled.

1. INTRODUCTION

Methods and models for information and data processing are the main focus of *information systems* (IS) (see, e.g., [21, 56]). Within the discipline, the support of managerial decision making has a long tradition and dates back to - at least - *Gorry and Morton's* well known framework for structuring different types of information systems [26]. Roughly speaking, computer-based tools for decision aiding provide access to heterogeneous sources of information and the functionality required for filtering, relating and aggregating such information to gain insight into business processes, identify moderators of process efficiency and effectiveness, and eventually form adept business decisions. Respective systems are commonly referred to as analytical information systems (AIS) or business intelligence systems (see, e.g., [12, 22, 23, 38]).

Data Mining is part of the AIS-family and stipulates a machine-centric approach towards decision support [5]. Specifically, the core of Data Mining consists of a set of methods, each of which is designed for a particular analytical task. For example, association rule mining identifies co-occurrences of frequent items in transactions, such as products commonly purchased together in shopping transactions, whereas cluster analysis facilitates an autonomous categorization of objects into subgroups, a common task in marketing to segment a heterogeneous market into more homogeneous segments [7, 8, 20]. A somewhat more guided approach is taken in classification analysis. Here, a functional relationship between a discrete variable of interest (i.e., a class label) and a set measurement is inferred from past data with the objective to employ the resulting function for prediction.

Applications of classification in corporate settings are manifold and include, e.g., process and quality control in manufacturing, the screening of loan applications in the financial service industry, detecting fraudulent transactions in the telecommunication or insurance business as well as marketing decision support in response modeling or customer attrition analysis (see, e.g., [42, 51]). Several authors have argued that the predictive accuracy of classification models is imperative in such applications (see, e.g., [41, 44, 53]). This view can be understood by noting that the number of predictions in large-scale corporate applications is massive. Consequently, even marginal gains in accuracy may translate into substantial financial returns [36].

Therefore, the role of IS in the quest for improved decision quality comprises monitoring methodological advancements in computational learning and statistical inference. In fact, it has been argued that a key responsibility of IS as scientific discipline consists of bridging the gap between method-centric domains such as computer sciences or statistics and application-supplying fields like business administration [37, 39]. A similar argument may be put forward from the perspective of innovation management (see, e.g., [29]). According to the *technology-push* hypothesis, advancements in basic research are a key contributor towards innovation and economic growth (see, e.g., [21, 40]), which emphasizes the importance of IS as an integrative discipline that matches business requirements with technological opportunities.

In hunting for ever more accurate prediction models, a vast number of different classification methods have been considered and evaluated in corporate applications (see [35] for a survey). The present study is basically in line with these endeavors and examines the potential of transductive learning [54] to aid managerial decision making. However, as will become clear in the remainder of the paper, transductive learning is more than a new *method*. Transductive learning is a different philosophy towards constructing predictive classification models from data. Compared to the classi-

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

cal inductive paradigm, transductive inference (TI) pays a price in terms of universality of application, but holds the promise of increased robustness and higher predictive accuracy in particular settings. Specifically, TI stipulates a direct estimation of class memberships to simplify the overall modeling task and circumvents the detour of building a general model to predict specific cases, characteristic to classical statistical inference.

The efficacy of TI has been evidenced in several studies (see, e.g., [6, 31, 46]). However, experiences in corporate environments are yet lacking. The objective of this paper is thus to introduce the novel approach to a business audience and disclose planning tasks that could benefit from its application. The potential of TI to increase decision quality in the settings identified must be sought in its ability to deliver more accurate class predictions. Consequently, the influence of TI's specific features on forecasting performance is appraised to clarify how accuracy gains may be attained and under which circumstances they should be particularly exposed. To complement the conceptual analysis of TI's potential, an empirical study is undertaken. Using data from the field of risk-management in consumer lending, the performance of transductive techniques is compared to their inductive counterparts in different scenarios. The results observed confirm the general superiority of the former and provide valuable insight into the prerequisites for its success.

The remainder of the paper is organized as follows: Section 2 provides a brief introduction into the theory of transductive learning and its background in statistical learning. Subsequently, specific algorithms embodying this learning paradigm are introduced. Section 3 examines the implication of TI's characteristics to discern factors governing its applicability and effectiveness in business applications. To verify the importance of the identified concepts, an empirical case study is undertaken whose results are presented in Section 4. Section 5 concludes the paper with a summary of key findings and an outlook to future research.

2. BACKGROUND

2.1 Classification

Given a set of u objects $\{x_j^\star\}_{j=1}^u$, the aim of classification is to predict corresponding group memberships $\{y_j^\star\}_{j=1}^u$. The objects are characterized by a set of n attributes, which are believed to determine an object's class. Thus, all $x_j^\star$ are vectors in $\mathcal{R}^n$. For example, the objects could represent assembled products, which are to be categorized into the groups functional or defective; similarly, a marketing objective could be to distinguish customers who are responsive to direct mail from those who aren't, on the basis of demographical and transactional customer attributes. Without loss of generality [1], we focus on such binary classification problems, whereby the two possible classes are encoded as -1 and $+1$ in the following.

To perform the categorization, a classification model or classifier is derived from an example set $D = \{(x_i, y_i)\}_{i=1}^l$ of l observations with known class memberships. D is referred to as learning or training set. Keeping in mind the goal of classification, the overall model building process and the use of D in particular need to be organized in a way so as to maximize accuracy when classifying novel objects not contained in the training set. Different learning paradigms can be distinguished according to their philosophy to pursue this objective.

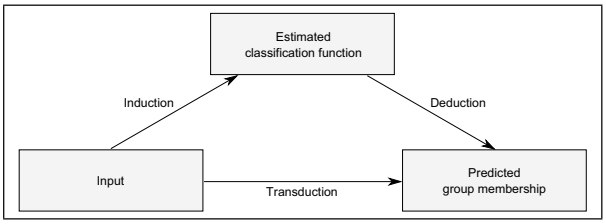


Figure 1: Inductive-deductive vs. transductive classification [54]

2.2 Inductive vs. Transductive Inference

Inductive inference and TI are different means towards building classification models from empirical data. The classical approach consists of two steps, induction and deduction. First, a classification function $f : x \mapsto y \ \forall x \in \mathcal{R}^n$ is derived from D, e.g., by minimizing the mismatch between model-estimated ($f(y_i)$) and actual class labels (y_i) over D. This model is subsequently employed for predicting class memberships for (arbitrary) novel objects. Note that the construction of a general model from a particular sample of data is a challenging undertaking, which holds several pitfalls (see, e.g., [16, 50]). Roughly speaking, representativeness of D for the whole problem space is essential, but may not always be taken for granted.

TI grounds on the observation that the complexity associated with building a global model from a limited sample can and should be avoided in settings where the objects to be classified are known in advance. That is, only the group memberships of some given objects are unknown and need to be predicted [54]. Whenever this requirement is met, the intermediate step of building a global model is unnecessarily complex and dispensable. Class labels should better be estimated in a direct manner [14, 43]. The conceptual difference between the two learning philosophies of transductive and inductive learning is illustrated in Figure 1.

In the transductive setting, class predictions are only sought for a clearly defined, a priori known set of objects, the *working set* $\{x_j^\star\}_{j=1}^u$. In this case, TI holds the promise to increase the accuracy and robustness of class predictions. To achieve this, the working set is considered alongside the ordinary training data during classifier construction. In other words, a classification rule is inferred from data comprising both, labeled and unlabeled examples. This way, a transductive classifier has access to the working set and the additional information contained therein. In other words, it is designed to handle precisely (and only) these objects. This differs notably from an inductive setting where the objects to be classified remain unknown until the (global) model is built. Therefore, a transductive classifier solves a much simpler estimation task and is thus less vulnerable to distributional discrepancies between training data and the working set. Consequently, it can be expected to be more accurate in classifying the points of interest (i.e., the working set) [14, 15, 17].

2.3 Support Vector Machines

Support Vector Machines (SVMs) are a popular method to construct inductive classification models. However, they can be extended to perform TI in a straightforward manner [31]. Therefore, SVMs facilitate an unbiased comparison between the two principles of inductive and transductive classification and will serve as representatives of both learning paradigms within the empirical evaluation. The respective procedures are sketched in the following, whereas a comprehensive introduction into the theory of SVMs can be found in textbooks like [17] or [28].

2.3.1 Inductive Support Vector Machines

The concept of inductive SVMs (iSVMs) is illustrated in Figure 2. Given a training set of examples together with their class membership, a hyperplane H is constructed so as to separate objects of adjacent classes with large margin. The concept of maximal margin is key to the SVM approach and has been shown to increase the accuracy of class predictions for novel objects [6, 54]. The hyperplane represents an inductive classification model, which classifies novel instances according to their relative position (above or below) to H. The model's parameters (normal and intercept) are estimated by solving a convex quadratic program for the training set (see, e.g., [28]).

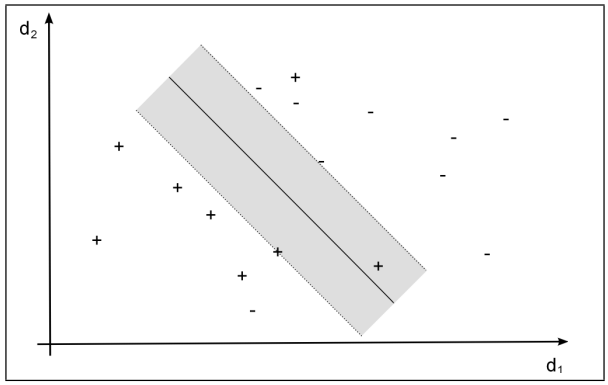


Figure 2: Inductive SVM in a linearly non-separable case in $\mathcal{R}^2$. $+,-$ represent training set objects $\{x_i\}_{i=1}^{l}$ and their classes ($y_i \in \pm 1$), respectively. The gray rectangle depicts the margin between the two classes. New data points $x_j^\star$ are classified according to their position, below ($f(x_j^\star) = y_j^\star = -1$) or above ($f(x_j^\star) = y_j^\star = +1$) the plane H.

2.3.2 Transductive SVM

A transductive SVM (tSVM) implements a similar strategy and differs only in the approach to determine the location of the separating hyperplane. Having access to the working set and thus information about the location of the points to classify, additional constraints concerning the orientation of H can be imposed: Objects with known class (i.e., the training set) should again be separated with large margin to increase the classifier's ability to generalize (see above). However, this principle cannot be applied to working set objects directly, since their class is, by definition, unknown. In order to approximate margin maximization over the working set, the hyperplane is pushed into a region with low data density. This approach is motivated by the *cluster-assumption*, which claims that data points are likely to be of the same class, if they are *close* to each other, i.e., in the same cluster of the space [13]. Consequently, pushing the hyperplane away from the unlabeled points can be expected to maximize the margin of separation over all data, the training set and the working set [17]. More specifically, it is reasonable that tSVM achieves larger margin, and thus higher accuracy, on working set examples, because these are considered during model fitting. To see this, consider Figure 3 that continues the previous example, but also depicts a possible set of points to classify.

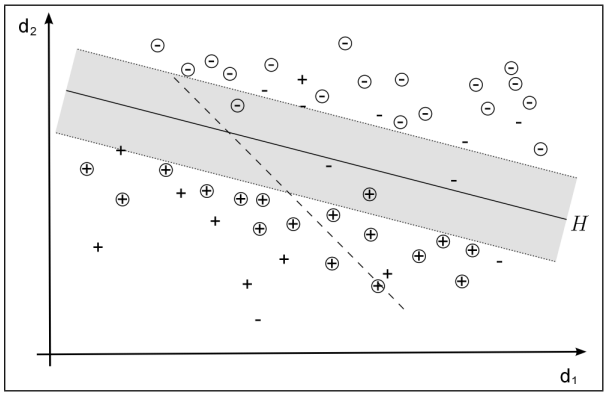


Figure 3: Transductive SVM in a linearly non-separable case in $\mathcal{R}^2$. $\oplus,\ominus$ represent the working set $\{x_j^\star\}_{j=1}^{u}$ and their class labels $\{y_j^\star\}_{j=1}^{u}$, as estimated by the transductive classifier (solid line). The dashed line equals the inductive classifier of Figure 2.

In appraising Figure 3, it is important to remember that the true class membership of working set objects is unknown. However, by visual inspection, one would assume that the class boundary of tSVM mimics the true relationship between object attributes and class membership more closely than those of iSVM since less points of the working set fall into the margin, i.e., the region around the estimated class boundary. This can be expected to provide more accurate class predictions over working set instances.

In order to take account of the working set during model fitting, the mathematical program underlying SVMs needs to be extended. The tSVM formulation used in this work has been proposed by [31]. It involves solving a mixed-integer program. Therefore, standard SVM training algorithms are no longer applicable. In order to eliminate factors influencing classification accuracy other than the embodied learning principle, we develop a novel metaheuristic that facilitates solving different variants of the SVM learning problem in a unified way. The procedure can be characterized as a *memetic* algorithm (see, e.g., [27]). It should be noted that the heuristic computes the equation of the optimal hyperplane (i.e., the global model) as byproduct of the learning phase. Although not the aim of transduction, H could be employed to classify all other $x \in \mathcal{R}^n$.[1]

[1]This principle, learning from both labeled and unlabeled examples to build global models, is called Semi-Supervised

3. APPLICABILITY AND EFFECTIVENESS OF TI IN BUSINESS CLASSIFICATION

In order to appraise the managerial utility of TI, it is essential to fully understand how its specific features create value (i.e., increase predictive accuracy), in which circumstances they are most effective, and how they affect TI's applicability in corporate contexts. In particular, a key characteristic and requirement of TI concerns the availability of the working set; in addition to a training dataset - always needed in predictive modeling - all objects which are to be classified need to be known in advance. The following sections clarify upon the implication of this particularity with respect to applicability and effectiveness of TI.

3.1 Moderators of TI Applicability

Considering typical business applications of classification (see, e.g., [35, 51]), it is easy to find examples where the requirement of an a priori given working is met. Consider, e.g., the task of targeting customers in direct marketing (see, e.g., [9, 32]). A mail-order company is well aware of the clients it could possibly solicit. In particular, the actual decision task is to select from a set of present customers those who are most likely to respond. Consequently, the objects to be classified are known in advance. A similar situation occurs in churn prediction (see, e.g., [41]): From the set of all current customers, a marketer wishes to identify those with highest attrition risk. Again, the objects to be classified are known in advance. On the other hand, decision support in credit scoring (see, e.g., [48]), also a popular application for classification in business, can be considered a counterexample. Here, loan approval models are employed to assess incoming credit applications, which have not been observed at the time the classification model was built.

The previous examples differ in terms of the number of objects that are classified. In credit scoring, decisions are sought for individual loan applications, whereas the marketing examples are associated with classifying a set of customers. We refer to this construct as decision scope and propose it as moderator of TI's applicability in business contexts. Specifically, TI benefits from wider decision scope, which involves handling a larger number of objects.

The previous assertion follows directly from the observation that if a decision concerns a (large) set objects, a working set containing these objects must be available. However, does this imply that singular decisions such as those in credit scoring prohibit use of TI? It seems that this is not necessarily the case. Consider for example the task of fraud detection in the insurance industry. Classification models are employed to screen incoming claims, e.g., to decide whether a claim can be settled immediately or whether it requires a closer examination by a human expert (i.e., appears suspicious). This task shares similarities with the credit approval example in that decision objects (insurance claims and loan applications, respectively) arrive periodically. However, as opposed to credit decisions which loan applicants will want to obtain almost instantly, insurance holders may not expect claims to be decided immediately, but accept some processing time. This delay will inevitably result in a queue of claims to be processed. Especially if the number of incoming claims is large, the amount of objects to classify (i.e., the working set size) may well be sufficient to justify a transductive approach. Therefore, less urgency in making decisions may facilitate use of TI, even if decision scope is narrow. Consequently, it seems well justified to consider decisions' urgency as an additional moderator of TI's applicability.

Learning (SSL) and resides somewhat between inductive and transductive classification [13].

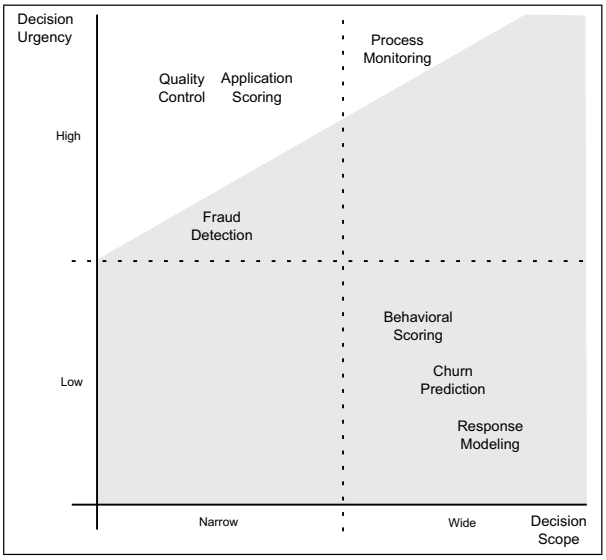


Figure 4: Decision problem characteristics and corresponding applications; the grey area depicts the region of TI-feasibility

The previous arguments are summarized in Figure 4. The grey zone within the decision urgency/decision scope portfolio delineates the general application area of TI. An ordinary scale has been employed to distinguish between two states for the identified moderators. In addition, some specific business classification tasks are given to exemplify applications suitable and unsuitable for TI.

Clearly, the location of classification tasks in Figure 4 is debatable and should be taken as a proposition. This is especially true for decision urgency, so that the actual applicability of TI for, e.g., insurance fraud detection depends on several factors such as company size, type of claim, etc. Therefore, it needs to be scrutinized on a case by case basis. In that respect, the ultimate objective of Figure 4 is to provide a tool for structuring and systematizing such an appraisal.

We refrain from discussing every single positioning decision in detail. However, the following considerations may help to motivate our proposition: Planning decisions in response modeling, churn prediction and behavioral scoring[2] involve a campaign approach where a model is built at a particular point in time and employed to obtain scores for a collection of customers. Differences in terms of decision scope among these three may exist because of variations in the frequency with which campaigns are launched. However, in comparison to other tasks such as credit scoring, the com-

[2]Behavioral scoring is related to application (credit) scoring but employed at a different stage within the customer lifecycle (see, e.g., [47, 49]). In particular, once credit has been granted, financial institutions have an interest in estimating the likelihood of default to take preemptive actions for high risk borrowers.

pany faces fewer constraints in timing the task since customer expectations in terms of process time do not exist. Consequently, these three tasks' urgency is considered as low. Quality control involves classifying assembled products according to their compliance with predefined quality indicators. This task shares many similarities with credit scoring in the sense that objects are classified on a one by one basis and that these decisions are relatively time critical. For instance, products of a particular lot may not be sold until quality has been verified. Contrary to quality control, process monitoring refers to the surveillance of a whole (manufacturing) system. Classification aids this endeavor by processing large amounts of data gathered by various process monitors to determine individual components' reliability, which, in turn, allows conclusions regarding the whole system's soundness to be drawn. This suggest wider decision scope. The positioning of fraud detection has been elaborated above.

3.2 Moderators of TI Effectiveness

In order to provide more accurate class predictions than an inductive model, a transductive classifier needs to extract additional information from the working set, over and above those already contained in the training set. This becomes possible if the latter is not well representative for the objects to be classified, i.e., if the distribution between the two sets differs in some fashion. Consequently, factors governing distributional discrepancy between training and testing data can be expected to influence the effectiveness of TI.

In previous work, TI has been found to be most useful in settings where unlabeled data is easily available, whereas obtaining labeled data is associated with high cost (see, e.g., [13, 31]). Such scenario appears well representative for many business applications. For example, mail-order companies have (or can easily gain) access to enormous amounts of customer data. However, whether a particular customer has previously responded to a catalog mailing (i.e., the class label) is known only if the customer has ever been solicited. In other words, obtaining the customer's label requires sending a catalog and is thus costly.

The fact that TI does work well in the case of an imbalance between the amounts of labeled and unlabeled data can be explained by distributional discrepancy. In particular, for an imbalance to exists, the size of the training dataset has to be small, relative to the working set, and a small sample may not represent well the overall distribution. This is especially so in corporate settings where the training data is commonly not a random sample. In particular, the objects whose class labels are known (i.e., the ones that can be employed for training) have themselves been selected by a previous classification. For example, loans are exclusively granted to applicants with low default probability, which, in turn, has been estimated by a credit scoring model. Since the information whether a debtor eventually defaults becomes available only for such preselected applicants, the data that can be employed for model building is restricted to low-risk applicants and thus a biased sample of the distribution of all applicants.[3] Therefore, representativeness of training data must not be taken for granted and may well be limited in some corporate applications.

[3]This problem is known as sample-selection-bias and elaborated in the context of reject inference (see, e.g., [4]).

Besides statistical consideration, mathematical arguments suggest that the effectiveness of TI depends upon an imbalance between labeled and unlabeled data (or the lack thereof). As explained in the case of tSVMs, training examples and unlabeled cases are employed jointly when constructing a transductive classifier. Consequently, the effect of the (possibly less representative) labeled instances on the classifier will be excessive unless sufficient unlabeled objects are available.

In view of the previous reasoning, we propose that the effectiveness of TI increases if the imbalance between labeled and unlabeled data (i.e., the ratio l/u) decreases.

In addition to being of insufficient size or sampling issues, another reason for discrepancy between training data and the cases to classify is change in the processes that provide these datasets. Mail-order companies, for example, commonly solicit customers with catalog mailings on a quarterly basis. The needs and preferences of customers, and also their affinity towards direct-mail, are not fixed but undergo a constant change. Therefore, the rules once inferred by a classifier to identify responsive customers may loose sharpness over time and require updating to take account of more recent developments in, e.g., customer behavior. In the case of credit scoring or fraud detection, this issue is even more severe since applicants/fraudsters may deliberately attempt to alter their characteristics so as to circumvent decision support models and obtain favorable classifications [10]. Clearly, such action - if successful - will also diminish the appropriateness of a previously derived classifier.

In summary, the relevance of model updating depends upon the stationarity of the data generation process, whereby regular changes (less stationarity) dictate more frequent model updates. By classifying working set objects directly and thus circumventing the construction of a general classification model, TI can be interpreted as an extreme approach in terms of model updating: A new model is constructed for every decision. Consequently, process stationarity can be identified as another moderator of TI's effectiveness, which leads to the portfolio shown in Figure 5.

The grey area illustrates scenarios, which appear particularly suited for TI, i.e., can be expected to benefit from its ability to classify working set instances directly. On the contrary, employing unlabeled data during model building seems less effective if the mechanisms that govern the relationship between class labels and object characteristics remain stable over time or the ratio of l/u is large.

Concerning the positioning of example applications, we motivate our choices as follows: It is reasonable to distinguish business applications that involve classifying customers from decision tasks referring to machines or manufactured products. The former change their behavior far more rapidly than, e.g., an assembly line its operation. Consequently, the stationarity of data generation processes is decreased in human-centric settings. Assuming that the size of available training data depends mainly upon company size, working set size becomes the main discriminator between example applications in terms of imbalance. The horizontal position of applications is then obtained by employing decision scope (see Figure 4) as proxy for amount of objects to classify.

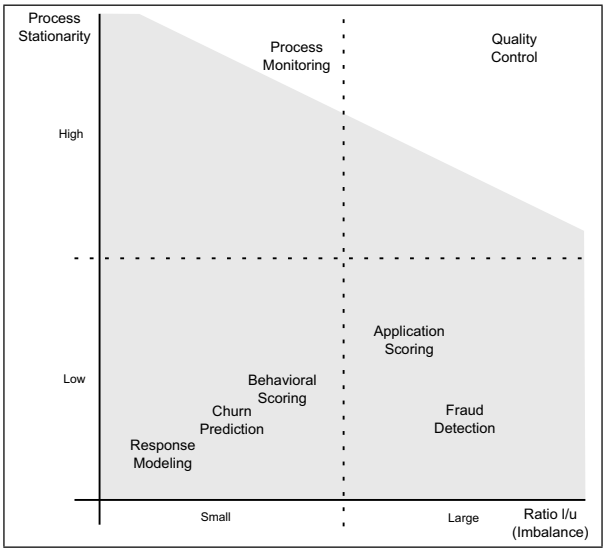


Figure 5: The marked area shows combinations of data imbalance and process stationarity promising an effective utilization of TI.

Overall, the managerial insight gained from the conceptual analysis of TI can be summarized as follows: When facing a classification problem, decision makers are well advised to examine the task's compliance with TI's requirements concerning decision characteristics. Problems that do exhibit wide scope and/or less urgency are candidate applications of TI and should be evaluated according to the imbalance between labeled and unlabeled objects and how likely changes in underlying data generation processes appear.[4] The procedures leading to such an appraisal have been exemplified in conjunction with seven business classification tasks and illustrate how promising candidate application can be identified. For the examples considered above, TI can be expected to be particularly effective in behavioral scoring, churn prediction and response modeling. It could also be considered in fraud detection applications, whereas application scoring, process monitoring and quality control remain an area for inductive classification.

4. EMPIRICAL EVALUATION

As pointed out by Hevner et al. [30], a careful and rigorous evaluation is crucial to ensure that an IT artifact is compatible with corporate requirements and successfully solves the problem it is meant to solve. The managerial utility of TI must be sought in its potential to give more accurate class predictions. As explained above, accuracy is often imperative in corporate settings and may offer substantial monetary rewards [41, 44, 53]. Whereas theoretical arguments in favor of TI's superiority over classical inductive learning have been put forward in the previous sections[5], additional empirical evidence is desirable to complement the evaluation of TI and confirm its effectiveness to deliver more accurate predictions.

[4]It should be noted that the problem of distributional change in data generation processes is examined in detail in the field of concept drift learning (see, e.g., [34]) and that dedicated procedures have been developed to cope with this modeling challenge (see, e.g., [33, 45]). However, these endeavors commonly concentrate on applications where new data arrives continuously (e.g., in streams) and is processed instantly. This differs notably from candidate applications of TI, where wide decision scope and low decision urgency cause/enable a batch-processing of the (working-set) instances to be classified.

4.1 Data

In order to verify the potential of TI for managerial decision support, inductive and transductive SVMs are evaluated on two real-life datasets: Australian (AC) and German Credit (GC). Both datasets originate from the domain of credit scoring and are publicly available in the UCI Machine Learning Repository [2]. They have been employed in numerous studies on classification (see, e.g., [3, 36, 52, 55]) and serve as examples for business classification problems in this study. Specifically, the binary target variable indicates whether a debtor x_i fulfills his obligation of repaying a loan ($y_i = -1$) or defaults ($y_i = +1$). Although this classification task belongs to the field of application scoring, which, due to narrow decision scope and high urgency, is less suitable for TI (see above), the variables provided to perform the classification are also relevant for behavioral credit scoring (see, e.g., [47]), i.e., a potential application area. Therefore, it seems appropriate to consider these datasets for the evaluation. Their characteristics are summarized in Table 1.

	$\|dataset\|$	n	A priori probability of class +1
AC	690	14	0.449
GC	1000	24	0.300

Table 1: Characteristics of credit scoring datasets.

4.2 Scenarios

A key objective of the comparison of inductive and transductive classifiers is to confirm the impact of imbalance and process stationarity on TI's effectiveness. To that end, two experimental scenarios are developed.

The impact of imbalance is examined by means of varying the ratio l/u, the cardinalities of the training and working set, respectively. Specifically, iSVM and tSVM models are built and compared in terms of their predictive accuracy for five settings of $l/u = 2, 4, 6, 8,$ and 10%. These ratios may be considered conservative; even more dramatic inequalities can be found in the literature [6, 15, 31].

To appraise the influence of distributional change, five discrete points in time are defined and the similarity between training data and working set data is decreased between consecutive time points. Specifically, distributional discrepancy is artificially introduced through systematical manipulation of the test set. Roughly speaking, the attributes of the test set examples are modified by adding a constant direction vector in general and a normal distributed variable individually. This is necessary since both sets are random samples drawn from the original dataset (i.e., AC or GC) and therefore equivalent in terms of their underlying data distribution.

[5]A formal mathematical discussion in terms of transductive and inductive techniques' ability to minimize bounds of generalization error can be found in [18].

4.3 Performance Measurement

Numerous approaches exist to measure a classifier's predictive power. We decided to assess performance by an indicator which grounds on the basis of a discrete categorization of predictions into four groups: false positive, false negative, true positive and true negative. The following contingency table depicts this principle.

		estimated class: -1	estimated class: +1
real class	-1	true negative (tn)	false positive (fp)
	-1	false negative (fn)	true positive (tp)

The motivation for considering discrete class predictions as opposed to probabilistic or confidence based measures such as AUC (see, e.g., [19]) is that TI is designed to generate crisp classifications.

To measure the accuracy of classifications, we calculate the F_1-Score (FSC), a widely used metric in Information Retrieval, which is defined as the harmonic mean between precision $\frac{tp}{tp+fp}$ and recall $\frac{tp}{tp+fn}$ (see, e.g., [11])[6]:

$$F_1 := \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}.$$

For each of the settings (imbalances or time points), a ten-fold split-sample setup is employed. That is, the original dataset (AC or GC) is randomly partitioned into training and working set (i.e., labeled and unlabeled data) to construct and assess classification models. The resulting F_1-scores are then averaged to obtain a final performance estimate.

4.4 Empirical Results

4.4.1 Imbalance

Figure 6 summarizes the results obtained from experiments with increasing imbalance between the amount of labeled and unlabeled data. Individual settings are represented by stems, whose height measures ΔF_1-values, the difference between the F_1-Score of tSVM minus the F_1-Score of iSVM. Thus, a positive value implies superior performance of the transductive approach.

Overall, the observed results are in line with theory: tSVM consistently achieves higher performance, with the least imbalanced setting on GC being the only exception. Moreover, there is a clear trend of tSVM outperforming iSVM with increasing margin as imbalance increases. In other words, the more extreme the ratio between labeled and unlabeled data, the more accurate are the class predictions of tSVM compared to those of iSVM. This dependency is confirmed by a linear regression of imbalance setting (1,2,...,5) on ΔF_1, which gives an R^2 of 0.528 and is significant at the two-percent level. Therefore, one may speculate that the superiority of transductive methods over their inductive counterparts will be even larger, if more extreme ratios of $\frac{l}{u}$ are present (see also [31]).

In view of the fact that the availability of labeled data will often be closely related to the cost incurred by querying class labels (e.g., sending a catalog to a potential customer), the overall conclusion emerging from this experiment is that TI is most effective in settings where labeling costs are high. This will commonly be the case in corporate settings, especially if obtaining actual class memberships involves human experts. Exemplary classification tasks include, e.g., an assessor examining insurance claims or a quality manager appraising the functionality of manufactured products. In that respect, TI appears to be a promising alternative to established approaches for business classification.

[6]Alternative performance metrics have also been considered, but their effect on the results was found to be negligible, they showed similar tendecies as the used measure. Therefore, the presentation is restricted to the F_1-Score.

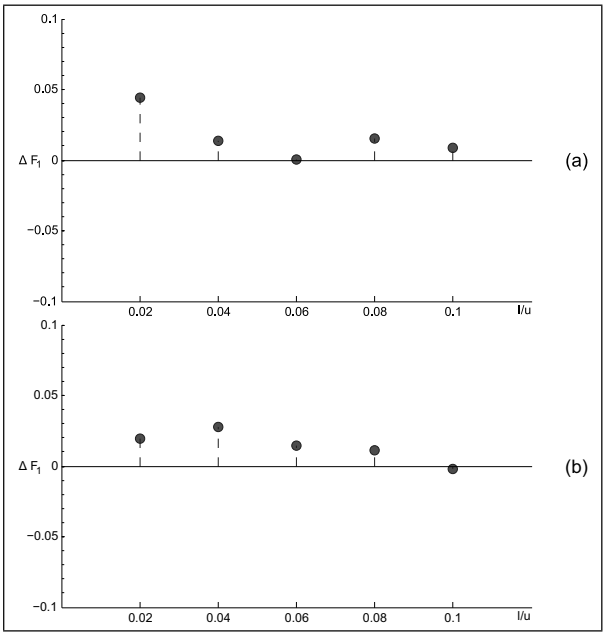


Figure 6: Results of the scenario "Imbalance" on AC (a) and GC (b)

4.4.2 Distributional Change

Results of the comparison of tSVM and iSVM under distributional change are presented in Figure 7. Here, the advantage of tSVM is even more exposed than in the previous experiment. The magnitude of the difference increases substantially with training data becoming less representative for the cases to classify. The results of a linear regression, $R^2 = 0.902$ and a p-value of the F statistic $< 0.01\%$, verify this finding. Therefore, the results confirm and reemphasize the view that the use of unlabeled data is most beneficial, if - for whatever reason - labeled examples collected in the past no longer reflect the present drivers of class membership. To further clarify upon the dominance of tSVM over iSVM in this experiment, a more detailed view on the empirical results is given in Figure 8. It depicts the raw F_1-Scores of the competing classifiers for AC and their development over the five time points with increasing discrepancy between the training and working set data distributions. Higher values once more indicate better performance.

In the first period, training and working set data are both random samples drawn from AC. Although the training data is thus well representative for the working set, a minor ad-

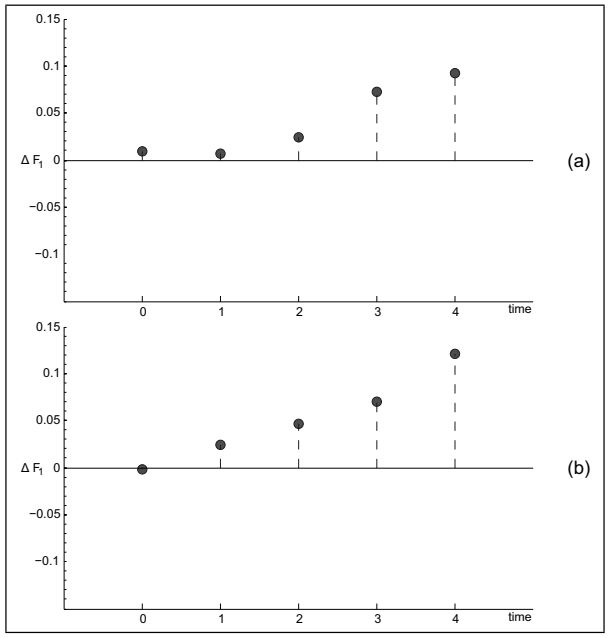


Figure 7: Results of the scenario "Change of distribution" on AC (a) and GC (b)

vantage of tSVM can be observed. This may be taken as empirical confirmation for the view that TI involves solving a simpler problem; i.e., class labels can be estimated in a direct manner if the instances to be classified are known. In the focal case, this approach indeed yields higher predictive accuracy.

Furthermore, Figure 8 illustrates the dramatic decline of iSVM's performance when training data becomes less representative. Although the transductive classifier's access to labeled data is also restricted to outdated training examples, it succeeds in distilling additional information from the unlabeled working set and, thereby, maintains its level of accuracy. For example, in the most extreme setting, the predictive ability of iSVM has declined by 10.85% , whereas tSVM is only 0.88% below its performance peak. Therefore, the results provide strong evidence for TI being indeed highly robust towards changes in data distributions.

5. SUMMARY AND CONCLUSIONS

Classification is a well established approach to support various decision making tasks. Due to the large number of decisions and thus classifications, the accuracy of classification models is commonly considered pivotal in business applications. Therefore, we aimed at exploring the potential of TI, a novel approach towards building predictive classification models, for corporate planning. A key characteristic of the novel paradigm involves a direct estimation of class labels to increase predictions' accuracy and robustness. The theoretical background of this principle has been reviewed and an analysis of the implications of TI's requirement has led to the identification of two factors which govern its applicability in corporate planning: decision scope and decision urgency. In a similar way, the factors imbalance and process stationarity have been proposed as major determinants of TI's expected

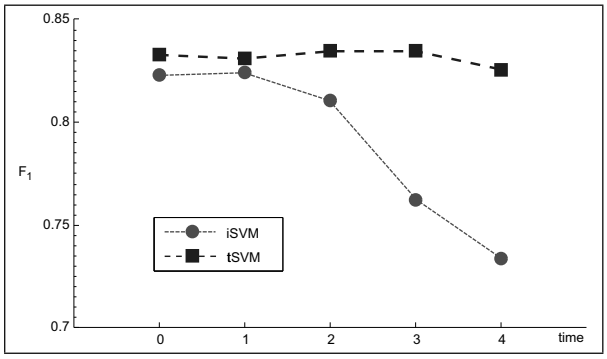


Figure 8: Change of distribution: Developing of F_1-scores

utility. These four concepts provide a framework for evaluating whether business decision tasks may benefit from an application of TI. An empirical case study has been carried out to complement the valuation of the novel approach. Overall, TI compares favorably to inductive classification and provides more accurate class predictions in most settings. More specifically, although TI's specific requirements in terms of working set availability constrain its applicability in general, TI has been shown to be well competitive if not superior to conventional techniques whenever these requirements are fulfilled. In particular, the performance of tSVM was at least comparable to those of iSVM throughout all experiments, and substantially better in most cases.

Clearly, empirical findings are restricted to the employed datasets and may not generalize to other applications. Although the data has been drawn from the domain of corporate planning, there is no guarantee that similar results can be observed in, e.g., marketing or manufacturing problems. For example, there is a reasonable risk that the relatively small size of our datasets have granted TI an advantage. This follows directly from the results of the imbalance experiment. Therefore, a careful evaluation of TI in other corporate planing domains is an important area for future research. Our work supports this undertaking by identifying the moderators of TI's applicability and effectiveness and proposing a methodology for appraising a decision task's fit with the novel paradigm. Respective procedures have been discussed in the context of typical corporate applications and concrete examples of promising decision problems have been provided. Moreover, the experimental scenarios developed for TI's evaluation may prove useful in subsequent studies.

On the other hand, there may be no need for being overly pessimistic when appraising the present results' external validity. Drawing inspiration from typical modeling challenges in business classification, all experiments have been carefully designed to assess particular features of TI, which theory would predict to be beneficial. In other words, the encouraging findings of the imbalance and distributional change experiment can well be explained with TI's statistical and mathematical underpinnings.

In view of the overall insights gained during the course of this evaluation, a general conclusion might be that TI represents a novel decision support tool which has the potential to complement or even replace established (i.e., inductive) techniques, if its particular requirements are fulfilled.

However, one may object that this view centers too drastically on technology, rather than decision makers' requirements. Alternatively, TI could be characterized as a planning tool offering higher task-technology fit [24, 25] in specific circumstances (i.e., when seeking class predictions for a known working set of objects) and is, in this sense, preferable - and supposedly superior - to standard methods like inductive classification. In that respect, TI may be considered an example, how advancements in data analytical techniques can and should be geared towards concrete business needs. That is, instead of being forced to match a given decision problem to some standardized procedures (e.g, a standard data mining algorithm for classification, clustering or association), dedicated planning tools should be devised that take account for application specific requirements. This is maybe the most promising avenue for future research in corporate decision support, with TI being a very first step towards the long term objective of a requirement-driven data mining.

6. REFERENCES

[1] E. L. Allwein, R. E. Schapire, and Y. Singer. Reducing multi-class to binary: a unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141, 2000.

[2] A. Asuncion and D. Newman. UCI Machine Learning Repository, 2007.

[3] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635, 2003.

[4] J. Banasik and J. Crook. Reject inference, augmentation, and sample selection. *European Journal of Operational Research*, 183:1582–1594, 2007.

[5] U. Bankhofer. Data Mining und seine betriebswirtschaftliche Relevanz. *Betriebswirtschaftliche Forschung und Praxis*, 56:395–412, 2004.

[6] K. P. Bennett and A. Demiriz. Semi-supervised Support Vector Machines. In *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*, volume 2, pages 368–374, 1999.

[7] M. J. A. Berry and G. Linoff. *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*. Wiley, New York, 2 edition, 2004.

[8] N. Bissantz and J. Hagedorn. Data Mining (Datenmustererkennung). *Wirtschaftsinformatik*, 51(1):139–144, 2009.

[9] A. Bodapati and S. Gupta. A direct approach to predicting discretized response in target marketing. *Journal of Marketing Research*, 41(1):73–85, 2004.

[10] F. Boylu, H. Aytug, and G. J. Köhler. Induction over strategic agents. *Information Systems Research*, 21(1):170–189, 2010.

[11] R. Caruana and A. Niculescu-Mizil. Data Mining in Metric Space: an Empirical Analysis of Supervised Learning Performance Criteria. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 69–78, New York, 2004.

[12] P. Chamoni and P. Gluchowski. *Analytische Informationssysteme*, pages 3–22. Springer, Berlin Heidelberg, 3 edition, 2006.

[13] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-Supervised Learning*. The MIT Press, London, 2006.

[14] O. Chapelle, V., and J. Weston. Transductive inference for estimating values of functions. *Advances in Neural Information Processing Systems*, 12:421–427, 1999.

[15] Y. Chen, G. Wang, and S. Dong. Learning with Progressive Transductive Support Vector Machine. In *International Conference on Data Mining*, pages 67–74, 2002.

[16] V. Cherkassky and Y. Ma. Another look at statistical learning theory and regularization. *Neural Networks*, 22(7):958–969, 2009.

[17] V. Cherkassky and F. M. Mulier. *Learning from Data: Concepts, Theory, and Methods*. Wiley & Sons, New Jersey, 2 edition, 2007.

[18] S. Decherchi, P. Gastaldo, S. Ridella, and R. Zunino. Explicit overall risk minimization transductive bound. *Atlas Conferences*, 2008.

[19] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[20] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, pages 37–45, 1996.

[21] A. Fink, G. Schneidereit, and S. Voß. *Grundlagen der Wirtschaftsinformatik*. Physica-Verlag, Heidelberg, 2 edition, 2005.

[22] P. Gluchowski, R. Gabriel, and C. Dittmar. *Management Support Systeme und Business Intelligence: Computergestützte Informationssysteme für Fach- und Führungskräfte*. Springer, Berlin, 2 edition, 2008.

[23] P. Gluchowski and H.-G. Kemper. Quo vadis business intelligence? *BI-Spektrum*, 1:12–19, 2006.

[24] D. L. Goodhue. Understanding user evaluations of information systems. *Management Science*, 41(12):1827–1844, 1995.

[25] D. L. Goodhue and R. L. Thompson. Task-technology fit and individual performance. *MIS Quarterly*, 19(2):213–236, 1995.

[26] G. A. Gorry and S. Morton. A framework for management information systems. *Sloan Management Review*, 13(1):55–70, 1971.

[27] W. E. Hart, N. Krasnogor, and J. E. Smith. *Recent Advances in Memetic Algorithms*. Springer, Berlin Heidelberg, 1 edition, 2005.

[28] T. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, 2009.

[29] J. Hauschild. *Innovationsmanagement*. Vahlen, München, 3 edition, 2004.

[30] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design science in information systems research. *MIS Quarterly*, 28(1):75–105, 2004.

[31] T. Joachims. Transductive Inference for Text Classification Using Support Vector Machines. In *Proceedings of the 16th International Conference on*

Machine Learning, pages 200–209, 1999.

[32] Y. S. Kim, W. N. Street, G. J. Russell, and F. Menczer. Customer targeting: a neural network approach guided by genetic algorithms. *Management Science*, 51(2):264–276, 2005.

[33] J. Z. Kolter and M. A. Maloof. Dynamic weighted majority: A new ensemble method for tracking concept drift. *Journal of Machine Learning Research*, 8:2755–2790, 2007.

[34] L. I. Kuncheva. Classifier Ensembles for Chaning Environments. In F. Roli, J. Kittler, and T. Windeatt, editors, *Proceedings of the 5th International Workshop on Multiple Classifier Systems*, volume 3077 of *Lecture Notes in Computer Science*, pages 1–15, Cagliari, Italy, 2004. Springer.

[35] S. Lessmann and S. Voß. Supervised Classification for Decision Support in Customer Relationship Management. In A. Bortfeldt, J. Homberger, H. Kopfer, and R. G. Pankratz R. Strangmeier, editors, *Intelligent Decision Support*, pages 231–253. Gabler, Wiesbaden, 2008.

[36] S. Lessmann and S. Voß. Unterstützung kundenbezogener Entscheidungsprobleme - Eine Analyse zum Potenzial moderner Klassifikationsverfahren. *Wirtschaftsinformatik*, 52(2):79–93, 2010.

[37] P. Mertens. Geschichte und ausgewählte Gegenwartsprobleme der Wirtschaftsinformatik. *Wirtschaftswissenschaftliches Studium*, 27:170–175, 1998.

[38] P. Mertens. Business Intelligence - Ein Überblick. *Information Management & Consulting*, 22:65–73, 2002.

[39] H. Müller-Merbach. Die ungenutzte Synergie zwischen Operations Research und Wirtschaftsinformatik. *Wirtschaftsinformatik*, 34(3):334–339, 1992.

[40] G. F. Nemet. Demand-pull, technology-push, and government-led incentives for non-incremental technical change. *Research Policy*, 38(5):700–709, 2009.

[41] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason. Defection detection: measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research*, 43(2):204–211, 2006.

[42] E. W. T. Ngai, L. Xiu, and D. C. K. Chau. Application of data mining techniques in customer relationship management: a literature review and classification. *Expert Systems with Applications*, 36(2, Part 2):2592–2602, 2009.

[43] S. Pang and N. Kasabov. Inductive vs transductive inference, global vs local models: SVM, TSVM, and SVMT for gene expression classification problems. *Neural Networks*, 2:1197–1202, 2004.

[44] F. F. Reichheld and W. Sasser. Zero defections: quality comes to service. *Havard Business Review*, 68(5):105–111, 1990.

[45] M. Scholz and R. Klinkenberg. Boosting classifiers for drifting concepts. *Intelligent Data Analysis*, 11(1):3–28, 2007.

[46] M. M. Silva, T. T. Maia, and A. P. Braga. An Evolutionary Approach to Transduction in Support Vector Machines. In *Proceedings of the 5th International Conference on Hybrid Intelligent Systems*, pages 329–334, 2005.

[47] L. C. Thomas. A survey of credit and behavioral scoring; forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16:149–172, 2000.

[48] L. C. Thomas, J. Crook, and D. Edelman. *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

[49] L. C. Thomas, R. Oliver, and D. J. Hand. A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society*, 56(9):1006–1015, 2005.

[50] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. Winston, Washington, D.C., 1977.

[51] K. L. Tsui, V. C. P. Chen, W. Jiang, and Y. A. Aslandogan. Data Mining Methods and Applications. In H. Pham, editor, *Springer Handbook of Engineering Statistics*, pages 651–669. Springer, London, 2006.

[52] B. Twala. Multiple classifier application to credit risk assessment. *Expert Systems with Applications*, 37(4):3326–3336, 2010.

[53] D. Van den Poel and B. Lariviere. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1):196–217, 2004.

[54] V. Vapnik and S. Kotz. *Estimation of Dependences Based on Empirical Data*. Springer, New York, 2 edition, 2006.

[55] G. Wang, J. Hao, J. Ma, and H. Jiang. A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 2010.

[56] Wissenschaftliche Kommission Wirtschaftsinformatik. Profil der Wirtschaftsinformatik. *Wirtschaftsinformatik*, 36(1):80–81, 1994.

Increasing Global Reach: Using Social Network Sites for Employer Branding

Franziska Brecht
Humboldt University, Berlin
Spandauer Str. 1
10178 Berlin, Germany
(+49) 30 2093 5730
Franziska.Brecht@wiwi.hu-berlin.de

Ksenia Koroleva
Humboldt University, Berlin
Spandauer Str. 1
10178 Berlin, Germany
(+49) 30 2093 1347
Koroleks@wiwi.hu-berlin.de

Oliver Guenther
Humboldt University, Berlin
Spandauer Str. 1
10178 Berlin, Germany
(+49) 30 2093 5743
Guenther@wiwi.hu-berlin.de

ABSTRACT

Employer branding by means of Social Network Sites (SNS) has been inspired by two recent developments: the increase of user-generated content with the spread of Web 2.0 and the lack of specialized human resources. However, this phenomenon is yet not very well understood: Which companies use SNS for employer branding? What are the motivations behind it and what are the necessary success strategies? In this paper, we uncover main goals companies pursue, strategies that are employed and the possible benefits that can be attained with an SNS presence. Our results are based on an evaluation of a quantitative survey and a qualitative analysis of interviews with company representatives. We find that, if an SNS presence is established strategically, companies can increase their global reach, improve employer brand awareness, achieve positive user bonding, gather valuable feedback as well as recruit for vacant positions.

Keywords

Social network sites, employer branding, recruitment

1. INTRODUCTION

Social Network Sites (SNS) became a mass phenomenon in 2005 when Facebook (FB) opened up to people outside the university network [6]. Today, FB claims to have over 500 million active users worldwide [8], making it the largest SNS. Its global orientation as well as heterogenous forms of use in private and professional contexts, make it a promising platform for companies and potential employees. The widespread usage of private SNS in a professional context began when FB users eventually launched their careers and started to add professional contacts [18]. On a B2B-level, however, there is still no globally used SNS. Instead, purely professional SNS have evolved within certain geographical boundaries: most of LinkedIn's 75 million members are located in English-speaking countries [12], whereas Xing is mainly popular in German-speaking countries [26].

Nowadays, companies are intensively using SNS for company branding and product branding [17]. However, only very few companies seem to apply SNS for employer branding, a term coined by Ambler and Barrow [2] and later completed by Barrow and Mosly [4]. They define employer branding as "… the package of functional, economic and psychological benefits provided by employment and identified with the employing company". Whereas company branding and product branding are usually realized by marketing departments, HR departments are in charge of employer branding. As these departments usually pursue different goals, they each require a separate SNS presence. The main role of employer branding is to " […] improve recruitment, retention and commitment" [4]. With the raging *war for talents* [16], characterized by the increasing lack of specialized workforce in advanced economies, it is becoming more important for companies to distinguish themselves and portray themselves as attractive employers in order to attract new talents and retain existing ones. Yet, the potential of employer branding via SNS is not fully exploited and needs to be further explored.

Therefore, the study at hand aims at investigating the state of deployment of SNS for employer branding by companies, and uncovering the specific goals and success strategies for building up an SNS presence. For these purposes, we analyzed a survey conducted by *squeaker.net* (an SNS specialized in online recruiting) concerning the adoption of SNS by companies, graduates and students. In order to gain deeper insights into motivations, we conducted eight interviews with professionals who carry recruiting responsibility in a major German company. Application of grounded theory to the gathered qualitative data resulted in the formulation of a conceptual framework of employer branding via SNS.

2. BACKGROUND AND RELATED WORK

Among the many existing definitions of SNS, the most suitable one for our purposes is the definition by Boyd and Ellison [6]. According to them, SNS are '*web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system*.' This means that SNS belong to the so-called "social media" applications. While the contacts to other users are essential for all SNS, other social media applications (e.g. video-platforms such as YouTube and online social games) concentrate on content and entertainment [9].

The functionalities of SNS and their user base vary considerably. The user's network can consist of friends and contacts via bi-directional confirmations, or of fans and followers via uni-directional confirmations, depending on the SNS respectively the type of account [5]. SNS generally support the maintenance of pre-existing social networks, whereas some even help strangers to connect. Communities are normally based on common interests, or on shared racial, sexual, cultural or other identities [6].

Professional SNS generally also provide a template to insert the user's CV, but mostly neglect personal preferences such as hobbies and favorite movies. A main feature of most SNS is the visualization of the user's direct connections to other users. This allows a user to approach strangers through a mutual acquaintance, a functionality that is particularly useful for recruiters [18]. However, professional networks are generally not suitable for students, since users have to give their current position and industry (ibid.).

Most research on the adoption of SNS is focused on private use, notably on issues such as the formation of *social capital*, a broad term reflecting the value arising from individual's relationships with others [7]. Its sub-category, the so-called B*ridging social capital* refers to potential benefits resulting from loose connections between heterogenous individuals, such as enhanced access to a broader set of material resources and exchange of useful information (ibid.). Applied to our context, increased interaction in SNS between companies and applicants may build up bridging social capital with potential benefits on both sides. The applicant can present herself to the company prior to the actual application process and obtain useful information for her application, whereas the company receives more applications from a wider range of candidates and gets a first impression of the applicants.

In the IS discipline, the influence of SNS in a professional context does not seem to be sufficiently explored. Research in this area is generally reduced to specific topics: e.g, Wirtz et al. [24] examine the implications of Web 2.0 and SNS for business models, challenges caused by privacy issues are addressed by Wang and Kobsa [22]. Some articles deal with the issue of company-internal communication via SNS, especially with interaction patterns as a signal to closeness between colleagues [25], or with the relationship between organizational social capital and the use of an internal SNS [20]. Thus the importance of the topic for IS seems to be recognized, but a comprehensive study on the goals and the impact of using SNS in a professional context is missing.

Moreover, very few articles specifically investigate employer branding and recruiting strategies of companies on SNS. Martin et al. [14], for example, deal with human resource management inside the company by means of SNS, whereas Backhaus [3] investigates the presentation of companies in online recruiting portals. Backhaus [3] finds that companies generally do not choose the information they are presenting in their online profiles carefully enough in order to attract applicants and distinguish themselves from competitors. Another important contribution is the study by Martin et al. [13], which confirms the importance of employer branding in general and discusses its impact. There are also specific studies concerning social media background checks: e.g., Kluemper and Rosen [10] confirm their usefulness for improved employment selection decisions based on the personal information of an applicant that is currently available in SNS; Skeels and Gruding [18] advise to be cautious when using information retrieved from SNS, as it may bias the selection decision. Here again we can note the increasing importance of SNS for the recruiting process.

However, no articles could be found on the specific topic of employer branding and recruiting by means of social media or more specifically by means of SNS. Our study intends to close this gap, and to contribute to IS research by conducting an exploratory study of company usage of SNS for employer branding. In order to uncover the importance of using SNS for employer branding, we aim to answer the following research questions: Why and how are companies using SNS for employer branding and recruiting? What benefits can be achieved?

3. METHODOLOGY

We used methodological triangulation to explore the adoption of SNS in employer branding (survey) and the motivations behind the usage of SNS (interviews), as indicated in Table 1.

Table 1. Methodology overview

Method	Sample	Goal	Analysis
Survey	(n= 409)	Determining adoption rate of SNS	Quantitative: descriptive statistics
Semi-structured interviews	Professionals with recruiting responsibility (n = 8)	Individual in-dept analysis	Qualitative: Grounded Theory

3.1 Survey

The invitations to take part in the survey were posted on FB, Twitter and sent via e-mail. Out of 588 participants, 409 completed all questions. The participants can be divided into three groups: students/graduates (76% of the survey participants), professionals without recruiting responsibility (11%), and professionals with recruiting responsibility (13%). The questions of the two latter groups (professionals) were slightly adapted for the first group (students/graduates). The questions concerned the adoption of SNS, social media background checks and the motivations behind the engagement on SNS. The adoption of SNS was evaluated based on the existence of at least one account in the following SNS: private SNS (FB, VZ-Networks, G Buzz), professional SNS (Xing, LinkedIn), specialised recruiting SNS (squeaker.net) or microblogging services (Twitter).

Almost all participants from the student/graduate sub-sample came from Germany, Austria or Switzerland (80%) and almost as many female (42%) as male (58%) survey participants replied. 67% of the survey participants were between 20-25 years old, 28% were between 26-30, and 5% were above 30 years old.

98% of all professionals (with or without recruiting responsibility) came from German-speaking countries; 73% were male and 27% were female. Most survey participants were between 26-30 (45%) and 31-40 (35%) years old. Only 5% of the professionals were younger than 25, and 15% were over 40 years old.

3.2 Interviews

In order to explore the company usage of SNS for employer branding, we used grounded theory due to its ability to analyze qualitative data systematically, to uncover the underlying relationships and to generate a theory based on them [21]. Our choice to use grounded theory was furthermore justified by the absence of systematic research on recruiting or employer branding via SNS. We pursued the 'Straussian' line of grounded theory, which requires an absence of an a-priori theory and emphasizes the usage of a paradigm [21]. In the process of constant comparison of emerging concepts against collected data, we inductively constructed a conceptual model of employer branding via SNS.

We conducted eight semi-structured in-depth interviews of 30 minutes with recruiters of several companies (for details, see Table 2). Because all recruiters were German native speakers, we conducted the interviews in German, as recommended by Wengraf [23]. The recruiters were asked questions such as: "*What role do SNS play for recruiting? What value can be attained by using SNS for recruiting?*". Two interviews were held with employees of companies doing social media consulting, which will be subsequently referred to as "experts" (in *italics* in Table 2).

All interviews were tape-recorded, transcribed and subsequently analyzed with the software tool "atlas.ti". The analysis was carried out in three distinct stages: open, axial and selective coding.

Table 2. Interviewees

Industry sector	Number of employees	Age of company (years)	Intensity of SNS usage
consulting (tax and audit)	40 (in Germany)	> 10	low
consulting (strategy and management)	14.500 (globally)	> 50	low
media	103.000 (globally)	> 130	high
online retailing	4100 (in Germany)	> 50	medium
electronics	116.000 (globally)	> 120	medium
rail transport	240.000 (globally)	> 15	high
consulting (social media)	*10 (in Germany)*	*> 5*	*expert*
consulting (social media)	*15 (in Germany)*	*> 5*	*expert*

During open coding, categories and properties are identified by looking for patterns in the data and comparing them with each other [21]. Accordingly, we identified 58 categories in total, each possessing at least one property and respective dimensions. To illustrate the process of open coding, consider the following example from one of the interviews: "*As regards XING we want to create an account* (category: build up presence, property: social network, dimension: professional), *but have not done it yet* (category: adoption rate of SNS, property: usage, dimension: intention). *We currently use it in order to find someone in the pool* (category: looking for people, property: source, dimension: wide), *who is interesting to us as a potential employee* (category: recruiting, property: certainty of outcome, dimension: moderate)" (Q).

During axial coding, the emerging categories were grouped into families by uncovering the relationships between them. The paradigm model by Strauss and Corbin [21], which includes causal and intervening conditions, strategies and actions, as well as consequences, served as a basis to identify these relationships. The process of selective coding allowed us to focus on the most relevant categories to describe the attitude towards employer branding via SNS. The result of our analysis, the conceptual framework (Figure 2), describes the context in which employer branding occurs on SNS.

4. RESULTS

Our interviews and our survey data confirm that companies increasingly use SNS for employer branding and recruiting. As one of the interviewees puts it: "*In comparison to the year 2000, online recruiting has become more important in our company: be it on professional networks such as XING, or on private networks such as Facebook, Twitter, …*" (Q). Our survey confirms these results: 53% of the participating companies stated that they already use SNS for employer branding and recruiting, or that they are planning to do so (see Figure 1). Xing and FB were named as the most commonly employed SNS, with adoption rates of 95% and 83%, respectively. Twitter (77%) was more popular than LinkedIn (58%), which we found quite surprising. The relatively low adoption rate of LinkedIn in comparison to Xing could be explained by the dominance of Xing in German-speaking countries, where 98% of the surveyed companies come from. At the same time, 99% of the surveyed students/graduates use at least one SNS. Whereas FB has similar adoption rates for companies and students/graduates, the adoption rates for Xing differ with 95% (professionals) and 57% (students/graduates) respectively. Here we can note that students tend to avoid using professional SNS, which is confirmed by the adoption rate of LinkedIn: whereas 58% of companies use this network, only 26% of students/graduates do. Again, the adoption rate for Twitter seems surprising: only 29% of the students/graduates report using it.

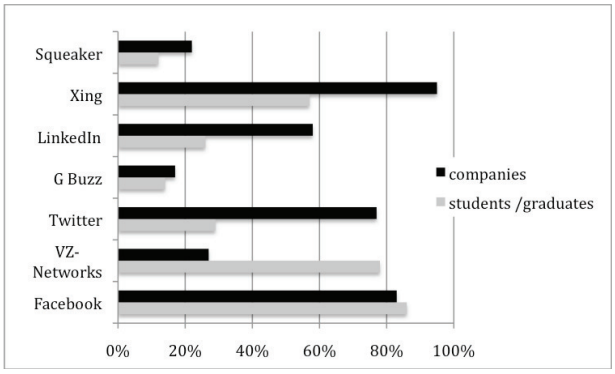


Figure 1. Adoption rate of SNS

Based on extensive qualitative data analysis, we formulate a conceptual model of employer branding and recruiting via SNS (Figure 2). The model aims to reflect the overall attitude of companies towards the adoption of SNS for employer branding and recruiting. The model differentiates between the strategic and operational *goals* that companies pursue on SNS; an *external environment* that can have an impact on the goals; the *strategies and actions* they need to undertake in order to achieve the goals; various short-term and long-term *consequences* as a result of implemented strategies; as well as *intervening conditions* that can either have an amplifying or softening effect on the strategies and/or consequences. The causal relationships (represented by the arrows in Figure 2) in the model are rather generic. They show

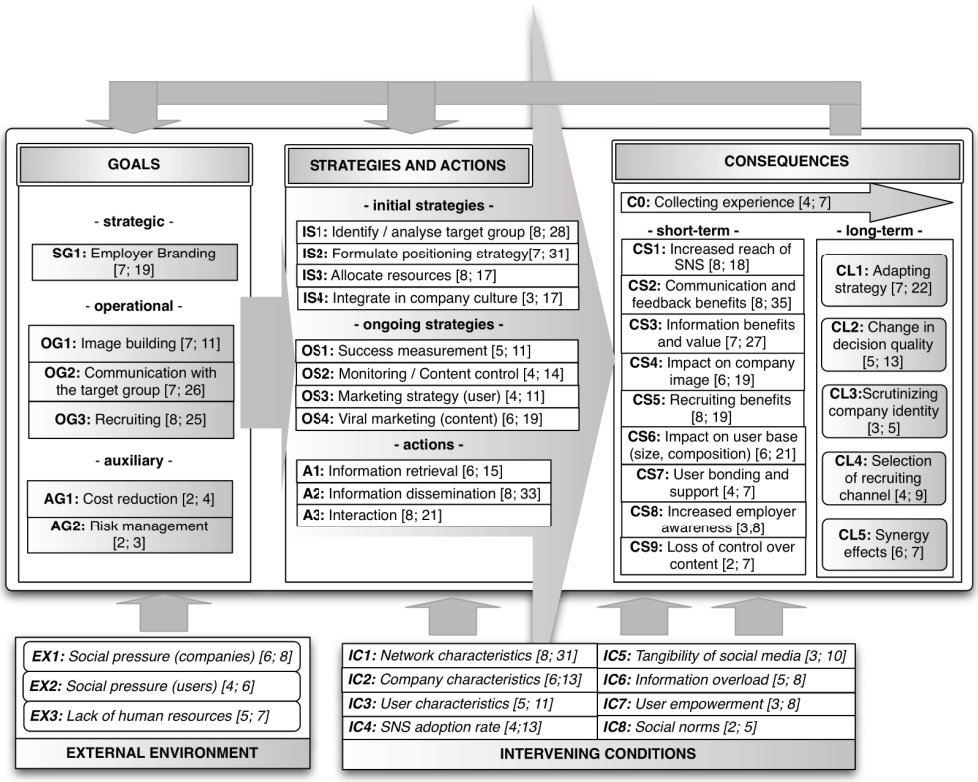


Figure 2. Conceptual Model

that the choice of strategies depends on the aspired goals, possibly influenced by the external environment. The realization of strategies, in turn, can lead to certain consequences, which possibly can be mediated by intervening conditions. The model does not provide specific recommendations, i.e. that a certain strategy necessarily leads to a certain consequence or is carried out in response to a certain goal, but rather shows possible goal-strategy-consequence combinations. The model is circular, reflecting the adjustment behavior of companies: the adaptation of strategies in the short-term and goals in the long-term in response to certain consequences.

The relative importance of each construct in the overall model can be assessed by the number of times respondents mentioned them. The first number in square brackets next to each category in Figure 2 reflects the number of respondents mentioning this category (max. 8), whereas the second one reflects the number of times this category was mentioned in all interviews. In the following sections we describe the major components of the model (goals, strategies and actions, consequences) in detail. Intervening conditions and external environment are considered across the sections. Additionally, as we progress to describe model elements, we provide examples to illustrate the relationships between the model components. The reader can trace the model elements in the text with the help of the abbreviations deinfed in Figure 2.

Goals

Company motivations behind initiating an SNS presence differ in their strategic orientation. Employer branding is usually the overarching goal, whereas image building, communication with the target group and recruiting are subordinate goals of an SNS presence. *Employer branding* (SG1) refers to positioning oneself as an employer for a broad target group as opposed to *recruiting* (OG3) a specific candidate: "*If I say recruiting, that means I want the person XY for this or that position and contact him directly. And employer branding is to generally address a specific target group, say computer scientists, because we have a lot of positions of the kind, and to make them aware of us as an employer*" (Q). The lack of specialized workforces (EX3) intensifies the desire to employ SNS for recruiting and employer branding: *"I would say that social networks are complementary, especially for positions that are not easily found on the general market" (Q).*

On the operational level, employer branding can help build up the company image (OG1): *"We want to be present on these networks, so that many potential candidates know us and think about us … in order to increase the reach of our employer brand" (Q).* Moreover, SNS offer a unique possibility for companies to reach a specific target group and initiate a dialogue with it (OG2): *"We want to create a page on Facebook, so that we open up as a company and have a possibility to communicate with our target group" (Q).* In addition to the strategic and operational goals, companies are motivated to reduce their costs (AG1) for the search of appropriate employees by entering SNS: *"...* [via SNS] *we want to reach more people for less money and reach more people from those that you would like to reach with less cost" (Q).*

Other reasons for maintaining an SNS presence concern external factors, most importantly *social pressure from other companies* (EX1), as well as from users. As to the former, other companies such as suppliers (upstream), B2B-customers (downstream), or companies being in the same interest group can exert a certain pressure. This may encourage the company to imitate and create an SNS presence without having any particular goal in mind: "*We*

are in many employer branding interest groups in Germany in which social network presences are debated a lot, and we decided to give it a try and see how it works" (Q). From a user or a (potential) employee perspective, company characteristics such as the company's industry (IC2) can make an SNS presence of a company a must: "*If I were an applicant I would expect from a media company to be present in these networks and to provide a lot of videos and so on*" (Q).

Another reason for being present on SNS is the response of companies to the growth of user-generated content on the Web. Users increasingly engage in discussions about companies, not least about their (justified or unjustified) negative aspects. In order to avoid scandals and damage to a company's image, it has to be able to react to critical content, which can be done by establishing an SNS presence. For example, quite recently a well-known sweets manufacturer felt compelled to create an online presence, after a popular user-generated video on YouTube accused them of environmental crimes [1]. Thus, in response to *social pressure from users* (EX2), companies are building up their online presence in order to control the dissemination of information and to minimize future risks: *"After such a scandal or a crisis, many companies realize that they have to become active in these networks, in order to prevent such things from happening in the future" (Q)*. Thus, an SNS presence can aid the company in its risk-management strategy (AG2): *"If a scandal like this happens, we will be ready for this and use SNS for communication and risk mitigation" (Q)*. Establishing such a presence, a company can interfere with public Web 2.0 opinion, as well as the public opinion in general, in a credible manner.

Strategies and Actions

In order to achieve the above-mentioned goals, companies employ a variety of strategies supported by a set of actions that can be carried out on SNS. The strategies defined in our model can be subdivided into those that are usually done prior to an SNS entry (initial) and those that should be carried out continuously (ongoing).

4.1.1 Initial Strategies

A purposeful SNS entry normally starts with the *identification and analysis of the target group* (IS1). Quite often, this results in a focus on users characterized by their technical, business and/or Internet affinity: "*Our target group is heterogeneous concerning age or level of experience, but rather homogeneous concerning their technical affinity: we want to reach people who linger in these networks, who have a certain connection with the medium*" (Q). However, the global reach of many SNS provides a unique possibility to reach other target groups, such as people possessing rare skills, who are passively willing to change their job or who will serve as multipliers for those who are in search of a new position: "*... other companies and industries can present themselves successfully and recruit through social media*" (Q). Thus, defining a target group can help achieve a company's goals, for example *recruiting* (OG3) for certain positions that are difficult to staff otherwise: "*The general mass market is only conditionally interesting, as we do not have recruiting problems for, say, clerical positions, but we are interested in people who are able to take on responsibility from the start*" (Q), helping to combat the lack of human resources in some areas (EX3).

Formulating a positioning strategy (IS2) is crucial in order to create a purposeful presence on SNS and includes many facets and principles. First and foremost, companies need to be authentic in defining their brand image and support this image by actions. Otherwise they run the risk of facing undesirable consequences: "*There are companies that don't succeed, because they try to position themselves in the same way as they do in traditional media - by having a classical brand image, for example: BMW - Sheer Driving Pleasure. In social media they better be 'Sheer Driving Pleasure' otherwise they would be taken apart by their audience*" (Q). User empowerment (IC7) is a peculiar characteristic of many SNS that renders authentic communication necessary. Using a *trial-and-error* 'strategy' when building up the SNS presence, i.e. not connecting goals to strategies, can endanger the entire SNS presence: "*...Twitter is hype, so we should also do something on Twitter, without any connection to our topics or any other online presences - then this would be without results, and the presence dies away*" (Q).

At the same time, any SNS presence needs a careful *allocation of resources* (IS3) and a clear assignment of responsibilities for initiation and maintenance: "*There are three people responsible for Twitter, and if users access our account, they can immediately see who they are dealing with*" (Q). The motivation behind naming particular persons is the principle of personality, which aims at presenting the company as a distinct group of people that users can identify with. Companies report that considerable time investments are necessary for SNS maintenance, ranging from several minutes to hours a day, with posting generally requiring less resources than monitoring and reacting: "*Everyone thinks that social media is totally cheap, because it does not cost anything. But it costs a lot of time, and pays off only because it creates attention and gives us a unique chance to initiate a dialogue with potential employees that we don't have otherwise*" (Q). This example clearly illustrates the relationship between goals, strategies and consequences: if companies allocate sufficient resources (IS3), they can receive feedback (CS2) and increase the company image (CS4) with the aim to enhance communication with their target group (OG2).

The *integration of an SNS presence into a company's culture* (IS4) is very important as well. A bridge between current employees and the desired target group can be build by promoting SNS adoption within the company: "*There are representatives of this target group in the company, and no big effort is necessary to convince people to take part in social media, you just have to find a right match between people sharing the same interests*" (Q). This "candidate-to-colleague" approach supports the principles of personality and authenticity and helps to disseminate the necessary content and to decentralize the maintenance process. Thus, this approach can lead to increased communication (CS2) and user bonding (CS7) for those striving to build up their company image (OG1): "*Personnel 2.0 means finding the candidates and linking them to employees, and through this personal contact creating a bond that has a positive impact on the company*" (Q). However, this is rather a long-term process requiring a change in the entire company regarding perception and adoption of SNS (IC4).

4.1.2 Ongoing Strategies

In the short-term, ongoing strategies are adopted in order to reach certain goals. This includes identifying the appropriate success measures from the start, and continuously *measuring performance*

(OS1) in order to justify the invested resources. Defining quantitative success measures helps to achieve desired outcomes, such as an increase in the user base (CS6), recruiting success (CS5) or image enhancement (CS4): "*Be it recruiting or image, the more quantitative results we provide, the more we can say that we employed someone via social media, the more we can justify the time we invested*" (Q). However, the success of an SNS presence is difficult to measure quantitatively, due to intangibility of social media (IC5) and the inability to directly attribute a certain consequence solely to the SNS: "*… maybe people were attracted to our company through SNS, but have placed their application through the website, and thus the direct traceability is really hard to implement*" (Q). Recruiting (OG3), however, is usually just one of the pursued goals of a company on SNS and thus qualitative success metrics, that measure the achievement of more intangible goals, such as image building (OG1), should be considered as well: "*My goal is not only to recruit, but to spread information in order to make the company better known as an employer. Thus, each article that is written about us can already be considered a success*" (Q).

Our findings suggest that if companies spend significant resources on presenting themselves on SNS, they should invest as much to *monitor* (OS2) the developments on SNS. Monitoring should be carried out continuously. However, content can only be controlled to a limited degree, as these manipulations are restricted by the ability of users to verify the content (IC7) and may lead to undesired consequences: "*Something that you can manipulate when you are in social media is the content of the message and the time of posting - but both have to be authentic, otherwise everything can blow up*" (Q). Thus, the companies should try to disseminate information respecting the principles of authenticity and constistency, i.e. present themselves in a positive light, but at the same time be transparent about negative aspects and try to mitigate them in an open dialogue with the users. This can help to achieve user bonding (CS7), which usually comes as an unexpected consequence for the company: "*Before, users were ready to tear the company apart if it admitted that it made a mistake. Nowadays if companies accept their problems, users can't help reacting with a certain degree of understanding*" (Q).

According to our survey results, 75% of the students/graduates stated that they already have noticed information about employers on SNS. Out of these, 62% of the students/graduates found this information useful. Moreover, we asked about the kind of information that is posted by companies (offer of information), and which information potential applicants would find interesting (demand of information), as indicated in Figure 3. As expected, only very few companies (2%) were said to provide information concerning salaries, but almost 40% of the graduates/students would find such kind of information useful. Another mismatch between offer and demand was revealed with regard to experience reports from insiders (23% offer vs. 52% demand) and insights about typical jobs (33% vs. 46%). In general, the information desired the most often was about job postings (63%), which could be an indication for the migration of online recruiting services to SNS in the future.

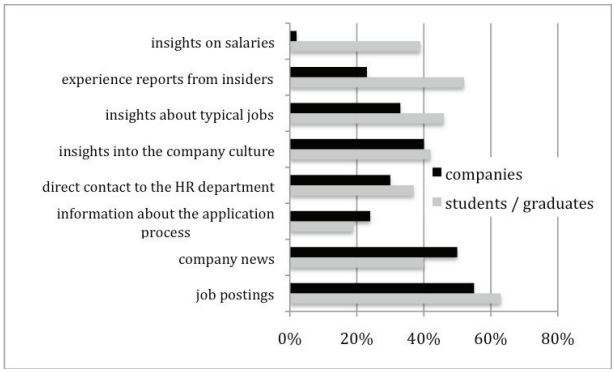


Figure 3. Offer and demand of specific content on SNS

For successful employer branding, companies engage in *marketing campaigns* (OS3) in order to promote their SNS presence and generate an initial audience. High concentration on such campaigns, however, can overshadow some of the operational goals, such as recruiting (OG3): "*We are not yet satisfied with the number of followers and we would like to double them through a campaign, because the goal is to have 700 until the end of the year*" (Q). The question arises whether this 'blind' generation of followers helps to achieve the desired goals, as the number of followers itself does not directly translate to the number of recruited employees (CS5), nor does it indicate that users are attentive to the posted content (CS2): "*There are companies that sell 'likers', that means they guarantee that the page will have a certain amount of 'likers', by presenting the required target group with the necessary advertisements on the network*" (Q).

Another peculiarity of SNS lies in the dynamics of its content as opposed to a static 'classical' website. This peculiarity is exploited by the strategy of *viral marketing* (OS4), which is employed by companies in SNS in order to increase the reach of their content (CS1): "*Today we told 14 people, and they will tell it to their friends. And it works pretty well, because I don't know where else I can post it so that everyone finds out about it*" (Q). Thus, the aim should not only be to reach the target group directly and generate large follower numbers, but also to try to generate good 'transmitters' in the follower base. For example, employment companies, university professors and online marketplaces could act as multipliers: "*From our perspective, Twitter is mostly necessary just to generate traffic on other platforms*" (Q).

As networks of users grow to include new friends and fan pages, *information overload* (IC6) has been found to decrease attention levels and activity on the network [11]. However, the presence of a so-called 'social filter' inherent in SNS helps to reduce information overload, in that it usually delivers the information to the required recipient, without much effort: "*What we do is just throw in a lot of job offers into the network and hope that one friend of a friend of a friend knows someone else who is looking for a job like this one and recommends it to him. And when the job reaches its recipient, and that is exactly the job he was looking for, then it was worth it*" (Q). Thus, the personal character of the message, and the ability of users to know the preferences of their friends better than any other marketing tool, increases the relevance of information and aids companies to achieve their goals.

4.1.3 Actions

The exact actions that companies usually undertake in order to implement the strategies described in our model depend on the functionality of the network (IC1). They can generally be subdivided into three levels, each of which describes a current status of adoption of SNS (Figure 4).

On the first level, *information retrieval* (A1) can deliver valuable insights about specific users, general opinions, current topics or market trends. This can be achieved by regularly monitoring ongoing discussions, by browsing the pages of other companies or users, as well as searching directly for specific users or types of information. One of the most commonly retrieved data from SNS is additional information about potential employees, known as 'background check': "*... some party pictures somewhere on Facebook – that's not interesting for me, I don't have the time to look at them, but with regard to persons who want to work in the social media context we do search for specific indicators of medium affinity - whether they are writing a blog or something*" (Q). Our survey confirms this development: 70% of the professionals confirmed having checked information about applicants on the Internet already. Applicants seem to have adjusted to this trend. 65% of the students/graduates hesitate when posting personal data on SNS, fearing that potential employers could find it. This concern is confirmed by Kluemper and Rosen [10], who state that one can reliably and accurately evaluate personality, intelligence and performance based on information available through SNS.

On the second level, companies can *disseminate desired information* (A2) by means of posting own content, reposting content from someone else, or banning certain content in an attempt to control the information flow. This requires identifying content, frequency and channel for every message.

On the third level, *interaction* (A3) on SNS usually includes discussing certain topics, reacting to questions or directly contacting candidates. This bi-directional communication with users, however, still remains a goal of many companies: "*I have a feeling that this dialogue between users and companies is not yet expected at the moment. I go to the fan page of the company to discuss things with other users, rather than with the company itself*" (Q). The efforts of companies should be directed towards encouraging the dialogue with users and potential employees, as it helps to gather valuable feedback (CS2), get to know their audience better (CS3) and create considerable bonding ties (CS7).

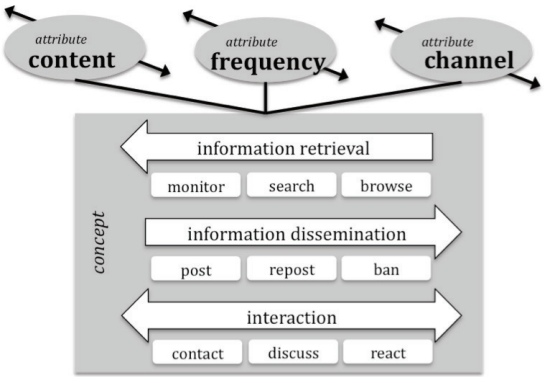


Figure 4. Categorization of possible actions on SNS

Consequences

When companies implement certain strategies, this can lead to the consequences depicted in our framework. These consequences can be subdivided into short-term and long-term ones, both of which can have a recurring impact on the goals. The former usually only have a considerable impact on the adaptation of SNS strategies and thus help achieve goals on the operational level. The latter usually only emerge after a certain time and experience with SNS. They have an impact not only on the presence strategy, but rather on the company itself and thus help fulfill the strategic goals. The variety of effects of a social media presence, which can translate into benefits of SNS under specific circumstances, are presented in the table below.

Table 3. Consequences

Short-term	Communi-cation and feedback benefits	(feedback) "*I ask these questions on Facebook, for example whether people want to see job offers or whether they say it is not the right time...*(A1) *and we are happy to get this feedback* (CS2)*, as our goal on the network is communication* (OG2)." (communication) "*This one-to-one dialogue will take place more often as opposed to this one-to-many, which was done by companies through other means before*."
	Information benefits and information value	(user perspective) "*… they can understand what kind of employer we are, how we work and who works here, what we do every day and how we communicate with each other… - exactly what is interesting for those looking for a job*." (company perspective) "*We can understand the applicant's expectations in relation to the job he is looking for, whether he is open to get job offers and the like*."
	Impact on user base (size and compo-sition)	(composition) "*Who is following us? With a high probability, these are other recruiters or companies and just 10% from the real target group. You stay among yourselves*" (size) "*And our Facebook community is growing, in October we had just about 1000 people, and now they are 2000 and I have a feeling that it grows exponentially*."
	Impact on company image	"*… this open communication and authenticity makes us more tangible and yet more vulnerable, and leads to a sense of presence and a sense of understanding among users of what this company really is*"
	Increased reach of media	"*Facebook has certainly the highest reach because of the hype and the amount of people on that network*."
	Recruiting benefits	"*We get more applications from the candidates we are interested in*."
	User bonding and support	"*To communicate with the target group* (OG2) *we connect the candidates with our colleagues and through this personal connection* (IS4) *we create a kind of bonding, a positive stimulus towards the company* (CS7)."

	Loss of control over content	*"Those are the people who are afraid to loose control … I send a message 20.000 times via advertising - and it will be accepted or not, but it does not change as opposed to social media."*
	Increased attention	*"The difference is whether 100 applications come in online or we have just two people that contacted us through Twitter, they are more present, they stand out."*
Long-term	Adapting strategy	*"If I understand that what I post is not accepted, then I can either not care about it or try to change it."*
	Change in decision quality	*"Yes, with Twitter it can be that someone is very active and posts smart things and thus it has a positive influence on our decision."*
	Selection of recruiting channel	*"What is qualitatively better with social media is the first contact, but after that the recruiting process does not change much."*
	Scrutinizing company identity	*"Today you have to define what is our value proposition and what distinguishes us from others in the eyes of the candidate."*
	Synergy effects	*"I think there is a convergence-trend, that everything grows together, and I wonder whether in 5 years we will need a separate career Webpage."*

One of the major advantages of SNS as opposed to other means of information dissemination is the personal aspect that is involved in the transmission process, which usually results in increased information value (CS3) for the participant: "*The big advantage of social media is that it automatically filters information, finding out what is relevant for me from all the things offered there, … irrespective of the incoming amount of information, the information that I really need will reach me*" (Q). This is quite intuitive, as the information received through a friend or via any other type of personal connection is usually regarded the most valuable and the most trusted. In this way, social norms (IC8) prevailing on the networks can increase the quality of information: "*And if someone I trust in tells me that this is important, I totally believe it, even though this information might not have been as important to me before*" (Q). Thus, what matters on SNS for successful employer branding (SG1) is not only the reach of media (CS1), but its personal value (CS3). Companies can easily exploit this property by engaging in viral marketing campaigns (OS4). Thus, SNS allow for a situation where the information is sent and received without much effort, but with a higher relevance than in traditional recruiting or marketing channels.

Our results show that the realization of one consequence can dependent upon another one: "*We have noticed through discussions with other companies that interaction usually only starts* (CS2) *after they have attracted a certain amount of likers* (CS6)" (Q). Moreover, short-term consequences can lead to the long-term ones as time goes by. For example, companies can adapt their presence strategy (CL1, IS2) in response to obtained feedback (CS2): "*... potential candidates can react by saying that they find it interesting or they find it boring, and based on this we can adapt our posting strategy*" (Q). At the same time, recruiting successes (CS5) can determine the future selection of the recruiting channel (CL4) and favor SNS over other means: "*This is an additional channel, which we use as much as it allows us to come in contact with more candidates we are looking for*" (Q).

This strengthens the gradual development and the circularity of the proposed conceptual model that allows to *accumulate the experience with SNS* (C0) as time goes by and gradually establish themselves as an employer on SNS: "*…we need to see what works well and then adjust the strategy accordingly over time*" (Q). This long-term orientation concerns in particular certain types of goals, such as encouraging interaction and communication with the target group (OG2), as it requires certain changes in the perception of the users and depends on the adoption rate of social media (I4): "*This dialogue orientation is not there yet, both from the side of companies and users, and I think it needs generations of applicants that have been brought up with and can act naturally in this set-up…*" (Q).

The success of the implemented strategies, i.e. the consequences, can be influenced by intervening conditions. For example, network functionality (IC1) can help to achieve synergy effects (CL5) between the networks: "*Things are posted on Twitter, and then they are discussed on Facebook, as the functionality of Twitter allows only for a limited number of words, but has the advantage of a wider audience* " (Q). Additionally, intervening conditions can have an indirect impact on the achievement of certain goals. For example, smaller companies get an equal possibility to present themselves on SNS (OG1) and thus gain even more in terms of image increases (CS4) as compared to their big competitors: "*Through these social media projects, the smaller and less known companies got the possibility to better position themselves*" (Q).

SNS can also cause unexpected consequences in the long-term. For example, in order to present themselves authentically and distinguishably from others (OG1), many companies are forced to incur many changes in order to adjust their company identity: "*Another advantage for the companies is the possibility to find and re-define themselves* (CL4). *In an attempt to formulate a unique value proposition the companies are having a hard time to admit their problems in certain areas and making decisions to change many things in the company, before presenting it to the general public*" (Q). Thus, employer branding via SNS can scrutinize company identity (CL4) and encourage sustainable improvement.

5. MANAGERIAL IMPLICATIONS

In order to build up their presence on SNS, we recommend companies to formulate an 'SNS enter strategy' that consists of two steps, as shown in our conceptual model. The first step (initial strategies, company-internal) includes defining the goals of the SNS presence, determining the amount of resources spent, allocating responsibilities to employees, identifying employees matching the desired applicant ('colleagues-to-candidate' approach), and mobilising them to actively contribute to the SNS, possibly by installing an incentive system.

Companies should carefully choose the SNS where they want to be present. This should be done according to the target group that they want to address, as very specialized candidates are as difficult to find in SNS as through traditional media channels. Thus, if a company is targeting young people (students and graduates), they should consider building up presences in private SNS such as FB in the first place, as young people are mostly active in private SNS (cf. Figure 1). Companies targeting more experienced applicants should predominantly focus on a

"colleague-to-candidate" approach in professional SNS such as XING.

In the second step (ongoing strategies, company-external), the company in question should first monitor the Web 2.0, i.e. it should analyze its existing image in social media in order to find out how to reach its target group (where are they active, which topics are relevant, which questions are discussed). This allows for slowly building up a long-term presence on SNS, disseminating desired information and initiating a dialogue with its target group (see figure 4 for possible actions in SNS).

In order to be successful, SNS strategies should take into consideration many parameters, reflected in our model by the intervening conditions. First, a company should address different strategic questions related to their SNS presence, such as: How experienced is the company in dealing with social media? Are there strategies employed for company or product branding, and should they be aligned with the employer branding strategy? Second, the company characteristics should be considered: the type of business model, the industry and size of the company, which can exert a significant influence on the success strategy as well. It can be suggested that employer branding is generally more effective for companies in consumer industries, such as food or electronics, as their clients are already familiar with their products via general marketing channels. Moreover, smaller companies can achieve a higher leverage in reaching their target groups as compared to traditional media channels. Compared to bigger companies, they might lack the resources to maintain an SNS presence, but at the same time may profit from easily decentralizing the maintenance process.

Moreover, it seems advisable to keep in mind the principles that support their SNS presence, as mentioned above: authenticity (content should be truthful), actuality (content should be regularly updated), consistency (content should be reliable), personality (certain persons should be responsible for maintenance) and peculiarity (companies should define an exclusive value proposition). Similarly, companies should not forget about the dynamics inherent in SNS or the possibility of the content to develop uncontrollably, and find ways to monitor it continuously, as well as interfere.

Our study shows that building and maintaining SNS presences requires a lot of experience. This causes an increased demand for social media consulting services. Companies who consider using SNS for employer branding should carefully deliberate about whether to get consulting or not, as an inability to establish a successful presence may cause damage to the company's image. In any case, SNS presence strategies should be defined before building up SNS presences.

6. CONCLUSION

This study identifies main goals companies pursue for their SNS presence, strategies they apply for employer branding, and possible consequences of establishing themselves on SNS. Based on the analysis of a survey and qualitative in-depth interviews with company representatives, we found that most of our interviewees already use SNS for employer branding or are planning to do so. The global reach of many SNS attracts companies, as various target groups can be found in these networks. Furthermore, we found that many companies do not have explicit strategies for employer branding in SNS and create online presences just for the sake of 'trying out'. However, if companies expertly follow strategies for building and maintaining SNS presences, they could reach their target groups even more efficiently. If companies succeed in this, they can increase their employer brand awareness, retrieve valuable feedback, establish connections with potential employees, as well as successfully keep existing staff and recruit new employees.

7. LIMITATIONS

In the field of qualitative data analysis, especially in the analysis of interviews, results can be biased. This may be due to the choice of questions, their formulation, or the interpretation of the interviewee's answer by the researcher. Although we analyzed the interviews independently, and formulated the interview questions as open and as careful as possible, we assume that we could not eliminate this bias completely.

The translation can be a cause for further biases. The interviews were conducted in German and had to be translated in order to be presented in this study. The choice of the mother tongue for conducting interviews is justified by Wengraf [23].

Moreover, the numbers in our conceptual model (Figure 2) that indicate the frequency of occurrence of a particular category may be strongly biased. Nevertheless, we indicated them as they can provide additional insights on the relative importance of a particular category in the overall conceptual model. We also report numbers indicating the number of respondents who mentioned a respective category, which can be considered a more reliable indication of importance as they avoid double-counts.

A weakness inherent in qualitative data analysis is the limited sample size. Our interviewees are from a small number of German companies in specific B2C industries. Furthermore, six out of eight interviews were done with only one company representative, whose statements do not necessarily have to be representative for the company.

8. FUTURE WORK

We aim to conduct further interviews in order to extrapolate the insights gained in this study. Moreover, the follow-up study will explore the user perspective in adoption of SNS: motivations behind SNS usage in a professional context and expectations users have regarding SNS presences of companies.

Future research could provide qualitative success measures and parameters to assess the impact of an SNS presence on employer brand awareness. Other interesting topics include the alignment of company-internal and company-external employer branding as well as the interaction between company branding, product branding and employer branding.

9. ACKNOWLEDGMENTS

We would like to thank Stefan Menden from squeaker.net, Marie-Luise Klaus and the company representatives who kindly agreed to be interviewed. This research was funded by the German Federal Ministry of Education and Research under grant number 01IA08001E as part of the Aletheia project (http://www.aletheia-projekt.de/). The responsibility for this publication lies with the authors.

10. REFERENCES

[1] Allfacebook.com. 2010. The Facebook Nestle Mess: When Social Media Goes Anti-Social. http://www.allfacebook.com/the-facebook-nestle-mess-when-social-media-goes-anti-social-2010-03, retrieved on August 15, 2010.

[2] Ambler, T. and Barrow, S. 1996. The employer brand. *Journal of Brand Management* 4, 185-206.

[3] Backhaus, K. 2004. An exploration of corporate recruitment descriptions on Monster.com. *Journal of Business Communication* 41, 2, 115-136.

[4] Barrow, S. and Mosley, R. 2005. The Employer Brand. John Wiley & Sons, Chichester.

[5] Boyd, D. M. 2006. Friends, Friendsters, and MySpace Top 8: Writing Community Into Being on Social Network Sites. *First Monday* 11, 12 (Dec. 2006) http://www.firstmonday.org/issues/issue11_12/boyd/index.html

[6] Boyd, D. M. and Ellison, N. B. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 1. http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[7] Ellison, N.B., Steinfield, C. and Lampe, C. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 12, 4.

[8] Facebook. 2010. Facebook.com Statistics. http://www.facebook.com/press/info.php?statistics, retrieved on August 16, 2010.

[9] Kaplan, A. M. and Haenlein, M. 2010. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons* 53, 1, 59-68.

[10] Kluemper, D. H. and Rosen, P. A. 2009. Future employment selection methods: evaluating social networking web sites. *Journal of Managerial Psychology* 24, 6, 567-580.

[11] Koroleva, K., Krasnova, H. and Guenther, O. 2010., 'STOP SPAMMING ME!' - Exploring Information Overload on Facebook. In *Proceedings of the 16th Americas Conference on Information Systems* (Lima, Peru, August 12-15, 2010) AMCIS 2010. http://aisel.aisnet.org/amcis2010/447

[12] LinkedIn. 2010. LinkedIn About. http://press.linkedin.com/about, retrieved on August 15, 2010.

[13] Martin, G., Beaumont, P., Doig, R. and Pate, J. 2005. Branding: A New Performance Discourse for HR? European Journal of Management 23, 1 (Feb. 2005), 76-88.

[14] Martin, G., Reddington, M., Kneafsey, M.B., and Sloman, M. 2009. Scenarios and strategies for Web 2.0. *Education + Training* 51, 5/6, 370-380.

[15] Matavire, R. and Brown, I. 2008. Investigating the use of "Grounded Theory" in information systems research. In *Proceedings of the 2008 Annual Research Conference of the South African institute of Computer Scientists and information Technologists on IT Research in Developing Countries: Riding the Wave of Technology* (Wilderness, South Africa, October 06 - 08, 2008). SAICSIT '08, vol. 338. ACM, New York, NY, 139-147. DOI= http://doi.acm.org/10.1145/1456659.1456676

[16] Michaels, E., Handfield, J. and Axelrod, B. 1997. The War for Talent. Havard Business School Press, Boston.

[17] Nielsen. 2009. Global Faces and Networked Places. http://blog.nielsen.com/nielsenwire/global/social-networking-new-global-footprint/, retrieved on August 15, 2010.

[18] Skeels, M. M. and Grudin, J. 2009. When Social Networks Cross Boundaries: A Case Study of Workplace Use of Facebook and LinkedIn. In *Proceedings of Group* (Sanibel Island, Florida, USA, May 10-13, 2009). GROUP'09.

[19] Smith, W. P, and Kidder, D. L. 2010. You've been tagged! (Then again, maybe not): Employers and Facebook. *Business Horizons* (forthcoming).

[20] Steinfield, C., DiMicco, J. M., Ellison, N. B., and Lampe, C. 2009. Bowling online: social networking and social capital within the organization. In *Proceedings of the Fourth international Conference on Communities and Technologies* (University Park, PA, USA, June 25 - 27, 2009). C&T '09. ACM, New York, NY, 245-254. DOI= http://doi.acm.org/10.1145/1556460.1556496

[21] Strauss, A. and Corbin, J. 1998. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Sage Publications Inc.

[22] Wang, Y. and Kobsa, A. 2009. Privacy in Online Social Networking at Workplace. In *Proceedings of the IEEE Int'l Conference on Computational Science and Engineering* (Vancouver, Canada, August 29-31, 2009), 975-978.

[23] Wengraf, T. 2001. Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods. Sage Publication, Thousands Oaks, CA.

[24] Wirtz, B.W., Schilke, O. and Ullrich, S. Strategic Development of Business Models: Implications of the Web 2.0 for Creating Value on the Internet. *Long Range Planning* 43, 2-3, 272-290. DOI= http://portal.acm.org/citation.cfm?doid=10.1016/j.lrp.2010.01.005

[25] Wu, A., DiMicco, J. M., and Millen, D. R. 2010. Detecting professional versus personal closeness using an enterprise social network site. In *Proceedings of the 28th international Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA, April 10 - 15, 2010). CHI '10. ACM, New York, NY, 1955-1964. DOI= http://doi.acm.org/10.1145/1753326.1753622

[26] Xing. 2010. Xing Press Release. http://corporate.xing.com/index.php?id=112&tx_ttnews[tt_news]=1144&tx_ttnews[pointer]=0&tx_ttnews[backPid]=0&cHash=2&cat=0&L=1, retrieved on August 10, 2010.

Semantically Reconnecting Fragmented Information through User Activity Monitoring

Hinnerk Brügmann
Institute of Information Systems II
University of Erlangen-Nuremberg
Nuremberg, Germany
hinnerk.bruegmann@wiso.uni-erlangen.de

Matthias Kurz Institute of Information Systems II
University of Erlangen-Nuremberg
Nuremberg, Germany
matthias.kurz@wiso.uni-erlangen.de

Dieter Sauer Institute of Information Systems II
University of Erlangen-Nuremberg
Nuremberg, Germany

ABSTRACT

Today information items on user's workstations are usually stored in separate collections depending on their format. This results in a disconnect between information systems and user needs leading to high lookup times during task related information retrieval. This paper presents an approach to reduce document based information fragmentation by semantically reconnecting electronic documents to each other without imposing additional training or tagging workload on the user. To this end the actions knowledge workers perform on their desktop are transparently monitored to analyze the user's interaction with his computer system. These action metadata are further clustered by superordinate activities performed by the user. Finally documents attached to window instances within the identified activity clusters are semantically to each other related reducing the fragmentation of their contained information. This allows a subsequent associative information discovery navigating from one document instance to other related document instances. A prototypical implementation and evaluation in a small scale testing setup indicates the validity of the approach.

1. INTRODUCTION

Due to the high importance computer-mediated communication and IT systems have in the execution of today's business processes, enterprises are a main driver in the overall growth of information by generating a rapidly increasing amount of information as by-product of their business activities [38]. Lyman and Varian estimate the annual growth of newly created information in enterprises to be 30% [25]. According to studies of Merrill Lynch and Ferris Research only 20% - 40% of this information is stored in a structured, semantically described form in electronic databases or structured business applications (e.g. SAP R/3) [13, 5]. The other 60% - 80% are contained in the unstructured form of electronic documents[1] (e.g., Microsoft Office files, e-mails, images, multimedia files, or web based content). This is even more dramatic as the annual growth rate of newly created unstructured content is considerably higher than that of structured information.

Historically unstructured information was and in most scenarios still is managed in an application specific way. Due to different business requirements, enterprises employ task-specific applications which leads to the storage and processing of inherently similar information in different places in the enterprise information architecture. A single task or workflow will therefore require the use of multiple applications resulting in different interaction techniques and information representations to support the user [27]. Tungare et al. labeled this situation *Information Fragmentation* where a user's data are tied to different formats, distributed across multiple locations, manipulated by different applications and reside in a generally disconnected manner [34]. Most often information items are stored in separate collections depending on their formats: documents are saved in a documents' folder hierarchy (e.g. in the My Documents folder), e-mails in a separate mailbox hierarchy, and bookmarks to favorite web sites in another browser hierarchy [19]. This hierarchy separation has several negative outcomes: It leads to potential redundancy and an increased required effort to locate document based knowledge. In addition to being time consuming, managing three (or more) different hierarchies generates cognitive load when trying to maintain a certain degree of consistency between the hierarchies and in using multiple different applications with inconsistent interaction designs [20]. Since users associate information objects with their projects and tasks rather than document formats [4] this represents a potential disconnect between information systems and a user's needs.

Such problems multiply in a multi user enterprise environment with numerous competing repositories, document management systems, shared network drives, and local file systems for each distinctive user. On one hand a persons ability to overlook the available number of information

[1]In the following the term document is understood as relating to electronic files containing unstructured information.

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

sources diminishes as the amount of documents in numerous different repositories grows. On the other hand a multi user environment with local file system repositories lacks the central coordinating instance of an individual user [29]. As a result individual knowledge workers in such multi user enterprise environments spend an even greater part of their working hours looking for the correct and most up-to-date information needed in workflow steps or tasks.

The research contribution presented in this paper is a method to reduce document based information fragmentation by semantically reconnecting electronic documents to each other without imposing additional training or tagging workload on the user. To this end the actions knowledge workers perform on their desktop are transparently monitored to analyze the user's interaction with his computer system. These action metadata are further clustered by superordinate activities performed by the user. Actions are clustered if their associated desktop operations can be related to each other by combining multiple relationship building algorithms. Finally documents attached to window instances within the identified activity clusters are semantically related to each other reducing the fragmentation of their contained information.

The paper is organized as follows: Section 2 presents a review of current research. This is followed by the description of the presented approach in section 3 consisting of the capturing of semantic user activity context, the subsequent clustering of these activity contexts, and the final inference of semantic interrelations between documents. Section 4 shows the results of evaluating a prototypical implementation of the approach in a small scale testing setup. The paper is concluded with a summary and an outlook to ongoing research and open questions.

2. RELATED WORK

Previous research approaches to establish and utilize relationships between documents can be separated into two groups. The first group makes use of the content of documents to detect similarity or references in documents, while the second group of work relies on analyzing user activity involving documents usage to infer interrelations.

While some techniques applied by the first group of approaches (e.g. Named Entity Recognition) are already quite advanced, more sophisticated parsing of document content (e.g. Natural Language Processing) is still problematic due to high manual training requirements of the applied algorithms [8]. Additionally content based approaches are often limited to text documents (e.g., multimedia files or files in proprietary formats) [6].

The second group of approaches focuses on the activities of a user around individual documents. Relying on tasks as the principle means of document relationship discovery requires identifying the connections of one piece of information in one application to task-related information in the same or other applications. The project *Stuff I've Seen* [12] focuses on document metadata and in part document content to create a context-snapshot of the time when a document was accessed. A user might then at a later time query the *Stuff I've Seen* datastore with arbitrary keywords to be presented with a list of documents he accessed earlier. The research prototype *Phlat* [9] combines keyword and property-value search with faceted browsing functionality. The tagging of documents which serves as the basis for *Phlat*'s superimposed search UI has to be manually performed by the user though. Both *Stuff I've Seen* as well as *Phlat* aim at enhancing information retrieval but do not reconnect previously fragmented information contained in different documents. The task-centered tools *UMEA* [21], *TaskMaster* [2] and *TaskTracer* [11, 33] address this to some degree by capturing the desktop context of project or task related user actions. All three approaches require the user to manually enter the specific task or project on which he is working at any current time. This necessity is avoided by the IBM prototype *Activity Explorer* which makes use of document-centric user activities within a collaborative document management system to add relations between document instances [14]. This approach relies on a specifically customized DMS environment though and does not detect user activity outside this proprietary work context.

The research projects *SWISH* [28], *Smart Desktop* [24], *Dyonipos* [17], *APOSDLE* [23], and *UICO* [31] are all engaged in automated task prediction and task switch detection within personal desktop environments. The first four approaches rely on textual metadata of surveyed desktop resources to distinguish task contexts. On top of that the UICO prototype also takes chronological metadata of contextual activities into account. While all five prototypes show high detection accuracy this comes at the cost of initially training the employed machine learning mechanisms to specific application domains [30]. Besides this initial training UICO requires an additional step to model the ontology classes.

A limiting factor to the adoption of Knowledge Management Systems (KMS) is often seen in lacking user acceptance [26]. Steve Bailey (Senior Adviser for Records Management issues at JISC infoNet) summarizes the issues employees have with the overhead generated by IT systems to improve the management of unstructured content [1]: *"As far as the average user is concerned, the EDRMS [(Electronic Document and Records Management System)] is something they didn't want, don't like and can't use. As such, its no wonder that so few users accept them. As one person once said to me 'making me use an EDRMS is like asking a plasterer to use a hammer!"* This low user acceptance results mostly from knowledge workers feeling burdened by additionally workload caused by e.g., manual tagging or categorization in KMS. Incidentally the described approaches to the management of unstructured information require some initial *knowledge investment* before a critical mass is reached which can be used to endow user benefit (e.g. a DMS only showing relevant search results after a certain amount of manual tagging on contained documents has been performed). The additional effort is felt by the employees to be distracting from the subjectively more important *real work* and is therefore often circumvented or altogether ignored [15].

3. CONSENSE APPROACH

The presented approach is set within the encompassing research context of the ConSense project[2] which aims to gather document related metadata within enterprise environments to deduct semantic relations between documents and business domain specific entities (e.g., documents, persons, projects or products) to allow a semantic exploration of the enterprise information landscape [6]. The presented approach forms one of several pillars to capture semantic metadata on client workstations. While reasoning within the semantic network is a crucial part of the encompassing project it occurs at a later stage after sensor data has been gathered and aggregated[3] and is therefore not in the focus of this paper.

3.1 Semantic Activity Context

According to Kuutti, user activities can be considered as having three hierarchical levels: activity, action, and operation, which correspond to motive, goal, and conditions of the task for which the activity is performed [22]. An activity may be achieved through a variety of actions. Similarly, operations may contribute to a variety of actions. In the following a task is regarded as containing one or more distinctive activities which in turn consists of various actions which all relate to the same object or motive (see figure 1). Actions in turn consist of any number of (desktop) operations with an action always having a particular goal within the context of the task-induced activity motive [35]. As will be shown for the purpose of the presented approach it is sufficient to focus on the lower three levels of activity, action, and operation to establish relations between documents containing fragmented information. The encompassing level of tasks spanning potentially multiple days is therefore not in further scope.

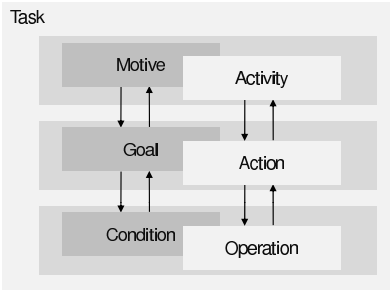


Figure 1: Tasks, activities, actions, and operations [22]

When evaluating user desktop activity each corresponding action has a context consisting of the sum of all other actions taken by a user as well as the respective workplace environment during the user action. The individual elements comprising such context can be separated into two dimensions: time and scope. The dimension of time is split into context elements which happened before, during, and after the user action. The dimension of scope differentiates among user actions targeting a specific document and the general desktop environment visible to the user. This desktop environment in the context of a document access consists of all opened and visible desktop and web applications as well as all content and documents contained in these applications.

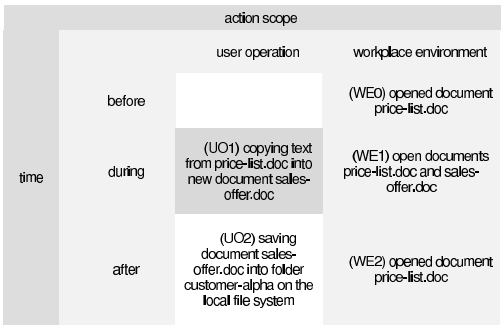


Figure 2: Exemplary context of electronic document access

To give a simple business scenario example: A user in the sales department copies some textual content from document PRICE-LIST.DOC into a new document SALESOFFER.DOC and saves it in the local file system folder CUSTOMER-ALPHA. Figure 2 shows the context of this action. The user operation (in this case copying) allows the assumption (not certain knowledge) of an existing semantic relation between the documents PRICE-LIST.DOC and SALES-OFFER.DOC. The metadata gained from observing user operation UO1 might further be influenced by the corresponding workplace environment WE1. So could for example a window positioning showing both documents next to each other without overlap strengthen the assumption of an existing relationship. Part of the action context is the second user operation UO2. This in turn can be used to establish a new (or strengthen an existing) semantic relation of SALES-OFFER.DOC and (to a lesser extend) PRICE-LIST.DOC to other documents contained in folder CUSTOMER-ALPHA. From a modeling perspective UO2 has its own context, which would then contain UO1 as an operation that happened shortly before UO2.

3.2 Capturing Semantic Activity Context

For the purpose of capturing the semantic context of user actions a client-side sensor plugin is installed on a user's workstation. The sensor takes a snapshot of window metadata whenever changes are occurring within the workspace environment either manually triggered by the user (e.g., opening or moving window instances) or automatically triggered by the system (e.g., a notification window being shown). Events being considered changes to the workplace environment are changes to a window's X/Y/Z position as well as to its size. Additional snapshots are taken upon user operations triggering new documents being opened within these window instances. Table 1 lists the recorded metadata for each window during a monitoring snapshot.

With multi monitor setups becoming more common [10] the sensor detects and logs changes in the workstation's monitor configuration including the monitor's resolution, physical position in relation to other monitors and the start and end

[2]www.consense-project.com
[3]See [6] on the subjects of reasoning as well as dealing with computational complexity in the ConSense project.

Table 1: Extracted basic window metadata

Data type	Metadata
int	X position of the top left corner
int	X position of the top left corner
int	Z axis position
int	Monitor id on which the top left corner is positioned
int	Height
int	Width
string	Unique window identifier
string	Unique parent window identifier
string	Window title
bool	Indicator if the window has UI focus

time of this specific setup. This allows the later calculation of the available screen estate to the user for any given time.

In order to analyze these raw action metadata effectively they are transformed into a hierarchical, chronologically ordered, data structure as depicted in figure 3. The main object *Working Item* contains all workspace related metadata between the start and end of a user session as indicated by the operation system's session start and session end events[4]. Within a *Working Item* window metadata for each snapshot are contained within a *Time Span* describing periods during which no changes to the workplace environment occurred. These *Time Spans* contain the window metadata of all windows visible to the user during the duration of the *Time Span*. Visibility in this regard is determined by calculating the relative square size to which a window is visible to the user using the window's X/Y/Z position and size metadata. Due to study results indicating a user preference to perform a high percentage of window switching on one primary monitor [18] windows being displayed on a secondary monitor as indicated by the operating system have their relative visibility reduced by 50%. For windows stretching across monitor boundaries this reduction is applied according to their partial visibility on secondary monitors. Only metadata for windows having a computed relative visibility higher than 0.2[5] are included in the *Time Stamp* recordings.

The raw window metadata captured in this way offer only limited leverage to identify substantive relationships between fragmented document based information. For this reason it is enriched by metadata relating to the respective application responsible for the window's creation as well as relating to document(s) opened within the window instances. Existing studies regarding the usage of document and application types during typical task execution patterns [12, 9] show e-mail as the most common type opened (~75%), followed by web pages (~15%), Microsoft Word files (~6%), Microsoft Powerpoint files (~3%), and various media types including pictures and audio files (~4%). Over 6% of all

[4]The logical session start and session end events are comprised of the physical start/stop, sleep/hibernate/stand by/wake up, user login/logout and remote connection initiated/terminated events.

[5]According to [18] windows hidden by more than ~75% do not impart any informational value to the user and are therefore treated as having no direct relation to the immediate action context.

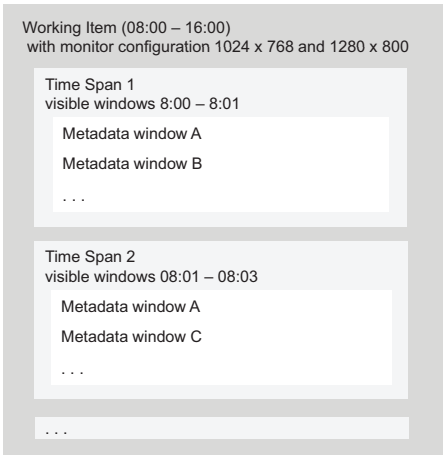


Figure 3: Structure of captured workspace metadata

items were e-mail attachments. Accordingly the sensor plugin's metadata extraction routines focus on the Microsoft Office Suite as well as the major web browser families Microsoft Internet Explorer and Mozilla Firefox. In particular the Microsoft Outlook element types e-mail, calendar entry, contact item, and files attached to these elements are included in the sensor range. Table 2 lists the (upon availability) recorded additional application and document metadata for each window during a monitoring snapshot.

Table 2: Extracted extended window metadata

Data type	Metadata
string	Unique id of the application responsible for the creation of the window
string	Name of the application responsible for the creation of the window
string	System path of the application responsible for the creation of the window
string[]	System path(s) of opened file based document(s)
string[]	Unique id(s) of opened Microsoft Outlook element(s)
string[]	Web page URI(s) opened in web browser tab(s)

Three categories of documents are treated in more depth using different content extraction approaches. These categories are web based HTML documents, elements within Microsoft Office Outlook, and lastly all other documents. In the case of web content it is first checked whether the target is the web address of a HTML file. If this is the case the sensor accesses the web browser tab's HTML source and temporarily stores it on the local machine. In the case of a Microsoft Outlook element the unique Outlook id initially extracted is used to capture further metadata for these elements. Additionally the respective Outlook element is evaluated regarding attached files. If found these are recursively processed in the same way as the originally identified documents. Lastly for the content extraction from other files the indexing service inherently contained within the Microsoft

Windows operation system is used to access and extract document metadata. If the file system in question is of type New Technology File System (NTFS), which is commonplace in modern Microsoft Windows operating systems, additional metadata from the file system are available for extraction by the sensor plugin.

Because the naming conventions for different types of metadata attached to files vary depending on the respective application responsible for the creation of the specific file types, a common semantic as well as syntactic metadata description is necessary. This underlying metadata description is provided by the ConSense ontology store [6] containing both custom ontology descriptions in the Web Ontology Language (OWL) format as well as a subset of the Friend-of-a-Friend (FOAF) and Dublin Core ontologies. In the case of Dublin Core the original RDF Schema notation has been transformed into OWL mapping *rdf:Property* predicate types to *owl:AnnotationProperty* types. This allows the resulting Dublin Core OWL to retain an expressivity of OWL-DL compared to the computationally more complex OWL-Full superset. The combined expressivity of these ontologies stays within OWL-DL corresponding to $\mathcal{SHOIN}^{(\mathcal{D})}$ using description logic naming conventions.

This ontology set forms the ground for a static mapping of raw file attributes to the gathered (RDF encoded) semantic metadata. Additionally the ontology set accessed by the sensor plugin could be further extended to account for e.g., application types beyond the abovely mentioned Microsoft Office suite. The sensor plugin validates gathered context input against the common ontologies (without resorting to expensive inference reasoning) and persists it in the form of semantic networks in a local Resource Description Framework (RDF) triple-store.

3.3 Clustering Semantic Activity Context

Utilizing the semantic metadata described in section 3.2 to relate fragmented document based information to each other is limited by the underlying assumption that the whole captured context belongs to a single user activity. Due to multitasking behavior being predominant within all kinds of office environments and scenarios though this does not hold true for most real world applications. Knowledge workers spend a great deal of time engaged in multiple tasks at the same time [32]. So do for example employees of an information-technology company spent an average of only 3 minutes per task before switching to another task [16]. As Salvucci et al. point out there exist multiple types of multitasking. On one hand there is *concurrent multitasking* where a person changes between different tasks back and forth several times a second giving the impression that they are performing the different tasks simultaneously. On the other hand there exists *sequential multitasking* where a task is carried out over minutes or hours until a secondary task interrupting the first (e.g., a person reading a book while cooking, briefly stirring the sauce, and then reading again for several minutes in the book). The basic procedure of human multitasking is always the same regardless of the kind of multitasking though as shown in figure 4. An important point in changing from one task to the next is the time delay which can take quite long depending on the complexity of the task. In the context of the presented approach a special subset of multitasking is of particular interest, namely *intermittent processing* [3]. Within this scenario several tasks are executed on a computer system in parallel though the user can put his attention only on one activity at any time resulting in a de facto sequential multitasking pattern.

Accordingly the sensor plugin unravels the previously captured semantic window action metadata and clusters them according to their superordinate activities. To this end multiple algorithms are employed to establish semantic relationships between the observed window instances. Initially, the simplest type of relationship is considered, namely whether a window instance can be identified as existing across multiple *Time Spans* using the unique window identifier as well as positioning metadata. The second scenario being resolved by the sensor plugin is the existence of multiple window instances visible within the same *Time Spans* belonging to the same application. Such situations often occur when dialog boxes (e.g. the application asking for confirmation of a specific user action) are displayed by an application. This is confirmed if the captured metadata show one of the window instances containing as parent id the unique handle of the other. In both scenarios a semantic statement of predicate *belongsToSameActivity* interrelating the window instances is added to the semantic network.

In the next step relationships between windows of different applications are uncovered which indicate user operations within the same action and therefore the same superordinate activity. This is initially done for all windows within a *Time Span* using the extracted textual metadata both from the window instances as well as from their associated documents (opened and attached). As the starting point for such a comparison the *title* metadata of windows and documents is utilized which is then enriched by the textual metadata of contained or attached documents.

When employing a basic string comparison algorithm (e.g. Jaro-Winkler distance [37]) different spelling of keywords or the presence of textual clutter can taint the relevance of resulting similarity indices as illustrated in the following two exemplary window titles (1) *The term Business Activity Monitoring (BAM) describes a collection of analyses and presentations - Windows Internet Explorer* and (2) *C:/Interest/Analyses/BAM - Windows Explorer*.

In order to increase the relevance of further processing results, the sensor plugin sanitizes the textual metadata. In a first step the application name is stripped from the window title if it is contained[7]. In the case of more extensive textual metadata as shown in (1) the removal of the application name alone is not sufficient. As the string contains a full English sentence it comes with the usual language specific expletives. To this end the sensor plugin normalizes spelling variations using the Porter Stemming algorithm [36] and separates compound strings into a list of individual words. Subsequently further splitting is performed on dashes to separate system paths into single directory name

[6]Hosted at http://ontology.consense-project.com

[7]Within the Microsoft Windows operation system each application registers its own name within the registry hive.

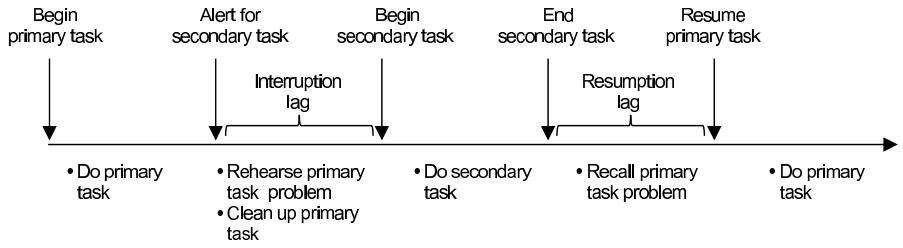


Figure 4: Multitasking flow [32]

parts to account for cases as illustrated in example (2). The resulting list of normalized strings is then cleaned off language specific stop words. As mentioned in section 1 the presented approach is set to avoid task interruption to the knowledge worker making a manual selection of language contexts suboptimal. To this end the sensor plugin employs a generic n-gram comparison which utilizes word frequency classes of the textual metadata to determine the language of a given (partial) text. According to the detected language, stop words are then removed from the word list using provided static stop word lists for the languages Dutch, English, French, Spanish, Italian, and German.

Finally the word lists for two window instances are compared using a Weighted Tag Similarity algorithm [7] resulting in a numeric value in the interval [0,1] indicating the assumed similarity of the two windows. To reduce the necessary amount of background processing window instances are only compared if their respective *Time Spans* lie within 10 minutes of each other.

Any *belongsToSameActivity* relation concluded by this similarity value can only be treated as an assumption with a certain amount of unreliability. To this end a semantic reification statement is added to the *belongsToSameActivity* relation indicating its reliability as calculated from the similarity value using a fuzzifying function (see figure 5):

$$rel = \begin{cases} 2sim^2 & sim < 0.5 \\ 4sim(1 - \frac{sim}{2}) - 1 & sim \geq 0.5 \end{cases}$$

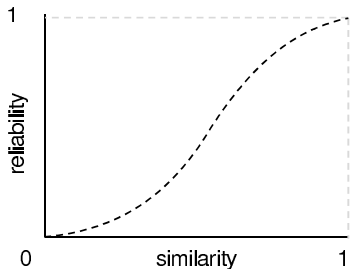


Figure 5: Fuzzification of textual similarity to reliability of statement

Figure 6 shows an exemplary semantic network of activity cluster relations being generated by evaluating window instances across two adjacent *Time Spans*. Lastly the previously generated relationships between window instances for all *Time Spans* within the current *Working Item* are condensed into a single clustering graph as depicted in figure 7 forming the final activity clusters. The reified reliability is shown on the graph edges.

3.4 Relating Documents based on Activity Clusters

Based on the identified activity clusters the sensor plugin relates documents associated to window instances within the same cluster. To this end a weighted degree centrality C for each window instance w is calculated[8] with W being all window instances, and $rel\,(w, w')$ being the attached reliability of the *belongsToSameActivity* relation between windows w and w':

$$W'_w = \left\{ w' \in W : \exists rel\,(w, w') \right\}$$

$$C_w = \frac{\sum_{w' \in W} (rel\,(w, w'))}{|W| - 1} \left| W'_w \right|$$

For every two window instances p and q within the activity cluster a semantic statement with predicate *hasRelationTo* is inserted into the semantic network. Additionally a reification statement signifying the reliability of the *hasRelationTo* statement with value $C(p) * C(q)$ is added.

Finally in addition to the reliability further reification statements are added containing provenance information to the statements relating documents to each other. These consist of the current time stamp, the semantic metadata which were input for the relation inference, and the identifier for the presented approach designating it as the source of the inferred relationship.

As the *ConSense* research project underlying the presented approach [6] is focused on multi user enterprise environments semantic document-relating metadata with relevance beyond that of the local user (that is higher-level metadata relating to domain-specific business entities) is then submitted to an enterprise-wide central semantic triple store. This central metadata repository, forming a virtualized view of assumptions on real-world relations among business entities and documents, can in turn be queried by a user or knowledge manager to visualize and cluster relationship types according to his task-specific discovery needs. So can for example an enterprise legal department query for documents

[8]Relations (edges) between window instances are treated as bidirectional for this purpose.

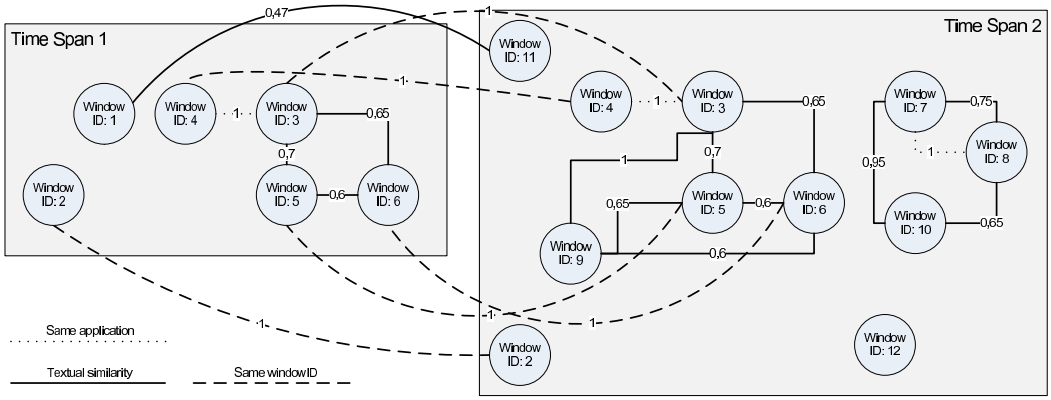


Figure 6: Example of window instances related across time spans

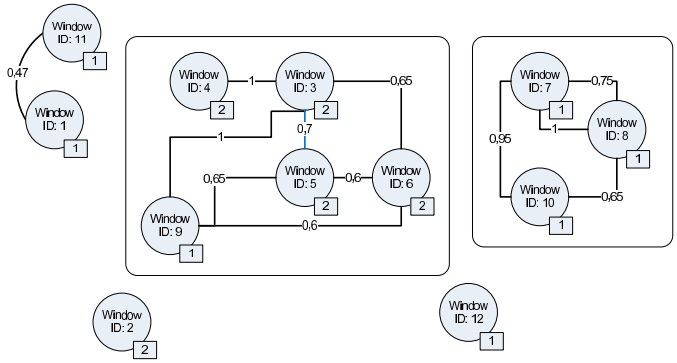


Figure 7: Example of a resulting activity cluster

having a relation to a product line being the subject in a legal low-product-quality complaint by a client. The query parameters could then be narrowed to documents having additional relations to the internal Quality Management process. Alternatively a rule or heuristic based electronic service can directly access the semantic aggregation layer and query or manipulate the semantic network.

4. EVALUATION

To test the viability of the presented approach the prototypically implementated sensor plugin was installed on the client desktops of four knowledge workers. During the evaluation 15 *Working Items* with durations ranging from 5 minutes to 5 hours were generated containing activity patterns of singular desktop activity, sequential desktop activity and multitasking activity as described in section 3.3. The goal of these experiments was mainly to to obtain information on the accuracy of the clustering algorithm both regarding false-positives (actions included in the clusters) as well as false-negatives (actual window activity relationships not detected by the sensor plugin). The test subjects were later asked to confirm or deny each of the individual *belongsToSameActivity* relationships concluded by the sensor plugin, denials resulting in a false-positive result for that match. In a second step the subjects were presented with the unconnected window instances and asked to cluster them according to their own remembrance of the performed activities resulting in false-negative errors where *belongsToSameActivity* relationships stated by the test subjects were missing in the sensor results.

In addition to the results shown in table 3 the document interrelations being identified as false-positives showed a lower average attached reliability (0.31 compared to 0.53) than the correctly identified relationships. Especially when considering that the reliabilities attached to document interrelations are not discarded but made available to further processing within the resulting semantic network a reconnection of the previously fragmented document based information can be achieved using the presented approach. So while the number of participants in this evaluation was limited the error rate lies in an acceptable range.

5. CONCLUSION AND OUTLOOK

It has been shown that the fragmentation of document based information in enterprise environments slows down workflow related information retrieval and poses a significant challenge to knowledge management initiatives. The paper demonstrated a new method to relate such unstructured documents containing similar information to each other without imposing additional tagging or training workload on the individual user. The method was prototypically implemented and evaluated in a small scale testing setup.

Table 3: Evaluation Results

Test	Subject	Window Instances	False Negatives	False Positives	Error Rate
1	A	62	1	6	11.29%
2	A	6	0	0	0.00%
3	A	12	0	1	8.33%
4	A	70	1	0	1.43%
5	B	12	0	1	8.33%
6	B	52	0	3	5.77%
7	B	28	0	0	0.00%
8	B	33	0	1	3.03%
9	B	21	0	2	9.52%
10	B	71	2	0	2.82%
11	C	51	0	0	0.00%
12	C	67	1	1	2.99%
13	D	16	0	0	0.00%
14	D	24	1	1	8.33%
15	D	60	0	1	1.67%
Total		585	6	17	3.93%

The amount of combined errors during the performed evaluation were in a noticeable yet – for the purpose of the presented approach – acceptable range. While the arguably more sophisticated existing task switching detection approaches as described in section 2 show a slightly higher rate of error when detecting task and activity switches this has to be explained by the steeper requirements applied by them. For the purpose of interrelating documents which does not require the detection of long running tasks contexts taking possibly multiple days to complete the presented approach is sufficient though. Furthermore the presented approach has the advantage of not causing user workload by imposing manual tagging or training requirements. Lastly the generated semantic relationships as well as the attached provenance and reliability statements can be further utilized by third party semantic Personal Information Management systems or Enterprise Content Management solutions.

In future research the mentioned heuristics to form activity clusters will have to be further tested and improved to reduce the number of false-positives. To this end domain specific heuristics might prove to be valuable. Also legislative aspects, especially privacy concerns of employees, have to be considered. Semantically white- and/or blacklisting named business entities as well as excluding specific window types or file system paths from the initial monitoring might be feasible to specifically include business process relevant documents only or to exclude documents and communication of sensible parties from the context readings.

6. REFERENCES

[1] S. Bailey. Has EDRMS been a success? The case for the prosecution. In *Opening Statement to the RMS Conference 2008, Edinburgh*, 2008.

[2] V. Bellotti, N. Ducheneaut, M. Howard, and I. Smith. Taking email to task: the design and evaluation of a task management centered email tool. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 352. ACM, 2003.

[3] R. Benbunan-Fich, R. Adler, and T. Mavlanova. Towards new metrics for multitasking behavior. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4039–4044. ACM, 2009.

[4] O. Bergman, R. Beyth-Marom, and R. Nachmias. The project fragmentation problem in personal information management. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, page 274. ACM, 2006.

[5] R. Blumberg and S. Atre. The problem with unstructured data. *DM REVIEW*, 13:42–49, 2003.

[6] H. Brügmann. Assisting the Discovery and Reuse of Document- based Knowledge using Semantic Metadata. In M. Schumann, L. M. Kolbe, M. H. Breitner, and A. Frerichs, editors, *Multikonferenz Wirtschaftsinformatik 2010*, pages 27–38, Göttingen, 2010. Universitätsverlag Göttingen.

[7] D. Buttler. A short survey of document structure similarity algorithms. *Proceedings of the 5th international conference on internet computing (IC2004)*, pages 191–208, 2004.

[8] P. Castells, M. Fernandez, and D. Vallet. An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 19(2), 2007.

[9] E. Cutrell, D. C. Robbins, S. T. Dumais, and R. Sarin. Fast, Flexible Filtering with Phlat? Personal Search and Organization Made Easy. *Microsoft Research*, 2006.

[10] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. Robertson, and G. Starkweather. Toward characterizing the productivity benefits of very large displays. In *Human-computer interaction: INTERACT'03; IFIP TC13 International Conference on Human-Computer Interaction, 1st-5th September 2003, Zurich, Switzerland*, page 9. Ios Pr Inc, 2003.

[11] A. Dragunov, T. Dietterich, K. Johnsrude, M. McLaughlin, L. Li, and J. Herlocker. TaskTracer: a desktop environment to support multi-tasking knowledge workers. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 75–82. ACM, 2005.

[12] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've seen: A system for personal information retrieval and re-use. *Annual ACM Conference on Research and Development in Information Retrieval*, 2003.

[13] Ferris Research. Industry Statistics 2009, http://www.ferris.com/research-library/industry-statistics/, 04/07/2010.

[14] W. Geyer, M. J. Muller, M. T. Moore, E. Wilcox, L.-T. Cheng, B. Brownholtz, C. Hill, and D. R. Millen. Activity explorer: activity-centric collaboration from research to product. *IBM Systems Journal*, 45(4):713, 2006.

[15] M. R. Gilbert, K. M. Shegda, D. Logan, K. Chin, T. Bell, L. Latham, R. E. Knox, and J. Lundy. Key Issues for Enterprise Content Management, 2006. Technical Report G00139515, Gartner Inc., 2006.

[16] V. M. Gonzalez and G. Mark. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 113–120. ACM Press, 2004.

[17] M. Granitzer, G. Granitzer, K. Tochtermann, S. Lindstaedt, A. Rath, and W. Groiß. Automating Knowledge Transfer and Creation in Knowledge Intensive Business Processes. In *Proceedings of the First Workshop on Business Process Management and Social Software BPMS208 in conjunction with 6th International Conference on Business Process Management*, pages 1–10. Springer Berlin / Heidelberg, 2008.

[18] D. R. Hutchings, G. Smith, B. Meyers, M. Czerwinski, and G. Robertson. *Display space usage and window management operation comparisons between single monitor and multiple monitor users*. ACM Press, New York, New York, USA, 2004.

[19] W. Jones and H. Bruce. A Report on the NSF-Sponsored Workshop on Personal Information Management, 2005.

[20] W. Jones, S. Dumais, and H. Bruce. Once found, what then? A study of keeping behaviors in the personal use of Web information. *Society for Information*, 2002.

[21] V. Kaptelinin. UMEA: User-monitoring environment for activities. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, volume 4, page 1, Ft. Lauderdale, 2002. ACM Press.

[22] K. Kuutti. *Activity theory as a potential framework for human-computer interaction research*, chapter 2, pages 17–44. Context and consciousness: Activity theory and human-computer interaction. MIT Press, 1996.

[23] R. Lokaiczyk and M. Goertz. Extending Low Level Context Events by Data Aggregation. *Proceedings of IKNOW Š08 and IMEDIA 08*, pages 118–125, 2008.

[24] D. Lowd and N. Kushmerick. Using salience to segment desktop activity into projects. In *IUI '09: Proceedings of the 14th International Conference on Intelligent User Interfaces*, page 463, New York, New York, USA, 2009. ACM Press.

[25] P. Lyman and H. R. Varian. How Much Information 2003. Technical report, University of California - Berkeley's School of Information Management and Systems, 2003.

[26] W. Money and A. Turner. Assessing knowledge management system user acceptance with the Technology Acceptance Model. *International Journal of Knowledge Management*, 1(1):8–26, 2005.

[27] D. Oard and J. Kim. Modeling information content using observable behavior. In *Proceedings of the 64th Annual Conference of the American Society for Information Science and Technology*, pages 481–488, 2001.

[28] N. Oliver, G. Smith, C. Thakkar, and A. C. Surendran. SWISH: semantic analysis of window titles and switching history. In *Proceedings of the 11th international conference on Intelligent user interfaces*, volume 06, page 201. ACM, 2006.

[29] M. Perez-Quinones, M. Tungare, P. Pyla, and M. Kurdziolek. A Special Topics Course on Personal Information Management, 2006.

[30] A. S. Rath. *User Interaction Context - Studying and Enhancing Automatic User Task Detection on the ComputerDesktop via an Ontology-based User Interaction Context Model*. Dissertation, Graz University of Technology, 2009.

[31] A. S. Rath, D. Devaurs, and S. N. Lindstaedt. UICO: An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop. In *Proceedings of the 1st Workshop on Context, Information and Ontologies - CIAO '09*, pages 1–10, New York, New York, USA, 2009. ACM Press.

[32] D. D. Salvucci, N. A. Taatgen, and J. P. Borst. *Toward a unified theory of the multitasking continuum*. ACM Press, New York, New York, USA, 2009.

[33] J. Shen, J. Irvine, X. Bao, M. Goodman, S. Kolibaba, A. Tran, F. Carl, B. Kirschner, S. Stumpf, and T. Dietterich. Detecting and correcting user activity switches: Algorithms and interfaces. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 117–126. ACM, 2009.

[34] M. Tungare, P. S. Pyla, M. Sampat, and P.-Q. nones. M.: Syncables: A framework to support seamless data migration across multiple platforms. In *IEEE International Conference on Portable Information Devices*, 2007.

[35] L. Uden, P. Valderas, and O. Pastor. An activity-theory-based model to analyse Web application requirements. *Information Research*, 13(2):1, 2008.

[36] C. J. Van Rijsbergen, S. E. Robertson, and M. F. Porter. New models in probabilistic information retrieval. In *British Library Research and Development Report no 5587*, 1980.

[37] W. E. Winkler. The state of record linkage and current research problems. *Statistical Research Division US Bureau of the Census Wachington DC*, 1999.

[38] E. Zadok, J. Osborn, A. Shater, C. Wright, and K.-K. Muniswamy-Reddy. Reducing Storage Management Costs via Informed User-Based Policies, 2004.

Kundenbindung durch Gratis-Musikdownloads: Eine empirische Untersuchung am Beispiel des Klassikmarktes in der Schweiz

Marcel Grüter
Institut für Wirtschaftsinformatik
Engehaldenstrasse 8
3012 Bern
+41 31 631 47 85
marcel.grueter@iwi.unibe.ch

Simone von Burg
Institut für Wirtschaftsinformatik
Engehaldenstrasse 8
3012 Bern
+41 31 631 47 85
simone.vonburg@iwi.unibe.ch

Thomas Myrach
Institut für Wirtschaftsinformatik
Engehaldenstrasse 8
3012 Bern
+41 31 631 47 85
thomas.myrach@iwi.unibe.ch

ZUSAMMENFASSUNG

Das Internet hat einen Wandel in der Musikindustrie bewirkt. Insbesondere CD-Verkäufe werden durch die zunehmende Verbreitung von Gratis-Musikdownloads (GMD) konkurrenziert. Im Zuge der Anpassung an diese neue Situation sind Musikanbieter gezwungen, neue Marketingmaßnahmen zur Bewerbung ihrer Musik zu ergreifen. Dabei bergen GMD nicht nur Gefahren sondern auch ein gewisses Potenzial als Marketinginstrument. Die vorliegende Studie untersucht am Beispiel des Klassikmarktes, wie GMD als Kundenbindungsinstrument eingesetzt werden können. Mit Hilfe der Kundenbindungstheorie wurde ein konzeptuelles Modell zu Voraussetzungen und Wirkungen von GMD in der Klassik entwickelt, das anhand einer experimentellen Umfrage mit 2x2-faktoriellem Design getestet wurde. Die Ergebnisse zeigen, dass die Tonqualität sowie die Länge des Stückausschnitts von GMD das wahrgenommene Wissen wie auch die Einstellung zur Musik eines Orchesters positiv beeinflussen. Darüber hinaus können GMD auch Verhaltensabsichten bezüglich des Kaufs von Tonträgern, des Besuchs eines Konzerts oder der Weiterempfehlung von Musik bewirken. Die Ergebnisse zeigen, dass sich GMD in Form eines Samplings als Instrument zur Kundenbindung verwenden lassen.

Keywords

Kundenbindung, Kulturmarketing, Gratis-Musikdownloads

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

1. EINLEITUNG

Die technologischen Möglichkeiten des Internets ermöglichen es, Musik als Datei zum Download anzubieten oder zu streamen. Dies führt für die Musikindustrie aber auch für die einzelnen Musikschaffenden zu Chancen und Risiken. Dabei stehen die Potenziale zur Distribution von Musikprodukten vielfach im Vordergrund der Betrachtung. Galt früher der Verkauf von Tonträgern als einzige Möglichkeit, die Musik auch außerhalb von Konzerten einem Publikum anzubieten, kann sie heute relativ einfach aus dem Internet heruntergeladen werden. Dabei stellt vor allem die unerwünschte Weitergabe von digitalen Inhalten ein erhebliches Problem dar. So werden insbesondere Raubkopien aber auch Gratis-Musikdownloads (GMD) häufig mit den sinkenden Umsätzen von CD-Verkäufen in Verbindung gebracht [26, 37]. Dieser Umstand wird damit begründet, dass mit der Möglichkeit zum unentgeltlichen Bezug der Musik der Anreiz zum Kauf aufgelöst wird [37]. Aufgrund einer solchen Argumentation werden Musikdownloads von der Musikindustrie nach wie vor kritisch betrachtet [34].

Neben den Potenzialen zur Distribution von Musikprodukten bieten sich durch GMD aber auch neue Möglichkeiten zur Bewerbung von Musik. In diesem Zusammenhang lassen sich GMD verschiedene positive Auswirkungen zuschreiben. Der Download lässt sich als eine Art Sample betrachten, mit dem eine auf einfache Weise mehr oder minder eingehende Produktinspektion stattfinden kann, welche anderweitig nicht oder nur umständlich in ausgewiesenen Fachgeschäften möglich wäre [36]. Dies soll den Kunden schließlich zum Kauf von CDs und Konzerttickets animieren [7, 24]. In diesem Sinne wirken GMD wie ein Kundenbindungsinstrument.

Der Musikmarkt für Klassik kämpft mit diversen Strukturproblemen. Nach wie vor besteht ein hohes Angebot an Musikschaffenden, die regional und überregional nach der Aufmerksamkeit des Publikums trachten. Dem steht eine sinkenden Nachfrage nach klassischer Musik gegenüber, wozu nicht zuletzt das Fehlen junger Interessenten beiträgt [3, 17]. Für klassische Künstler und Orchester stellen GMD daher eine interessante Alternative zu den bisherigen Marketingmaßnahmen dar. Über sie kann ein großes Publikum sehr kosteneffizient und unabhängig von geografischen Grenzen angesprochen werden, wobei sich aufgrund der Internet-Demografie jüngere, gegenüber

den neuen Medien aufgeschlossene Menschen gut erreichen lassen [3, 34].

Bislang lassen die vorhandenen Forschungsresultate noch keinen Schluss über die genauen Auswirkungen von GMD auf die Kundenbindung zu. Es ist unklar, ob und wie das Verhalten und die Einstellung eines Konsumenten zur Musik eines Interpreten oder Orchesters durch GMD beeinflusst werden können. Zudem fehlen Erkenntnisse darüber, welche Voraussetzungen GMD erfüllen müssen, um die gewünschten Effekte zu erzielen. Desweiteren gibt es noch keine Untersuchungen zu GMD im Bereich der klassischen Musik. Da sich das Klassikpublikum sowohl in der demografischen Struktur als auch in den Ansprüchen an Tonqualität vom Publikum anderer Musikrichtungen unterscheidet, bedarf es hier einer spezifischen Untersuchung [29].

Ziel der vorliegenden Studie ist es, herauszufinden, ob und wie GMD als Kundenbindungsinstrument in der Klassik eingesetzt werden können. Dazu wird basierend auf der Kundenbindungstheorie untersucht, wie GMD auf die Einstellung sowie das Verhalten eines Musikkonsumenten wirken. Ein besonderes Augenmerk gilt dabei der Rolle der Tonqualität und der Stücklänge des Downloads. Im Folgenden wird zunächst ein Überblick über die Literatur zu GMD sowie zur Kundenbindung geschaffen. Basierend auf der Theorie zur Kundenbindung werden in einem zweiten Schritt verschiedene Hypothesen zur Wirkung von GMD hergeleitet. Im dritten Kapitel wird der Aufbau der experimentellen Umfrage beschrieben, aufgrund derer die Hypothesen anschließend mittels ANOVA und PLS getestet werden. Im letzten Kapitel erfolgt eine Diskussion der Resultate, Limitationen werden aufgezeigt und Implikationen für Forschung und Praxis beschrieben. Der Artikel schließt mit einem kurzen Fazit sowie einem Ausblick auf die weitere Forschung im Gebiet der GMD.

2. THEORETISCHER RAHMEN

2.1 Gratis-Musikdownloads

Begünstigt durch die spezifischen Produkteigenschaften von Musik, wie z.B. deren Digitalisierbarkeit, deren tiefe Reproduktionskosten oder deren hoher Anteil an Erfahrungseigenschaften, hat sich sowohl das Angebot als auch die Nutzung von GMD stark verbreitet [4, 16, 34]. Die Auswirkungen, welche GMD mit sich bringen, werden jedoch sehr kontrovers diskutiert.

Verschiedene Studien, die allerdings nicht zwischen diversen Musikgenres unterscheiden, betrachten insbesondere den Einfluss von GMD auf CD-Käufe und kommen dabei zu unterschiedlichen Resultaten. Wie [26] erläutert, verwenden Konsumenten GMD als Substitute für CDs, was die sinkenden Verkaufszahlen im Zusammenhang mit vermehrten Gratis-Downloads erklärt. Auch [37] findet einen signifikant negativen Zusammenhang zwischen dem Ausmaß des unbezahlten Download-Angebots und den Verkaufszahlen von CDs.

Andere Studien betrachten demgegenüber GMD eher als Komplementärgut denn als Substitut für den CD-Kauf [28, 33]. Diese haben u.a. gezeigt, dass Nutzer von GMD eine höhere Kaufbereitschaft von CDs aufweisen [14]. Der Download dient in diesem Fall als Informationsmedium, um das Musikstück vor dem Kauf besser beurteilen zu können. Damit werden GMD als eine Art Sampling und Informationsquelle betrachtet, um die Unsicherheit vor dem CD-Kauf zu reduzieren [16]. Die Komplementärfunktion von Gratis-Downloads wird durch Qualitätsunterschiede legitimiert. So wird Musik auf einer CD oft als qualitativ hochwertiger eingestuft als beispielsweise in einer mp3-Datei [15]. Andere Autoren hingegen finden keinen signifikanten Unterschied in der Beurteilung der Tonqualität von CD-Aufnahmen und mp3-Dateien, was das Argument, GMD würden aufgrund der schlechteren Tonqualität keine Gefahr für CDs darstellen, in Frage stellt [6]. Neben der Tonqualität wird ein zum Download bereitgestelltes Musik-Sample auch durch die Länge des Musikausschnittes definiert [36]. Welchen Einfluss diese beiden Determinanten auf das Kaufverhalten von Konsumenten haben, wird in der Literatur bisher allerdings noch nicht eindeutig beantwortet.

GMD werden schließlich nicht nur als Sampling verwendet, um die Musikkäufe zu steigern, sondern auch, um die Konzertbesuche zu erhöhen [13, 24]. Im Sinne eines Samplings erhöhen GMD die Bekanntheit des Künstlers und steigern dadurch die Verkaufszahlen von Konzerttickets. Die steigenden Verkäufe von Konzerttickets werden wiederum mit der steigenden Komplementarität von CD und Konzert durch GMD begründet [24]. So entscheiden sich heute immer mehr Konsumenten nicht mehr entweder für eine CD oder für ein Konzert sondern kaufen sowohl CDs und besuchen Konzerte [7]. Da Konzertbesuche ebenfalls einen hohen Anteil an Erfahrungseigenschaften aufweisen, können Gratis-Downloads als Informationsquelle dienen, um die Entscheidungsunsicherheit zu reduzieren. Neben den Auswirkungen auf CD-Käufe und Konzertbesuche wurden in der bisherigen Forschung kaum weitere Effekte von GMD betrachtet.

In der vorliegenden Studie soll untersucht werden, welche Wirkungen GMD im Klassikmarkt mit sich bringen. Damit sind in erster Linie solche GMD gemeint, die vom Künstler oder dem Orchester selber in Form von Hörproben auf der eigenen Webseite oder auf speziellen Portalen angeboten werden. Die Bereitstellung dieser GMD kann vom Anbieter selbst aktiv beeinflusst und kontrolliert werden, im Gegensatz zu Downloads auf P2P-Netzwerken. Klassische Orchester und Künstler können GMD nutzen, um einerseits neue Hörer zu gewinnen und andererseits die einmal gewonnenen an sich zu binden. Dadurch sollen schließlich auch die Einnahmen und der Gewinn erhöht werden. Aus theoretischer Sicht lassen sich diese Bestrebungen durch das Konzept der Kundenbindung abbilden, welches im folgenden Abschnitt erläutert wird.

2.2 Theorie der Kundenbindung

Kundenbindung beschreibt den Sachverhalt, dass der Konsument über längere Zeit dem Anbieter treu bleibt [11, 22]. In der Forschung wird zwischen der einstellungsorientierten und der verhaltensorientierten Kundenbindung unterschieden (Abbildung 1). Die einstellungsorientierte Kundenbindung äußert sich zum einen in einem Zustand des „Nicht-Wechseln-Wollens“ und wird auch mit dem Begriff der Verbundenheit umschrieben [11]. Der Kunde hat eine positive Haltung sowie ein Gefühl der inneren Verpflichtung gegenüber dem Anbieter [27]. Demgegenüber steht die Gebundenheit, ein Zustand des „Nicht-Wechseln-Könnens“, weil ein Wechsel z.B. aus vertraglichen oder technischen Gründen nicht möglich ist oder weil er nur mit unverhältnismäßig hohem Aufwand erreichbar wäre [27]. Die verhaltensorientierte Kundenbindung bezeichnet dagegen das beabsichtigte sowie das tatsächliche Verhalten eines Kunden bezüglich Größen wie der Kauf-

intensität, dem Wiederkauf, dem Cross-Buying oder Weiterempfehlungen [25]. Ein verbundener Kunde mit einer positiven Einstellung weist meist auch eine hohe Zufriedenheit mit dem Anbieter auf, was wiederum im Verhalten der Kunden sichtbar wird [30]. Auf diese Weise kann Verbundenheit zu nachhaltigem ökonomischem Erfolg führen [18].

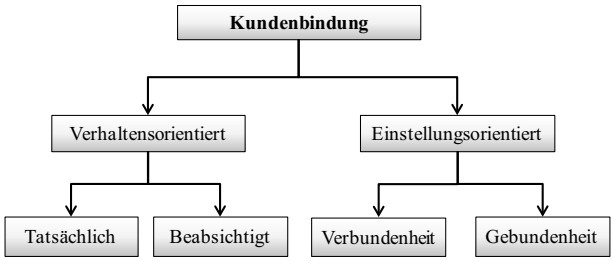


Abbildung 1. Formen der Kundenbindung.

Um sich eine Meinung zu einem Produkt oder einem Angebot bilden zu können, muss der Konsument überhaupt einmal mit dem Anbieter in Kontakt kommen [18]. Dies kann bspw. mittels Sampling erreicht werden, indem dem Kunden gratis eine Probe des Produkts angeboten wird [23]. GMD sind eine Möglichkeit für ein solches Sampling [5]. Auf diese Weise können sich Konsumenten kostenlos von der Musik eines Künstlers oder eines Orchesters überzeugen. Wie GMD schließlich konkret als Instrument für das Marketing eines Künstlers oder Orchesters genutzt werden können, wird im Folgenden durch ein konzeptionelles Modell erklärt, welches sowohl Voraussetzungen von GMD wie auch Auswirkungen auf die Einstellung und das Verhalten von Konsumenten beschreibt.

2.3 Voraussetzungen und Wirkungen von Gratis-Musikdownloads

2.3.1 Explorative Vorstudie

Wie bereits erläutert, sind in der Literatur bisher noch kaum Untersuchungen zum Thema GMD im Zusammenhang mit Kundenbindung und in der Anwendung auf die Klassik zu finden. Die Erklärung der Auswirkungen und der möglichen Anwendungsformen von GMD für das Marketing werden deshalb aus anderen Theorien abgeleitet. Als Ergänzung dazu wurden im Rahmen einer explorativen Vorstudie sechs qualitative Interviews mit Nutzern von GMD in der Klassik durchgeführt. Dabei beschäftigten sich drei Probanden beruflich und drei in ihrer Freizeit mit klassischer Musik. Des Weiteren unterschieden sich die Probanden bezüglich ihres Alters und ihres Einkommens. Im Folgenden werden für die Entwicklung des Modells zu GMD und seiner Hypothesen neben den theoretischen Überlegungen auch die Aussagen der Probanden hinzugezogen.

2.3.2 Modell und Hypothesen

Im Sinne des Stimulus-Organism-Response-Modells von [19] wird angenommen, dass der Einsatz von GMD (Stimulus), im Konsumenten einen Prozess zur Veränderung der Einstellung (Organism) in Gang setzt und in einem bestimmten Verhalten resultiert (Response). Anhand der GMD wird ein Produktwissen aufgebaut, das die Beurteilung des Produktes erlaubt [16]. Grundsätzlich sollen GMD einem Konsumenten die Gelegenheit geben, sich so weit zu informieren, dass ein späterer Kaufentscheid, z.B. für eine CD oder ein Konzertticket, genügend abgestützt werden kann [31]. Durch die Informationen, welche durch GMD gewonnen werden, kann eine Verhaltensänderung ausgelöst werden. Beeinflusst durch das wahrgenommene Wissen bildet sich zunächst die einstellungsorientierte Kundenbindung, die hier mit der Variable Einstellung ausgedrückt wird. Diese bezeichnet die Verbundenheit mit dem Orchester oder Künstler. Nach der Theory of Planned Behaviour (TPB) determinieren Einstellungen schließlich die Verhaltensabsichten [1]. Dementsprechend folgt auf die einstellungsorientierte die verhaltensorientierte Kundenbindung, welche durch das beabsichtigte Verhalten, einen Tonträger zu kaufen, ein Konzert zu besuchen oder den Künstler bzw. das Orchester weiter zu empfehlen abgebildet wird. Abbildung 2 stellt das konzeptionelle Modell der Kundenbindung durch GMD grafisch dar.

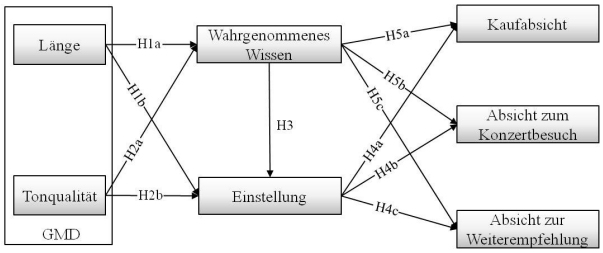


Abbildung 2. Konzeptionelles Modell der Kundenbindung durch GMD.

Die Länge des bereitgestellten Musikausschnitts sowie die Tonqualität bilden zwei wichtige Voraussetzungen für GMD, welche sich vom Anbieter aktiv gestalten lassen [36]. So können Orchester oder Künstler mit GMD entweder verschieden lange Ausschnitte eines Musikstücks oder aber das ganze Stück bereitstellen. Kann ein Konsument vor dem Kauf das ganze Musikstück gratis hören, so wird die Informationsgrundlage verbessert und die Kaufunsicherheit verringert [16]. Gerade bei klassischer Musik spielt die Länge des Stückausschnitts bei GMD eine wichtige Rolle, da sich das Hauptthema der Musikstücke nicht ständig wiederholt und zudem erst nach einer gewissen Zeit erkennbar wird. Dies wurde auch in den Interviews deutlich: *„Meistens hat man ja nur so einen 10-Sekunden-Ausschnitt zur Verfügung und bis bei einem klassischen Stück überhaupt nur der Anfang vorüber ist, ist die Hörprobe schon vorbei.“* Je länger also der gehörte Musikausschnitt ist, desto mehr wesentliche Informationen erhält ein Konsument über das Musikstück und desto besser kann er die Qualität der Musik beurteilen. Diese Annahmen führen zu der folgenden Hypothese:

H1a: Je länger der Musikausschnitt von GMD ist, desto besser fühlt sich der Downloader über die Musik des Orchesters/Künstlers informiert.

Wie bereits im Zusammenhang mit dem CD-Kaufverhalten erläutert, kann auch die Tonqualität von GMD einen entscheidenden Einfluss auf die Kundenbindung haben [15]. Die Tonqualität des GMD sollte so gut sein, dass sie den Prozess zur Bewertung der Musik nicht störend beeinträchtigt und damit die Informationsgrundlage für eine Beurteilung schmälert.

H2a: Je besser die Tonqualität von GMD ist, desto besser fühlt sich der Downloader über die Musik des Orchesters/Künstlers informiert.

Die Länge des Musikausschnittes und die Tonqualität von GMD beeinflussen neben dem wahrgenommenen Wissen auch die Einstellung zur Musik des Orchesters oder des Künstlers. Nach dem Technology Acceptance Model (TAM) führt eine hohe wahrgenommene Nützlichkeit einer bestimmten Technologieanwendung zu positiveren Einstellungen gegenüber der Technologie, in diesem Fall gegenüber von GMD [10]. Es ist anzunehmen, dass dies auch die Einstellung zur Musik positiv beeinflusst. Sind Stücklänge und Tonqualität von GMD für den Konsumenten in einem ausreichenden Maß vorhanden, wird der GMD für die Beurteilung der Musik als nützlich empfunden, womit auch die Einstellungen zur Musik positiv beeinflusst werden sollte. Ist dies nicht der Fall, so ist auch die Einstellung weniger gut, wie die folgende Aussage aus den Interviews zeigt: *„In der Regel war ich mit den Downloads eigentlich nicht so zufrieden, da die Aufnahmen häufig schlecht sind."* Aufgrund dieser Annahme lassen sich zwei weitere Hypothesen festhalten:

H1b: Je länger der Musikausschnitt von GMD ist, desto positiver ist die Einstellung des Downloaders gegenüber der Musik des Orchesters/Künstlers.

H2b: Je besser die Tonqualität von GMD ist, desto positiver ist die Einstellung des Downloaders gegenüber der Musik des Orchesters/Künstlers.

Die Einstellung zur Musik des Orchesters bzw. Künstlers wird nicht nur direkt durch den GMD sondern auch indirekt über das wahrgenommene Wissen beeinflusst. Die Verarbeitung der gewonnen Informationen äußert sich gemäß dem S-O-R-Modell in der Einstellung gegenüber dem Produkt und dem Anbieter, im vorliegenden Fall also gegenüber dem Orchester bzw. Künstler und seiner Musik [19]. Beurteilt ein Konsument die erhaltenen Informationen als ausreichend, so hat er ebenfalls ein ausreichendes Ausmaß an Informationen, um das Produkt beurteilen und sich ein Zufriedenheitsurteil bilden zu können [30]. Fällt dieses positiv aus, so wirkt sich dies positiv auf die Einstellung aus [11, 30]. Dies macht auch die folgende Aussage aus den Interviews ersichtlich: *„Der Download hat mir […] Anhaltspunkte geliefert, um mir eine Meinung über die Musik zu bilden. Sie haben es gut gemacht. Ich habe dann gedacht, […] die können das."* Unter der Bedingung, dass die Musik eines Orchesters oder Künstlers dem Konsumenten gefällt, lässt sich seine diesbezüglich positive Einstellung auf das ausreichende Wissen über das Musikstück zurückführen, welches erst ein Urteil über die Musik ermöglicht.

H3: Je besser sich der Downloader über die ihm gefallende Musik eines Orchesters/Künstlers informiert fühlt, desto positiver ist seine Einstellung gegenüber der Musik des Orchesters/Künstlers.

Das Ziel des Einsatzes von GMD als Kundenbindungsinstrument ist neben der Verbesserung der Einstellung auch eine Veränderung im Konsumentenverhalten und damit eine Verbesserung des ökonomischen Erfolgs. Die Betrachtung von Verhaltensabsichten als Auswirkung der vorgängig geformten Einstellungen, welche anschließend zu tatsächlichem Verhalten führen, wird ebenfalls durch die TPB unterstützt [1]. Das durch die Einstellung beeinflusste Verhalten äußert sich nach der verhaltensorientierten Kundenbindung im Kauf- oder Wiederkaufverhalten, im Cross-Buying-Verhalten sowie im Weiterempfehlungsverhalten [25]. Im Zusammenhang mit GMD lassen sich diese drei Verhaltensweisen ebenfalls abbilden. Wird aufgrund von GMD z.B. eine CD oder ein bezahlter Musik-Download gekauft, so bezieht sich das typischerweise auf das Kauf- oder Wiederkaufverhalten. Besucht ein Konsument nach dem Download von GMD ein Konzert des entsprechenden Orchesters bzw. Künstlers, betreibt er damit eine Art Cross-Buying. Auch das Weiterempfehlungsverhalten ist in der Musikbranche sehr wichtig, da Musik ein Erfahrungsgut ist und Weiterempfehlungen zur Unsicherheitsreduktion bei anderen Konsumenten beitragen [7]. Da eine empirische Beobachtung des tatsächlichen Verhaltens nur schwer zu realisieren ist, liegt der Fokus in dieser Studie auf dem beabsichtigten Verhalten. Nach der TPB werden die Verhaltensabsichten eines Konsumenten also von dessen Einstellung zum Orchester bzw. Künstler determiniert.

H4a-c: Je positiver die Einstellung eines Downloaders gegenüber der Musik des Orchesters/Künstlers ist, desto größer ist seine Absicht (a) eine Tonaufnahme dieses Orchesters/Künstlers zu kaufen, (b) ein Konzert dieses Orchesters/Künstlers zu besuchen und (c) das Orchester bzw. den Künstler weiter zu empfehlen.

Verhaltensabsichten werden nicht nur durch affektive sondern auch durch kognitive Größen beeinflusst [30]. Verhaltensentscheide werden in erster Linie aufgrund der vorhandenen Informationen über ein Produkt gefällt [19]. Ein Konsument wird sich nur für ein bestimmtes Verhalten entscheiden, wenn er die Informationsgrundlage zur Beurteilung des Produkts als ausreichend empfindet. GMD müssen dem Konsumenten genügend Informationen liefern, damit er die Musik des Orchesters bzw. Künstlers beurteilen kann. Die folgende Aussage aus den Interviews drückt diesen Sachverhalt ebenfalls aus: *„Wenn ich etwas downloade, dann […] kann ich schauen, ob mir diese Aufnahme entspricht, bevor ich sie kaufe."* Daraus ergeben sich die drei folgenden Hypothesen:

H5a-c: Je besser sich der Downloader über die ihm gefallende Musik eines Orchesters/Künstlers informiert fühlt, desto größer ist seine Absicht (a) eine Tonaufnahme dieses Orchesters / Künstlers zu kaufen, (b) ein Konzert dieses Orchesters / Künstlers zu besuchen und (c) das Orchester bzw. den Künstler weiter zu empfehlen.

Das konzeptionelle Modell soll nun mittels einer experimentellen Umfrage getestet werden. Im Folgenden werden zunächst der Aufbau sowie die Durchführung der Umfrage beschrieben und anschließend die Auswertung und die Ergebnisse präsentiert.

3. EMPIRISCHE UNTERSUCHUNG

3.1 Untersuchungsdesign

Die Umfrage weist ein 2x2-faktorielles Design auf. Als Stimulus wurde jeweils zufällig einer von vier Downloads bereitgestellt. Diese vier GMD unterschieden sich einzig bezüglich der Faktoren Tonqualität (gut, schlecht) und bereitgestellte Stücklänge (lang, kurz). Als Musikstück ausgewählt wurde der erste Satz von Schuberts fünfter Symphonie, gespielt von einem renommierten Schweizer Berufsorchester. Von dem Stück in CD-Qualität wurden mp3-Dateien mit unterschiedlicher Kompression generiert. Der Download in guter Tonqualität beinhaltete eine Bitrate von 320 Kbps, was derzeitig der Stand der Technik für hochwertige Musik-Downloads bei mp3 ist. Der Download in schlechter Tonqualität hatte eine Bitrate von nur 16 Kbps, was einer unüblich hohen Kompression entspricht. Dieser Wert wurde deshalb so tief angesetzt, um auch bei schlechten Abhörbedin-

gungen direkt über den Computer eine deutliche Qualitätsunterscheidung zu ermöglichen. Beim Download mit der langen Stücklänge wurde das Musikstück in voller Länge (7 Minuten und 8 Sekunden) bereitgestellt. Beim Download mit der kurzen Stücklänge waren dagegen nur die ersten 30 Sekunden des Musikstücks zu hören, wie dies in der Praxis für einen Sample häufig der Fall ist [36].

Die Umfrage wurde mittels eines Online-Fragebogens durchgeführt. Dabei wurden die Probanden per Email angeschrieben, in welchem sie einen Link zugestellt bekamen, der sie zu der Umfrage mit einem der vier eben beschriebenen GMD führte. Die Umfrage wurde online durchgeführt. Dies erscheint hier angemessen, da sich auch Downloads nur über das Internet herunterladen lassen und sich die Umfrage dementsprechend an ein technikaffines Publikum richtet.

Mit dem Online-Fragebogen wurde zuerst ein Pretest durchgeführt. Dieser zeigte, dass die Probanden beim Download der 16 Kbps Musikdatei die Ursache der schlechten Tonqualität bei ihren qualitativ minderwertigen Boxen sahen. Um diesen störenden Einfluss kontrollieren zu können, wurde das Untersuchungsdesign modifiziert. Zu Beginn des Online-Fragebogens mussten die Probanden nunmehr zunächst einen ersten, für alle identischen Download herunterladen und anhören. Dieser Download beinhaltete den vierten Satz von Schuberts fünfter Symphonie in voller Länge und in guter Tonqualität (320 Kbps). Dieser erste Download sollte den Probanden vor Augen führen, wie gut ein Musikstück auf ihrer Anlage klingen kann und somit sicherstellen, dass allfällige Qualitätsunterschiede beim eigentlichen Test-Download auch wirklich auf die Tonqualität des Downloads und nicht auf die Abhörbedingungen zurückgeführt wurden. Nach den Downloads wurden die Probanden schließlich gebeten, verschiedene Fragen bezogen auf den zweiten Musikdownload (Stimuli) zu beantworten.

3.2 Stichprobenauswahl

Die experimentelle Umfrage wurde am Beispiel des Klassikmarktes in der Schweiz durchgeführt. Der Schweizer Klassikmarkt ist einerseits relativ kleinstrukturiert mit einem hohen Anteil an Amateurorchestern sowie einer großen Dichte von Orchestern auf kleinem geografischen Raum [12]. Andererseits lässt er sich im Hinblick auf die Konsumenten durch einen hohen Anteil an über 60-Jährigen sowie einem vorwiegend städtischen, gut ausgebildeten und eher weiblichen Publikum beschreiben [3]. Bedingt durch diese beiden Eigenschaften verzeichnen diverse Orchester einerseits Probleme mit einem fehlenden Zuhörernachwuchs sowie mit sinkenden Einnahmen, da sich sowohl der Konsum als auch die Sponsoring- und Subventionsgelder auf eine relativ große Anzahl an Orchestern verteilen [17]. GMD bieten deshalb für den Schweizer Klassikmarkt eine interessante Möglichkeit, um für ihre Angebote zu werben und zudem auch die vorwiegend jüngeren Nutzer als interessantes Zielpublikum zu erreichen [3, 37].

Um den Effekt auf die Kundenbindung abschätzen zu können, sollten mit der Untersuchung vor allem bereits bestehende Kunden des Klassikmarktes angesprochen werden. Bezogen auf diesen Personenkreis geht es weniger darum, neue Kunden für klassische Musik zu begeistern, als bestehende Klassikliebhaber für ein ihnen zuvor unbekanntes Orchester zu interessieren. Vor allem bei Klassikliebhabern ist zudem damit zu rechnen, dass sie auf die Feinheiten der Interpretation achten und dementsprechend Untersuchungsparametern wie Stücklänge und Tonqualität besondere Beachtung schenken.

Der Link zum Fragebogen wurde an mehrere Schweizer Amateurorchester und -chöre sowie an verschiedene Schweizer Musikinstitutionen gesendet, mit der Bitte, den Fragebogen an ihre Mitglieder weiterzuleiten. Zudem konnte der Newsletter-Verteiler eines Schweizerischen Berufsorchesters zur Verteilung des Fragebogens genutzt werden. Auf diese Weise sowie durch die weitere Verteilung des Fragebogens mittels Schneeballtechnik wurde versucht, eine Stichprobe von vorwiegend jungen, internetaffinen Klassikhörern aus der Schweiz zu erreichen.

3.3 Messvariablen

Die Variablen aus dem konzeptionellen Modell wurden mit bereits vorhandenen Skalen aus der Literatur operationalisiert und messbar gemacht. Dabei mussten sie jeweils ins Deutsche übersetzt und an den Untersuchungskontext angepasst werden.

Die Variable des *wahrgenommenen Wissens* wurde anhand einer 7-Punkte-Likertskala (stimme voll zu/stimme überhaupt nicht zu) adaptiert nach [35] mit vier Items („Ich habe das Gefühl, genügend Kenntnisse über das Spiel des Orchesters gewonnen zu haben“; „Wenn Freunde sich über das Spiel des Orchesters informieren möchten, könnte ich ihnen darüber genügend Auskunft geben“; „Der Download genügt mir, um das Spiel des Orchesters beurteilen zu können“; „Ich brauche keine oder nur noch wenige zusätzliche Hörproben, um zu entscheiden, ob ich noch weitere Aufnahmen dieses Orchesters hören möchte“) gemessen. Die *Einstellung* zur Musik des Orchesters wurde adaptiert nach [9] auf einer 7-Punkte-Skala durch ein semantisches Differenzial mit vier Items (gut/schlecht, angenehm/unangenehm, gefällt mir/gefällt mir nicht, mag ich/mag ich nicht) erfasst. Ebenfalls in Anlehnung an [9] wurden jeweils durch eine 7-Punkte-Likertskala (stimme voll zu/stimme überhaupt nicht zu) mit je drei Items die *Kaufabsicht* („Sehr wahrscheinlich würde ich Tonaufnahmen wie CDs oder bezahlte Musikdownloads dieses Orchesters kaufen“; „Ich würde gerne weitere Tonaufnahmen wie CDs oder bezahlte Musikdownloads dieses Orchesters ausprobieren“; „Ich würde es in Betracht ziehen Tonaufnahmen wie CDs oder bezahlte Musikdownloads dieses Orchesters zu kaufen“), die *Konzertbesuchsabsicht* („Sehr wahrscheinlich würde ich ein Konzert dieses Orchesters besuchen“; „Ich würde gern ein Konzert dieses Orchesters besuchen“; „Ich würde es in Betracht ziehen ein Konzert dieses Orchesters zu besuchen“) und die *Weiterempfehlungsabsicht* („Sehr wahrscheinlich würde ich das Orchester bei anderen loben“; „Ich würde das Orchester gern weiterempfehlen“; „Wenn ein Freund meinen Rat zu klassischer Musik möchte, würde ich ihn auf dieses Orchester verweisen“) abgebildet. Auf sämtlichen 7-Punkte-Skalen bildete 1 jeweils den positiven und 7 den negativen Pol.

Um möglichst viele Störgrößen ausschließen zu können, wurden neben den Variablen aus dem konzeptionellen Modell Kontrollvariablen erfasst. Zum einen sollte die Einstellung zur Musik des Orchesters nicht dadurch beeinflusst werden, dass das Musikstück einem Probanden nicht gefällt. Der *Gefallen* des Musikstücks wurde dementsprechend mit der Frage „Wie gut gefällt Ihnen dieses Musikstück an sich“ auf einer 7-Punkte-Likertskala (ausgezeichnet/überhaupt nicht) gemessen. Zum anderen sollte die Internet- und die Klassikaffinität der Probanden möglichst

konstant sein, damit die Ergebnisse nicht durch diese Variablen verzerrt werden. Adaptiert von [32] wurden zur Messung der *Internetaffinität* zwei Fragen („Wie oft benützen Sie das Internet"; „Wie gut kennen Sie sich mit dem Internet aus") mit je einer 7-Punkte-Likertskala (sehr oft/nie, sehr gut/sehr schlecht) eingesetzt. Genau so wurden auch zur Messung der *Klassikaffinität* zwei Fragen („Wie oft hören Sie klassische Musik"; „Wie gut kennen sie sich mit klassischer Musik aus") mit der entsprechenden Antwortskala gestellt.

4. ERGEBNISSE

4.1 Stichprobenbeschreibung

Insgesamt haben 189 Probanden den Fragebogen vollständig ausgefüllt. Um Verzerrungen in der Stichprobe zu vermeiden, sollten alle Probanden gewissen Anforderungen genügen: Sie müssen eine Affinität zum Internet als auch zur klassischen Musik haben und das bereitgestellte Musikstück muss ihnen gefallen. Für die Auswertung wurden daher nur diejenigen Probanden berücksichtigt, deren Klassikaffinität, Internetaffinität und Gefallen am Musikstück jeweils einen Mittelwert (MW) von tiefer als 5 betrug. Von den 189 Fragebögen konnten so schließlich 174 ausgewertet werden.

Die Stichprobe setzt sich aus 45.4% männlichen und 54.6% weiblichen Probanden zusammen. Über 54.6% der Probanden sind weniger als 30 Jahre alt und nur knapp 22.4% sind 50 Jahre alt oder älter. Bezüglich Ausbildungsstand konnte zu 49.4% ein Publikum mit Hochschulabschluss erreicht werden, gefolgt von 31.6%, die einen Maturitäts- oder Lehrabschluss haben. Des Weiteren weisen die Probanden eine hohe Klassik- (MW=2.26) sowie eine hohe Internetaffinität (MW=2.06) auf. In Bezug auf das Downloaden von Musik sind die Probanden jedoch nicht sehr erfahren (MW=4.93). Noch weniger Erfahrung haben sie mit dem Downloaden von klassischer Musik (MW=5.71). Diese tiefe Downloaderfahrung entspricht jedoch den Erwartungen, da das Angebot an klassischen GMD in der Schweiz bisher noch relativ eingeschränkt ist. Diese Daten belegen, dass tatsächlich wie geplant eine Stichprobe mit eher jungen, klassik- und internetaffinen Probanden mit wenig Erfahrung im Bezug auf Musikdownloads gewonnen werden konnte.

4.2 Messmodell

Bevor die Hypothesen getestet werden, soll zuerst die Güte des Modells und der einzelnen Variablen überprüft werden. Tabelle 1 zeigt eine Übersicht zu den Güterkriterien der Variablen des Modells. Als Reliabilitätsmaße werden das Cronbach's Alpha, die Konstruktreliabilität sowie die durchschnittlich erfasste Varianz (DEV) betrachtet. Das Cronbach's Alpha bildet die interne Zuverlässigkeit der Skalen ab und sollte einen Wert von 0.8 nicht unterschreiten [8]. Während die Konstruktreliabilität die Eignung eines Faktors zur Erklärung seiner reflektiven Indikatorvariablen misst und einen Schwellenwert von 0.7 erreichen sollte, beschreibt die DEV mit einem Mindestwert von 0.6, wie hoch der durch den Faktor erklärte Varianzanteil manifester Variablen in Relation zum nicht erklärten Varianzanteil ist [21]. Wie in der Tabelle 1 zu sehen ist, werden sämtliche Mindestwerte deutlich überschritten.

Im Zusammenhang mit den beiden letzteren Größen werden auch die Faktorladungen betrachtet, welche nicht in der Tabelle enthalten sind. Diese sollten mindestens einen Wert von 0.6 überschreiten und im Idealfall höher sein als 0.8 [20]. Bis auf ein Item der Variable Wissen weisen sämtliche Items eine Faktorladung von über 0.8 auf. Immerhin lädt dieses zweite Item noch mit einer Ladung von über 0.7 auf die Variable Wissen.

Tabelle 1. Übersicht Gütekriterien

Variablen	Cronbach's α	Konstrukt Reliabilität	DEV	Stone-Geissers Q^2
Wissen	0.85	0.90	0.69	0.57
Einst.	0.95	0.97	0.88	0.76
Kaufabs.	0.86	0.91	0.78	0.53
Konzertabs.	0.91	0.94	0.85	0.64
Empf.Abs.	0.92	0.95	0.87	0.68

Ein wichtiges Validitätsmaß ist die Vorhersagevalidität, welche über Stone-Geissers Q^2 abgebildet wird. Die Vorhersagevalidität gibt Auskunft darüber, wie gut eine Rekonstruktion der latenten Variablen durch ihre Indikatoren möglich ist, wobei der kritische Wert von 0 überschritten werden sollte [21]. Auch diese Anforderung kann jede Variable des Modells erfüllen.

Ein weiteres Validitätsmaß stellt die Diskriminanzvalidität dar. Sie zeigt, ob die gemeinsame Varianz zwischen der latenten Variablen und ihren Indikatoren größer ist als die gemeinsame Varianz mit anderen latenten Variablen. Dabei sollte die DEV des latenten Konstrukts jeweils größer sein, als jede quadrierte Korrelation mit einem anderen Konstrukt [2].

Tabelle 2. Diskriminanzvalidität

Variablen	Wissen	Einst.	Kaufabs.	Konzertabs.	Empf.Abs.
Wissen	0.83	-	-	-	-
Einst.	0.21	0.94	-	-	-
Kaufabs.	0.27	0.40	0.88	-	-
Konzertabs.	0.30	0.38	0.58	0.92	-
Empf.Abs.	0.47	0.39	0.56	0.66	0.93

Tabelle 2 zeigt auf der Diagonalen jeweils die Wurzel der DEV für die latenten Variablen. In den nicht diagonalen Feldern sind die Korrelationen zwischen den latenten Variablen dargestellt. Aus der Tabelle ist ersichtlich, dass das Kriterium für die Diskriminanzvalidität bei jeder Variable erfüllt ist. Aufgrund der Erfüllung sämtlicher dieser Anforderungen, kann das Messmodell als reliabel und valide bezeichnet werden.

4.3 Hypothesentest

Die Überprüfung der Hypothesen wird im Folgenden mit einer Analysis of Variance (ANOVA) sowie mit Partial Least Squares (PLS) durchgeführt. Dabei werden insbesondere die Hypothesen H1a, H1b, H2a und H2b mit einer ANOVA überprüft, wobei zur Berechnung der ANOVA die Stichprobe anhand der beiden manipulierten Variablen Länge des Musikausschnitts und Tonqualität in vier Gruppen eingeteilt werden kann. Die Verteilung der Stichprobe auf die einzelnen Gruppen ist in der Tabelle 3 zu sehen.

Tabelle 3. Verteilung des Samples

Variablen		Tonqualität		
		gut	schlecht	**TOTAL**
Länge	lang	42	48	**90**
	kurz	41	43	**84**
	TOTAL	**83**	**91**	**174**

4.3.1 Experimentelles Design

Bevor die ANOVA berechnet wird, muss zunächst überprüft werden, ob die Manipulation der Variablen erfolgreich war und damit das experimentelle Design gelungen ist. Dazu wird ein Manipulation Check durchgeführt. Dieser zeigt, ob die unterschiedlichen Tonqualitäten und Stücklängen von den Probanden tatsächlich als unterschiedlich wahrgenommen wurden. Im Fragebogen wurde jeweils gefragt, wie die Tonqualität beurteilt und ob die Stücklänge als ausreichend lange wahrgenommen wird. Der gruppenweise Mittelwertvergleich weist signifikante ($p<0.01$) Unterschiede bezüglich der wahrgenommenen Tonqualität zwischen der Gruppe mit der guten (MW=3.00) und derjenigen mit der schlechten Tonqualität (MW=5.08) aus. Auch die wahrgenommene Stücklänge in der Gruppe mit dem langen Musikausschnitt (MW=3.01) unterscheidet sich signifikant ($p<0.01$) von der wahrgenommenen Stücklänge in der Gruppe mit dem kurzen Musikausschnitt (MW=5.01).

Weiter konnten in der Befragung allfällige Störvariablen wie das Gefallen des Stücks, die Klassikaffinität und die Internetaffinität relativ konstant gehalten werden, was in Tabelle 4 zu sehen ist.

Tabelle 4. Confounding Check

Gruppe	1	2	3	4	Signifikanz
Gefallen	2.62	2.50	2.54	2.60	0.905
Internetaffinität	2.14	1.94	2.24	1.94	0.299
Klassikaffinität	2.50	2.02	2.27	2.29	0.142

Der entsprechende t-Test zeigt, dass sich die Mittelwerte der drei Variablen in den vier Gruppen nicht signifikant ($p>0.01$) voneinander unterscheiden. Ein verzerrender Einfluss dieser Störvariablen kann demnach ausgeschlossen werden. Das experimentelle Design ist damit als gelungen anzusehen.

4.3.2 Einfluss der Stücklänge und der Tonqualität

Das Ziel der ANOVA ist, zu überprüfen, ob unter dem Einfluss einer variierten Tonqualität und der unterschiedlichen Länge des gehörten Ausschnitts eine Veränderung des Wissens und der Einstellung bezüglich der Musik des Orchesters festgestellt werden kann. Wie Tabelle 5 zeigt, weist die Variable Wissen sowohl bezüglich der Tonqualität als auch der Stücklänge signifikante Unterschiede in den Mittelwerten der verschiedenen Gruppen auf.

Tabelle 5. ANOVA wahrgenommenes Wissen

Wissen					
		Tonqualität			
		gut	schlecht	**TOTAL**	MW Differenz
Länge	lang	3.79	4.47	**4.13**	-0.37
	kurz	4.30	4.70	**4.50**	
	TOTAL	**4.05**	**4.59**		
	MW Differenz	-0.54			
Signifikanzniveau der Tonqualität		0.002			
Signifikanzniveau der Stücklänge		0.046			

Bezüglich der Stücklänge erweist sich der Mittelwert der Gruppe mit dem ganzen Stück als signifikant tiefer und damit besser als bei der Gruppe, die nur einen Ausschnitt gehört hat. Damit kann die Hypothese H1a mit einem Signifikanzniveau von $p<0.05$ bestätigt werden. Weiter weisen die beiden Gruppen mit einer guten Tonqualität einen tieferen und damit besseren Mittelwert auf als die Gruppen mit einer schlechten Tonqualität. Dieser Unterschied ist auf höherem Niveau signifikant ($p<0.01$) und bestätigt die Hypothese H2a.

Für die Variable Einstellung wurde ebenfalls eine ANOVA durchgeführt. Aus Tabelle 6 ist ein deutlicher Unterschied ($p<0.01$) zwischen den Mittelwerten der Gruppen mit einer guten und derjenigen mit einer schlechten Tonqualität festzustellen. Damit wirkt eine bessere Tonqualität positiv auf die Einstellung zur Musik des Orchesters. Hypothese H2b kann somit bestätigt werden.

Tabelle 6. ANOVA Einstellung

Einstellung					
		Tonqualität			
		gut	schlecht	**TOTAL**	MW Differenz
Länge	lang	2.26	2.90	**2.58**	0.05
	kurz	1.92	3.15	**2.53**	
	TOTAL	**2.09**	**3.02**		
	MW Differenz	-0.93			
Signifikanzniveau der Tonqualität		0.000			
Signifikanzniveau der Stücklänge		0.808			

Nicht bestätigt wird hingegen die Hypothese H1b. Es lässt sich kein signifikant positiver Einfluss der Stücklänge auf die Einstellung nachweisen. Der Mittelwert für die Gruppe mit dem Musikausschnitt ist tendenziell sogar etwas tiefer als für die Gruppe mit dem ganzen Stück. Dieser Unterschied ist aber nicht signifikant und wird deshalb anderen Einflussfaktoren zugeordnet.

4.3.3 Berechnung des Strukturgleichungsmodells

Die Hypothesen H3, H4a-c und H5a-c wurden als Strukturgleichungsmodell mit PLS getestet. Die Resultate der Tests sind der Tabelle 7 zu entnehmen.

Tabelle 7. Pfadkoeffiziente und T-Statistik

Variablen	Wissen			Einstellung		
	β	T-Wert	Signifikanz	β	T-Wert	Signifikanz
Einst.	0.21	2.81	✓	0.00	0.00	-
Kaufabs.	0.20	2.66	✓	0.36	4.84	✓
Konzertabs.	0.23	3.08	✓	0.33	4.76	✓
Empf.Abs.	0.41	6.11	✓	0.30	4.33	✓

Sämtliche Pfadkoeffizienten β sind positiv und signifikant ($p<0.01$). Dies zeigt zum einen, den positiven Einfluss der Variablen Wissen und Einstellung auf sämtliche Verhaltensvariablen wie die Kaufabsicht von CDs/bezahlten Musikdownloads als auch auf die Absicht zu Konzertbesuchen und die Absicht zu Weiterempfehlungen. Mit β=0.21 beeinflusst das wahrgenommene Wissen neben den Verhaltensabsichten auch die Einstellung zur Musik eines Orchesters positiv.

Da sämtliche Pfadkoeffizienten auf einem Niveau von $p<0.01$ signifikant sind, können die Hypothesen H3, H4a-c und H5a-c bestätigt werden. Abbildung 3 stellt die Ergebnisse zusammenfassend dar, wobei ***=$p<0.01$ bedeutet. Die mittels der ANOVA signifikant erhobenen Einflüsse von Stücklänge und Tonqualität auf das wahrgenommene Wissen und die Einstellung sind in der Grafik durch + gekennzeichnet.

In der Abbildung sind ebenfalls die Determinationskoeffizienten (R^2) der Verhaltensabsichten angegeben. Dabei erklären die beiden Variablen Einstellung und Wahrgenommenes Wissen mit 31,1% am besten die Varianz der Weiterempfehlungsabsicht. Die Varianz der übrigen Verhaltensabsichten kann nur knapp zu 20% durch das Wissen und die Einstellung erklärt werden, was als relativ schwacher Wert anzusehen ist.

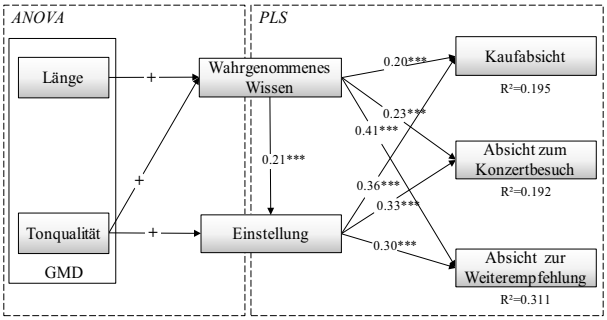


Abbildung 3. Modell der Kundenbindung durch GMD.

5. DISKUSSION

5.1 Theoretische Implikationen

Die Untersuchung hat gezeigt, dass GMD auch positive Effekte mit sich bringen. So sind sie im Stande, über das wahrgenommene Wissen wie auch über die Einstellung zur Musik Konsumenten in ihren Absichten positiv zu beeinflussen. GMD bestärken neben der Kaufabsicht auch die Absicht, ein Konzert zu besuchen sowie das Orchester oder den Künstler weiter zu empfehlen. Gerade die positiven Effekte, die GMD auf die Kaufabsicht haben können, schwächen bisherige Annahmen, dass ein Konsument keine Tonträger mehr kaufen wird, wenn er das Musikstück in guter Qualität und zudem als Ganzes im Internet downloaden kann. Wie die Determinationskoeffizienten allerdings gezeigt haben, kann die Weiterempfehlungsabsicht besser erklärt werden als die anderen Verhaltensabsichten. Eine mögliche Erklärung dafür könnte sein, dass sich die Absicht zur Weiterempfehlung viel unverbindlicher und ohne finanzielle Konsequenzen gestaltet als die Absicht zum Kauf oder Konzertbesuch.

Die Länge des bereitgestellten Musikausschnitts und die Tonqualität erweisen sich als zwei wichtige Gestaltungsfaktoren von GMD. So kann das wahrgenommene Wissen über die Musik eines Orchesters bzw. Künstlers durch einen längeren Stückausschnitt wie auch durch eine bessere Tonqualität von GMD gesteigert werden. Entgegen der getroffenen Annahmen wird die Einstellung zur Musik lediglich durch die Tonqualität beeinflusst. Ist diese nicht ausreichend vorhanden, lässt sie die musikalische Leistung des Orchesters oder des Künstlers schlechter erscheinen als sie wirklich ist. Aufgrund einer schlechten Tonqualität werden bspw. Dynamik und Klangfarbe der Musik reduziert, was schließlich zu einer schlechteren Einstellung führt.

Die Einstellung zur Musik wird direkt kaum durch die Länge des bereitgestellten Musikausschnitts beeinflusst. Es scheint, als könne diese auch bei einem kurzen Musikausschnitt positiv ausfallen. Allerdings wird die Einstellung indirekt über das wahrgenommene Wissen beeinflusst. Ein kürzerer Musikausschnitt schmälert die Wissensgrundlage zur Beurteilung der Musik. Wenn die Wissensgrundlage als nicht ausreichend empfunden wird, kann dies die Einstellung zur Musik beeinträchtigen. Bezüglich der Wissensbildung sollten GMD also möglichst das ganze Musikstück in einer guten Tonqualität umfassen.

Hinsichtlich der aufgeführten Effekte ist allerdings zu berücksichtigen, dass sich die Downloadzeit von GMD verlängert je besser die Tonqualität und je länger der bereitgestellte Musikausschnitt ist. Hier besteht eine Trade-Off Situation in Bezug auf die Usability. Ist die Downloadzeit zu lange, wird der Konsument den Download abbrechen, was nicht die gewünschten Verhaltensabsichten mit sich bringen würde.

Die Resultate bestätigen die grundsätzliche Eignung von GMD als Marketing- und Kundenbindungsinstrument sowie die Wirkung der beiden Gestaltungsgrößen Tonqualität und Ausschnittslänge. Für das Marketing eines klassischen Orchesters oder Künstlers weisen sie deshalb große Potenziale auf, welche im nächsten Abschnitt genauer erläutert werden.

5.2 Praxisimplikationen

Wie die vorangehend präsentierten Resultate und Interpretationen gezeigt haben, bieten GMD Orchestern und Künstlern ein effizientes Marketinginstrument zur Bindung von Klassikhörern. Es beeinflusst nicht nur die Bildung einer Wissensbasis und positiveren Einstellung, sondern fördert auch die für ein Orchester sehr wichtigen Verhaltensabsichten wie die Kauf- und Konzertbesuchsabsicht sowie die Absicht zur Weiterempfehlung der Musik.

Dabei sollten Orchester oder Künstler ihre Musik jeweils in ausreichend guter Tonqualität zur Verfügung stellen, um so eine positive Einstellung und das gewünschte Verhalten zu fördern. Bezüglich Stücklänge empfiehlt es sich im Hinblick auf die Wissensbildung ebenfalls, ein ganzes Stück zum Download anzubieten. Dadurch wird die Bildung des wahrgenommenen Wissens unterstützt. Dies insbesondere auch im Hinblick darauf, dass die Konsumenten so selber bestimmen können welchen Teil des Stücks sie hören möchten und nicht nur einen vorbestimmten Ausschnitt erhalten. In der Klassik erscheint dies besonders wesentlich, da die Stücke komplex sind und das Thema nicht bereits in den ersten Sekunden vorkommt. Die normalerweise in der Praxis bereitgestellten kurzen Ausschnitte des Stückanfangs werden diesem Problem also nicht gerecht. Zur Optimierung der Wirkung werden deshalb am besten einige Stücke ganz und in guter Qualität zur Verfügung gestellt. Fürchtet man, durch die Bereitstellung des ganzen Stückes in guter Qualität die Musikverkäufe zu kanibalisieren, so könnte allenfalls ein Musikstück auch in zwei Downloads, nämlich einmal ganz in schlechter Qualität und einmal ein Ausschnitt in guter Qualität, angeboten werden [36].

5.3 Limitationen

Eine wichtige Limitation der Untersuchung betrifft ihre eingeschränkte Repräsentativität. Aufgrund der Verteilung des Fragebogens durch verschiedene Gruppierungen und Institutionen mittels Schneeballprinzip, kann nicht auf eine bestimmte Grundgesamtheit geschlossen werden. Tendenziell wurden eher junge bereits bestehende Kunden des Klassikmarktes erreicht. Damit konnte der Nutzen der GMD zur Bindung dieser Kunden aufgezeigt werden. Zwar wird vermutet, dass sich durch GMD auch neue Klassikhörer gewinnen lassen. Aufgrund der für diese Studie verwendeten Stichprobe können dazu aber keine Schlüsse gezogen werden. Zudem wurden die erhobenen Zusammenhänge lediglich für die klassische Musikbranche in der Schweiz untersucht. Ob GMD anderer Musiksparten dieselben Wirkungen auf Konsumenten haben, müsste gesondert geprüft werden.

Im Rahmen der experimentellen Umfrage wurde eine künstliche Situation geschaffen, indem die Probanden gebeten wurden, einen bestimmten Musikdownload auszuprobieren. Um die Wirkung der Tonqualität auch bei schlechten Abhörbedingungen demonstrieren zu können, wurde eine unrealistisch hohe

Kompressionsrate gewählt und diese zudem noch durch einen vorgängigen Download hoher Qualität betont. Dieses artifizielle Design ist dadurch zu rechtfertigen, dass mit der Untersuchung herausgefunden werden sollte, ob Tonqualität bei der Wissens- und Einstellungsbildung eine Einflussgröße sein kann. Bei real angebotenen Downloads sind allerdings bessere Qualitäten üblich, wie etwa 96 oder 120 kbits, und eine Qualitätsreferenz ist auch nicht gegeben. Hier wäre schon fraglich, ob Benutzer beim Abhören über Computer mit angeschlossenen Boxen überhaupt Qualitätsunterschiede gegenüber der CD-Qualität wahrnehmen. Diese würden sich allenfalls offenbaren, wenn man den Download auf eine Hifi-Musikanlage übertrüge, was derzeit wohl die wenigsten Musikliebhaber praktizieren. Die Frage nach der ausreichend guten Qualität für einen Sample-Download kann somit durch die Untersuchung nicht beantwortet werden. In abgeschwächter Form gilt dies auch für die Ausschnittlänge, wofür ebenfalls zwei Extremwerte gewählt wurden.

Weiterhin wurde der Download nicht aktiv gesucht, wodurch keine Schlüsse über die Motivation zum Download gezogen werden können. Neben möglichen Motiven hätten auch noch weitere Variablen im Modell berücksichtigt werden können, was die gemessenen Werte der Determinationskoeffizienten vielleicht noch erhöht hätte. Beispiele dafür sind die bereits angesprochene Downloadgeschwindigkeit, der verwendete Internetbrowser oder die Zugriffszeit der Konsumenten auf die GMD.

Schließlich muss der Vollständigkeit halber noch erwähnt werden, dass in der vorliegenden Studie lediglich Verhaltensabsichten erhoben wurden. Mit einem solchen Untersuchungsdesign lässt sich nicht mit letzter Sicherheit auf das tatsächliche Verhalten der Konsumenten schließen. Wenn z.B. ein Konsument nach einem GMD einen entsprechenden Tonträger kaufen oder sich ein Konzert anhören möchte, kann er diese Absicht vielleicht nicht realisieren, da es seine ökonomische Situation nicht zulässt oder da das Konzert viel zu weit entfernt stattfindet.

6. SCHLUSSBETRACHTUNGEN

6.1 Fazit

In der vorliegenden Studie wurden erstmals Voraussetzungen und Effekte von GMD im Bereich der klassischen Musik untersucht. Dabei wurde nicht wie in bisherigen Untersuchungen nur der Einfluss von GMD auf das Kauf- und Konzertbesuchsverhalten sondern auf verschiedene weitere Wirkungsgrößen überprüft. Neben diesen Auswirkungen konnten auch relevante Faktoren von GMD aufgezeigt werden, die erfüllt sein müssen, um die entsprechenden Effekt zu erzielen. Wo andere Studien GMD bisher in einem eher allgemeinen Kontext betrachtet haben, bestand das Ziel der vorliegenden Untersuchung darin, das Potenzial von GMD zur Kundenbindung im Klassikmarkt aufzuzeigen.

Die Ergebnisse des Experiments zeigen, dass GMD von klassischer Musik als Instrument zur Kundenbindung eingesetzt werden können. Sie beeinflussen beim Konsumenten nicht nur das Wissen über und die Einstellung zur Musik eines Orchesters oder Interpreten positiv sondern auch dessen Absichten zum Kauf, Konzertbesuch oder zur Weiterempfehlung der Musik. Dementsprechend können GMD Größen der einstellungs- wie auch der verhaltensorientierten Kundenbindung determinieren. Dabei sind besonders die Länge des bereitgestellten Musikausschnittes und die Tonqualität des GMD zu beachten.

6.2 Ausblick

In weiteren Studien wäre nun das tatsächliche Verhalten von Konsumenten nach GMD zu untersuchen. Dies könnte z.B. mittels einer Befragung von Kunden eines Online-Portals geschehen, wo klassische Musik zum einen als Sample zum Download bereitgestellt wird und zum anderen direkt gekauft werden kann. Speziell bezüglich dem Verkauf von CDs wäre es zudem sinnvoll zu wissen, in welcher Form GMD gesamthaft angeboten werden sollen. Ist es z.B. besser Ausschnitte von sämtlichen Musikstücken anzubieten oder sollten eher nur zwei, drei Musikstücke in voller Länge bereitgestellt werden? Dabei sind stets auch die Besonderheiten von klassischer Musik an das Marketing eines Orchesters bzw. Künstlers zu beachten. Wie schon erwähnt, suchen Laien in der Klassik häufiger nach bestimmten Stücken, denn nach Künstlern oder Orchestern, was bezüglich der Auffindbarkeit von GMD spezielle Anforderungen an das Marketing stellt. Solche und weitere Unterschiede gegenüber anderen Musikgenres gilt es in weiteren Forschungsarbeiten zu beachten.

Um die Ergebnisse dieser Studie zu verallgemeinern, ist das Modell im Kontext von zusätzlichen Musikmärkten zu testen. Dabei sind besonders auch technische Einflussfaktoren wie die Downloadgeschwindigkeit, die Internetverbindung oder der verwendete Browser zu berücksichtigen. Je besser die Funktionsweise und die Wirkungen von GMD schließlich verstanden werden, desto besser lässt sich deren Potenzial nutzen und so vom internetbedingten Wandel der Musikindustrie profitieren.

7. ACKNOWLEDGMENTS

Die Autoren bedanken sich bei der Camerata Bern für die hilfreiche und tatkräftige Unterstützung der Studie.

8. REFERENCES

[1] Ajzen, I. 1991. The Theory of Planned Behaviour. *Organizational Behaviour and Human Decision Processes* 50, 2 (Dez. 1991), 179-211.

[2] Barclay, D., Higgins, C., Thompson, R. 1995. The Partial Least Squares (PLS) Approach to Causal Modeling: Personal Computer Adoption and Use as an Illustration. *Technology Studies* 2, 2 (1995), 285-309.

[3] BfS, Bundesamt für Statistik 2008, Kulturverhalten in der Schweiz: Musik. DOI= http://www.bak.admin.ch/aktuelles/03026/03038/index.html?lang=de.

[4] Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R. 2003. Economic of Online Music. In *Proceedings of the 5th Annual International Conference on Electronic Commerce* (Pittsburgh, PA, Sept. 30 - Okt. 03, 2003).

[5] Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R. 2003. No More Shadow Boxing with Online Music Piracy: Strategic Business Models to Enhance Revenues. In *Proceedings of the 36th Annual Hawaii International Conference on System Science* (Big Island, HI, Jan. 06 - 09, 2003).

[6] Bhattacharjee, S., Gopal, R. D., Sanders, G. L. 2003. Digital Music and Online Sharing: Software Piracy 2.0? *Communications of the ACM* 46, 7 (Jul. 2003), 107-111.

[7] Bounie, D., Bourreau, M., Waelbroeck, P. 2005. *Pirates or Explorers? Analysis of Music Consumption in French Graduate Schools*. Working Paper Nr. EC-05-01. Telecom ParisTech, Paris, Frankreich.

[8] Brosius, F. 2008. *SPSS 16: Das mitp-Standardwerk*. MITP, Heidelberg.

[9] Coyle, J., Thorson, E. 2001. The Effect of Progressive Levels of Interactivity and Vividness in Web Marketing Sites. *Journal of Advertising* 30, 3 (2001), 65-77.

[10] Davis, F. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 13, 3 (Sept. 1989), 319-340.

[11] Dick, A. S., Basu, K. 1994. Customer Loyalty: Toward an Integrated Conceptual Framework. *Journal of the Academy of Marketing Science* 22, 2 (Mrz. 1994), 99-113.

[12] EOV, Eidgenössischer Orchesterverband 2009, Mitgliederstatistik. DOI=http://www.eov-sfo.ch/de/start/index.htm.

[13] Farchy, J. 2004. P2P and Piracy: Challenging the Cultural Industry's Financing System. *Review of Economic Research on Copyright Issues* 1, 2 (Dez. 2004), 55-69.

[14] Findahl, O. 2006. *Thieves or Customers? File-sharing in the Digital World*. Working Paper. Upsala University, Upsala, Schweden.

[15] Fox, M. 2004. E-Commerce Business Models for the Music Industry *Popular Music and Society* 27, 2 (Jun. 2004), 201-220.

[16] Gopal, R. D., Bhattacharjee, S., Sanders, G. L. 2006. Do Artists Benefit from Online Music Sharing? *Journal of Business* 79, 3 (Mai 2006), 1503-1534.

[17] Hamann, T. K. 2005. Die Zukunft der Klassik. *Das Orchester* 9 (Sept. 2005), 10-19.

[18] Homburg, C., Bruhn, M. 2008. Kundenbindungsmanagement - Eine Einführung. In: Bruhn, M., Homburg, C. (Hrsg.), *Handbuch Kundenbindungsmanagement - Strategien und Instrumente für ein erfolgreiches CRM*. 6. Auflage, Gabler, Wiesbaden, 5-37.

[19] Howard, J. A., Sheth, J. H. 1969. *A Theory of Buyer Behaviour*. John Wiley & Sons Inc., New York, NY.

[20] Herrmann, A., Huber, F., Kressmann, F. 2006. Varianz- und kovarianzbasierte Strukturgleichungsmodelle – Ein Leitfaden zu deren Spezifikation, Schätzung und Beurteilung. *Zeitschrift für betriebswirtschaftliche Forschung* 58, 2 (2006), 34-66.

[21] Huber, F., Herrman, A., Meyer, F., Vogel, J., Vollhardt, K. 2007. *Kausalmodellierung mit Partial Least Squares*. Gabler, Wiesbaden.

[22] Jäggi, N., Zaugg, A. D. 2006. Channel-Specific Consequences of Customer Loyalty on Information Search Behaviour. In *Proceedings of 7th International We-B Conference* (Melbourne, Australia, Nov. 29 - Dez. 01, 2006).

[23] Kreutzer, R. T. 2003. Konzeption und Positionierung des Couponing im Marketing. In Hartmann, W., Kreutzer, R. T., Kuhfuss, H. (Hrsg.), *Handbuch Couponing*. 1. Auflage, Gabler, Wiesbaden, 3-25.

[24] Krueger, A. B. 2005. The Economics of Real Superstars: The Market for Rock Concerts in the Material World. *Journal of Labor Economics* 23, 1 (2005), 1-30.

[25] Kumar, V., Shah, D. 2004. Building and Sustaining Profitable Customer Loyalty for the 21st Century. *Journal of Retailing* 80, 4 (2004), 317-329.

[26] Liebowitz, S. J. 2006. File Sharing: Creative Destruction or Just Plain Destruction? *Journal of Law and Economics* 49, 1 (Apr. 2006), 1-28.

[27] Meffert, H. 2008. Kundenbindung als Element moderner Wettbewerbsstrategien. In Bruhn, M., Homburg, C. (Hrsg.), *Handbuch Kundenbindungsmanagement - Strategien und Instrumente für ein erfolgreiches CRM*. 6. Auflage, Gabler, Wiesbaden, 159-180.

[28] Molteni, L., Ordanini, A. 2003. Consumption Patterns, Digital Technology and Music Downloading. *Long Range Planning* 36, 4 (Aug. 2003), 389-406.

[29] Nyffeler, M. 2009. Beethoven im weltweiten Netz. *NZZ Online*. DOI= http://www.nzz.ch/nachrichten/kultur/aktuell/beethoven_im_weltweiten_netz_1.2284445.html.

[30] Oliver, R. L. 1999. Whence Consumer Loyalty? *Journal of Marketing* 63, 4 (Okt. 1999), 33-44.

[31] Peitz, M., Waelbroeck, P. 2004. *The Effect of Internet Piracy on CD-Sales: Cross-Section Evidence*. CESifo Arbeitsbericht Nr. 1122. Universität Mannheim, Mannheim.

[32] Roehm, M. L., Sternthal, B. 2001. The Moderating Effect of Knowledge and Resources on the Persuasive Impact of Analogies. *Journal of Consumer Research* 28, 2 (Sept. 2001), 257-272.

[33] Sandulli, F. D. 2007. CD Music Purchase Behaviour of P2P Users. *Technovation* 27, 6-7 (Jun.-Jul. 2007), 325-334.

[34] Sinha, R. K., Mandel, N. 2008. Preventing Digital Music Piracy: The Carrot or the Stick? *Journal of Marketing* 72, 1 (Jan. 2008), 1-15.

[35] Smith, D., Park, W. 1992. The Effects of Brand Extensions on Market Share and Advertising Efficiency. *Journal of Marketing Research* 29, 3 (Aug. 1992), 296-313.

[36] Tu, Y., Lu, M. 2006. An Experimental and Analytical Study of On-Line Digital Music Sampling Strategies. *International Journal of Electronic Commerce* 10, 3 (2006), 39-70.

[37] Zentner, A. 2008. Online Sales, Internet Use, File Sharing, and the Decline of Retail Music Specialty Stores. *Information Economics and Policy* 20, 3 (Sept. 2008), 288-300.

Kontextsensitive Service-Infrastruktur für die mobile Nutzung von Home-IT

Marc Herbrechter
Wirtschaftsinformatik und neue Medien
Universität Siegen
marc.herbrechter@uni-siegen.de

Benedikt Ley
Wirtschaftsinformatik und neue Medien
Universität Siegen
benedikt.ley@uni-siegen.de

Martin Stein
Wirtschaftsinformatik und neue Medien
Universität Siegen
martin.stein@uni-siegen.de

ZUSAMMENFASSUNG

Die Digitalisierung der Medienlandschaft und die damit einhergehenden Veränderungen für die Home-IT ermöglichen eine Vielzahl neuer Dienste. Beispielsweise sind Fernsehgeräte nicht mehr nur reine Informations- und Unterhaltungsmedien sondern bieten durch Netzwerkfähigkeit bereits heute eine Reihe von interaktiven Zusatzfunktionalitäten. Um solche Dienste den Nutzern sinnvoll zugänglich zu machen wurde im Rahmen dieser Arbeit ein Framework entwickelt, das eine dynamische Service-Infrastruktur für Home-IT Umgebungen bereitstellt und die es Benutzern ermöglicht lokal verfügbare Dienste zu finden und diese über ihr mobiles Endgerät zu nutzen.

SCHLÜSSELWÖRTER

Home-IT, Kontextsensitive Systeme, Serviceorientierte Architekturen, Komponentensbasierte Architekturen, OSGi

1. EINLEITUNG

Die Digitalisierung der Medienlandschaft bringt eine Vielzahl an Veränderungen für die häusliche Anwendungs- und Informationstechnik (Home-IT) mit sich. Das Fernsehen ist weiterhin immer noch eines der am stärksten genutzten Informations- und Unterhaltungsmedien in der deutschen Gesellschaft. Gleichzeitig aber gewinnt das Internet als Unterhaltungs- und vor allem als Kommunikationsmedium immer mehr an Bedeutung [17]. Bereits heute integrieren viele Fernsehgeräte, häufig in Kombination mit Media-Center PCs oder Set-Top Boxen, eine Netzwerkschnittstelle und ermöglichen somit zahlreiche internetbasierte Zusatzfunktionen wie Webradio, On-Demand Videodienste oder den Zugang zu sozialen Netzwerken. Die damit verbundene Veränderung des Nutzungsverhaltens stellt allerdings gleichzeitig neue Anforderungen an die Gestaltung von Benutzerschnittstellen um solche interaktiven Mehrwertangebote sinnvoll nutzen zu können. Wir haben ein Framework konzipiert, das eine dynamische Service-Infrastruktur für Home-IT Umgebungen bereitstellt und die Nutzung von Diensten erleichtert. Über ein Smartphone oder ein anderes mobiles Endgerät können die Benutzer erschließen, welche Dienste verfügbar sind und die für sie Interessanten auf dem Endgerät installieren, um z.B. eine Funktion auf einem lokalen Media Center PC nutzen zu können. Damit die Benutzer nicht mit einer zu großen Auswahl an Diensten überfordert werden, verfolgen wir darüber hinaus einen kontextsensiblen Ansatz, so dass nur solche Dienste angezeigt werden, die in der unmittelbaren Umgebung des Anwenders verfügbar und nutzbar sind. Das Framework basiert auf einer komponentenbasierten Architektur, die eine plattformunabhängige Nutzung sowie das dynamische Hinzufügen und Entfernen von Diensten ermöglicht. Gleichzeitig werden aber auch die für eine serviceorientierte Architektur (SOA) typischen Kommunikationsmuster zwischen Service-Consumer, Service-Broker (Service-Registry) und Service-Provider implementiert, um den Benutzern dynamisch neue oder aktualisierte Dienste anzubieten. Eine Softwarekomponente entspricht hierbei einem Service, der jedoch für die Nutzung lokal auf dem Endgerät installiert wird, so dass auch native Funktionalitäten der Endgeräte genutzt werden können.

2. STAND DER FORSCHUNG

Existierende Arbeiten haben gezeigt, dass Fernsehen keine isolierende Aktivität ist, sondern Gesprächsthemen bereitstellt und damit soziale Interaktion fördert, die dank Informationstechnologie auch trotz großer Distanzen unterstützt werden kann [10]. Es wurde außerdem deutlich, dass auch eine parallele Nutzung von Internetdiensten und Fernsehen von Interesse ist, beispielsweise bei Chat-Anwendungen [14].

Die wachsende Verfügbarkeit von Diensten auf immer mehr Endgeräten erfordert in Verbindung mit der Erwartung der Benutzer nach immer stärker personalisierten Lösungen, die deren Wünsche und Bedürfnisse widerspiegeln [24], hoch dynamische Strukturen. Es wurden verschiedene Paradigmen entwickelt, die Anpassung und Erweiterung bestehender Anwendungssysteme erleichtern. Geeignete Ansätze bieten beispielsweise serviceorientierte Architekturen oder Komponentenmodelle, die durch Abstraktion von Services bzw. Komponenten gut in heterogenen und sich wandelnden Infrastrukturen eingesetzt werden können. Komponentenmodelle basieren auf Softwarekomponenten, welche gemäß eines Kompositions-Standards mit anderen Komponenten verknüpft und ausgeführt werden können, ohne dass Änderungen an den Komponenten vorgenommen werden müssen [11]. Demnach besitzen Softwarekomponenten definierte Schnittstellen über die mit anderen Komponenten

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

kommuniziert werden können. In welcher Form sich eine solche Komponente darstellt, hängt vom darunterliegenden Komponentenmodell ab. Sollten Komponenten nicht vollständig unabhängig ausgeführt werden können, so werden Abhängigkeiten mit benötigten zusätzlichen Komponenten explizit durch die Komponenten angegeben. Im Vergleich dazu sind SOA ein Paradigma für die Verteilung von Anwendungen, wobei einzelne Funktionalitäten (Services) bei unterschiedlichen Anbietern liegen können. SOA spezifizieren einheitlich, wie solche Funktionalitäten bzw. Dienste angeboten und gefunden werden können, wie mit diesen interagiert werden kann und wie diese genutzt werden können [3].

Eine Integration von serviceorientierten Ansätzen und komponentenbasierten Modellen wird bereits im „Gravity" Projekt verfolgt [4]. Die Autoren beschreiben ein Komponentenmodell bei dem die Kommunikation zwischen den Komponenten über Services abläuft. Abhängig von der Verfügbarkeit bestimmter Services kann so eine dynamische Anpassung der Anwendung ermöglicht werden. Eine weitere Arbeit von Sánchez-Nielsen et al. beschreibt ein Framework auf der Basis von SOA, welches speziell für mobile Endgeräte entwickelt wurde und die dynamische Integration von Services sowie das dynamische Auffinden neuer Services unterstützt [22].

Ansätze in denen komponentenbasierte Frameworks mit Multimediaanwendungen verknüpft wurden, finden sich beispielsweise in der Arbeit von Redondo et al. Sie beschreiben eine kooperative Umgebung zwischen OSGi (Open Service Gateway Initiative) als komponentenbasierte Plattform und MHP (Multimedia Home Platform). Die entwickelte Plattform ermöglicht auf Basis von OSGi die Nutzung von Diensten verschiedener Geräte im Haushalt innerhalb von MHP Anwendungen [21]. Auch Gu et al. verwenden eine OSGi-Plattform. In ihrer Arbeit stellen sie eine Infrastruktur für kontextsensible Anwendungen vor. Das dynamische Verhalten dieser Anwendung basiert dabei auf einem formalen ontologiebasierten Kontextmodell [9].

Passen sich Systeme an den jeweiligen Kontext an so spricht man von kontextsensiblen Systemen. Nach Dey handelt es sich um ein kontextsensibles System, „wenn es den Kontext benutzt, um dem Nutzer relevante Informationen und/oder Dienste zur Verfügung zu stellen, wobei die Relevanz aufgabenabhängig ist"[7]. Um den Kontext bewerten zu können, müssen Informationen des Umfeldes abstrahiert und formalisiert werden. Dabei werden nach Schilit et al. und Pascoe folgende Aspekte betrachtet: Wo ist der Nutzer, welche Personen sind in der Nähe und welche Ressourcen sind verfügbar [19,23]? In unserer Arbeit wird der Kontext auf der Basis von ortsbezogenen und infrastrukturellen Informationen gemessen.

Die Erfassung ortsbezogener Kontextinformationen kann im Freien heute problemlos mit GPS realisiert werden. Die Lokalisierung innerhalb geschlossener Gebäude hingegen stellt immer noch ein Problem dar. Cheng stellt in seiner Arbeit einen möglichen Lösungsansatz vor, der eine Ortung auf der Basis des Funknetzstandards ZigBee realisiert [5]. Auch das System „Locawe" setzt bei der Ortung innerhalb von Gebäuden auf ZigBee-Netzwerkinformationen auf, jedoch zusätzlich kombiniert mit WLAN-Netzwerkinformationen [15]. Eine andere Möglichkeit wurde von Betz et. al entwickelt. Sie beziehen die Kontextinformationen über RFID-Tags, die an Gegenständen befestigt werden und zu diesem Informationen beinhalten. Die ortsbezogene Kontextinformation die durch das Auslesen eines Tags erfasst wird stellt hierbei keine absolute Ortsangabe dar, da die ortsbezogene Information an den getaggten Gegenstand gebunden ist. Wird dieser bewegt, ändert sich auch die Position des Tags [2].

Neben der Identifizierung des räumlichen Kontexts existiert das Problem der Darstellung der im Kontext verfügbaren Dienste. Es bestehen verschiedene Möglichkeiten solche Dienste zu visualisieren, jedoch muss die Darstellung zum einem generisch genug sein um für Geräte bzw. Dienste unterschiedlichster Art spezifiziert werden zu können, aber zum anderem nicht zu abstrakt sein, um für den Nutzer gebrauchstauglich zu bleiben. Ein Ansatz zur Realisierung einer solchen generischen Visualisierung wird in [16] dargestellt. Die Autoren stellen eine geräteunabhängige Sprache zur Spezifizierung von Oberflächen vor, die zudem die Möglichkeit bietet, Oberflächen gemäß den Anforderungen des Nutzers zu individualisieren. Die Sprache eignet sich zur Beschreibung der Funktionalität von über 30 verschiedenen Gerätearten, wie Fernsehgeräten, Videorecordern und Kopierern, aber auch zur Steuerung von Microsoft Powerpoint und des Windows Media Players. Als Konzept für den Einsatz der Beschreibungssprache stellen die Autoren folgenden Ablauf vor:

1. Das zu steuernde Gerät überträgt eine Beschreibung seiner Funktionen in der vorgestellten Beschreibungssprache.
2. Das mobile Endgerät generiert aus dieser Beschreibung eine Oberfläche, die vom Benutzer individuell angepasst werden können.
3. Durch Betätigen der generierten Steuerelemente werden, die durch die Beschreibung assoziierten Steuersignale an das zu bedienende Gerät übertragen.

Die Arbeit stellt einen Ansatz zur Steuerung von stark heterogenen Umgebungen dar. Im Vergleich dazu nutzen wir für unseren Ansatz ein ähnliches Konzept zur Beschreibung von Benutzerschnittstellen, die jedoch nicht den Kernfokus dieser Arbeit bilden und daher nur rudimentär erarbeitet wurden.

Problematisch bei der Entwicklung mobiler Anwendungen sind die begrenzten Ressourcen mobiler Endgeräte (Rechenleistung, Speicherkapazität, Akkuleistung etc.). Es ist daher sinnvoll, dass Anwendungen nur dann laufen, wenn dies auch wirklich erforderlich ist. Um dies zu erleichtern werden u.a. „Push"-Dienste eingesetzt. Sie ermöglichen es, dass ein mobiles Endgerät bzw. eine Anwendung erst dann aktiv wird, wenn bestimmte Umstände eintreten, wie bspw. die Aktualisierung von Daten oder der Erhalt einer Nachricht. Hierfür stellen verschiedene Plattformanbieter mobiler Endgeräte Push-Server bereit, über die Anwendungsentwickler Benachrichtigungen an die jeweiligen Endgeräte schicken können, ohne dass die entsprechende Anwendung auf dem Gerät gestartet sein muss [8][1].

3. HOME-IT FRAMEWORK

Im folgenden Abschnitt wird auf das grundlegende Konzept für die Plattform und unsere prototypische Umsetzung eingegangen. Als grundlegendes Gerüst für die Implementierung wurde das

OSGi Framework ausgewählt, welches später kurz erläutert werden soll. Auf dieser Grundlage wurden verschiedene Komponenten entwickelt, die zum einen grundlegende Dienste zur Steuerung, Kommunikation und Bedienung bereitstellen und zum anderen die Entwicklung darauf aufbauender Pakete strukturieren und erleichtern.

3.1 KONZEPT

In dieser Arbeit stellen wir einen integrierten Ansatz zur Nutzung und Steuerung von Home-IT Technologie auf mobilen Endgeräten vor. Hierbei sollen zum einen den individuellen Anforderungen heterogener Home-IT Umgebungen Rechnung getragen werden und zum anderen durch ein plattformunabhängiges Design beliebige mobile Endgeräte unterstützt werden. Die Anpassung an verschiedene Umgebungen zur Laufzeit erfordert eine hohe Dynamik, da ein Umgebungswechsel des Benutzers häufig mit einer vollständigen Veränderung der vorhandenen technischen Infrastruktur und den damit verbundenen Diensten und Funktionalitäten einhergeht. Die hohe Komplexität heterogener Umgebungen verhindert, dass ein Service oder eine Funktion (z.B. zur Steuerung eines bestimmten Fernsehgerätes) in beliebigen Home-IT Umgebungen genutzt werden kann. In unserer Arbeit haben wir ein Konzept für ein Framework entwickelt, das umgebungsspezifische Dienste und Anwendungen den Benutzern auf deren mobilen Endgeräte zur Nutzung bereitstellt. Durch die Beschränkung auf lokal verfügbare Dienste wird vor allem die Komplexität der Visualisierung verfügbarer Services vermindert, so dass der Benutzer kontextrelevante Dienste schneller auffinden und nutzen kann.

Für die Konzeption eines solchen Frameworks bedarf es einer hoch flexiblen Architektur die es ermöglicht, dass Dienste und Anwendungen zur Laufzeit plattformunabhängig den Anwendern zur Nutzung bereitgestellt werden können. Bei unserem Ansatz setzen wir auf ein Konzept, das das Rollenmodell serviceorientierter Architekturen mit den Vorteilen komponentenbasierter Ansätze integriert. Dienste werden demnach nicht wie z.B. Webservices entfernt aufgerufen, sondern lokal auf dem mobilen Gerät des Benutzers installiert. Dies hat den Vorteil, dass auch native Sensoren und Funktionen des Endgeräts (z.B. Bewegungssensoren, Vibration, Kamera, Mikrofon) von der Anwendung genutzt werden können.

Analog zum Rollenkonzept serviceorientierter Architekturen wird zwischen drei verschiedenen Rollen differenziert [3]. Services sind dabei als einzelne Komponenten realisiert. Zunächst gibt es die Rolle des Service-Nutzers, welcher Services auswählt und nutzt. Eine Auswahl von Services wird vom Service-Broker bereitgestellt. Der Service-Broker hält dazu ein Verzeichnis von Komponenten bereit, welches verschiedene Metadaten wie Name, Version und Ort, zu den jeweiligen Softwarekomponenten beinhaltet. Der Service-Provider ist, der im Brokerverzeichnis spezifizierte Ort bzw. Server der Komponente. Der konzeptuelle Aufbau einer beispielhaften Home-IT Umgebung ist in **Abbildung 1** dargestellt. Betritt demnach ein Benutzer mit seinem mobilen Endgerät eine solche Umgebung, werden Informationen über verfügbare Dienste von lokalen Service-Brokern (z.B. TV-Gerät, Hi-Fi-Anlage) auf das Endgerät gepusht (1). Der Nutzer hat nun die Möglichkeit aus der dargestellten Liste die für ihn relevanten Dienste (z.B. Fernbedienung für TV-Gerät) auszuwählen und zu installieren. Dabei kann die zu installierende Komponente entweder direkt beim Service-Broker liegen oder aber auf einem beliebigen externen Server (2). Wurde eine Komponente erfolgreich installiert, so kann der Benutzer diese z.B. zur Steuerung des lokalen TV-Geräts verwenden. Beim Verlassen der Home-IT Umgebung werden die installierten Anwendungen wieder entfernt bzw. ausgeblendet, so dass dem Benutzer immer nur kontextrelevante Dienste angezeigt werden. Das Installieren und entfernen von Dienste erfolgt dabei zu Laufzeit, ohne dass ein Beenden oder Neustarten des Geräts erforderlich wird.

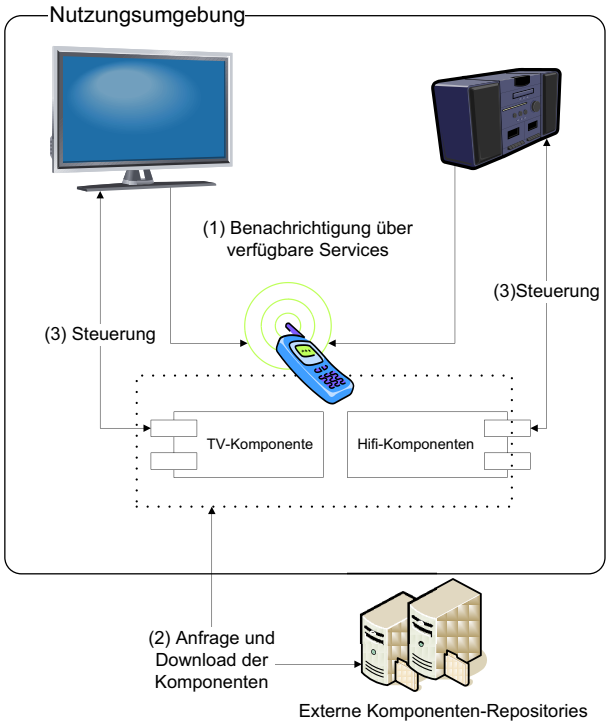


Abbildung 1: konzeptueller Aufbau der Serviceinfrastruktur für Home-IT Umgebungen

3.1.1 Architektur

Für die Realisierung eines solchen hochdynamischen, kontextsensitiven Frameworks bedarf es einer architektonischen Basis, die die Konzepte von SOA und komponentenbasierten Architekturen integriert und die erforderliche Flexibilität bereitstellt. Daher haben wir uns bei der Konzeption und der späteren prototypischen Realisierung des Home-IT Frameworks für den Einsatz von OSGi entschieden

OSGi ist ein auf Java basiertes Framework für dynamische, komponentenbasierte Softwareentwicklung. Es erlaubt das dynamische Laden so genannter Bundles zur Laufzeit, also ohne Neustart der Umgebung oder des Geräts. Weiter erlaubt es eine entfernte Wartung und Konfiguration sowie das entfernte

Installieren, Starten, Stoppen und Updaten von Bundles. Die Pakete werden von der Laufzeitumgebung implizit einem Lebenszyklus unterstellt und können nebenläufig oder interaktiv ausgeführt werden. OSGi als Komponentenmodell findet Anwendung in vielen renommierten Open-Source Projekten wie Eclipse. Zahlreiche namhafte Unternehmen wie IBM und SAP haben sich in der OSGI Alliance zusammengeschlossen, um die Standardisierung voran zu treiben.

Das OSGi Framework basiert auf einer mehrschichtigen Architektur, welche in Abbildung 2 dargestellt ist. Zur Bereitstellung von Sicherheitsaspekten setzt OSGi auf die Java 2 Security Architecture in Verbindung mit spezifischen OSGi Sicherheitsmechanismen. Hierzu können Bundles auf Basis einer Public Key Infrastructure mittels X.509 Zertifikaten gesichert werden um Aspekte wie Authentizität und Integrität zu gewährleisten. Die Sicherheitsschicht zieht sich Vertikal durch das gesamte Framework. Als Ausführungsumgebung kommen insbesondere Java Umgebungen wie die Java Standard Edition in Frage, wobei ein bestimmter Umfang der Java API vorausgesetzt wird. In der Lebenszyklusschicht wird eine API für Funktionen wie das Installieren, Starten, Stoppen, Updaten und Deinstallieren von Bundles bereitgestellt welche auch über entfernte Aufrufe erfolgen können. In der Serviceschicht werden Methoden für das Publizieren, Finden und Nutzen von Services bereitgestellt. Als serviceorientierte Architektur betrachtet ist diese Schicht mit einem Service Broker vergleichbar, wobei zusätzlich Persistenzdienste zur Verfügung stehen [18].

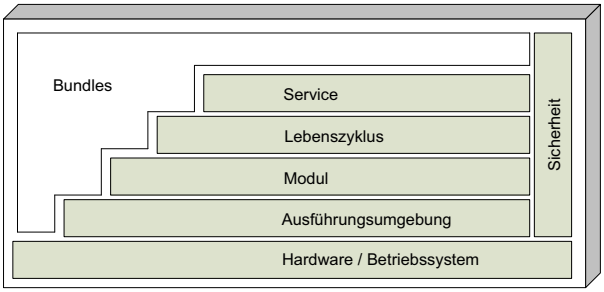


Abbildung 2: Schichten des OSGi-Frameworks

OSGi Bundles können unabhängig voneinander genutzt und ausgeliefert werden können. Dies ermöglicht den Entwicklern einen hohen Grad an Modularisierung, das insbesondere bei häufig vorkommenden partiellen Änderungen an Anwendungssystemen hilfreich sein kann. Durch einen hohen Grad an Wiederverwendbarkeit kann zudem der Implementierungsaufwand für zukünftige Module verringert und durch die Einbindung von standardisierten Kommunikationsformen die Interoperabilität – z.B. mit Webservices – gefördert werden.

Für das auf der OSGi Plattform aufbauende Home-IT Framework wurde eine nachrichtenorientierte Kommunikation gewählt, die auf dem Internetprotokollstapel aufbaut (s. Abbildung 3). Hierdurch wird der ereignisorientierte Nachrichtenaustausch abstrahiert, um auf verschiedenen Wegen stattfinden zu können. Für die Kommunikation zwischen mobilem Endgerät und Basisstation wurde eine einfache Implementierung auf Socket Basis gewählt während für die Kommunikation zwischen Media Center und Basis-Station Webservices zum Einsatz kommen.

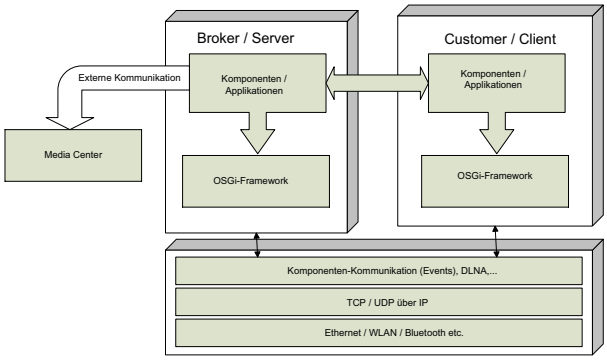


Abbildung 3: Aufbau von Server und Client und deren Kommunikation untereinander und mit externen Systemen

Durch den hohen Grad der Modularisierung des Gesamtsystems ist es wichtig, einen einheitlichen Weg der Ereignisverarbeitung zu schaffen, um komponentenübergreifend und transparent auf Ereignisse zu reagieren. Um dies sicherzustellen wurde das Konzept der komponentenbasierten Softwareentwicklung mit Konzepten der nachrichtenorientieren Kommunikation kombiniert. Ereignisse die innerhalb einer Komponente des Gesamtsystems auftreten, werden von dieser an eine zentrale Komponente weitergegeben, die aufgrund der Informationen innerhalb der Nachricht entscheidet, über welchen Kommunikationsweg und an welche Komponente das Ereignis weitergegeben werden soll.

3.1.2 Generische Oberflächenbeschreibung

Da Komponenten plattformunabhängig auf unterschiedlichen Geräten ausgeführt werden sollen, ist es erforderlich, dass die Beschreibung der grafischen Oberflächen der Anwendungen als Bestandteil der Komponente im Bundle mit abgelegt werden, so dass diese auf den jeweiligen Geräten interpretiert und nativ dargestellt werden können. Hierfür verwenden wir eine XML-basierte Beschreibungssprache (ähnlich XUL).

3.2 PROTOTYPISCHE UMSETZUNG

Die Kommunikation zwischen den Hauptkomponenten Home-IT Geräte, Basis Station und mobilem Endgerät wurde nach dem Client- und Server Paradigma realisiert wobei unterschiedliche Techniken zum Einsatz kommen. Um die Kommunikation zwischen Basis und mobilem Endgerät möglichst leichtgewichtig zu halten wird diese über Socketverbindungen auf Basis von TCP und UDP umgesetzt. Die Kommunikation zwischen Basis und Media Center wird aufgrund der einfacheren Handhabbarkeit durch Webservices realisiert (Basis und Home-IT Gerät können hierbei auch in einem Gerät vereint sein). Über diese Kommunikationskanäle werden ereignisorientiert Nachrichten versandt, welche im Framework selber generiert und ausgewertet werden können. Der Einfachheit halber werden diese Nachrichtenobjekte als JSON Repräsentationen auf den jeweiligen Kanälen übertragen [6].

Für die Umsetzung auf den mobilen Endgeräten wurde Google Android als exemplarische Plattform gewählt sowie die OSGi

Implementierung der Firma Prosyst GmbH [20]. Diese Kombination stellt zum Zeitpunkt der Veröffentlichung den größten Funktionsumfang zur Verfügung und erlaubt die schnelle und einfache Implementierung verschiedener Pakete für die exemplarischen Nutzungsfälle.

Da Google Android eine besondere Form der Java Umgebung bereitstellt, sind einige Besonderheiten zu beachten und diverse Änderungen an den Bundles vorzunehmen. Durch die bei der Android Plattform verwendete Dalvik Java VM muss der Bytecode vor dem Deployment umgewandelt werden um auf dem Zielsystem lauffähig zu sein. Weiter fehlen der Java Umgebung auf der Android Plattform diverse Pakete im Vergleich zur Java Standard Edition, wie etwa Java RMI. Dies erschwert die Benutzung von externen Bibliotheken oder macht diese gänzlich unmöglich. Aufgrund dessen wurde die Implementierung verschiedener Module, bei denen die Re-Implementierung bestehender Bibliotheken durch zu hohen Aufwand nicht sinnvoll wäre, nach dem Paradigma des Thin Client vorgenommen, um diese möglichst einfach und performant zu halten. Andere Pakete implementieren zudem gewisse Teile der Anwendungslogik im mobilen Endgerät, insbesondere um eine Unabhängigkeit von der Basis-Station zu erreichen.

Für die Umsetzung der Basis-Station wurde die Java Standard Edition in Verbindung mit dem Equinox OSGi Framework verwandt. Diese Kombination bietet einen großen Funktionsumfang und eine große Community. Durch die Verwendbarkeit von externen Bibliotheken zur Anbindung bestehender Dienste wie Twitter und Facebook auf der Basis Station, nicht jedoch auf den mobilen Endgeräten, bot sich eine Implementierung der Dienste dort an, um die Schnittstellen möglichst komfortabel nutzen zu können. Dies führt zu einer deutlich gesunkenen Einstiegshürde bei der Implementierung neuer Module und beschleunigt das Anfertigen von Prototypen. Inhalte dieser Module werden durch das Nachrichtensystem des Frameworks in komprimierter Form an das mobile Endgerät weitergereicht, das diese ohne weitere Verarbeitung sofort darstellen kann.

Auf der Basis dieses Frameworks wurde anschließend das grundlegende Bundle entwickelt das es ermöglicht, dass andere Komponenten (OSGi-Bundles) dynamisch installiert und genutzt werden können. Dabei ist die Komponente selbst durch zwei OSGi-Bundles realisiert. Eine Komponente dient nur zur Erzeugung und Darstellung der Oberfläche und die andere stellt die Fachlogik der Service-Komponente dar. Dadurch wird die Plattformunabhängigkeit gewährleistet.

Bei der Realisierung der Kontextsensitivität ist die Granularität der Umgebung durch das verwendete Übertragungsmedium (WLAN, Bluetooth, etc.) begrenzt. Die Kommunikation erfolgt über TCP/IP- oder UDP/IP-Protokolle. Im Verzeichnis des Brokers kann spezifiziert werden, welche Komponenten speziell in der aktuellen Umgebung nutzbar sind und ob Komponenten installiert werden müssen (Realisierung der Push-Semantik). Befindet sich der Nutzer bereits in einer Umgebung, während der Push einer Komponenten erfolgt, so wird dieser über spezielle Framework-Ereignisse darüber informiert und veranlasst die Installation. Zusätzlich können ortsunabhängige bzw. umgebungsunabhängige Komponenten angeboten werden, die ebenfalls installiert und genutzt werden können (Realisierung Pull-Semantik initiiert durch den Client).

Beim Verlassen der Umgebung werden ortsabhängigen Dienste wieder entfernt. Bei der Konzipierung des Prototyps wurde vorgesehen, dass Komponenten lediglich gestoppt und nicht wieder vollständig entfernt werden, jedoch hat sich während der Implementierung herausgestellt, dass ein Entfernen die momentan praktikabelste Lösung darstellt. Dies hängt u.a. damit zusammen, dass die frühe Version des OSGi-Frameworks, das verwendet wurde, noch keine ausgereiften Mechanismen für eine Versionskontrolle der Komponenten angeboten hat. Um hieraus resultierende Inkonsistenzen zu vermeiden, werden Komponenten vollständig deinstalliert und grundsätzlich neu installiert. In späteren Versionen wäre die zunächst geplante Realisierung vorteilhaft um unnötige Übertragungen zu vermeiden.

Broker in einer Umgebung geben sich durch Broadcast-Nachrichten bekannt. Betritt ein Client die Umgebung so registriert sich der Client beim Broker und speichert den Broker zur Kennzeichnung der Umgebung ab. Der Broker sendet daraufhin dem Client sein Verzeichnis von Komponenten (in XML-Repräsentation). Dieser führt anschließend die vorgeschriebenen Installationen und ggf. optionale Installationen durch.

Beim Verlassen der Umgebung (oder bei Unerreichbarkeit des Brokers) wird durch einen Time-Out die Deinstallation der ortsabhängigen Komponenten veranlasst. Der Service-Broker kann gleichzeitig Service-Provider sein, d.h. er hält die zu installierenden Komponenten bereit. Unterscheiden sich Broker und Provider, so ist sicherzustellen, dass der Provider zum Zeitpunkt der Installation verfügbar ist. Neben der Möglichkeit die Installation von Services vom Broker initiieren zu lassen (Service Push), können weitere Broker manuell hinzugefügt werden.

Abbildung 4: Verzeichnis, Quellenauswahl und Dialoges zum Hinzufügen von Quellen

4. ANWENDUNGSFALL: TVitter

Um den entwickelten Prototypen evaluieren zu können, wurde ein Beispieldienst implementiert der in Verbindung mit einem TV-Gerät in Verbindung mit einem MediaCenter PC die einfache Einbindung externer Social-Network Dienste zeigen soll. Dies wurde exemplarisch am Beispiel des Microblogging Dienstes Twitter umgesetzt.

Die Applikation besteht aus drei Bundles, die zum einen die Anbindung an die reguläre Twitter API und zum anderen die

Anbindung an eine Streaming API bereitstellen, die Informationen der Twitter Timeline in einem simulierten Push-Verfahren in Echtzeit zur Verfügung stellt. Das dritte Bundle dient der Steuerung auf dem mobilen Endgerät auf dem Beiträge verfasst und versandt werden können. Dieses Bundle orientiert sich explizit am Thin-Client Paradigma, um die Plattformunabhängigkeit gewährleisten zu können. Die Interaktion mit dem Webdienst Twitter geschieht allein auf der Basis-Station, die die entsprechenden Twitter-Bibliotheken nutzt und die notwendige Funktionalität bereitstellt. Auf dem Media Center PC wird zudem eine Applikation benötigt, die die Informationen der Basisstation annehmen und entsprechend darstellen kann. Dazu wird von der Basis-Station ein vorher spezifizierter Webservice aufgerufen, wobei die Implementierung für die Basis-Station keine Rolle spielt. Die Basis-Station kann auf dem gleichen Rechner laufen auf dem auch die Media Center Software läuft, so dass in diesem Fall beide Komponenten in einem Gerät vereint sind.

Ein beispielhafter Ablauf für eine Nutzung der Komponente könnte sich wie folgt darstellen: Der Anwender betritt den Raum des Media Centers und nutzt die TV Funktion um die Tagesschau anzusehen. In dieser wird über das Thema Steuererhöhungen berichtet, wodurch beim Anwender ein Interesse nach der Meinung Anderer geweckt wird, die die Sendung ebenfalls verfolgen. Dazu startet er auf seinem mobilen Endgerät die Anwendung TVitter, die sich bei der Streaming API mit dem aktuellen Sender als Suchbegriff registriert und die aktuellen Beiträge auf dem Bildschirm darstellt. Dies kann entweder in Form eines Overlay oder als separates Fenster erfolgen. Nun kann der Anwender sowohl die Meinung anderer Nutzer verfolgen als auch eigene Beiträge beisteuern, die den anderen Teilnehmern sowie ihm sofort angezeigt werden.

Auf technischer Ebene erzeugt das Starten der TVitter Komponente ein Ereignis auf dem mobilen Endgerät. Dieses Ereignis wird durch eine zentrale Routingkomponente über eine Socket-Verbindung an die Basis Station übergeben, die daraufhin über eine Webservice Schnittstelle die aktuellen Senderdaten der TV Applikation abruft. Diese werden ebenfalls als Nachricht auf dem umgekehrten Weg an die TVitter Applikation zurückgegeben, die daraufhin dem Benutzer kontextuelle Vorschläge zur Nutzung der Twittersuche unterbreitet. Weiterhin kann der Benutzer automatisch die Senderinformationen zur Annotation seiner Beiträge nutzen, was bei Twitter üblicherweise semi-strukturiert über so genannte Hash-Tags geschieht (bspw. #Tagesschau). Wechselt der Benutzer den Sender, wird dies von der Basisstation an das mobile Endgerät übermittelt und entsprechend neue Vorschläge unterbreitet.

5. EVALUATION

Der entwickelte Prototyp wurde abschließend unter kontrollierten Bedingungen im Labor evaluiert. Aufgrund der noch fehlenden Dienste und Geräte war es in dieser Phase noch nicht möglich die Evaluation unter realweltlich-alltäglichen Bedingungen durchzuführen. Außerdem war es nicht möglich ein reales Setting nachzustellen, da auch hierfür eine minimale Anzahl an Services und unterschiedliche Home-IT Umgebungen notwendig gewesen wären. Für die Evaluation wurden daher bewusst Lead-User nach [13] ausgewählt, also Benutzer die bereits ein grundlegendes technisches Wissen und Erfahrungen im Kontext neuer Unterhaltungs- und Kommunikationsmedien vorweisen können. Im Fokus der Evaluation stand demnach auch nicht die Verwendung des Systems im Anwendungskontext sondern mehr das Verständnis und die Sinnhaftigkeit des umgesetzten Konzepts sowie die Untersuchung des TVitter Dienstes.Durchführung

Die Evaluation wurde mit sechs Teilnehmern durchgeführt. Die Partizipanten wurden dazu in einem mit Fernseher und daran angeschlossenem Notebook ausgestatteten Raum interviewt und bei der Durchführung verschiedener Nutzungstests beobachtet. Die Befragung erfolgte auf der Basis semi-strukturierter Interviews mittels einem vorgegebenem Leitfaden, von dem auch abgewichen werden konnte. Um die Evaluation möglichst kontextnah zu gestalten, wurde diese in einem Labor durchgeführt, das einem Wohnzimmer nachempfunden ist (s. Abbildung und Bestandteil eines Living Lab-Konzepts an der Universität Siegen ist [12]. Alle Tests wurden für die spätere Analyse mit Audiogeräten aufgezeichnet sowie von einem passiven Beobachter protokolliert

Abbildung 5: Umgebung für Interviews und Evaluation

Der Test gliederte sich in drei Teile:

1. Pre-Test Interview
2. Nutzungstest des Prototyp
3. Post-Test Interview

Vor dem Test des Prototyps wurden demographische Informationen der Teilnehmer erfasst und diese zu ihren medialen Nutzungsgewohnheiten befragt. Hierbei lag der Fokus darauf, einen guten Überblick über die derzeitige Nutzungspraxis der Teilnehmer im Kontext von Home-IT und Medien zu gewinnen.

Nachfolgend wurde die prototypische Umsetzung präsentiert. Hierbei handelte es sich insbesondere um die Umsetzung der Direct Service Push Applikation und des TVitter Use-Case. Die Teilnehmer wurden mit dem jeweils in den Ausgangszustand versetzten Mobiltelefon gebeten den Raum und damit die Reichweite des WLANs zu verlassen, um kurz darauf wieder hereinzutreten. In diesem Moment wurden die zur Verfügung stehenden Services vorgeschlagen und installiert, im konkreten Nutzungstest ausschließlich die TVitter Anwendung. Diese wurde im weiteren Verlauf ausführlich erläutert und von den Teilnehmern genutzt. Dazu wurde beispielhaft eine Fernsehsendung eingeschaltet, die von anderen Twitter

Benutzern kommentiert wurde und zu der die Benutzer einen eigenen Beitrag verfassen sollten. Passend zu Sendung wurde mit der Wizzard-of-Oz Methode geeignete Suchbegriffe an das Mobiltelefon des Nutzers gesandt. Die Teilnehmer konnten nun entscheiden diese zu übernehmen, zu erweitern oder ganz zu verwerfen um einen eigenen Suchbegriff zu verwenden. Durch das Absenden der Suchanfrage wurden entsprechende Twitter-Beiträge unter dem Fernsehbild in einer transparenten Applikation dargestellt. Während der Nutzung wurde stets mit den Teilnehmern kommuniziert, um zu erfahren wie diese im Moment interagieren, wie das Konzept und dessen Umsetzung empfunden wird und ob Hilfestellung benötigt wird.

Im Anschluss an den Nutzungstest wurden die Teilnehmer ein zweites Mal befragt. Hierbei lag der Fokus auf dem Grundkonzept der Infrastruktur und der konkreten Implementierung der medialen Integration am Beispiel von Fernsehen und Twitter. Die Teilnehmer wurden gebeten ihre Haltung bzgl. der gezeigten Anwendung, Usability Aspekte und alternative Anwendungsszenarien zu erörtern. Hierbei wurden konkrete Vorstellungen und Änderungswünsche der Teilnehmer herausgearbeitet und detailliert besprochen.

5.1 Teilnehmer

Insgesamt wurden sechs Teilnehmer während der Evaluation befragt (s. Tabelle 1). Bei der Auswahl der Teilnehmer wurden keine besonderen Kriterien berücksichtigt, jedoch wurden bewusst Personen ausgewählt die eine hohe Technik-Affinität sowie Erfahrungen im Kontext neuer Unterhaltungs- und Kommunikationsmedien vorweisen konnten, weshalb die hier betrachtete Gruppe von Nutzern als Lead-User bezeichnet werden kann [13].

Tabelle 1: Anonymisierte Übersicht der Teilnehmer

Name	Beruf	Alter
Achim	Student	23
Nikolas	Student	25
Tobias	Doktorand	30
Andre	Doktorand	30
Kristofer	Student	24
Andrea	Fachleiterin	34

Die Teilnehmer waren im Durchschnitt 27,7 Jahre alt, besaßen alle einen hohen Bildungsgrad und waren an neuen Technologien und an Medien interessiert. Die Probanden arbeiten zum Teil auch in der Forschung oder Entwicklung, waren aber nicht an der vorliegenden Arbeit beteiligt. Alle sehen regelmäßig fern und nutzen unabhängig davon verschiedene soziale Netzwerke wie Facebook, Twitter oder Wer-Kennt-Wen.

5.2 Ergebnisse

Die Ergebnisse lassen sind in die folgenden drei Kategorien einteilen:

1. Anwendungsbezogen
2. Interfacebezogen
3. Soziologisch

Bei den anwendungsbezogenen Ergebnissen standen Usabilityaspekte im Vordergrund. Diese wurden durchweg als unzureichend und unbefriedigend empfunden. Grundlegende Kriterien der Dialoggestaltung wurden bemängelt und die Implementierung als sehr technisch empfunden. Vor dem Hintergrund, hier einen eher technisch orientierten Prototyp zum Test gestellt zu haben, waren diese Ergebnisse zu erwarten. In kommenden Versionen werden diese berücksichtigt und erneut überprüft bevor der Prototyp unter realen Bedingungen evaluiert wird.

Die zweite Ergebniskategorie bezieht sich auf Bedienung und Darstellung. Hierbei wird zwischen Ein- und Ausgabeinterface unterschieden, insbesondere weil am Beispiel der TVitter-Anwendung eine dichotome Trennung vorliegt. Auf der Eingabeseite wurden insbesondere die folgenden drei Aspekte angesprochen:

1. Textuelle vs. sprachliche Eingabe
2. Mobiltelefon als Mittler
3. Formfaktor für textuelle Eingabe

Ein Teilnehmer stellte gleich zu Beginn der Evaluation das Konzept der textuellen Eingabe in Frage, da diese im Kontext des Fernsehkonsums für sein Empfinden vermutlich nicht die optimale Form der Interaktion darstelle. Statt einer textuellen Eingabe wurde eine sprachgesteuerte Interaktion vorgeschlagen, abhängig vom Stand der Technik im Bereich Spracherkennung. Diese würde dem Teilnehmer als geringere Nutzungshürde erscheinen.

„…deshalb weiß ich nicht ob tippen hier die richtige Eingabeform ist, weil es zu lange dauert und das Fernsehen zu schnell ist…“ - Tobias

Das Mobiltelefon wurde weitgehend als Eingabemedium akzeptiert. Durch die ständige Verfügbarkeit erschien dieses Gerät als ideale Verknüpfung zu lokalen Technologien. Auch die Eingabeform mithilfe von Tastatur oder Handydisplay wurde als akzeptabel, wenn auch nicht ideal bewertet. Ein Trade-Off zwischen ständiger Verfügbarkeit und komfortabler Eingabe wurde von den Teilnehmern beschrieben und als solcher akzeptiert.

„…gerade wenn man das Gerät immer bei sich haben will, weil es eben location-basiert ist, denke ich ist ein Handy die beste Wahl…“ - Andre

Alternativ wurden von den Teilnehmern Eingabegeräte wie eine erweiterte Fernbedienung, klassische Tastaturen und Tablet Computer genannt. In allen Fällen wurden diese Eingabeformen als optional gesehen und stellten für die Teilnehmer eine Optimierung dar.

„Da habe ich nicht den Overkill einer normalen Tastatur.“ – Achim

Bei der Ausgabeform wurde wiederholt auf die Intrusivität der Einblendung von Informationen hingewiesen. Die Benutzer empfanden die Vorstellung der Überlagerung des Fernsehbildes durch die Nachrichten als störend und würden selber Steuern wollen in welcher Form diese Nachrichten angezeigt werden. Dabei sollten Aspekte wie Art, Dauer und Intensität der Darstellung durch den Benutzer selbst einstellbar sein.

„…wenn ich etwas sehe was mich interessiert, dann will ich glaube ich keine Einblendungen darin haben…es hängt dann glaube ich viel davon ab, wie intrusiv das Ganze ist…" – Tobias

„… die Informationen kurz ein- und dann wieder ausblenden…wenn es die ganze Zeit bleibt stört es glaube ich." – Kristofer

Abbildung 5: Fernseher mit laufender Sendung, Twitter Stream und Debugging-Ausgabe auf Notebook

Bei der dritten Kategorie handelt es sich um soziologische Aspekte insbesondere im Kontext der sozialen Netzwerke. Aspekte wie Datenschutz, Privatsphäre und bisherige Nutzungspraxis stellen wichtige Kriterien für die Nutzung einer Plattform wie der hier vorgestellten dar. Für einen der Teilnehmer waren insbesondere die Verwendung der von ihm bereitgestellten Daten ein wichtiger Aspekt. Hier stand die Sorge um eine kommerzielle Nutzung der Daten im Vordergrund. Weitere Aspekte wie die Zugänglichkeit der Beiträge wurden ebenfalls als wichtig beschrieben, jedoch nicht als K.O. Kriterium. Bei den befragten Teilnehmern zeichnete sich ein deutliches Bewusstsein für die Öffentlichkeit von sozialen Netzwerken wie Facebook oder Twitter ab, wodurch auch bei fehlenden Einstellungsmöglichkeiten zur Privatsphäre die Selbstreflektion als Selektionsmechanismus genutzt würde.

„Fernsehen ist für mich eine ganz private Sache. … Ich müsste wissen, dass das jetzt nicht genutzt wird um Benutzerprofile zu erstellen." - Andre

„Ich wäre nicht bereit dafür viel Geld auszugeben, deshalb hätte ich die Vermutung wenn es umsonst angeboten würde, dass die Unternehmen mit Werbung Geld verdienen möchten. Das wäre es mir vielleicht dann nicht wert." - Andre

Als ein ebenfalls wichtiger Aspekt stellte sich die Privatsphäre heraus, die von allen Teilnehmern angesprochen wurde. Hierbei standen die Sichtbarkeit der Beiträge und die Definition der Gruppe im Vordergrund sowie die bisherige Nutzungspraxis bzw. Zugehörigkeit zu sozialen Netzwerken. Die Steuerung der Sichtbarkeit der Beiträge wurde regelmäßig als ein zu bedenkender Faktor genannt, jedoch nicht als Ausschlusskriterium empfunden. Einige Teilnehmer sahen einen öffentlichen Raum als ideal an, andere sahen einen Nutzen primär im bestehenden sozialen Netzwerk.

„…grundsätzlich würde ich immer versuchen so anonym wie möglich zu sein wenn es denn öffentlich ist, aber für mich liegt der Mehrwert tatsächlich im persönlichen Bereich…" - Andrea

Die Teilnehmer wurden auch gebeten, losgelöst vom konkreten Anwendungsfall, also der Integration von Fernsehen und Twitter im Wohnbereich, weitere Szenarien darzustellen in denen sie das zu Grunde liegende Konzept der dynamischen Home-IT als nützlich empfinden würden. Neben diversen Hinweisen die stark im Zusammenhang mit klassischen Home-Automation Ansätzen lagen, wurden die zwei folgenden Szenarien aufgezeigt. Zum einen wurde eine Verwendung im semi-öffentlichen Bereich vorgeschlagen, also beispielsweise einer Messe. Hier könnte einem geschlossenen Benutzerkreis in einem vordefinierten Raum durch dynamische Erkennung der Location ein Mehrwert geboten werden. Ein anderes Szenario zielte auf die Integration weiterer Medien wie Musik, die dem Benutzer im Haus folgen könne.

„…die Musik folgt mir sozusagen. Das fände ich sehr praktisch." - Tobias

5.3 Methodische Reflektion

Für die Evaluation wurden insgesamt sechs Teilnehmer persönlich befragt und während der Nutzung des Prototyps beobachtet. Die Auswahl der Teilnehmer war nicht an spezifische Eigenschaften oder Kenntnisse gebunden. Um die Evaluation möglichst realitätsnah zu gestalten wurden diese in einem wohnzimmerähnlichen Raum durchgeführt. Durch die technische Orientierung des Prototyps wurde im zweiten Teil der Evaluation, der Benutzer stark geführt, um die Funktionalität zu verdeutlichen und Missverständnissen vorzubeugen.

Die Benutzer können aufgrund ihres Wissens und ihrer Erfahrung als Lead-User bezeichnet werden und so sind die Ergebnisse der Studie entsprechend zu interpretieren. Aus einer Genderperspektive betrachtet, wurden 5 Männer und lediglich eine Frau befragt. Insbesondere die unterschiedliche Nutzung von sozialen Netzwerken sollte hier berücksichtig werden. Durch die Evaluation unter Laborbedingungen sind evtl. andere Arten der Nutzung beobachtet worden, als sie im tatsächlichen häuslichen Umfeld der Teilnehmer beobachtet worden wäre. Weiter wurde im konkreten Fall keine Vorstudie der tatsächlichen häuslichen Nutzungspraxis vorgenommen, sondern nur eine Befragung hierzu. Bei zukünftig geplanten Studien innerhalb des Siegener Living Labs können diese Aspekte berücksichtigt werden. Beim konkreten Test der TVitter Implementierung wurden die Teilnehmer stark geführt und so möglicherweise in ihrer Nutzungspraxis zu stark beeinflusst, weshalb viele Beobachtungen aus dieser Phase der Evaluation nicht Einzug in diese Studie halten.

5.4 Fazit

In der hier vorliegenden Studie der prototypisch implementierten dynamischen Home-IT Infrastruktur konnten wichtige Erkenntnisse zur Nutzung und Empfindung erfasst werden. Es wurden grundlegende Informationen zu technischen und soziologischen Aspekten des Konzeptes und der konkreten Implementierung gesammelt, die für weitere Forschung in diesem Feld wichtig und hilfreich sind. Grundlegend wurde das Konzept der dynamischen Home-IT positiv bewertet, ebenso wie die konkrete Implementierung des Anwendungsfalls TVitter.

„Ich finde das sehr gut, denn es kommt meinem Bedürfnis grundsätzlich entgegen, in dem Sinne das ich ja das Notebook normalerweise an habe." – Andrea

Bei der Auswertung der Interviews zeigten sich insbesondere die folgenden Punkte als wichtige Kriterien für die Weiterentwicklung:

1. Usability
2. Ein- und Ausgabe
3. Privatsphäre und Datenschutz

Diese Aspekte sollen bei einer weitergehenden Implementierung des Frameworks berücksichtigt und in weiteren Tests erprobt werden.

6. AUSBLICK

Das von uns entwickelte Framework bietet die Möglichkeit Dienste in Home-IT Umgebungen bereitzustellen und verschiedene Medien zu integrieren. Bei der Implementierung haben wir auf einem komponentenbasierten Modell aufgesetzt, orientieren uns aber bei der Interaktion an serviceorientierten Architekturen. Durch die Dynamik beider Paradigmen ist das Framework in der Lage komplexere Funktionalität in Abhängigkeit von Ort bzw. vorhandener Infrastruktur anzubieten und medienübergreifend nutzbar zu machen. Gleichzeitig wird auf diese Weise eine Plattformunabhängigkeit gewährleistet, wodurch der Einsatz in heterogenen Landschaften erleichtert wird. Exemplarisch wurde die Komponente TVitter entwickelt, um das Framework im konkreten Anwendungsfall vorstellen und testen zu können. Im Hinblick auf die Evaluation des Frameworks sollen in anschließenden Arbeiten neben der TVitter Anwendung weitere Komponenten entwickelt werden, um eine sinnvolle Nutzung des Frameworks zu ermöglichen. Zusätzlich zu der hier durchgeführten Evaluation soll anschließend in Testhaushalten eine Langzeitevaluation durchgeführt werden. Hierzu wurden 16 Haushalte, die ebenfalls Teil des Siegener Living Labs sind, ausgewählt und mit Media Center PCs und Smartphones ausgestattet. Über eine Projektlaufzeit von etwa zwei Jahren sollen kontinuierlich weitere Komponenten entwickelt und über das Framework den Testhaushalten bereitgestellt werden.

7. DANKSAGUNG

Wir danken Stefan Hering für die Mitarbeit bei der Entwicklung des Prototyps sowie Joachim Ritter von Prosyst für seine Unterstützung bei Fragen bezüglich des OSGi Frameworks.

Diese Arbeit wurde gefördert vom Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes NRW mit Mitteln der Europäischen Union sowie der DFG.

8. LITERATUR

[1] Apple Inc. About Apple Push Notification Service. *About Apple Push Notification Service*, 2010. http://support.apple.com/kb/HT3576.

[2] Betz, M., Ley, B., Pipek, V., and Wulf, V. Folksonomies for real things: Tagging objects with RFID as a source for context-awareness. *Informatik 2008 - Beherrschbare Systeme dank Informatik*, Springer Verlag (2008), 787-794.

[3] Brown, P.F., Metz, R., and Hamilton, B.A. Reference model for service oriented architecture 1.0. (2006).

[4] Cervantes, H. and Hall, R.S. Autonomous Adaptation to Dynamic Availability Using a Service-Oriented Component Model. *Proceedings of the 26th International Conference on Software Engineering*, IEEE Computer Society (2004), 614-623.

[5] Cheng, Y.-M. Using ZigBee and Room-Based Location Technology to Constructing an Indoor Location-Based Service Platform. *Intelligent Information Hiding and Multimedia Signal Processing, International Conference on*, IEEE Computer Society (2009), 803-806.

[6] Crockford, D. JSON. *RFC-4627 JSON*, 2006. http://www.ietf.org/rfc/rfc4627.txt?number=4627.

[7] Dey, A.K. Understanding and Using Context. *Personal Ubiquitous Comput. 5*, 1 (2001), 4-7.

[8] Google. Android Cloud to Device Messaging Framework - Google Projects for Android. *Google Projects for Android: C2DM*, 2010. http://code.google.com/intl/de-DE/android/c2dm/index.html.

[9] Gu, T., Pung, H.K., and Zhang, D.Q. Toward an OSGi-Based Infrastructure for Context-Aware Applications. *IEEE Pervasive Computing 3*, 2004, 66-74.

[10] Harboe, G., Massey, N., Metcalf, C., Wheatley, D., and Romano, G. The uses of social television. *Comput. Entertain. 6*, (2008), 8:1–8:15.

[11] Heineman, G.T. and Councill, W.T. *Component-Based Software Engineering: Putting the Pieces Together*. Addison-Wesley Longman, Amsterdam, 2001.

[12] Hess, J. and Ogonowski, C. Steps toward a living lab for socialmedia concept evaluation and continuous user-involvement. *Proceedings of the 8th international interactive conference on Interactive TV&Video*, ACM (2010), 171–174.

[13] Hippel, E. von. Lead Users: A Source of Novel Product Concepts. *MANAGEMENT SCIENCE 32*, 7 (1986), 791-805.

[14] Kiesler, S., Zhang, H., Ren, Y., Kraut, R.E., Konstan, J.A., and Weisz, J.D. Watching together: integrating text chat with video. 877.

[15] Luimula, M. and Kuutti, K. Locawe: a novel platform for location-aware multimedia services. *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia*, ACM (2008), 122-129.

[16] Nichols, J. and Myers, B.A. Creating a lightweight user interface description language: An overview and analysis of the personal universal controller project. *ACM Trans. Comput.-Hum. Interact. 16*, 4 (2009), 1-37.

[17] Oehmichen, E. and Schröter, C. Zur Differenzierung des Medienhandelns der jungen Generation. *Media Perspektiven 2009*, 2009.

[18] OSGi. OSGi Alliance. *OSGi Specification*, 2010. http://www.osgi.org/Specifications/HomePage.

[19] Pascoe, J. Adding Generic Contextual Capabilities to Wearable Computers. (1998), 92--99.

[20] Prosyst. *Prosyst OSGi Framework*. Prosyst, 2008.

[21] Redondo, R.P.D., Vilas, A.F., Cabrer, M.R., and Pazos, J.J. Exploiting OSGi capabilities from MHP applications. *Journal of Virtual Reality and Broadcasting 4*, (2007), 16.

[22] Sanchez-Nielsen, E., Martin-Ruiz, S., and Rodriguez-Pedrianes, J. An open and dynamical service oriented architecture for supporting mobile services. *Proceedings of the 6th international conference on Web engineering*, ACM (2006), 121-128.

[23] Schilit, B., Adams, N., and Want, R. Context-Aware Computing Applications. *IN PROCEEDINGS OF THE WORKSHOP ON MOBILE COMPUTING SYSTEMS AND APPLICATIONS*, (1994), 85--90.

[24] Schumacher, J., Feurstein, K., and Pitkänen, O. *European Living Labs: A new approach for human centric regional innovation*. wvb Wissenschaftlicher Verlag Berlin, 2008.

Partizipieren statt Konsumieren, oder: Lohnt sich Engagement in Q&A-Communities?

Philipp Herrmann
Universität Paderborn
Warburger Straße 100
33098 Paderborn
0049 5251 60 5511
philipp.herrmann@wiwi.uni-paderborn.de

Dennis Kundisch
Universität Paderborn
Warburger Straße 100
33098 Paderborn
0049 5251 60 5533
dennis.kundisch@wiwi.uni-paderborn.de

ZUSAMMENFASSUNG

In den letzten Jahren haben sich online Question & Answer Communities wie Yahoo! Answers oder CosmiQ für viele Internetnutzer als bedeutende Quelle von Informationen etabliert. Es ist jedoch fraglich, welche Faktoren die Qualität der Informationen, die ein Nutzer in einer solchen Community erhält, beeinflussen. Zur Beantwortung dieser Frage nutzen wir die Tauschtheorie und die Uncertainty Reduction Theory. Anhand dieser Theorien entwickeln wir ein Modell zur Vorhersage der Qualität der Antworten, die ein Nutzer einer solchen Community auf seine Fragen erhält. Dieses Modell testen wir anhand von 295.500 Fragen, die in den Jahren 2007 und 2008 in einer deutschsprachen Question & Answer Community gestellt wurden. Die Ergebnisse zeigen, dass das allgemeine Interesse an einer Frage, vorausgegangenes, qualitativ hochwertiges Engagement des Fragestellers in der Community sowie die Preisgabe bestimmter persönlicher Informationen einen positiven Einfluss auf die Qualität der Antworten haben.

Schlüsselwörter

Soziale Netzwerke, Q&A-Community, Sozialkapital, Tauschtheorie, Uncertainty Reduction Theory

1. Einleitung

Immer mehr Menschen nutzen das Internet, um ihren Informationsbedarf zu befriedigen. Dabei stellen Suchmaschinen einen Hauptzugang zu den im Internet verfügbaren Informationen dar. Neben diesen haben sich in den letzten Jahren verschiedene *Question & Answer Communities* (Q&A-Communities) als wichtige Quelle von Informationen etabliert [25]. Die Anzahl der Seitenaufrufe der Top 5 Q&A-Communities ist zwischen 2007 und 2009 um fast 900% gestiegen [24] und Yahoo! Answers (answers.yahoo.com) ist heute, nach Wikipedia, die zweit populärste Internetseite in der Kategorie *Reference*[1] [2]. In Q&A-Communities stellen Millionen registrierter Nutzer weltweit täglich zehntausende von Fragen [16] aus den unterschiedlichsten Lebensbereichen und erhalten – meist öffentlich einsehbare – Antworten auf diese Fragen.

Harper et al. [16] definieren eine Q&A-Community als „[…] a Web site […] purposefully designed to allow people to ask and respond to questions on a broad range of topics". Die Autoren unterteilen Q&A-Communities in *Digital Reference Services*, *Ask an Expert Services* und *Community Q&A Sites*. Dabei stellen Digital Reference Services die online Analogie zu traditionellen Bibliotheksauskunftsdiensten dar [23]. In diesen beantworten professionelle Rechercheure die Online-Anfragen von Bibliotheksbenutzern. Ask an Expert Services ermöglichen es den Nutzern Fachfragen direkt an Experten aus den jeweiligen Fachgebieten zu stellen, während Community Q&A Sites zur Beantwortung der Fragen auf alle Nutzer des Services zugreifen. Dabei bieten Community Q&A Sites neben der Formulierung und Beantwortung von Fragen weitere Funktionalitäten, wie z.B. Kommentar- und Bewertungsfunktionen zu Fragen und Antworten oder die Möglichkeit sich ein persönliches Profil anzulegen.

Die Literatur zu Community Q&A Sites lässt sich in *User-zentrierte* sowie in *Inhalts-zentrierte* Arbeiten einteilen. User-zentrierte Arbeiten untersuchen verschiedene Charakteristika der Nutzer von Community Q&A Sites (Shah et al. [25] geben einen guten Überblick über diese Arbeiten), während Inhalts-zentrierte Arbeiten meist der Frage nachgehen, welche Einflussfaktoren auf die Qualität der Antworten auf eine Frage wirken. Wir fokussieren uns im Folgenden auf diesen Inhalts-zentrierten Aspekt.

Harper et al. [16] untersuchen zum einen, ob sich Digital Reference Services, Ask an Expert Services und Community Q&A Sites in der Antwortqualität unterscheiden und zum anderen, wie ein Fragesteller die Qualität der Antworten auf seine Frage positiv beeinflussen kann. Dabei analysieren die Autoren den Einfluss der für die Beantwortung der Frage ausgesetzten Belohnung sowie der rhetorischen Qualität und des Themas der

[1] In dieser Kategorie werden Nachschlagewerke aller Art zusammengefasst. Diese umfassen unter anderem Enzyklopädien (z.B. wikipedia.org), Wörterbücher (z.B. dict.cc) und Bibliographien (z.B. ebscohost.com)

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Frage. Sie zeigen, dass die Qualität der Antworten nur durch die für die Beantwortung der Frage ausgesetzte monetäre Belohnung signifikant beeinflusst werden kann. Weiterhin zeigen die Autoren, dass die durchschnittliche Antwortqualität auf Community Q&A Sites im Vergleich zu den beiden anderen Community Typen am höchsten ist.

Während die Arbeit von Harper et al. [16] Antworten auf die Frage gibt, in welchem Community Typ ein Fragesteller besonders gute Antworten auf seine Fragen erhält und durch welche Handlungen er die Qualität der Antworten auf seine Fragen positiv beeinflussen kann, gehen die im Folgenden vorgestellten Beiträge der Frage nach, wie besonders hochwertige Inhalte auf Community Q&A-Sites identifiziert werden können. Diese Ergebnisse können unter anderem dazu dienen als hochwertig eingestufte Inhalte als passende Suchergebnisse für in Suchmaschinen eingegebene Fragen vorzuschlagen.

Agichtein et al. [1], Liu et al. [22] und Shah und Pomerantz [26] analysieren Fragen und die dazugehörigen Antworten aus der Community Q&A Site Yahoo! Answers. Dabei nutzen die Autoren direkt aus Fragen und Antworten ableitbare Maße (z.B. Rechtschreibfehler), Maße die sich auf den Fragesteller bzw. Antwortgeber beziehen (z.B. die Qualität der bisher gestellten Fragen bzw. gegebenen Antworten) sowie Maße, welche die Attraktivität einer Frage bzw. Antwort erfassen (z.B. wie oft wurde eine Frage bzw. Antwort angeklickt), um qualitativ hochwertige Fragen und Antworten [1] bzw. nur qualitativ hochwertige Antworten [22], [26] zu identifizieren.

Die Autoren wenden Data Mining Verfahren an, um wertvolle Inhalte auf der Community Q&A Site zu identifizieren. Dabei nutzen sie sowohl Informationen über die Frage und den Fragesteller als auch Informationen über die Antwort und den Antwortgeber. Dies erfolgt jedoch ohne theoretische Fundierung, sondern vielmehr aus einer explorativen Sichtweise. Gemein ist den genannten Studien zudem, dass diese nicht die Qualität der Antworten, die ein *bestimmter Fragender* in Zukunft erhalten wird, untersuchen. Vielmehr identifizieren die Autoren potenziell wertvolle, *schon vorhandene* Antworten für *andere* Nutzer mit vergleichbaren Fragen.

Die Ergebnisse von Agichtein et al. [1], Liu et al. [22] sowie Shah und Pomerantz [26] legen außerdem nahe, dass neben den von Harper et al. [16] untersuchten Einflussfaktoren weitere, im Fragesteller und in der Frage begründete Einflussfaktoren, auf die Qualität einer Antwort wirken.

An dieser Stelle setzen wir mit unserer Arbeit an. Wir entwickeln ein theoretisch fundiertes Modell, mit dem die Qualität der Antworten auf eine Frage anhand von in der Frage und im Fragesteller begründeten Charakteristika vorhergesagt werden soll. Dieses Modell testen wir anhand eines Datensatzes einer deutschen Community Q&A Site. Dabei beantworten wir die folgende Forschungsfrage:

Welche in der Frage und im Fragesteller begründeten Einflussfaktoren führen dazu, dass ein Nutzer einer Community Q&A Site besonders hilfreiche Antworten auf seine Fragen erhält?

Unsere Untersuchung leistet dabei drei Beiträge zur bestehenden Literatur: Erstens liefern wir eine theoretische Fundierung der, von uns und anderen, betrachteten Einflussfaktoren auf die Qualität der Antworten auf einer Community Q&A Site. Zweitens überprüfen wir diese Theorien empirisch und zeigen, welche Faktoren einen Beitrag zur Erklärung der Antwortqualität leisten. Drittens internationalisieren und erweitern wir die bestehende Forschung zu Community Q&A Sites, indem wir einen neuen Datensatz in die Literatur einführen.

Es zeigt sich, dass unterschiedliche Faktoren einen Einfluss darauf haben, ob ein Nutzer wertvolle Antworten auf seine Frage erwarten darf. Zum ersten spielt es eine Rolle, ob sich ein Nutzer insbesondere durch ein hochgeladenes Bild und der Angabe des Wohnsitzlandes freiwillig nicht vollständig anonym in der Community präsentiert. Zum zweiten zeigt sich, dass andere Nutzer nicht quantitatives sondern qualitatives Engagement belohnen. Wer in der Vergangenheit viele sehr hilfreiche Antworten gegeben hat, bekommt auch selbst mit einer höheren Wahrscheinlichkeit sehr wertvolle Antworten. Die Menge an bisher in der Community gegebenen wenig hilfreichen Antworten wirkt sich hingegen eher negativ aus. Zum dritten haben der Frageinhalt und die Fragestellung einen substanziellen Einfluss darauf, ob man wertvolle Antworten erwarten kann. Gut gestellte und für die Community-Teilnehmer interessante Fragen werden häufiger und qualitativ besser beantwortet als andere Fragen. Insgesamt finden sich Hinweise darauf, dass die Tauschtheorie und die Uncertainty Reduction Theory einen Beitrag zur Erklärung der Antwortqualität leisten.

Der Artikel ist wie folgt aufgebaut. In Abschnitt 2 stellen wir die Community Q&A Site vor. Die Hypothesen für das zu testende Modell werden in Abschnitt 3 entwickelt. Im vierten Abschnitt präsentieren wir den analysierten Datensatz. In Abschnitt 5 stellen wir die Ergebnisse vor und diskutieren diese im sechsten Abschnitt. In Abschnitt 7 fassen wir die Ergebnisse zusammen und zeigen Perspektiven für zukünftige Forschungsarbeiten auf.

2. Community Q&A Site

Die in diesem Beitrag untersuchte, kostenlose Community Q&A Site[2] ermöglicht es ihren Nutzern Fragen, die innerhalb eines festgelegten Zeitraums beantwortet werden können, an die Community zu stellen.

Jeder Nutzer der Seite hat die Möglichkeit, sich kostenlos in der Community zu registrieren und so Zugriff auf verschiedene weitere Funktionalität zu bekommen. Für jeden registrierten Nutzer existiert ein persönlicher Bereich, in dem er unter anderem Übersichten über die selbst gestellten und beantworteten Fragen, über Fragen die ihn möglicherweise interessieren könnten und über seine persönliche Linksammlungen erhält.

Weiterhin nimmt jeder registrierte Benutzer automatisch an einem Community-internen Punktesystem teil. In diesem wird zwischen Status- und Bonuspunkten unterschieden. Die Statuspunkte entscheiden über den sogenannten Rang eines Nutzers. Jeder Nutzer startet bei Registrierung im Rang eines Einsteigers und hat die Möglichkeit über das Sammeln von Statuspunkten im Rang aufzusteigen. Diese können durch folgende Aktivitäten gesammelt werden:

[2] Auf Wunsch des Praxispartners wird der Name der Community in diesem Beitrag nicht genannt.

- Durch das Stellen von Fragen, die durch andere Nutzer beantwortet oder als hilfreich bewertet werden.
- Durch die Beantwortung von Fragen mit als hilfreich oder als top bewerteten Antworten (eine Bewertung als Top-Antwort stellt die bestmögliche Bewertung durch einen Fragesteller dar).
- Durch die Bewertung von Antworten auf eigene Fragen innerhalb einer Frist von 14 Tagen nach Ablauf der für die Beantwortung der Frage eingestellten Frist.
- Durch das Hinzufügen von Links.
- Durch das Einladen neuer Nutzer in die Community sowie durch Rangaufstiege dieser eingeladenen Nutzern.

Neben den Statuspunkten erhält jeder registrierte Nutzer zu Beginn seiner Mitgliedschaft 1.000 Bonuspunkte. Diese kann er als Belohnung für die Beantwortung für von ihm gestellte Fragen einsetzen. Beantworten mehrere Antwortgeber eine Frage mit einer, mindestens als hilfreich bewerteten, Antwort, werden die ausgesetzten Bonuspunkte zwischen allen Antwortgebern, deren Antwort als mindestens hilfreich bewertet wurde, aufgeteilt. Dabei wird Top-Antworten ein überproportionaler Anteil der eingesetzten Bonuspunkte zugeteilt. Gibt ein Fragesteller innerhalb von 14 Tagen keine Bewertung für die Antworten auf seine Frage ab, werden diese automatisch alle als hilfreich bewertet und die eingesetzten Bonuspunkte werden zu gleichen Teilen auf die Antwortgeber verteilt.

Für jeden registrierten Nutzer wird außerdem ein persönliches Profil freigeschaltet. In diesem Profil hat jeder Nutzer die Möglichkeit, ein Bild hochzuladen sowie weitere persönliche Informationen (z.B. Adresse, Geschlecht, Email, Geburtsdatum, Land) anzugeben. Diese Informationen können für alle Nutzer, nur für Freunde, oder für niemanden im Netzwerk sichtbar gemacht werden. Neben diesen freiwilligen Angaben werden im Profil die Anzahl der gestellten Fragen und gegebenen Antworten, der Rang eines Nutzers, seine Statuspunkte, das Anmeldedatum, Themen, in denen ein Nutzer aktiv ist, sowie die Anteile von gegebenen Top-Antworten, hilfreichen Antworten und nicht-hilfreichen Antworten für alle anderen Nutzer sichtbar angezeigt.

Es werden weiterhin Funktionalitäten zum Aufbau und zur Pflege eines persönlichen Netzwerks angeboten. Ein registrierter Nutzer hat die Möglichkeit, andere Personen in die Q&A-Community allgemein und/oder in sein persönliches Netzwerk innerhalb der Community einzuladen. Zudem hat ein registrierter Nutzer unter anderem Zugriff auf Fragen von Personen aus seinem persönlichen Netzwerk und deren verwaltete Links. Schließlich gibt es auch die Möglichkeit, private Nachrichten an andere registrierte Nutzer zu verschicken.

3. Hypothesen

Zur Beantwortung der Forschungsfrage stützen wir uns auf unterschiedliche theoretische Erklärungsansätze, die wir im Folgenden vorstellen. Basierend auf diesen Theorien entwickeln wir Hypothesen, die wir im weiteren Verlauf dieses Beitrags empirisch überprüfen.

3.1 Sozialkapital und Tauschtheorie

In Anlehnung an Bordieu [6], Coleman [9] und Flap und de Graaf [10] wird *Sozialkapital* als aktuell und in Zukunft aus sozialen Beziehungen resultierende Ressourcen definiert, welche ein Akteur für eigene Zwecke nutzen kann. Allgemein beruht Sozialkapital dabei auf der Bereitschaft oder Verpflichtung verbundener Akteure einem Akteur eigene Ressourcen zur Verfügung zu stellen.

Für eine Bewertung von Sozialkapital bedarf es einer Betrachtung der individuellen Netzwerkbeschaffenheit eines Akteurs, der so genannten *Sozialstruktur*. Die Sozialstruktur beschreibt, wie ein Akteur innerhalb seines sozialen Netzwerks mit anderen Akteuren verbunden ist. Die sich aus der Sozialstruktur (potenziell) ergebenden Handlungsbegünstigungen werden als *soziale Ressourcen* bezeichnet. Diese Handlungsbegünstigungen beruhen zum einen darauf, dass ein Akteur befähigt wird auf Ressourcen, wie z. B. Kapital und komplexes Wissen anderer, zu ihm in Beziehung stehender Akteure zurückzugreifen. Zum anderen lassen sich aus den Beziehungen selbst Handlungsbegünstigungen, z. B. in Form von Information und Kontrolle über andere Akteure ableiten (vgl. [7], [8] und [11]).

Im Rahmen unserer Analyse fassen wir das gesamte soziale Netzwerk der Q&A-Community als Sozialstruktur auf und wollen untersuchen, ob aus dieser Sozialstruktur bei entsprechendem Verhalten des Akteurs soziale Ressourcen verfügbar gemacht werden können. Diese sozialen Ressourcen sind in unserem speziellen Anwendungsfall fachliche Informationen und Wissen, die einem Akteur in seinem privaten oder beruflichen Handlungskontext nützlich sein können.

Als theoretische Grundlage bedienen wir uns dabei der sozio-ökonomischen Tauschtheorie [9], denn die von uns genutzte Sozialkapitaldefinition impliziert ein tauschtheoretisches Verständnis, da sie explizit auf die Bereitschaft der Akteure abzielt, eigene Ressourcen zur Verfügung zu stellen.

Grundlage der Sozialkapitalbildung oder -realisation ist ein Ressourcentransfer vom Sozialkapitalgeber zum Sozialkapitalempfänger. Ein Sozialkapitalgeber wird eigene Ressourcen ggf. jedoch nur dann offerieren, wenn er dafür eine Gegenleistung erhält. Im Gegensatz zum Markttausch werden Ressourcen hier allerdings ohne Geld und ohne explizite Verträge getauscht. Die Gegenleistung ist im sozialen Tausch also in Form, Höhe und Zeitpunkt nicht spezifiziert und nicht einklagbar. [19]

In der Tauschtheorie werden Akteure als Nutzenmaximierer unterstellt, die rational ihre Eigeninteressen verfolgen. Der Beitrag einer Ressource zur Befriedigung der Bedürfnisse eines Akteurs determiniert dabei das Interesse des Akteurs an einer Ressource. Tauschprozesse führen in einem sozialen System zu einem Gleichgewichtszustand, in dem weitere Tauschhandlungen nicht mehr zu einer individuellen Verbesserung von Akteuren führen würden. Unter den Bedingungen des von Coleman [9] beschriebenen Basismodells entspricht die Ressourcenmacht eines Akteurs im Gleichgewicht seinem Sozialkapital. Ein Akteur kann in einem solchen System auf Ressourcen anderer Akteure in dem Maße zugreifen, in dem er selbst Ressourcen – hier also bspw. für andere Nutzer wertvolle Informationen und Wissen – kontrolliert. Tritt ein Akteur dabei in Vorleistung, dann kann er theoretisch Verpflichtungen in Höhe seiner Ressourcenmacht auf Seiten der anderen Akteure schaffen. [19]

Auf unsere Analyse übertragen leiten wir daraus folgende Hypothese ab:

H1) Je stärker sich ein Fragesteller in der Vergangenheit im Netzwerk engagiert hat, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Wir messen das vergangene Engagement eines Fragestellers anhand von acht Variablen. Da wir für jede Variable einen abnehmenden Grenznutzen unterstellen und gleichzeitig alle Variablen auch den Wert Null annehmen können, addieren wir zu jeder Variable eins hinzu und bilden im Anschluss den natürlichen Logarithmus des Ergebnisses.

- FREUNDE bezeichnet die Anzahl der registrierten Nutzer, mit denen ein Fragesteller zum Zeitpunkt der Erstellung seiner Frage befreundet ist.
- EINGELADEN bezeichnet die Anzahl der registrierten Nutzer, die ein Fragesteller zum Zeitpunkt der Erstellung seiner Frage in die Community eingeladen hat.
- ANTWORTEN bezeichnet die Anzahl der weder als top noch als hilfreich bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- HILFREICH bezeichnet die Anzahl der als hilfreich bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- TOP bezeichnet die Anzahl der als top bewerteten Antworten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage gegeben hat.
- VERSCHICKT bezeichnet die Anzahl der Nachrichten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage an andere registrierte Nutzer verschickt hat.
- ERHALTEN bezeichnet die Anzahl der Nachrichten, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage von anderen registrierten Nutzern erhalten hat.
- LINKS bezeichnet die Anzahl der Links, die ein Fragesteller bis zum Zeitpunkt der Erstellung seiner Frage erstellt hat.

Ähnliche Variablen zur Messung der früheren Aktivitäten des Fragestellers finden sich auch bei [26] sowie [22].

3.2 Uncertainty Reduction Theory

In direkter Face-to-Face Kommunikation führt das Vorhandensein von zusätzlichen, den Gesprächspartner betreffenden Informationen zu verstärkter positiver Aufmerksamkeit gegenüber diesem [27]. Diese Aussage stimmt mit den Vorhersagen der *Uncertainty Reduction Theory* von Berger und Calabrese [5] für erstmalig stattfindende Face-to-Face Kommunikation überein. Die Theorie besagt, dass zusätzliche Informationen über den Gesprächspartner Unbehagen verringern und die Vorhersagbarkeit sowie die Gewogenheit gegenüber dem Gesprächspartner erhöhen.

Tidwell und Walther [28] zeigen in ihrer Arbeit, dass die Theorie auch auf computer-mediierte Kommunikation übertragen werden kann. Dieses Ergebnis wird durch die Arbeiten von Zheng et al. [30] und Walther et al. [29] bestätigt. In beiden Arbeiten finden sich Hinweise darauf, dass das Vorhandensein von Bildern der Gesprächspartner einen positiven Einfluss auf das Vertrauen [30] bzw. die Gewogenheit und die soziale Attraktivität [29] der Gesprächspartner hat.

Interpretiert man die Formulierung und die Beantwortung von Fragen innerhalb der untersuchten Community Q&A Site als computer-mediierte Kommunikation, bieten sich den registrierten Nutzern verschiedene Möglichkeiten Informationen über sich preiszugeben. Registrierte Nutzer haben zum einen die Möglichkeit, ein Bild in ihr Profil einzufügen, und können zum anderen verschiedene Informationen wie z.B. ihren Namen, ihren Wohnort oder ihr Geschlecht angeben (vgl. Abschnitt 2). Gemäß der Uncertainty Reduction Theory sollte sich die Preisgabe von zusätzlichen Informationen positiv auf die Gewogenheit gegenüber einem Nutzer auswirken. Wir nehmen außerdem an, dass sich eine erhöhte positive Gewogenheit gegenüber einem Nutzer positiv auf die Qualität der Antworten auf seine Fragen auswirkt. Daraus leiten wir die folgende Hypothese ab:

H2) Je mehr Informationen ein Fragesteller über sich preisgibt, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Wir messen die Informationen, die ein Fragesteller über sich preisgibt, anhand von sieben binären Variablen:

- BILD nimmt den Wert eins an, wenn ein registrierter Nutzer ein Bild in sein Profil eingefügt hat.
- GESCHLECHT nimmt den Wert eins an, wenn ein registrierter Nutzer sein Geschlecht angibt.
- NAME nimmt den Wert eins an, wenn ein registrierter Nutzer seinen Vornamen angibt.
- POSTLEITZAHL nimmt den Wert eins an, wenn ein registrierter Nutzer seine Postleitzahl angibt.
- LAND nimmt den Wert eins an, wenn ein registrierter Nutzer sein Herkunftsland angibt.
- GEBURTSDATUM nimmt den Wert eins an, wenn ein registrierter Nutzer sein Geburtsdatum angibt.
- BERUF nimmt den Wert eins an, wenn ein Nutzer seinen Beruf angibt.

3.3 Allgemeines Interesse an einer Frage

Neben den in den Abschnitten 3.1 und 3.2 präsentierten Faktoren unterstellen wir analog zu [26], [22] sowie [1], dass das allgemeine Interesse an einer Frage einen weiteren wichtigen Einflussfaktor auf die Antwortqualität einer Frage darstellt. Hypothese 3 lautet daher:

H3) Je höher das allgemeine Interesse an einer Frage ist, desto qualitativ hochwertigere Antworten erhält dieser Fragesteller auf seine Fragen.

Dieses Interesse kann über vier unterschiedliche Variable gemessen werden. Da wir, wie auch für Hypothese H1, jeweils einen abnehmenden Grenznutzen der betrachteten Variablen unterstellen und gleichzeitig alle Variablen auch den Wert Null annehmen können, addieren wir wiederum zu jeder Variable eins hinzu und bilden den natürlichen Logarithmus des Ergebnisses.

- BONUSPUNKTE bezeichnet die für die Beantwortung einer Frage ausgesetzten Bonuspunkte.
- HITS bezeichnet die Anzahl der Aufrufe eine Frage.

- F_HILFREICH bezeichnet die Anzahl der Markierungen einer Frage als hilfreich.
- F_ANTWORTEN bezeichnet die Antworten, die für eine Frage erstellt wurden.

4. Datensatz

Der zur Verfügung stehende Datensatz beinhaltet alle Aktivitäten die in der Community Q&A Site zwischen dem 09.12.2005 und dem 06.06.2008 durchgeführt wurden.

In Tabelle 1 sind die Fragen und Antworten aggregiert für den gesamten Datensatz sowie für die einzelnen Jahre dargestellt. Die Tabelle zeigt, dass der Datensatz insgesamt 1.033.542 Fragen und 2.996.446 Antworten enthält. Auf jede Frage wurde also im Durchschnitt knapp dreimal geantwortet.

Tabelle 1: Fragen und Antworten

Jahr	# Fragen	# Antworten	# Antworten / # Fragen
Gesamter Datensatz	1.033.542	2.996.446	2,9
2005	343	411	1,2
2006	73.714	243.992	3,3
2007	580.717	1.662.719	2,9
2008	378.767	1.089.324	2,9

Tabelle 2 zeigt die Entwicklung der Nuterzahlen der untersuchten Community Q&A Site vom 09.12.2005 bis zum 06.06.2008. Von den ca. 460.000 identifizierbaren Nutzern haben sich ca. 20% in der Community registriert. Neben den registrierten Nutzern besteht die Community aus ungefähr 368.000 Gastnutzern sowie 70 Moderatoren. Für unsere Untersuchungen werden im Weiteren nur die durch registrierte Nutzer gestellten Fragen von Relevanz sein.

Da die Community erst am 12.01.2006 für die Öffentlichkeit zugänglich gemacht wurde und um mögliche Verzerrungen durch die relativ geringe Mitgliederanzahl im ersten Jahr der Community zu vermeiden, nehmen wir nur Fragen, die nach dem 01.01.2007 erstellt wurden in unsere Analyse auf.

Tabelle 2: Nutzertypen

Jahr	Nutzer	Davon reg. Nutzer	Davon Modera-toren	Davon Gäste
Gesamter Datensatz	461.321	93.291	70	367.960
2005	89	77	9	3
2006	36.551	12.901	20	23.630
2007	290.770	54.404	39	236.327
2008	133.911	25.909	2	108.000

Von diesen 959.484 Fragen (vgl. Tabelle 1), werden 590.972 Fragen entfernt, da die Antworten auf diese Fragen nicht durch den Fragesteller sondern durch das System bewertet wurden. Dies kann aus den beiden folgenden Gründen geschehen: Entweder handelt es sich bei dem Fragesteller um einen nicht registrierten Nutzer oder der Fragesteller ist ein registrierter Nutzer, der die Antworten auf seine Frage nicht innerhalb von 14 Tagen bewertet hat.

Weitere 3.659 Fragen werden entfernt, da diese Fragen nicht durch regulär angemeldete Nutzer sondern durch Moderatoren der Community gestellt wurden. Bei diesen Fragen könnten weitere, nicht für unsere Studie relevante Effekte die Bewertung der Antworten durch den Fragesteller beeinflussen.

Da eine stichpunktartige Textanalyse von Fragen mit besonders vielen Antworten ergeben hat, dass es sich bei diesen sehr häufig um tägliche Begrüßungen, Glückwünsche zum Statuswechsel oder Scherzfragen handelt, haben wir außerdem alle Fragen mit mehr als fünf Antworten von der Analyse ausgenommen (insgesamt 69.209). Von den verbleibenden 295.644 Fragen werden abschließend 144 Fragen aufgrund von offensichtlichen Fehlern im Datensatz gelöscht.

Der finale Datensatz besteht somit aus insgesamt 295.500 Fragen mit insgesamt 654.030 Antworten. Diese werden in zwei Kategorien eingeteilt. 145.475 Fragen, bei denen der Fragesteller keine der Antworten als Top-Antwort bewertet hat, ordnen wir der ersten Kategorie zu, wohingegen 150.025 Fragen mit mindestens einer Top-Antwort der zweiten Kategorie zugeordnet werden.

Tabelle 3 zeigt die Mittelwerte für alle metrisch skalierten Variablen sowie die Anteile der jeweiligen Ausprägung für die nominal skalierten Variablen für beide Kategorien. Anhand von χ^2-Tests für alle nominal skalierten Variablen sowie t-Tests für alle metrisch skalierten Variablen wird untersucht, ob sich die Mittelwerte bzw. die Anteile der jeweiligen Ausprägung zwischen den Kategorien signifikant unterscheiden. Obwohl es sich bei der untersuchten Datenbasis um die Grundgesamtheit aller im Betrachtungszeitraum erstellten Fragen handelt, berechnen wir die χ^2- und t-Statistiken sowie die dazugehörigen p-Werte.[3] Wir interpretieren diese jedoch analog zu Gefen und Carmel [12] als Goodness-of-Fit Indikatoren und nicht als Irrtums-Wahrscheinlichkeit dafür, dass die Mittelwerte voneinander verschieden sind. Die in Tabelle 3 präsentierten Ergebnisse dieser Tests zeigen, dass sich die Anteile der jeweiligen Ausprägungen beider Kategorien bzw. die Mittelwerte für alle betrachteten Variablen substantiell voneinander unterscheiden. Diese Ergebnisse geben einen ersten Hinweis darauf, dass die betrachteten Variablen die Qualität der Antworten auf eine Frage beeinflussen.

5. Ergebnisse der logistischen Regression

Die Hypothesen werden anhand einer binären, logistischen Regression empirisch überprüft.[4] Als abhängige Variable wird ein binärer Indikator gewählt, welcher den Wert eins annimmt, sobald

[3] Umfasst der untersuchte Datensatz die komplette Grundgesamtheit, entsprechen die geschätzten Regressionskoeffizienten genau den Parametern der Grundgesamtheit [15].

[4] Wir wählen die logistische Regression, da bei dieser Methode explizit dichotome abhängige Variablen berücksichtigt werden [14], [18].

ein Fragesteller mindestens eine der Antworten auf seine Frage als Top-Antwort bewertet. Wir nehmen an, dass dies der beste Indikator für eine vom Fragesteller als sehr gut empfundene

Tabelle 3: Deskriptive Statistiken

	Top-Antwort erhalten		
Variable	Nein	Ja	
Anzahl Beobachtungen	145.475	150.025	
	Panel A: Sozialkapital und Tauschtheorie		t-Statistic (p-Wert)
FREUNDE	1,15	1,59	87,86 (0,000)
EINGELADEN	0,02	0,03	20,60 (0,000)
ANTWORTEN	1,99	2,45	76,67 (0,000)
HILFREICH	3,88	4,63	101,45 (0,000)
TOP	2,67	3,41	108,37 (0,000)
VERSCHICKT	0,81	1,20	73,13 (0,000)
ERHALTEN	0,46	0,66	43,97 (0,000)
LINKS	0,84	1,23	65,48 (0,000)
	Panel B: Uncertainty Reduction Theory		χ^2-Statistic (p-Wert)
BILD	67,20%	81,23%	7619,37 (0,000)
GESCHLECHT	29,55%	35,13%	1048,49 (0,000)
NAME	15,67%	19,52%	657,74 (0,000)
POSTLEITZAHL	5,04%	5,62%	50,25 (0,000)
LAND	17,76%	23,82%	1643,08 (0,000)
GEBURTSDATUM	15,68%	18,32%	366,17 (0,000)
BERUF	18,98%	22,46%	544,80 (0,000)
	Panel C: Allgemeines Interesse an einer Frage		t-Statistic (p-Wert)
BONUSPUNKTE	4,57	4,98	28,96 (0,000)
HITS	3,86	4,29	115,12 (0,000)
F_HILFREICH	0,09	0,20	90,79 (0,000)
F_ANTWORTEN	0,88	1,21	189,90 (0,000)

Beantwortung seiner Frage ist. Die unabhängigen Variablen werden anhand des von Hosmer und Lemeshow [18] empfohlenen Grenzwertes von 0,25 für die in der vierten Spalte in Tabelle 3 präsentierten p-Werte ausgewählt. Da keiner der p-Werte größer als 0,25 ist, werden alle in Tabelle 3 aufgeführten Variablen in das logistische Regressionsmodell aufgenommen.

Tabelle 4 zeigt die Ergebnisse der logistischen Regression. In der Tabelle sind in der ersten Spalte die unabhängigen Variablen, in der zweiten Spalte die anhand unserer Hypothesen vermuteten Wirkzusammenhänge, in der dritten Spalte die Regressionskoeffizienten, in der vierten Spalte die sogenannten Effektkoeffizienten und in der fünften Spalte die p-Werte dargestellt. Analog zu Abschnitt 4 interpretieren wir diese wiederum als Goodness-of-Fit Indikatoren und nicht als Wahrscheinlichkeiten dafür, dass die Regressionskoeffizienten signifikant von Null verschieden sind. Der Tabelle kann man entnehmen, dass mit Ausnahme der Koeffizienten für die Variablen FREUNDE und EINGELADEN alle Koeffizienten einen guten Fit ausweisen. Weiterhin haben alle Koeffizienten, deren Betrag größer als 0,1 ist, das durch die Theorie vorhergesagte Vorzeichen.[5]

Anders als bei einer klassischen OLS Regression ist eine lineare Interpretation oder ein Vergleich der Regressionskoeffizienten untereinander bei einer logistischen Regression nicht möglich. Anhand der Regressionskoeffizienten kann nur auf die Wirkungsrichtung eines Zusammenhangs geschlossen werden [3]. Eine Angabe der Stärke des Wirkzusammenhanges ist jedoch mit Hilfe der Effektkoeffizienten (vgl. Spalte 4) möglich. Steigt z.B. die Variable TOP um eins, so wächst das Chancenverhältnis (Odds), d.h. der Quotient aus der Wahrscheinlichkeit, mindestens eine Top-Antwort zu erhalten, und der Gegenwahrscheinlichkeit um den Faktor 1,219. Ist der Effektkoeffizient dagegen kleiner als eins so sinkt dieses Chancenverhältnis. Ein Effektkoeffizient von genau eins hat keine Auswirkung auf das Chancenverhältnis.

[5] Alle Variablen deren Effektkoeffizienten um mehr als 0,1 von 1 verschieden sind werden in Tabelle 4 fett hervorgehoben.

Tabelle 4: Ergebnisse der logistischen Regression

Variable	Vermutete Wirkungsrichtung	*Regressionskoeffizient B*	Exp^B	p-Wert
		Panel A: Sozialkapital und Tauschtheorie		
FREUNDE	+	-0,005	0,995	0,309
EINGELADEN	+	0,009	1,009	0,779
ANTWORTEN	**+**	**-0,135**	**0,874**	**0,000**
HILFREICH	+	-0,029	0,971	0,001
TOP	**+**	**0,198**	**1,219**	**0,000**
VERSCHICKT	+	0,047	1,048	0,000
ERHALTEN	+	-0,044	0,957	0,000
LINKS	+	0,006	1,006	0,049
		Panel B: Uncertainty Reduction Theory		
BILD	**+**	**0,395**	**1,484**	**0,000**
GESCHLECHT	+	-0,030	0,971	0,001
NAME	+	0,065	1,067	0,000
POSTLEITZAHL	+	-0,073	0,929	0,000
LAND	**+**	**0,207**	**1,230**	**0,000**
GEBURTSDATUM	+	-0,064	0,938	0,000
BERUF	+	-0,032	0,968	0,000
		Panel C: Allgemeines Interesse an einer Frage		
BONUSPUNKTE	+	0,016	1,016	0,000
HITS	**+**	**0,100**	**1,106**	**0,000**
F_HILFREICH	**+**	**0,498**	**1,645**	**0,000**
F_ANTWORTEN	**+**	**1,173**	**3,231**	**0,000**

Hypothese H1 wird durch den relativ hohen Effektkoeffizienten für die Variable TOP unterstützt. Das negative Vorzeichen sowie der relativ hohe Effektkoeffizient für die Variable ANTWORTEN sprechen jedoch gegen die in Abschnitt 3.1 aufgestellte Hypothese. Die weiteren dieser Hypothese zugeordneten Variablen haben nur einen vernachlässigbaren Einfluss auf die Wahrscheinlichkeit, dass mindestens eine Antwort auf eine Frage als Top-Antwort bewertet wird.

Ein ähnliches Bild ergibt sich für Hypothese H2. Hier haben nur die Effektkoeffizienten für die Variablen BILD und LAND einen substantiellen Einfluss auf die Wahrscheinlichkeit, dass mindestens eine Antwort auf eine Frage als Top-Antwort bewertet wird. Der Einfluss der weiteren dieser Hypothese zugeordneten Variablen ist auch hier vernachlässigbar.

Den größten Einfluss auf die Wahrscheinlichkeit der Beantwortung einer Frage mit mindestens einer Top-Antwort haben die Variablen, die unserer dritten Hypothese zugeordnet sind. Hier sind insbesondere die Effektkoeffizienten für die Variablen F_ANTWORTEN und F_HILFREICH hervorzuheben.

Die Modellgüte wird in einem ersten Schritt anhand eines Likelihood-Ratio-Tests beurteilt. Bei diesem Test wird die Anpassung des vollständig spezifizierten Modells mit einem Modell, welches nur eine Konstante enthält, verglichen. Als Ergebnis dieses Tests kann die Nullhypothese, dass alle Regressionskoeffizienten im vollständig spezifizierten Modell gleich Null sind, abgelehnt werden. Somit verfügt das vollständig spezifizierte Modell über eine im Vergleich zum Nullmodell verbesserte Vorhersagekraft bezüglich des Auftretens mindestens einer Top-Antwort.

Mit Werten von 0,183 für das Nagelkerke-R^2 und 0,137 für das Cox & Snell R^2 liegen die Werte für die Pseudo-R^2 Statistiken leicht unter dem von der Literatur als akzeptabel empfohlenen Niveau [3].[6] Dies lässt darauf schließen, dass neben den im Modell betrachteten noch weitere unabhängige Variable existieren, die einen Einfluss auf die abhängige Variable ausüben. Hieraus ergibt sich jedoch kein Widerspruch zum Ziel unserer Studie, den Einfluss der betrachteten unabhängigen Variablen auf die Wahrscheinlichkeit mindestens eine Top-Antwort zu erhalten, zu untersuchen. Weiterhin zeigt der Vergleich mit anderen Studien, dass auch Modelle mit weitaus geringeren Werten für die

[6] Analog zum Bestimmtheitsmaß der klassischen, linearen Regression zeigen diese Werte, welcher Anteil der Variation der abhängigen Variablen durch das vollständig spezifizierte Modell erklärt wird. Anders als bei der linearen Regression wird jedoch nicht auf den Anteil der erklärten Varianz, sondern auf das Verhältnis der Likelihood-Werte zwischen dem vollständigen und dem Null-Modell zurückgegriffen.

Pseude-R^2-Statistiken als ausreichend eingestuft wurden. (vgl. z.B. [12] oder [13]).

Die Beurteilung der Klassifikationsergebnisse bietet weitere Hinweise auf einen guten Modellfit. Der Vergleich der Klassifikationsergebnisse des vollständig spezifizierten Modells (63,4% der Beobachtungen werden richtig klassifiziert) mit einer reinen Zufallsklassifikation (50,8% der Beobachtungen werden richtig klassifiziert) zeigt, dass unser Modell einen substantiellen Beitrag zur korrekten Klassifikation der Beobachtungen leistet. Ein berechneter Wert von 21.224 für Press's Q unterstützt dieses Ergebnis.[7] Wir verzichten auf die Durchführung eines Hosmer-Lemeshow-Tests, da dieser bei nicht perfekt fittenden Modellen sensitiv auf die Größe des Datensatzes reagiert [20].

Zusammengefasst deuten somit alle Maße auf eine akzeptable Güte des Modells hin.

6. Diskussion

Insgesamt können alle in Abschnitt 3 aufgestellten Hypothesen durch die empirische Analyse bestätigt werden, jedoch sind die Ergebnisse differenziert zu betrachten.

Tabelle 4 zeigt, dass nur für Hypothese H3 alle Regressionskoeffizienten einen guten Fit und gleichzeitig das erwartete Vorzeichen aufweisen. Die vergleichsweise hohen Effektkoeffizienten für die Variablen HITS, F_HILFREICH, und F_ANTWORTEN zeigen, dass der größte Teil der durch unser Modell erklärten Wahrscheinlichkeit, mindestens eine Top-Antwort zu erhalten, im Interesse an einer Frage begründet liegt. Dieses Ergebnis ist durchaus plausibel, da ein ausreichendes Interesse für eine Frage notwendige Bedingung dafür ist, überhaupt eine Antwort zu erhalten. Weiterhin signalisiert ein hohes Interesse an einer Frage eine tendenziell höhere Fragequalität, welche sich zusätzlich positiv auf die Qualität der Antworten auswirken könnte. Dagegen hat die Höhe der ausgesetzten Belohnung für die Beantwortung einer Frage im Gegensatz zu der Studie von Harper et al. [16] nahezu keinen Einfluss auf die Qualität der Antworten. Betrachtet man jedoch, dass es sich bei der untersuchten Belohnung bei Harper et al. [16] um reales Geld, in dieser Studie jedoch nur um quasi wertlose Bonuspunkte handelt, ist dieses Ergebnis weniger verwunderlich.

Die Hypothesen H1 und H2 müssen dagegen differenzierter betrachtet werden. Die Effektkoeffizienten der Hypothese H1 zugeordneten Variablen FREUNDE, EINGELADENE, VERSCHICKT, ERHALTEN und LINKS zeigen, dass nicht jede Form der Partizipation in der Community einen positiven Einfluss auf die Qualität der Beantwortung der eigenen Fragen hat. Dieses Bild wird durch den Effektkoeffizienten für die Variable ANTWORTEN noch bestätigt. Dieser zeigt, dass von anderen Nutzern eher negativ bewertete Antworten einen negativen Einfluss auf die Qualität der Antworten auf eine eigene Frage haben, wohingegen sich sehr gute Antworten auf Fragen Anderer positiv auf diese Qualität auswirken. Dieses Ergebnis kann dahingehend interpretiert werden, dass im Rahmen der untersuchten Community Q&A Site ein tatsächlicher Austausch von qualitativ hochwertigen Antworten stattfindet und dass es nicht möglich ist, andere Formen der Partizipation gegen qualitativ hochwertige Antworten einzutauschen.

Ein ähnliches Bild ergibt sich für die Variablen die unserer zweiten Hypothese zugeordnet werden. Nur das Vorhandensein eines Bildes und die Angabe des Herkunftslandes des Fragestellers haben einen substantiellen Einfluss auf die Qualität der Beantwortung einer Frage. Ein möglicher Erklärungsansatz für den nahezu nicht vorhandenen Einfluss der anderen Variablen könnte der in jedem Fall fragliche, zusätzliche Informationsgehalt dieser Variablen sein. Hierbei ist weiterhin zu berücksichtigen, dass die bei der Registrierung angegeben Daten nicht durch das Team der untersuchten Community Q&A Site verifiziert werden und somit insbesondere bei sensiblen, persönlichen Daten davon auszugehen ist, dass nicht immer die Wahrheit angegeben wird.

Zusammenfassend zeigen die Ergebnisse der logistischen Regression, dass ein Fragesteller, der eine qualitativ hochwertige Beantwortung seiner Frage wünscht, insbesondere auf eine qualitativ hochwertige und interessante Fragestellung Wert legen sollte. Durch Hinzufügen eines Nutzerbildes sowie der Angabe des Herkunftslandes ist es möglich, die Wahrscheinlichkeit einer qualitativ hochwertigen Beantwortung weiter zu steigern. Aktive Teilnahme in der Community wirkt sich hingegen nur bei sehr hoher Qualität von vergangenen Antworten des Fragestellers positiv auf diese Wahrscheinlichkeit aus.

Vor dem Hintergrund des vergleichsweise hohen Zeitaufwandes für die qualitativ hochwertige Teilnahme an einer solchen Community Q&A Site kann die Frage aus der Überschrift dieses Beitrags daher tendenziell mit Nein beantwortet werden. Allerdings zeigen die Ergebnisse von Krasnova et al. [21], dass neben einer reinen Nutzenüberlegung viele weitere Faktoren auf die Partizipation in Online Communities wirken. Diese könnten einen substantiellen Beitrag zur Erklärung der vielen aktiven Nutzer auf Community Q&A Sites liefern.

7. Zusammenfassung und Ausblick

"A major challenge for IS research lies in making models and theories that were developed in other academic disciplines usable in IS research and practice" [4]. In diesem Beitrag nutzen wir die Uncertainty Reduction Theory aus der Psychologie sowie die sozio-ökonomische Tauschtheorie, um Einflussfaktoren zu identifizieren, die in Q&A-Communities zu besonders wertvollen Antworten auf gestellte Fragen führen. Die Analyse wird erstmals mit einem umfangreichen Datensatz einer deutschsprachigen Community Q&A Site durchgeführt.

Die Ergebnisse dieser Studie sind sowohl aus praktischer als auch aus wissenschaftlicher Sicht von Interesse. Bspw. können Community-Betreiber das Community-Design und die Teilnahme-Incentivierungssysteme dergestalt anpassen, dass vermehrt mit hilfreichen Antworten von Nutzern gerechnet werden kann. Dies kann zu einer Steigerung der Attraktivität und des Wertes einer Community insgesamt führen. Ebenso lassen sich für den individuellen Nutzer Verhaltensempfehlungen ableiten, welche bei Befolgung die Chancen auf wertvolle Antworten erhöhen. Für die Wissenschaft finden sich Hinweise darauf, dass das Verhalten der Nutzer teilweise mit der Tauschtheorie und der Uncertainty Reduction Theory erklärt werden kann.

[7] Übersteigt Press's Q den kritischen Wert von 3,84, sind die Klassifikationsergebnisse signifikant ($\alpha = 0{,}05$) von einer zufälligen Zuordnung verschieden [3].

Es gibt noch zahlreiche Richtungen, diese Analyse zu verfeinern und zu erweitern. Zum ersten ist eine automatische Textanalyse der Fragen und Antworten geplant, um treffgenauer tägliche Begrüßungen, Glückwünsche zum Statuswechsel und Scherzfragen aus dem analysierten Datensatz zu entfernen. Hinweise zur Umsetzung finden sich z.B. bei [17]. Zum zweiten ist der Vergleich mit einer alternativen Modellspezifikation von Interesse. Als abhängige Variable wird statt der Bewertung von Antworten durch den Frager selbst die Bewertung von Antworten durch andere Nutzer verwendet. Zum dritten ist eine Separierung von Effekten des persönlichen Netzwerks von den Effekten der gesamten Community vorgesehen.

8. Literatur

[1] Agichtein, E., Castillo, C., Donato, D., Gionis, A. und Mishne, G. 2008. Finding High-Quality Content in Social Media. In *Proceedings of the 1st ACM Intl. Conf. in Web Search and Data Mining (WSDM)* (Palo Alto, California, February 11-12, 2008).

[2] Alexa, The Web Information Company. 2010. Top Sites > Reference. Zuletzt abgerufen: August 20, 2010, http://www.alexa.com/topsites/category/Reference.

[3] Backhaus, K., Erichson, B., Plinke, W. und Weiber, R. 2008. *Multivariate Analysemethoden – Eine anwendungsorientierte Einführung*. 12. Aufl., Springer, Berlin.

[4] Benaroch, M., Kauffman, R.1999. A Case for Using Real Options Pricing Analysis to Evaluate Information Technology Project Investments. *Information Systems Research*, 10, 1, 70-86.

[5] Berger, C. und Calabrese, R. 1975. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. *Human Communication Research*, 1, 2, 99-112.

[6] Bourdieu, P. 1983. Ökonomisches Kapital, kulturelles Kapital, soziales Kapital. In *Soziale Ungleichheiten, Soziale Welt*, R. Krekel, Ed. Schwartz, Göttingen, 183-198.

[7] Brass, D. 1995. A Social Network Perspective on Human Resources Management. In *Research in Personnel and Human Resources Management* Ferris, G. Ed. JAI Press, Greenwich, CT, 39-79.

[8] Coleman, J. 1988. Social Capital and the Creation of Human Capital. *American Journal of Sociology*, 94, 95-120.

[9] Coleman, J. 1990. *Foundations of Social Theory*. Harvard University Press. Cambridge, MA.

[10] Flap, H. und de Graaf, ND. 1986. Social Capital and Attained Occupational Status. *The Netherlands Journal of Sociology*, 22, 145-161.

[11] Gargiulo, M., und Bernassi, M. 1999. The dark side of social capital. In *Corporate social capital and liability* R. Leenders and S. Gabbay Eds. Kluwer, Boston, 298-322.

[12] Gefen, D. und Carmel, E. 2008. Is the World Really Flat? A Look at Offshoring at an Online Programming Marketplace. *MIS Quarterly*, 32, 2, 367-384.

[13] Haas, A. 2009. Kann zu viel Kundenorientierung nachteilig sein? Eine Analyse der Wirkung der Kundenorientierung von Verkäufern auf die Kaufentscheidung. *Zeitschrift für Betriebswirtschaft*, 79, 7-29.

[14] Hair, J., Andersion, R., Tatham, R. und Black, W. 1992. *Multivariate Data Analysis with Readings*. 4th Ed., Prentice Hall, Englewood Cliffs, NJ.

[15] Hand, D. 1999. Statistics and Data Mining: Intersecting Disciplines. *ACM SIGKKD Explorations*, 1,1, 112-118.

[16] Harper, M., Raban, D., Rafaeli, S. und Konstan, J. 2008. Predictors of Answer Quality in Online Q&A Sites. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05-10, 2008). ACM, New York, NY.

[17] Harper, M., Moy, D. und Konstan, J. 2009. Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (Boston, USA, April 4-9, 2009).

[18] Hosmer, D. und Lemeshow, S. 1989. *Applied Logistic Regression*. Wiley, New York.

[19] Iseke, A. 2009. Humankapital und Sozialkapital: Gibt es einen Matthäus-Effekt bezüglich der Sozialkapitalbildung von Nachwuchsführungskräften? *Zeitschrift für betriebswirtschaftliche Forschung*, 61, 3, 186-211.

[20] Kramer, A. und Zimmermann, J. 2007. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Critical Care Medicine*, 35, 9, 2052-2056.

[21] Krasnova, H., Hildebrand, T., Günther, O., Kovrigin, A. und Nowobilska, A. Why Participate in an Online Social Network: An Empirical Analysis. In *Proceedings of the 16th European Conference on Information Systems (ECIS)* (Galway, Ireland, June 09-11, 2008).

[22] Liu, Y., Bian, J., Agichtein, E. 2008. Predicting Information Satisfaction in Community Question Answering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008)* (Singapore, July 20-24, 2008).

[23] Pomerantz, J., Nicholson, S., Belanger, Y. and Lankis, D. 2004. The Current State of Digital Reference: Validation of a General Digital Reference Model through a Survey of Digital Reference Services. *Information Processing and Management*, 40, 2, 347-363.

[24] Rosenbaum, H. und Shachaf, P. 2010. A Structuration Approach to Online Communities of Practice: The Case of Q&A Communities. *Journal of the American Society for Information Science and Technology*, im Erscheinen.

[25] Shah, C., Oh, S. und Oh, J. 2009. Research Agenda for Social Q&A. *Library and Information Science Research*, 11, 4, 205-209.

[26] Shah, C. und Pomerantz, J. 2010. Evaluating and Predicting Answer Quality in Community QA. In *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010)* (Geneva, Switzerland, July 19 – 23, 2010).

[27] Storck, J. und Sproull, L. 1995. Through a Glass Darkly What do People Learn in Videoconferences? *Human Communication Research*, 22, 2, 197-219.

[28] Tidwell, L. und Walther, J. 2002. Computer-Mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations. *Human Communication Research*, 28, 3, 317-348.

[29] Walther, J., Slovacek, C. and Tidwell, L. 2001. Is a Picture Worth a Thousand Words? Photographic Images in Long Team and Short Team Virtual Teams. *Communication Research*, 28, 1, 105-134.

[30] Zheng, J., Veinott, E., Bos, N., Olson, J. and Olson, G. 2002. Trust without Touch: Jumpstarting Long-Distance Trust with Initial Social Activities. In *Proceedings of the SIGCHI Conference in Computing Systems* (Minneapolis, USA, April, 20-25, 2002) ACM, New York, NY.

Frühwarnsystem zur Identifikation kritischer Situationen der Meinungsbildung im Web 2.0

Carolin Kaiser, Sabine Schlick
Lehrstuhl Wirtschaftsinformatik II, Universität Erlangen-Nürnberg
Lange Gasse 20, 90403 Nürnberg
Deutschland
{carolin.kaiser, sabine.schlick}@wiso.uni-erlangen.de

ZUSAMMENFASSUNG

Im Web 2.0 schließen sich immer mehr Menschen zu Communities zusammen, in denen sie sich über die Produkte von Unternehmen austauschen. Dabei beeinflussen sie gegenseitig ihre Meinungen. Durch eine Beobachtung und Analyse der Meinungsbildung können Unternehmen Risiken frühzeitig erkennen.

In dieser Arbeit wird ein Frühwarnsystem zur Identifikation kritischer Situationen der Meinungsbildung im Web 2.0 vorgestellt. Das System ermöglicht es, den Meinungsbildungsprozess in Communities automatisch zu analysieren und zu beurteilen. Bei Eintritt einer kritischen Situation erfolgt eine Warnung an den zuständigen Marketing-Manager. Dieser kann durch rechtzeitiges Eingreifen negative Konsequenzen abwenden.

Das Frühwarnsystem wird als Neuro-Fuzzy-System konzipiert. Es kann anhand der Einschätzung früherer Situationen linguistische Regeln erlernen, welche die Beurteilung zukünftiger Situationen erlauben. Dadurch ermöglicht es eine einfache Wissensakquisition. Außerdem sind die Regeln aufgrund ihrer linguistischen Form gut interpretierbar. Anhand zweier Anwendungsfälle wird gezeigt, wie das konzipierte System zur Identifikation kritischer Situationen im Web 2.0 eingesetzt werden kann.

Schlüsselwörter

Frühwarnsystem, Neuro-Fuzzy-System, Opinion Mining, Soziale Netzwerkanalyse, Web 2.0

1. EINLEITUNG

In den letzten Jahren fand eine Verlagerung der sozialen Aktivitäten von der realen in die virtuelle Welt statt. So schließen sich im Web 2.0 immer mehr Kunden zu Communities zusammen, in denen sie Produkte bewerten und diskutieren. Dabei beeinflussen sie gegenseitig ihre Meinungen und somit ihre späteren Kaufentscheidungen. Laut einer Studie von TNS Infranet spielt für 43 % aller Deutschen, die das Internet zur Recherche vor dem Kauf nutzen, der interaktive Austausch über Produkte mit anderen eine wichtige Rolle, um so unterschiedliche Erfahrungen und Meinungen über Produkte zu erhalten [29]. Die diskutierten Meinungen wirken sich dabei nicht nur auf die Kaufentscheidungen derjenigen aus, die sich aktiv an den Diskussionen beteiligen, sondern auch auf die Kaufentscheidungen der Kunden, die das Internet zur Informationssuche nutzen.

Für die Unternehmen hat dies, je nachdem welche Meinungen vertreten werden, positive oder negative Folgen. Der Vorteil des Internets ist, dass Meinungsäußerungen in Echtzeit und kostenfrei beobachtet werden können. Dadurch besteht die Möglichkeit, Risiken frühzeitig zu erkennen und Gegenmaßnahmen rechtzeitig zu ergreifen. Durch richtiges Handeln können bestehende Gefahren abgewendet werden. Die Unternehmen sind somit in der Lage sich besser an den Kundenwünschen zu orientieren, was sich wieder positiv auf die Kundenzufriedenheit und die Kundenbindung auswirkt [2].

Die Herausforderung bei der Systementwicklung liegt in der Identifikation der kritischen Situationen. Es wäre mit unverhältnismäßig hohem Aufwand und Kosten verbunden auf jede einzelne negative Meinungsäußerung zu reagieren. Vielmehr muss die Situation als Ganzes gesehen werden. Eine besondere Bedeutung haben z.B. sogenannte Meinungsführer, d.h. Personen, die einen hohen Einfluss auf andere Personen haben [20]. Auch der Vergleich des eigenen Unternehmens mit den Konkurrenten spielt eine Rolle bei der Einschätzung einer Situation. Ein System ist erforderlich, mit dem kritische Situationen automatisch identifiziert und dem Unternehmen angezeigt werden können.

Es wird ein Frühwarnsystem zur Identifikation kritischer Situationen der Meinungsbildung im Web 2.0 vorgestellt, das auf einem Neuro-Fuzzy-Ansatz beruht. Das Frühwarnsystem lernt linguistische Regeln auf Basis der Einschätzung früherer Situationen. Diese Regeln ermöglichen es, zukünftige Situationen zu beurteilen. Aufgrund ihrer linguistischen Form sind die Regeln leicht für Marketing-Manager verständlich. Die Einflussfaktoren, die zur Klassifikation der Situationen erforderlich sind, werden automatisch aus Unternehmensdatenbanken und dem Internet mithilfe von Methoden der Informationsextraktion, des Text Mining und der sozialen Netzwerkanalyse gewonnen. Das Frühwarnsystem wurde zur Evaluierung auf zwei Szenarien angewandt: eine Bewertungsplattform, in der Meinungen zu Fußballschuhen ausgetauscht werden, und ein soziales Netzwerk, in dem Meinungen zu Computerspielen diskutiert werden. Die Ergebnisse dieser Anwendungsfälle werden dargestellt.

2. VERWANDTE ARBEITEN

Opinion Mining zielt darauf ab, Meinungen in Texten zu erkennen. Es wird zunehmend dazu verwendet, um Konsumentenmei-

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

nungen im Internet aufzudecken. In zahlreichen Forschungsarbeiten werden Text Mining Ansätze vorgestellt, die die automatische Identifikation von Meinungen zu Produkten und ihren Eigenschaften erlauben ([15], [8], [25], [26], [12], [19]). Diese Ansätze betrachten Meinungen jedoch nur zu einem bestimmten Zeitpunkt.

In anderen Arbeiten liegt der Fokus auf der dynamischen Entwicklung von Inhalten. Beispielweise befassen sich Viermetz et al. [34] mit einer Methode zur Beobachtung der Entwicklung von kurzfristigen und langfristigen Thementrends über die Zeit hinweg. Tong und Yager [30] beschreiben ein System zur automatischen Zusammenfassung der Inhalte von öffentlichen Online-Diskussionen. Dazu werden die Informationen aus diesen Diskussionen in Form von Zeitreihen abgebildet. Das System fasst diese zusammen und generiert daraus eine leicht verständliche linguistische Beschreibung. Huang et al. [14] beobachten die Entwicklung von Online-Communities. Sie untersuchen die Entstehung und Auflösung von Clustern in Folge von veränderten Interessen der Community-Mitglieder. Choudhury et al. [5] analysieren die Entwicklung von Online-Communities mithilfe sogenannter Schlüsselgruppen, deren Diskussionsthemen repräsentativ für die Community sind. Durch Beobachtung der Schlüsselgruppen über verschiedene Zeitpunkte hinweg kann auf die Dynamik in der gesamten Community geschlossen werden. Alle vier Ansätze ermöglichen Aussagen über die dynamischen Veränderungen in der Vergangenheit. Zukünftige Entwicklungen sind jedoch nicht Gegenstand der Untersuchungen.

Einige Wissenschaftler nutzen die Aktivität in Online-Umgebungen zur Prognose von zukünftigen Verkäufen. Gruhl et al. [13] untersuchen den Zusammenhang zwischen der Nennung verschiedener Büchern in Blogs und dem Sales Rank dieser Bücher bei Amazon. Mit ihrem Algorithmus können anhand der Nennung von Büchern in Blogs, Spitzen in den Sales Ranks automatisch vorhergesagt werden. Dhar und Chang [9] kommen zu dem Ergebnis, dass mithilfe von User-generated Content Aussagen über Musikverkäufe gemacht werden können. Eine Rolle spielen dabei die Häufigkeit der Nennungen der jeweiligen CDs in Blogs und die Networking-Intensität des Musikers. Die Networking-Intensität wird durch die Anzahl der Personen, die den jeweiligen Musiker als Freund auf ihrer eigenen MySpace-Seite hinzugefügt haben, gemessen. Ähnliche Untersuchungen finden sich bei Onishi und Manchanda [24] für die Vorhersage der Verkäufe von grünem Tee, Filmen und Mobilfunktelefonen. Sie kommen ebenfalls zu dem Ergebnis, dass Blogs prädiktiv für Verkäufe sind. Diese Ansätze stellen auf die Konsequenzen von Online-Aktivitäten ab, prognostizieren jedoch nicht zukünftiges Online-Verhalten.

Andere Arbeiten beschäftigen sich mit der Vorhersage der Gruppenaktivität, des Kommunikationsflusses oder der Meinungsbildung. Choudhury [4] befasst sich mit Gruppenaktivitäten (z.B. Einstellen neuer Inhalte) in sozialen Medien. Sie zeigt auf, wie Aktivitäten über die Zeit hinweg modelliert und prognostiziert werden können. Choudhury et al. [3] schlagen ein Verfahren zur Vorhersage des Kommunikationsflusses in sozialen Netzwerken vor. Anhand der Umgebung eines Users, d.h. seines sozialen Netzwerks, seines Kommunikationsverhaltens bezüglich des betrachteten Themas in der Vergangenheit und der Identität des Empfängers, wird der zukünftige Kommunikationsfluss vorhergesagt. Kaiser [16] stellt ein System vor, das Meinungen und Kommunikationsbeziehungen mithilfe von Text Mining automatisch extrahiert und mittels der sozialen Netzwerkanalyse Meinungsführer und Meinungstrends identifiziert. Dadurch kann die Meinungsbildung analysiert und prognostiziert werden. Diese Arbeiten betrachten jedoch nur einzelne Faktoren der Meinungsbildung.

Die vorgestellten Ansätze sind zur Frühwarnung nur bedingt geeignet. Bei den Methoden zur statischen und dynamischen Meinungsanalyse erfolgt die Warnung zu spät, da sie dann generiert wird, wenn negative Folgen bereits eingetreten sind. Die prognosebasierten Ansätze ermöglichen ein frühzeitiges Eingreifen, da sie sich mit der zukünftigen Meinungsentwicklung beschäftigen. Allerdings wird hier nicht die Gesamtsituation beurteilt. Das hier dargestellte Frühwarnsystem zielt darauf ab, kritische Situationen frühzeitig zu erkennen. Situationen werden hierbei aufgrund aller wichtigen Einflussfaktoren beurteilt. Die Beobachtung der Faktoren erfolgt über die Zeit hinweg. Bei kritischen Situationen werden Marketing-Manager gewarnt. Dadurch sind sie in der Lage geeignete Marketing Maßnahmen rechtzeitig einzuleiten.

3. ANSATZ

Aufgabe des Frühwarnsystems ist es, Marketing-Manager im Fall von kritischen Situationen zu warnen. Dabei werden solche Situationen als kritisch angesehen, in denen die Gefahr besteht, dass sich negative Meinungen verbreiten, die Image und Absatz des Unternehmens gefährden. Das Frühwarnsystem (siehe Abbildung 1) basiert auf einer Wissensbasis. Im Rahmen der Wissensakquisition werden Daten aus internen und externen Quellen extrahiert und in strukturiertes Wissen überführt. Hierbei werden Erfolgsfaktoren, wie zum Beispiel der Absatz des betrachteten Produktes, aus Unternehmensdatenbanken gewonnen. Darüber hinaus werden Konsumentenmeinungen aus dem Internet heruntergeladen und mittels Text Mining als positiv, negativ oder neutral klassifiziert. Desweiteren werden Beziehungen in sozialen Netzwerken des Internets analysiert, um Meinungsführer zu identifizieren und die Netzwerkstruktur zu charakterisieren. Auf Basis der Erfolgsfaktoren, der Meinungen aus dem Web und den Netzwerkcharakteristika beurteilt das Frühwarnsystem die Situationen. Hierzu wird ein Neuro-Fuzzy-System eingesetzt, das Regeln zur Unterscheidung zwischen kritischen und nicht kritischen Situationen erlernt. Die Regeln werden aus vergangen Situationen abgeleitet und ermöglichen die Bewertung zukünftiger Situationen. Wenn Situationen als kritisch eingeschätzt werden, erfolgt eine Warnung an die Marketing-Manager. Somit sind sie in der Lage, frühzeitig Marketing-Maßnahmen einzuleiten, um die Verbreitung von negativen Meinungen zu verhindern.

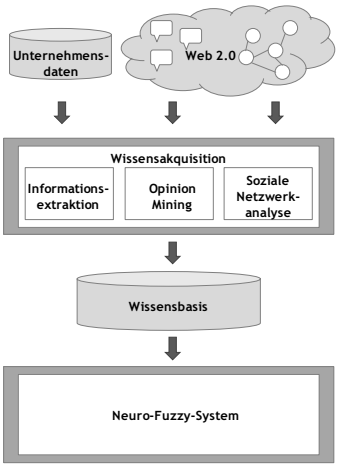


Abbildung 1: Frühwarnsystem

Das Frühwarnsystem wurde auf zwei Szenarien angewandt: die Bewertungsplattform „fussball-forum.de“, auf der Meinungen über Fußballschuhe ausgetauscht werden, und das soziale Netzwerk „gamestar.de“, in dem Erfahrungen zu Computerspielen diskutiert werden.

4. WISSENSAKQUISITION

4.1 Informationsextraktion

Die Zielsetzung der Informationsextraktion besteht darin, die gewünschten Erfolgsfaktoren für das betrachtete Produkt aus mehreren Quellsystemen zu gewinnen. Diese Aufgabe kann sehr zeitaufwendig und komplex sein, abhängig von den zugrunde liegenden Quellsystemen [7]. Relevante Daten sind in verschiedenen operativen Systemen von verschiedenen Organisationen gespeichert, wie zum Beispiel Online Shops, eigene Läden und unabhängige Kaufhäuser. Die Daten können in unterschiedlichen Formaten vorliegen, sodass eine Transformation erforderlich sein kann. Zum Beispiel kann es notwendig sein, Zeitangaben in ein standardisiertes Format zu überführen oder Absatzangaben in die gleiche Währung zu bringen. Darüber hinaus ist eine Aggregation durchzuführen. So müssen beispielsweise die Absatzzahlen verschiedener Läden zu einem Gesamtabsatz aggregiert werden. Dieser Prozess der Informationsextraktion wird zu Update-Zwecken periodisch durchgeführt.

4.2 Opinion Mining

Ziel des Opinion Mining ist es, die Polarität von Konsumentenmeinungen auf Basis ihrer Produktbewertungen in Reviews oder Posts zu erkennen. Hierbei werden drei Polaritätsklassen unterschieden: positiv, negativ und neutral.

Text Mining Methoden ermöglichen die automatische Erkennung von Meinungen. Der Prozess zur Meinungsidentifikation besteht aus zwei aufeinanderfolgenden Schritten [15]. Im ersten Schritt werden Merkmale aus den Bewertungen extrahiert. Im zweiten Schritt wird die Polarität der Bewertungen mithilfe der extrahierten Merkmale bestimmt.

Die Merkmalsextraktion basiert auf der statistischen und linguistischen Analyse der Bewertungen [36]. Der Text aus den Bewertungen wird zunächst in Wörter zerlegt. Nach Entfernung aller Stop-Wörter (z.B. Artikel) wird für alle verbleibenden Worte eine Wortstammreduktion durchgeführt. Anschließend werden diejenigen Wortstämme als Merkmale ausgewählt, die häufig in einer Polaritätsklasse, aber nur selten in den anderen beiden Polaritätsklassen auftreten.

Mithilfe dieser Merkmale werden die Bewertungen in die Klassen positiv, negativ oder neutral eingeteilt. Grundsätzlich können zur Klassifikation verschiedene Lernalgorithmen herangezogen werden, wie z.B. Hidden Marcov Models oder Maximum Entropy [36]. In diesem Projekt kommt die Stützvektormethode [6] zum Einsatz, da sie die Verarbeitung vieler Merkmale erlaubt und gute Klassifikationsergebnisse in ähnlichen Projekten geliefert hat [25].

Da die Stützvektormethode zur Klasse der überwachten Lernalgorithmen zählt, ist ein Trainingsdatensatz bestehend aus den Konsumentenbewertungen mit extrahierten Merkmalen und von Menschen zugeordneten Polaritäten erforderlich. Mithilfe dieses Trainingsdatensatzes können Parameter von binären Klassifikationsfunktionen erlernt werden. Zur Unterscheidung von drei Klassen werden drei Klassifikationsfunktionen erlernt, die positive von nicht positiven Bewertungen, negative von nicht negativen Bewertungen und neutrale von nicht neutralen Bewertungen trennen können. Bewertungen werden am Ende der Klasse zugeordnet, die die höchste Wahrscheinlichkeit aufweist. Im vereinfachten zweidimensionalen Fall lässt sich eine binäre Klassifikationsfunktion als Gerade visualisieren, die die Bewertungen in zwei Klassen einteilt (siehe Abbildung 2). Bewertungen, die auf der gleichen Seite der Geraden liegen, gehören zur gleichen Klasse.

Im Anschluss an die Klassifikation werden die Meinungen den Verfassern der Bewertungen zugeordnet. Für jede Person wird die durchschnittliche Meinung pro Zeiteinheit ermittelt. Die durchschnittliche Meinung wird auf einer Skala von -1 bis +1 gemessen. Während ein Wert von -1 angibt, dass eine Person nur negative Meinungen zum Produkt abgegeben hat, zeigt ein Wert von +1 an, dass eine Person sich ausschließlich positiv über ein Produkt geäußert hat.

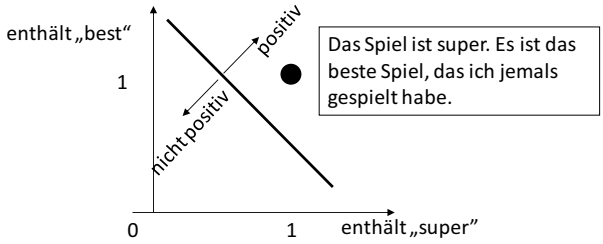


Abbildung 2: Klassifikation von Meinungen

4.3 Soziale Netzwerkanalyse

Gemäß der Diffusionstheorie [27] hat das soziale Umfeld von Personen eine große Auswirkung auf ihre Einstellungen. Personen bilden ihre Meinungen nicht unabhängig voneinander, sondern in Interaktion mit anderen Mitgliedern ihres sozialen Netzwerkes. Hierbei spielen zwei Faktoren bei der Meinungsbildung eine wichtige Rolle: die Meinungsführer und die Struktur des Netzwerkes.

Meinungsführer sind zentrale Personen in einem Netzwerk, die großen Einfluss auf die Meinung anderer Mitglieder des Netzwerks haben ([17], [33]). Bereits eine geringe Anzahl an Meinungsführern ist in der Lage, eine große Anzahl an Personen im Netzwerk zu beeinflussen [18]. Einzelne Meinungsführer können dabei einen unterschiedlichen Grad an Einfluss haben [27].

Zentralitätskennzahlen aus der sozialen Netzwerkanalyse ([35], [28]) ermöglichen die Identifikation von Meinungsführern und die Bestimmung ihres Einflusses. Diese Kennzahlen können Werte auf einer Skala von null bis eins annehmen. Während ein Wert von eins für maximalen Einfluss steht, steht ein Wert von null für minimalen Einfluss.

Die Kennzahl Degree Zentralität zeigt, wie bedeutend eine Person innerhalb ihrer Nachbarschaft ist. Degree Zentralität ist definiert als Anteil der Beziehungen, die eine Person hat, zu den Beziehungen aller Personen im Netzwerk. Personen mit hoher Degree Zentralität haben viele direkte Kontakte zu anderen Netzwerkmitgliedern und sind in der Lage die Meinung von anderen zu beeinflussen. Sie werden als lokale Meinungsführer bezeichnet.

Im Gegensatz zur Degree Zentralität zeigt die Closeness Zentralität, wie bedeutend eine Person innerhalb des gesamten Netzwerkes ist. Closeness Zentraliät errechnet sich als die inverse Summe der Distanzen einer Person zu allen anderen Personen. Personen mit hoher Closeness Zentralität haben eine geringe Distanz zu

allen anderen Personen im Netzwerk und verfügen somit über die Möglichkeit, die Meinung von vielen anderen zu beeinflussen. Auf Grund ihrer besonderen Position im Netzwerk gelten sie als globale Meinungsführer.

Betweeness Zentralität beschreibt wie viel Kontrolle eine Person über die Interaktion anderer Netzwerkmitglieder hat. Sie errechnet sich als Quotient aus den kürzesten Pfade, die über eine Person verlaufen, und allen kürzesten Pfaden im Netzwerk. Personen mit hoher Betweenness Zentralität liegen auf vielen Kommunikationspfaden zwischen Netzwerkmitgliedern und sind in der Lage, deren Informationsaustausch zu beeinflussen. Sie werden Intermediäre genannt.

Nicht nur zentrale Personen, sondern auch die Struktur des sozialen Netzwerkes hat Auswirkung auf die Meinungsbildung. Zentralisierung und Dichte sind wichtige Kennzahlen der sozialen Netzwerkanalyse ([35], [28]), die die Struktur eines Netzwerks näher charakterisieren.

Die Zentralisierung misst, wie straff ein Netzwerk um seine zentralsten Personen organisiert ist. Die Berechnung basiert auf der Abweichung der Zentralitäten aller Personen von der zentralsten Person. In einem sehr zentralisierten Netzwerk haben Meinungsführer einen hohen Einfluss, da es nur wenige Meinungsführer im Zentrum des Netzwerkes gibt und viele andere Netzwerkmitglieder am Rand des Netzwerkes liegen. Somit können sich die Meinungen der Meinungsführer leicht verbreiten [16].

Die Dichte spezifiziert die Verbundenheit eines Netzwerkes. Sie wird als Quotient aus der tatsächlichen Anzahl an Beziehungen in einem Netzwerk und der maximal möglichen Anzahl an Beziehungen in diesem Netzwerk errechnet. Je dichter ein Netzwerk ist, desto mehr Information kann zwischen den Netzwerkmitgliedern weitergegeben werden und desto schneller können sich Meinungen in einem Netzwerk verbreiten [16].

5. FRÜHWARNSYSTEM

5.1 Anforderungsanalyse

Vor der Konzipierung des Frühwarnsystems müssen die Anforderungen festgelegt werden, die dieses erfüllen muss. Dabei wird zwischen allgemeinen Anforderungen, die für alle Frühwarnsysteme gelten, und speziellen Anforderungen, die an das hier entwickelte Frühwarnsystem gestellt werden, unterschieden.

Allgemeine Anforderungen an Frühwarnsysteme

Alarmsensitivität

Eine Anforderung an das Frühwarnsystem ist die Identifikation kritischer Situationen und die rechtzeitige Warnung. Durch keine oder eine verzögerte Warnung entstehen Kosten [11]. In Bezug auf die vorliegende Problemstellung bedeutet dies, dass sich eine negative Meinung weiter verbreiten kann und zu höheren Umsatz- und Imageverlusten sowie Aufwendungen für die Beseitigung der negativen Konsequenzen führt. Die Warnung sollte daher für alle kritischen Situationen so früh wie möglich erfolgen [11].

Geringe Fehlalarmrate

Eine Warnung soll nur dann generiert werden, wenn tatsächlich ein Risiko besteht [11]. Auch Fehlalarme verursachen Kosten [10]. Zu den Kosten für einen Fehlalarm zählt z.B. die von einem Marketing-Manager aufgewendete Zeit, die er braucht, um einen Fehlalarm als solchen zu erkennen. Hat er bereits vor dieser Erkenntnis Gegenmaßnahmen ergriffen, zählen auch diese Aufwendungen zu den Kosten für den Fehlalarm. Außerdem führen zu viele Fehlalarme zu einer Desensibilisierung der Adressaten. Die Marketing-Manager reagieren dann im Zweifelsfall nicht mehr auf eine gegebene Warnung. Eine hohe Fehlalarmrate kann dazu führen, dass vollkommen auf das Frühwarnsystem verzichtet wird [31]. Ein Marketing-Manager wird keinen Sinn in einem Frühwarnsystem sehen, das häufig Fehlalarme generiert und so unnötig von anderen Aufgaben ablenkt. Eine weitere wichtige Anforderung an Frühwarnsysteme ist es daher, die Fehlalarmrate möglichst gering zu halten.

Richtiges Maß zwischen Alarmsensitivität und Fehlalarmrate

Zwischen Alarmsensitivität und Fehlalarmrate besteht ein Zielkonflikt [37]. Ist die Alarmsensitivität besonders hoch, wird eine Situation schnell als kritisch eingestuft. Dies kann allerdings dazu führen, dass viele Fehlalarme entstehen. Ist die Alarmsensitivität niedrig, sinkt zwar die Fehlalarmrate, aber die Warnung erfolgt später oder bleibt aus. Für den vorliegenden Fall bedeutet dies, dass das richtige Maß zwischen den Kosten für eine zu späte Warnung (z.B. höhere Kosten für notwendige Gegenmaßnahmen) und den Kosten für Fehlalarme (z.B. Kosten für irrtümlich eingeleitete Gegenmaßnahmen) gefunden werden muss. Daher ist die dritte Anforderung an Frühwarnsysteme, das richtige Maß zwischen Alarmsensitivität und Fehlalarmrate zu finden.

Angemessene Kosten

Eine Investition ist grundsätzlich nur dann sinnvoll, wenn die Kosten den Nutzen nicht übersteigen. Daher gilt auch für Frühwarnsysteme, dass die entstehenden Kosten (Anschaffung, Personalkosten, Instandhaltung, usw.) die durch das Frühwarnsystem vermiedenen Kosten (Umsatzeinbußen, usw.) nicht übersteigen dürfen. Die letzte allgemeine Anforderung ist daher die der angemessenen Kosten.

Spezielle Anforderungen an das entwickelte Frühwarnsystem

Lernfähigkeit

Es gibt eine Reihe von Einflussgrößen (durchschnittliche Gesamtmeinung, Meinung von Meinungsführern, usw.), die bei der Identifikation kritischer Situationen der Meinungsbildung berücksichtigt werden müssen. Diese Einflussgrößen weichen in den verschiedenen Branchen voneinander ab und es sind nicht immer alle Zusammenhänge zwischen den einzelnen Einflussgrößen von vorneherein bekannt. Daher soll das hier entwickelte System selbst dazu in der Lage sein, diese aus vorliegenden Beispieldaten zu erlernen.

Nutzung von a-priori Expertenwissen

Im Marketing gibt es Experten, die sich in ihrer langjährigen Berufserfahrung Wissen angeeignet haben. Marketing-Manager können sagen, wann sie in der Vergangenheit gerne gewarnt worden wären. Sie besitzen bereits Regelwissen über bestimmte Zusammenhänge, wie zum Beispiel „*Wenn der Meinungsführer eine negative Meinung hat und sein Einfluss hoch ist, dann Warnung"*. Daher ist eine weitere Anforderung an das Frühwarnsystem, die Nutzung dieses a-priori Expertenwissens. Es soll zum einen die Möglichkeit bestehen, das Wissen von vorneherein mit einzubringen und zum anderen spätere Anpassungen des Systems mithilfe dieses Wissens durchzuführen.

Interpretierbarkeit

Eine wichtige Anforderung ist auch die Interpretierbarkeit. Die Marketing-Manager sollen die Möglichkeit haben, die Methode zu verstehen, mit der das Frühwarnsystem arbeitet. Demzufolge darf das Frühwarnsystem keine Black-Box darstellen, die keine Einblicke über die Funktionsweise zulässt und nur das Ergebnis „War-

nung“ oder „keine Warnung“ ausgibt. Es muss nachvollziehbar sein, unter welchen Umständen eine Warnung zustande kommt.

5.2 Methode

Die Wahl der Methode für das Frühwarnsystem erfolgt unter Beachtung der speziellen Anforderungen. Das bedeutet, die Methode muss gewährleisten, dass das Frühwarnsystem *lernfähig* ist, *a-priori Expertenwissen nutzen* kann und *interpretierbar* ist. Ausgehend von diesen Anforderungen kommen grundsätzlich zwei Methoden aus dem Bereich des Soft Computing in Betracht: Künstliche Neuronale Netze und Fuzzy-Systeme. Künstliche Neuronale Netze erfüllen zwar die Anforderung der Lernfähigkeit, verhalten sich aber wie eine Black-Box, d.h. sie sind nicht interpretierbar. Fuzzy-Systeme hingegen ermöglichen die Nutzung von a-priori Expertenwissen und sind interpretierbar, aber nicht lernfähig. Neuro-Fuzzy-Systeme vereinen die Vorteile und minimieren die Nachteile von Künstlichen Neuronalen Netzen und Fuzzy-Systemen [1]. Das Frühwarnsystem wird daher als Neuro-Fuzzy-System konzipiert. So können die Marketing-Manager sowohl ihr a-priori Wissen in Form einer linguistischen Regelbasis einbringen, als auch dem System das Auffinden von Zusammenhängen überlassen und die Ergebnisse leicht interpretieren. Es gibt eine Reihe von Neuro-Fuzzy-Ansätzen. Hier wurde das NEFCLASS Modell (NEuro Fuzzy CLASSification) [1] gewählt, da es sowohl Fuzzy-Mengen und Regeln lernen, als auch manuell definierte linguistische Regeln verarbeiten und optimieren kann[1].

Das NEFCLASS Modell ist ein drei-schichtiges Perzeptron ([22], [23]). Die erste Schicht repräsentiert die Eingangsvariablen (Gesamtmeinung, Meinung des Meinungsführers, usw.), die Zwischenschicht die Fuzzy-Regeln und die dritte Schicht die Ausgangsvariablen. In der Ausgangsschicht gibt es für jede Klasse (*Warnung* und *keine Warnung*) jeweils ein Neuron. Die Fuzzy-Mengen (linguistische Terme) werden durch die Gewichte der Verbindungen zwischen den Eingabe- und Zwischenneuronen kodiert. Die Anzahl der Fuzzy-Mengen kann je nach Anwendung variieren. Es ist Aufgabe der Marketing Manager eine für den Anwendungsfall geeignete Anzahl zu bestimmen. Jede Regel ist genau einer Ausgabeeinheit (*Warnung* oder *keine Warnung*) zugeordnet [23].

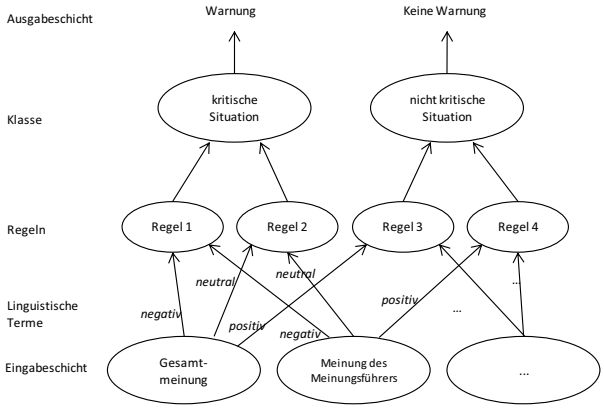


Abbildung 3: Aufbau des Fuzzy-Prezeptrons

[1] Bei der Anwendung des Neuro-Fuzzy-Systems wird die javabasierte NEFCLASS-Implementierung NEFCLASS-J verwendet. Diese kann auf der Webseite http://fuzzy.cs.uni-magdeburg.de/nefclass/nefclass-j/ heruntergeladen werden.

Das in Abbildung 3 dargestellte Perzeptron besteht aus vier Fuzzy-Regeln. Die erste Regel klassifiziert beispielsweise Situationen, in denen die Gesamtmeinung und die Meinung des Meinungsführers negativ sind, als kritisch und generiert eine Warnung (siehe Abbildung 4).

Regel 1:

Wenn die Gesamtmeinung *negativ* ist

und die Meinung des Meinungsführers *negativ* ist

dann Warnung.

Abbildung 4: Regel 1 des Fuzzy-Perzeptrons

Im Gegensatz zur klassischen Logik, in der ein Objekt entweder einer Menge angehört oder nicht, kann in der Fuzzy-Logik [38] ein Objekt mehreren Mengen mit verschiedenen Zugehörigkeitsgraden angehören. Fuzzy-Mengen spezifizieren mit welchen Zugehörigkeitsgraden Werte der Eingabevariablen zu linguistischen Termen gehören (siehe Abbildung 5). Dadurch werden auch unscharfe Aussagen möglich, wie z.B. „Meinung ist mit einem Zugehörigkeitsgrad von 0,8 negativ“. Auf Basis dieser Aussagen und den linguistischen Wenn-dann-Regeln können Fuzzy-Systeme die Klassenzugehörigkeiten von Objekten (*Warnung* oder *keine Warnung*) bestimmen.

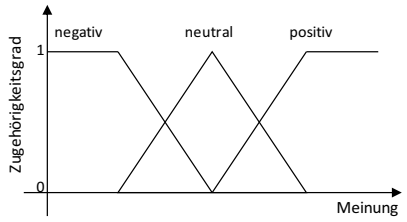


Abbildung 5: Fuzzy-Mengen

Vor ihrer Anwendung müssen die Regeln erstellt werden. Die Regeln können entweder von Experten formuliert oder von einem Lernalgorithmus auf Basis eines Trainingsdatensatzes, der Einschätzungen früherer Situationen enthält, automatisch generiert werden. Zur Erzeugung der Regeln durchläuft der Lernalgorithmus des Neuro-Fuzzy-Systems die Trainingsdatenmenge zwei Mal [1]. Beim ersten Durchlauf wird für jeden Datensatz der Trainingsmenge ein Regelantezedens („Wenn-Teil“) gebildet. Im Regelantezedenz wird jeder Eingabevariablen der linguistische Term zugeordnet, dessen Fuzzy-Menge den höchsten Zugehörigkeitsgrad liefert. Während des zweiten Durchlaufs ermittelt der Lernalgorithmus für jedes Regelantezedenz das beste Regelkonsequenz („Dann-Teil“). Durch Zusammensetzen der Regelantezedenzen und -konsequenzen entsteht eine Menge von Regelkandidaten. Aus dieser werden die besten Regeln selektiert und zu einer Regelbasis zusammengestellt. Anschließend modifiziert der Lernalgorithmus die Lage und Form der Fuzzy-Mengen, um die Klassifikationsgenauigkeit der Regeln zu erhöhen. Abschließend werden die Regeln gestutzt, um die Interpretierbarkeit und die Generalisierungsfähigkeit der Regelbasis zu optimieren.

6. ANWENDUNG

6.1 Bewertungsplattform

6.1.1 Trainingsdaten

Für den Zeitraum von Januar 2007 bis einschließlich Juni 2008, wurden 407 Posts zu den Schuhen „Mercurial“ und „Predator“ mit insgesamt 2095 Sätzen aus der Fussball-Bewertungsplattform „fussball-forum.de“ extrahiert. Auf Basis dieser Sätze wurden die Meinungen der User zu den Schuhen „Mercurial“ von Nike und „Predator“ von adidas mittels Text Mining identifiziert (Genauigkeit 73%) und zu einer Gesamtmeinung pro Monat aggregiert. Dadurch entsteht für beide Schuhe jeweils ein Datensatz mit 18 Monaten. Daneben wurden die Absatzzahlen für die beiden Schuhe aus den Unternehmensdatenbanken extrahiert.

Das Frühwarnsystem basiert auf insgesamt sechs Eingangsvariablen. Für jeden Monat wird die Gesamtmeinung zu den Fußballschuhen „Mercurial“ und „Predator“ sowie deren Absatzzahlen verwendet. Außerdem wird jeweils die Differenz der Gesamtmeinung zum Vormonat gebildet, um Aussagen darüber machen zu können, inwieweit sie im Vergleich zum Vormonat gefallen oder gestiegen sind. Jeder Variablen werden fünf linguistische Werte (siehe Abbildung 6) zugeordnet. Ausgabeparameter sind die zwei Klassen *Warnung* und *keine Warnung*.

Meinung zu Predator	Meinung zu Predator im Vergleich zum Vormonat	Absatz von Predator	Meinung zu Mercurial	Meinung zu Mercurial im Vergleich zum Vormonat	Absatz von Mercurial
negativ	gefallen	sehr gering	negativ	gefallen	sehr gering
eher negativ	etwas gefallen	gering	eher negative	etwas gefallen	gering
neutral	unverändert	mittel	neutral	unverändert	mittel
eher positiv	etwas gestiegen	hoch	eher positiv	etwas gestiegen	hoch
positiv	gestiegen	sehr hoch	positiv	gestiegen	sehr hoch

Abbildung 6: Input-Variablen für Review-Plattformen

Da das NEFCLASS-Modell überwacht lernt, sind Trainingsdaten erforderlich, die vergangene Situationen mit den zugehörigen Klassen (*Warnung* oder *keine Warnung*) enthalten. Es ist die Aufgabe der Marketing-Manager diese Klassen vor dem Training festzusetzen. Für den Ausschnitt an vergangenen Situationen im Trainingsdatensatz müssen sie entscheiden, ob sie gerne gewarnt worden wären oder nicht.

Zur Einteilung der Trainingsdaten in die Klassen *Warnung* und *keine Warnung* können sie beispielsweise die grafische Abbildung der Gesamtmeinungs- und Absatzkurven heranziehen. Anhand der Meinungskurven lässt sich neben der Meinung im jeweiligen Monat auch die Meinungsentwicklung im Vergleich zum Vormonat erkennen. Für jeden Monat werden alle vier Kurven gleichzeitig in die Überlegung mit einbezogen. Abbildung 7 zeigt die vier Kurven. Die Situation im Monat Juni 2007 wird beispielsweise als kritisch für adidas eingeschätzt, da hier die Meinung zum Eigenprodukt („Predator“) im Vergleich zum Vormonat gefallen ist, während die Meinung zum Konkurrenzprodukt eher unverändert ist und im nächsten Monat sogar ansteigt. Hinzu kommt, dass der Absatz des Eigenproduktes („Predator“) sehr gering ist. Für den Monat März 2008 ist keine Warnung für adidas erwünscht, obwohl das Konkurrenzprodukt („Mercurial“) hier sein Absatzhoch hat. Der Grund ist, dass auch das Eigenprodukt („Predator“) sein Absatzhoch hat und die Meinung des Eigenprodukts im Vergleich zum Vormonat gestiegen, während die Meinung zum Konkurrenzprodukt gefallen ist. Aus der Einteilung der Trainingsdaten lassen sich bereits Wenn-dann-Regeln ableiten, die später als apriori Wissen in die Regelbasis eingebracht werden können.

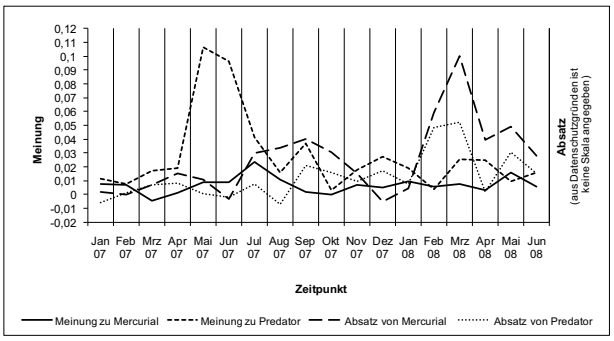


Abbildung 7: Entwicklung der Input-Variablen

6.1.2 Regeln

Die klassifizierten Trainingsdaten werden zum Regellernen herangezogen. Im vorliegenden Fall ergibt sich für beide Schuhe eine Regelbasis mit jeweils fünf Regeln. Jede der Regelbasen ist übersichtlich und leicht interpretierbar. Abbildung 8 stellt einen Auszug der Regelbasis für den Schuh „Predator“ dar. Regel 1 klassifiziert Situationen, die kritisch sind, d.h. in denen die Marketing-Manager von adidas gewarnt werden müssen. Diese Situationen stellen für das Unternehmen ein Risiko dar, da die Meinung zum Eigenprodukt („Predator“) im Vergleich zum Vormonat etwas gefallen ist und der Absatz des Schuhs „Predator“ bereits sehr gering ist. Der Absatz des Konkurrenzproduktes ist außerdem nur gering und nicht sehr gering. Es besteht die Gefahr, dass die gefallene Meinung zu weiteren Image- und Absatzeinbußen führt und sich die ohnehin schon ungünstige Lage weiter verschlechtert. In Situationen, die durch Regel 2 identifiziert werden, muss hingegen nicht gewarnt werden, obwohl in diesen Situationen der Absatz des Konkurrenzproduktes („Mercurial“) mittel und der Absatz des Eigenproduktes („Predator“) sehr gering ist. Der Grund dafür ist, dass hier die Meinung zu dem Schuh „Predator“ eher positiv und im Vergleich zum Vormonat sogar leicht gestiegen ist. Es kann davon ausgegangen werden, dass sich die Situation von adidas in Zukunft weiter verbessern wird. Ein Eingreifen seitens der Marketing-Manager ist nicht notwendig.

Regel 1:

Wenn die Meinung Predator im Vergleich zum Vormonat *etwas gefallen* ist

und der Absatz Predator *sehr gering* ist

und der Absatz Mercurial *gering* ist

dann Warnung.

Regel 2:

Wenn die Meinung Predator *eher positiv* ist

und die Meinung Predator im Vergleich zum Vormonat *etwas gestiegen* ist

und der Absatz Predator *sehr gering* ist

und der Absatz Mercurial *mittel* ist

dann keine Warnung.

Abbildung 8: Regeln für adidas

6.1.3 Klassifikationsergebnis

Für beide Klassifikatoren können gute Ergebnisse erzielt werden (siehe Tabelle 1). Der Klassifikator für den Schuh „Predator" klassifiziert 80% und der Klassifikator für den Schuh „Mercurial" 70% aller Situationen korrekt (10-fache Kreuzvalidierung). Der Klassifikator für den Schuh „Predator" erzielt eine Alarmsensitivität von 77,8% bei einer Falschalarmrate von 0%. Das heißt, es gibt keine Falschalarme. Für den Schuh „Mercurial" können alle kritischen Situationen identifiziert werden, allerdings sind 40% aller Alarme Fehlalarme. Es gilt abzuwägen, ob auf die Identifikation aller kritischen Situationen zu Gunsten einer geringeren Fehlalarmrate verzichtet werden soll. Dies hängt in erster Linie davon ab, ob die Kosten für die Fehlalarme oder für die Nichterkennung kritischer Situationen höher sind, und muss von den zuständigen Marketing-Managern individuell entschieden werden.

Tabelle 1: Klassifikationsergebnisse für adidas und Nike

Schuh	Anzahl an Trainingsdaten	Richtig klassifiziert	Alarm-sensitivität	Fehlalarmrate
Predator	18	80,0%	77,8%	0,0%
Mercurial	18	70,0%	100,0%	40,0%

6.2 Soziales Netzwerk

6.2.1 Trainingsdaten

Die Daten für den zweiten Anwendungsfall stammen aus dem sozialen Netzwerk „gamestar.de", in dem Mitglieder Meinungen zu Computerspielen diskutieren. Aus dem Zeitraum vom 5. Oktober bis zum 28. November 2008 wurden 3776 Posts für das Spiel „Fallout 3", 1350 Posts für das Spiel „Dead Space" und 1470 für das Spiel „Far Cry 2" extrahiert. Für jedes Spiel wurde eine Folge von tagesbasierten Netzwerken generiert, in denen die Mitglieder miteinander verlinkt wurden, die am gleichen Tag direkt vor oder nach einem anderem Mitglied Posts eingestellt haben. Dadurch entstehen für die Spiele „Dead Space", „Fallout 3" und „Far Cry 2" Trainingsdatensätzen, die 56, 45 bzw. 33 Diskussionsnetzwerke enthalten. Die Meinungen der Mitglieder wurden mittels Text Mining als positiv, negativ oder neutral klassifiziert (Genauigkeit 77%). Zusätzlich erfolgte eine Soziale Netzwerkanalyse, um Meinungsführer zu identifizieren und die Netzwerkstruktur zu charakterisieren.

Die Frühwarnung für alle drei Spiele basiert auf neun Eingangsvariablen. Sechs der Variablen beziehen sich auf die Meinungsführer. Es wird jeweils die Meinung und der Einfluss des lokalen Meinungsführers (Degree Zentralität), des globalen Meinungsführers (Closeness Zentralität) und des Intermediäres (Betweenness Zentralität) herangezogen. Zwei weitere Variablen stehen für die Wahrscheinlichkeit, dass sich die Meinung eines Meinungsführers durchsetzt (Zentralisierung), und die Geschwindigkeit, mit der sich die Meinung eines Meinungsführers verbreitet (Dichte). Darüber hinaus wird die Gesamtmeinung im sozialen Netzwerk betrachtet. Jeder Variablen werden drei linguistische Ausprägungen zugewiesen (siehe Abbildung 9). Ausgabeparameter sind die zwei Klassen *Warnung* und *keine Warnung*.

Einfluss des lokalen Meinungsführers	Meinung des lokalen Meinungsführers	Einfluss des globalen Meinungsführers	Meinung des globalen Meinungsführers	Einfluss des Intermediäres	Meinung des Intermediäres
gering	negativ	gering	negativ	gering	negativ
mittel	neutral	mittel	neutral	mittel	neutral
hoch	positiv	hoch	positiv	hoch	positiv

Geschwindigkeit, mit der sich die Meinung verbreitet	Wahrscheinlichkeit, dass sich die Meinung des Meinungsführers durchsetzt	Gesamtmeinung
gering	gering	negativ
mittel	mittel	neutral
hoch	hoch	positiv

Abbildung 9: Input-Variablen für soziale Netzwerke

Für das automatische Lernen der Regeln ist ein Trainingsdatensatz bereitzustellen, der zu Situationen aus der Vergangenheit die richtigen Einschätzungen (*Warnung* oder *keine Warnung*) erhält.

Abbildung 10 zeigt das Diskussionsnetzwerk zu einem Zeitpunkt, der nicht kritisch für die Marketing-Manager von „Dead Space" ist. Der lokale Meinungsführer und Intermediär ist USER 25. Er hat eine positive Meinung und einen mittleren Einfluss (Degree Zentralität 0,57, Betweenness Zentralität 0,48). Auch die Meinung des globalen Meinungsführers (USER 26) ist positiv und sein Einfluss ist hoch (Closeness Zentralität 0,65). Die Wahrscheinlichkeit, dass sich die Meinung der Meinungsführer durchsetzt, ist mittel (Zentralisierung 0,42). Die Geschwindigkeit, mit der sich die Meinung verbreitet ist eher gering (Dichte 0,18). Die Gesamtmeinung zu diesem Zeitpunkt ist mit 0,24 eher positiv bis neutral. Wie zu erkennen ist, haben auch andere zentrale Personen, wie z.B. USER 27 eine positive Meinung. Die Situation zu diesem Zeitpunkt ist daher nicht kritisch.

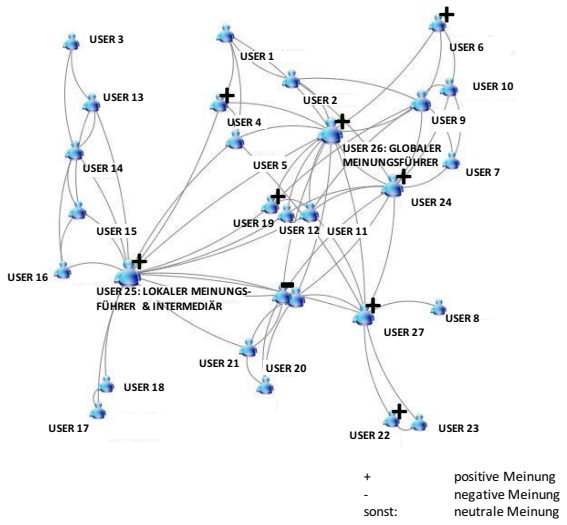


Abbildung 10: Nicht kritische Situation im sozialen Netzwerk

Abbildung 11 zeigt das Diskussionsnetzwerk in einer Situation, in der die Marketing-Manager des Spiels „Dead Space“ gewarnt werden müssen. Alle drei Meinungsführer haben hier eine neutrale Meinung. Der lokale Meinungsführer und Intermediär, USER 8, hat einen hohen (Degree Zentralität 0,75) bzw. mittleren (Betweenness Zentralität 0,41) Einfluss. Der Einfluss des globalen Meinungsführers, USER 9, ist auch eher hoch (Closeness Zentralität 0,57). Wie zu erkennen ist, haben die meisten anderen Personen zu diesem Zeitpunkt eine neutrale Meinung. Nur ein User hat eine positive und zwei User haben eine negative Meinung. Die Gesamtmeinung ist mit -0,11 als neutral bis leicht negativ einzustufen. Die Wahrscheinlichkeit, dass sich die Meinung eines Meinungsführers durchsetzt, ist hoch (Zentralisierung 0,47) und die Geschwindigkeit, mit der sich die Meinung verbreitet, ist mittel (Dichte= 0,36). Die Situation ist kritisch, da eine neutrale Meinung schwer einer negativen Meinung entgegenwirken kann. Sie wird weniger beachtet und nicht als Gegenmeinung wahrgenommen.

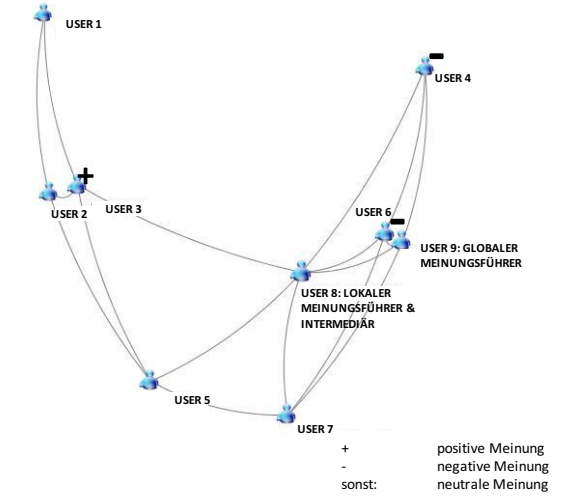


Abbildung 11: Kritische Situation im sozialen Netzwerk

6.2.2 Regeln

Die anhand der Trainingsdaten gelernten Regelbasen für die Spiele „Dead Space“, „Fallout 3” und „Far Cry2” sind mit acht, neun und drei Regeln übersichtlich und einfach interpretierbar. Abbildung 12 und Abbildung 13 zeigen einen Auszug aus der Regelbasis für das Spiel „Fallout 3“.

Regel 3 stuft Situationen als kritisch ein und führt somit zu einer Warnung. In diesen Situationen ist die Meinung des globalen und des lokalen Meinungsführers negativ und der Einfluss des lokalen Meinungsführers hoch. Außerdem ist die Wahrscheinlichkeit hoch, dass sich die Meinung des Meinungsführers verbreitet, und es besteht die Gefahr, dass die ohnehin schon negative Gesamtmeinung noch weiter sinkt. In diesen Situationen ist ein Eingreifen der Marketing-Manager unbedingt notwendig, um eine Verbreitung der negativen Meinung und der damit verbundenen Image- und Umsatzeinbußen zu vermeiden.

Regel 3

Wenn die Meinung des globalen Meinungsführers *negativ* ist

und die Meinung des lokalen Meinungsführers *negativ* ist

und der Einfluss des lokalen Meinungsführers *hoch* ist

und die Wahrscheinlichkeit, dass sich die Meinung des Meinungsführers verbreitet, *hoch* ist

und die Gesamtmeinung *negativ* ist

dann Warnung

Abbildung 12: Regel für Warnung

In Situationen, die durch Regel 4 abgedeckt sind, besteht hingegen kein Handlungsbedarf. Sie beschreibt Situationen, in denen der globale Meinungsführer eine positive Meinung hat. Die Wahrscheinlichkeit, dass sich die Meinung des Meinungsführers durchsetzt, ist zwar gering, allerdings ist die Gesamtmeinung bereits positiv. Es besteht keine Gefahr für das Unternehmen.

Regel 4

Wenn die Meinung des globalen Meinungsführers *positiv* ist

und die Wahrscheinlichkeit, dass sich die Meinung des Meinungsführers durchsetzt, *gering* ist

und die Gesamtmeinung *positiv* ist

dann keine Warnung

Abbildung 13: Regel für keine Warnung

6.2.3 Klassifikationsergebnis

Für die Spiele „Dead Space“ und „Fallout 3“ lassen sich sehr gute Klassifikationsergebnisse erzielen (10-fache Kreuzvalidierung) (siehe Tabelle 2). Der Klassifikator für das Spiel „Dead Space“ identifiziert 90,3% aller Situationen richtig und 100% der kritischen Situationen. Die Falschalarmrate liegt bei 11,8% und ist somit gering. Für das Spiel „Fallout 3“ können 93,5% aller Situationen richtig erkannt werden. Dabei werden 88,8% der kritischen Situationen richtig klassifiziert und es gibt mit 5,9% nur sehr wenige Fehlalarme. Für das Spiel „Far Cry 2“ können mit einer Klassifikationsgenauigkeit von 70,6% vergleichsweise weniger Situationen richtig zugeordnet werden. Der Grund dafür liegt hier in der relativ hohen Fehlalarmrate von 25%. Von den kritischen

Situationen werden dennoch 88,2% richtig erkannt. Insgesamt handelt es sich um sehr gute Klassifikationsgenauigkeiten.

Tabelle 2: Klassifikationsergebnisse für die Spiele

Spiel	Anzahl der Trainings-daten	Richtig klassifiziert	Alarm-sensitivität	Fehl-alarmrate
Dead Space	56	90,3 %	100,0%	11,8%
Fallout 3	45	93,5 %	88,8%	5,9%
Far Cry 2	33	70,6 %	88,2%	25,0 %

7. IMPLIKATIONEN FÜR DAS MARKETING

Durch die frühzeitige Warnung sind Marketing-Manager in der Lage, rechtzeitig Gegenmaßnahmen einzuleiten und die Meinungsbildung positiv zu beeinflussen. Eine negative Meinungsentwicklung und die damit verbundenen Image- und Umsatzeinbußen für das Unternehmen können somit minimiert oder vermieden werden.

Eine Möglichkeit Einfluss auf den Meinungsbildungsprozess zu nehmen ist beispielsweise, dass sich die Marketing-Manager selbst als Kunden ihres Unternehmens ausgeben und sich an den Diskussionen beteiligen. So können sie eine positive Meinung vertreten und einem negativen Meinungstrend entgegenwirken. Allerdings bekämpfen sie damit lediglich die Ausbreitung einer negativen Meinung, nicht aber deren Ursache. Darüber hinaus besteht die Gefahr, dass die Marketing-Manager als solche erkannt werden und der negative Meinungstrend weiter verstärkt wird.

Eine andere Möglichkeit ist die direkte und offene Ansprache der unzufriedenen Konsumenten [32]. Das heißt, der Marketing-Manager tritt offiziell als Vertreter seines Unternehmens auf und beteiligt sich an der laufenden Konversation. Dabei kann er auf die negativen Äußerungen eingehen und die Kunden auf ihre Wünsche ansprechen. Damit wird nicht nur die Verbreitung der anfänglichen negativen Meinung vermieden, sondern gleichzeitig die Verbreitung einer positiven Meinung gefördert, da den Konsumenten das Gefühl vermittelt wird, dass ihre Meinung ernst genommen wird [21].

Durch das Eingreifen in den Meinungsbildungsprozess hat der Marketing-Manager die Möglichkeit, diesen positiv zu beeinflussen. Für ein Unternehmen ist es allerdings von Interesse die Gründe für das Zustandekommen der negativen Meinungen zu kennen. So kann bei einer weiteren Betrachtung der einzelnen Diskussionsbeiträge herausgefunden werden, warum die Meinung zu einem Produkt schlecht ist. Das heißt, das Unternehmen hat die Chance die Kundenwünsche zu erkennen und die Produkte langfristig anzupassen [32]. Dadurch können der Umsatz erhöht und das Image des Unternehmens verbessert werden.

8. ZUSAMMENFASSUNG UND AUSBLICK

In dieser Arbeit wird ein Frühwarnsystem zur Identifikation kritischer Situationen der Meinungsbildung im Internet vorgestellt. Mithilfe des Frühwarnsystems können Marketing-Manager gewarnt werden, sobald sich abzeichnet, dass es in Zukunft zu einer negativen Meinungsentwicklung kommt und Umsatz- und Imageeinbußen drohen.

Das System wurde als Neuro-Fuzzy-System umgesetzt, wodurch es sowohl anhand von Beispielsdaten Klassenzugehörigkeiten lernen und a-priori Expertenwissen nutzen kann als auch für die Marketing-Manager einfach interpretierbar ist.

Um die Einsatzfähigkeit des Systems zu validieren, wurde das Frühwarnsystem beispielhaft für Bewertungsplattformen und soziale Netzwerke angewendet. Für alle Beispiele konnten gute bis sehr gute Ergebnisse hinsichtlich der Anforderung der richtigen Klassifikation, der Alarmsensitivität und der Fehlalarmrate erzielt werden.

Die Einzigartigkeit des konzipierten Frühwarnsystems liegt zum einen in der automatischen Analyse der Meinungsbildung und Beurteilung, ob die Gesamtsituation für ein Unternehmen kritisch ist oder nicht. Dabei spielen für jedes Unternehmen und für jedes Produkt andere Faktoren bei der Identifikation der kritischen Situationen eine Rolle, die durch das hier entwickelte Frühwarnsystem individuell berücksichtigt werden können. Zum anderen ermöglicht das Frühwarnsystem eine einfache Wissensakquisition und Nutzung für die zuständigen Marketing-Manager.

In zukünftigen Arbeiten wird das Frühwarnsystem unter Verwendung größerer Datenbestände weiter validiert. Darüber hinaus soll es dahingehend ausgebaut werden, dass es neben der Identifikation kritischer Situationen auch noch geeignete Handlungsempfehlungen für die Marketing-Manager erstellt.

9. LITERATURVERZEICHNIS

[1] Borgelt, C., Klawonn, F., Kruse R., Kruse; Nauck, D. 2003. *Neuro-Fuzzy-Systeme: Von den Grundlagen künstlicher Neuronaler Netze zur Kopplung mit Fuzzy-Systemen*. 3. Aufl. Vieweg Verlag, Wiesbaden.

[2] Bruhn, M. 2006. Das Konzept der Kundenorientierten Unternehmensführung. In: *Kundenorientierte Unternehmensführung: Kundenorientierung – Kundenzufriedenheit - Kundenbindung*. Hinterhuber, H. H., Matzler, K., Hrsg., Aufl. 5, Gabler, Wiesbaden, S. 35-65.

[3] Choudhury, M. D., Sundaram, H. John, A., Seligmann, D. D. 2007. Contextual Prediction of Communication Flow in Social Networks. In: *Proceedings of the IEEE/WIC/ACM international Conference on Web*. IEEE Computer Society, Washington, S. 57-65.

[4] Choudhury, M. D. 2009. Modelling and Predicting Group Activity over Time in Online Social Media. In: *Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia*. Torino.

[5] Choudhury, M. D., Sundaram, H., John, A., Seligmann, D. D. 2009. Which are the Representatative Groups in a Community? Extracting and Characterizing Key Groups in Blogs. In *Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia*, Torino.

[6] Cortes C., Vapnik V. N., 1995. Support Vector Networks. *Machine Learning* 20, pp. 273-297.

[7] Dasu T., Johnson T., 2003. *Exploratory data mining and data cleaning*, John Wiley & Sons, Hoboken.

[8] Dave, K., Lawrence, S., Pennock, D. M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, Budapest.

[9] Dhar, V., Chang, E. 2007. Does Chatter Matter? *The Impact of User-Generated Content on Music Sales*. Arbeitspapier, Leonard N. Stern School of Business, New York University.

[10] Fawcett, T., Provost, F. 1997. Adaptive fraud detection. In *Data Mining and Knowledge Discovery* 1, S. 291-316.

[11] Fawcett, T., Provost, F. 1999. Activity Monitoring: Noticing interesting changes in behavior. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. San Diego, S. 53-62.

[12] Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R. Tomokiyo, T. 2005. Deriving Marketing Intelligence from Online Discussion. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. Chicago, S. 419 – 428.

[13] Gruhl, D., Guha, R., Kumar, R., Novak, J., Tomkins, A. 2005. The Predictive Power of Online Chatter. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. Chicago, S. 78 - 87.

[14] Huang, Y., Liu, S., Wang, Y. 2007. Online Detecting and Tracking of the Evolution of User Communities. In *Proceedings of the Third International Conference on Natural Computation*. S. 681-685.

[15] Kaiser, C. 2008. Produkt-Mining im Web 2.0. In: Bichler, M; Hess, T.; Krcmar, H., Lechner, W.; Matthes, F.; Picot, A.; Speitkamp, B.; Petra W. (Hrsg.): *Tagungsband der Mulitkonferenz Wirtschaftsinformatik 2008*. GITO-Verlag, Berlin 2008, S. 229-240.

[16] Kaiser, C. 2009. Analyse von Meinungen in sozialen Netzwerken des Web 2.0. In: Hansen, H. R., Karagiannis, D., Fill, H.-G. (Hrsg.) 2009: *Business Services: Konzepte, Technologien, Anwendungen. 9. Internationale Tagung Wirtschaftsinformatik*. Österreichische Computer Gesellschaft, Wien.

[17] Katz E., Lazarsfeld P. F, 1955. *Personal influence, the part played by people in the flow of mass communication*, Glencoe, Free Press.

[18] Keller E. B., Berry J., 2003, *The influentials*, New York, Free Press.

[19] Kim, S.-M., Hovy, E. 2007. Crystal: Analysing Predictive Opinions on the Web. In *Proceedings of the 2007 Joint Conference on the Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Prague, S. 1056-1064.

[20] Kroeber-Riel, W., Weinberg, P. 2009. *Konsumentenverhalten*, 9. Aufl., Verlag Vahlen, München.

[21] Matzler, K., Stahl, H., Hinterhuber, H. H. 2006. Die Customer-based View der Unternehmung. In Hinterhuber, H. H., Matzler, K. (Hrsg.) 2006, *Kundenorientierte Unternehmensführung: Kundenorientierung – Kundenzufriedenheit - Kundenbindung*. Aufl. 5, Gabler, Wiesbaden, S. 35-65.

[22] Nauck, D., Kruse, R. 1994. A Fuzzy Perceptron as a Generic Model for Neuro-Fuzzy Approaches. In *Fuzzy-Systeme '94. Wissenschaftliche Veröffentlichung des zweiten Deutschen GI-Workshops*, München.

[23] Nauck, D., Kruse, R. 1995. NEFCLASS - A Neuro-Fuzzy Approach for the Classification of Data. In: George, K.M.; Carrol, Janice, H.; Deaton, Ed; Oppenheim, Dave; Hightower, Jim (Hrsg.): *Applied Computing 1995 - Proceedigns of the 1995 ACM Symposium on Applied Computing*, Nashville, ACM Press.

[24] Onishi, H., Manchanda, P. 2009. *Marketing Activity, Blogging and Sales*. Arbeitspapier, Ross School of Business, University of Michigan.

[25] Pang, B., Lee, L. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Philadelphia.

[26] Popescu, A.-M., Etzioni, O. 2005. Extracting Product Features and Opinions from Reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Vancouver, S. 339-346.

[27] Rogers E., 2003. *Diffusion of innovations*, 5th ed., New York, Free Press.

[28] Scott J., 2000. *Social Network Analysis – A Handbook*, London, SAGE.

[29] TNS Infratest 2010. *Kaufentscheidung fällt immer häufiger im Netz*. URL: www.tns-infratest.com/presse pdf/Presse/2010_01_12_TNS_Infratest_Kaufentscheidung_Internet.pdf, Abruf am 2010-01-15

[30] Tong, R. M.; Yager, R. R. 2004. Characterizing Attitudinal Behaviors in On-Line Open-Source. In *Proceedings of the Association for the Advancement of Artificial Intelligence Spring Symposium* 2004, Atlanta.

[31] Tränkler, H.-R., Derbel, F., Hensel, A., Horn, M., Oppelt, U. 2000. Smart Sensor Systems für die Branderkennung. In *Tagungsband der VdS-Fachtagung Gas Sensors for Fire Detection*, Köln, S. 15-16.

[32] Trommsdorff, V. 2009. *Konsumentenverhalten*. 7. Aufl., Verlag W. Kohlhammer, Stuttgart.

[33] Valente T. W., 1999. *Network Models of the Diffusion of Innovations*, Cresskill, Hampton Press.

[34] Viermetz, M., Skubacz, M., Ziegler, C.-N.; Seipel, D. 2008. Tracking Topic Evolution in News Environments. In *Proceedings of the 10th IEEE Conference on E-commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services*, Washington, S. 215-220.

[35] Wassermann, S., Faust, K., 1999. *Social Network Analysis – Methods and Applications*, Cambridge, Cambridge University Press.

[36] Weiss S., Indurkhya N., Zhang T., Damerau F., 2005. *Text Mining – Predictive Methods for Analyzing unstructured Information*, New York, Springer.

[37] Xu, L., He, W. 2005. Application of Fuzzy Neural Network to Fire Alarm System of High-rise Building. In: *Journal of Communication and Computer* 2 (9), 200, S. 18-21.

[38] Zadeh, Lotfi (1965): Fuzzy Sets. In: *Information and Control* 8 (3), 1965, Academic Press, San Diego, S. 338-353.

Using Reputation Information on Internet-of-Services Markets

Stefan König
University of Bayreuth
Chair of Information Systems Management
95440 Bayreuth, Germany
stefan.koenig@uni-bayreuth.de

Torsten Eymann
University of Bayreuth
Chair of Information Systems Management
95440 Bayreuth, Germany
torsten.eymann@uni-bayreuth.de

ABSTRACT

The paper identifies trusting problems between autonomous services in the Internet-of-Services (IoS). This scenario vision describes a general computational paradigm, which allows companies to procure computational resources externally. The arising conflicting interests between providers and consumers lead to strategic behaviour of single services. Usually trust and reputation models are proposed to set incentives for acting honestly. But when using Double Auctions to match buyers and sellers, these trust and reputation models fail to close this "trusting gap". This paper proposes a modified Double Auction protocol fulfilling the deducted requirements. Simulation experiments show that the usage of this modified protocol leads to increased trustworthiness for the participants.

1. INTRODUCTION

The increasing dynamic of markets leads for companies to the need of adapting their processes to continuously changing environment. For every-day business, the use of computationally intensive information technology (IT) seems essential to implement new flexible business models within a short time.

The *Internet-of-Services (IoS)* describes a general computational paradigm, which allows companies to procure computational resources externally and thus to save both internal capital expenditures and operational costs. The notion of Internet-of-Services follows the idea of consuming different services externally, provided by distinct Service Providers (SP), but from a blurred cloud of resources within a single business unit or even between different businesses [10]. As the interaction frequency is assumed to be very high and the volume of a single interaction is assumed to be very small, the whole process has to be fulfilled without human interaction. The process includes finding a suitable SP, negotiating with it, invoking the service and fulfilling some post-processing steps if necessary.

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

To match providers and consumers, an efficient allocation mechanism between service demand and supply is needed: a market [28]. But introducing a market will lead to other problems, for example asymmetrically distributed information between SPs and consumers. SPs usually have more information about quality or availability of the services they provide, than Service Consumers (SC). Further, the effectively provided functionality might differ from the promised functionality. This case of asymmetrically distributed information usually leads to suboptimal results due to the uncertainty on the consumer side, and thus to an inferior usage of the service environment in total. Contrarily, consumers have more information about their liquidity. In addition, both interaction participants deal with uncertainty caused by environmental factors (for example network failures).

One common way to overcome this asymmetrical information distribution is the usage of trust and reputation models [13]. The own experiences, experiences from other participants or just gossiping received from other participants about a target service can influence one's behaviour. The usage of trust and reputation models is quite common in decentralized environments, because no central entity has to be implemented that has central knowledge or even central control.

From Electronic Commerce research different payment models are "known": the *pay-before* model determines the payment before a service invocation takes place, the *pay-later* model works vice versa. The choice of the payment model directly impairs the direction of the aforementioned asymmetrical information distribution. These models are also assumed for IoS.

This paper will especially focus on the usage of Double Auctions to coordinate a future IoS. Coordinating sellers and buyers by a Double Auction means, that there is a central Auctioneer that receives the bids of sellers and buyers. The Auctioneer matches these bids following a known algorithm, which determines the price [22]. Following Streitberger et al. [28], a Double Auction represents an efficent mechanism to coordinate resource allocation in service systems.

With this paper we are going to investigate the question if it is possible to use reputation information in order to achieve trustworthy interactions in a Double Auction-coordinated IoS environment. Therefore, the paper is structured as follows: whereas section 2 presents foundations and related

work for this paper, section 3 explains the proposed design of the Double Auction protocol. Section 4 demonstrates and evaluates the proposed protocol in combination with the $AVALANCHE_{dec}$ reputation model. Finally, section 5 concludes the paper and presents future work.

2. FOUNDATIONS AND RELATED WORK

This section presents foundations of the work and related work. Foundations can be split into the *AVALANCHE* reputation model, requirements and hypotheses on Double Auctions that are deducted from literature. Related work comprises papers regarding trusting relationships between the actors Seller, Buyer and Auctioneer.

2.1 AVALANCHE Reputation System

Trust and reputation models have been discussed in research extensively in the past (see for example [25]). Due to the simple requirements a reputation model has to fulfil as soon as it is assumed that no cognitive acting humans are directly involved, we chose a simple mathematical model to discuss the usage in Double Auctions: "*AVALANCHE*". Discussing the generalizability of the findings for other models will be not part of this paper.

Padovan et al. [19, 20], the authors of the *AVALANCHE* system, denote their reputation coefficient as R_Y^X, whereas X denotes the identity of the rating software agent (Evaluator), and Y denotes the identity of the rated agent (Target). The reputation coefficient obtains values in between $0 \leq R_Y^X \leq 1$. $R_Y^X = 1$ denotes the best reputation (target agent seems to be a reliable agent), $R_Y^X = 0$ denotes a bad reputation (target seems to be non-cooperative). Per se the reputation coefficient represents a private value for each agent, such that two different agents might differ in their coefficient about target agent three: $R_3^1 \neq R_3^2$. "In general, the reputation coefficient is used to adapt the software agent's negotiation strategy according to its partner's expected cooperative behaviour." [19, p. 6]

To structure the trust and reputation mechanism, Padovan et al. distinguish four stages: obtaining the reputation coefficient, adapting this coefficient to the own negotiation strategy, rating the partner's behaviour after the transaction and last but not least the distribution of the reputation information.

2.1.1 Obtaining Reputation Information

Within the obtaining stage, Padovan et al. [20] identify three different cases: The target agent is unknown for all agents. In this case the agent has to estimate the risk of an interaction with the target. To do that, the authors propose as an alternative to a default value between 0 and 1 the average value of all known reputation values. If no reputation value is available (first interaction), a default value has to be chosen.

If the agent has already made personal experiences with the target, the agent can use its personal own information about the target's cooperative behaviour and eventually take further information from others, for example if this information is not reliable.

Table 1: Assessment of price and reputation [19]

I	Y_i	p_i	$R_{Y_i}^X$	$2 - R_{Y_i}^X$	p_i^*	rank
1	12	47	$R_{12}^X = 0.63$	1.37	64.39	2
2	3	52	$R_3^X = 0.65$	1.35	70.20	3
3	16	54	$R_{16}^X = 0.85$	1.15	62.10	1
4	5	56	$R_5^X = 0.44$	1.56	87.36	4

2.1.2 Adapting Reputation Information to Negotiation Strategies

At the beginning of the negotiation phase (after the information phase) the agent has received a list of potential interaction partners. These offers and the corresponding agents are ranked in by an assessed offer price $p_{Y_i}^*$ that is calculated based on the initial offered price p_{Y_i} and its reputation coefficient $R_{Y_i}^X$: $p_{Y_i}^* = p_{Y_i} + (R_{Y_i}^X * 0 + (1 - R_{Y_i}^X) * p_{Y_i}) = p_{Y_i} * (2 - R_{Y_i}^X)$

Table 1 illustrates one example with four offers. The agent X starts to negotiate with partner number 16. If this negotiation fails, X will negotiate with number 12 [19].

2.1.3 Rating Cooperative Behaviour

After each settlement phase of a transaction, the agents are able to rate each other. This value is denoted by the authors with r_j, whereas j represents the index of the transaction. Successful transactions are rated with $r_j = 1$ (best value), unsuccessful transaction are rated with $r_j = 0$ (worst value) [19].

2.1.4 Distribution and Updating

This obtained rating value r_j updates the reputation coefficients of the involved transaction partners X and Y. To emphasis latest ratings compared to older ones, the authors use an average weighting calculation with a weighting factor α. This weighting factor can be instantiated by the agent owner. Following, the new reputation value is calculated as follows: $R_{Y_i}^X = R_{Y_i-1}^X * (1 - \alpha) + r_j * \alpha$. If a global reputation agency is used (centralized model), a global value for α has to be defined. Based on the assumptions on future IoS environments we avoid using a central unit that coordinates the reputation values.

The main difference of a decentralized to a centralized reputation mechanism is not the concept itself. Instead the process of exchanging reputation information increases in its complexity. While in a centralized model, as *AVALANCHE* represents one, the reputation unit manages all published reputation information, in a decentralized model neighboured agents have to be requested for information on other participants. Each participant has to manage reputation information for its own. In order to extend the initial *AVALANCHE* model to the decentralized version $AVALANCHE_{dec}$, we just have to modify the reputation communication process.

For the remaining paper, we are assuming the decentralized version of the reputation model. That model has been validated against the original model within the SimIS [15] environment before conducting the simulation experiments. The used replication methodology follows the replication replica-

tion process model of Sansores and Pavon [27] and bases on experiments and corresponding simulation data of [19]. The result of this replication has been positive.

2.2 Hypotheses on Double Auctions and Reputation Usage

In order to define the requirements for the research artifact [12], we will present the hypotheses that have been deducted from literature review in König et al. [16] (see table 2). The table combines the negotiation roles Service Provider (SP), Service Consumer (SC) and Auctioneer (Auct.) with the two possible payment models *pay-before* and *pay-later*. Using one of the two payment models determines whether the SC or the SP role acts as trustee (has cheating possibility).

Agents acting strategically regarding the reputation model usage might have different interests regarding the process of generating and distributing ratings. This includes the rating of target agents and the memetic acting, that means participating in gossiping. This subsection considers the overlapping of the reputation roles describing the agents' participation in the overall reputation process.

In their work, Conte and Paolucci [11] define the four different agent roles as follows:

- The set M is a group of agents that sends information to other agents.
- Set E are all agents which evaluate a certain target T.
- Set T, on the other side, are the agents which are evaluated by E.
- Finally, set B is defined as a group of beneficiaries that benefit from the evaluations performed by the evaluators (set E) about the targets (set T) that can be spread through the memetic agents (set M). The beneficiaries benefit from it as they receive information about the degree to which the target conforms with the social norm. [11, p. 74 et seqq.]

As soon as two or more roles overlap, certain effects regarding the reputation model can occur [11]. For us, especially two hypotheses are of interest:

H1: $B \simeq E, B \cap T = E \cap T = \emptyset$: The overlapping of the sets B and E and the non-overlapping of T lead to working reputation system, because participants are motivated to provide their own experiences, the reputation system works well.

H2: $B \cap T, B \cap E = E \cap T = \emptyset$: As soon as all sets are disjunctive from each other, the participants have no incentives to participate at the reputation system.

Even if we implement a model like $AVALANCHE_{dec}$, it is not possible to gain an adequate system when using Double Auctions. The problem in this case is that neither the SC nor the SP are fulfilling the Beneficiary role. Instead, only the Auctioneer in terms of a Double Auction might benefit from correct ratings due to a higher fulfilment of its users' expectations.

Table 2: Reputational Configurations in Double Auctions (following [16])

Negot.	Trustee	SP	SC	Auct.	H1	H2
DA	SP	T	E	B	No	**Yes**
	SC	E	T	B	No	**Yes**
DA_{Rep}	SP	T	E,B		**Yes**	No
	SC	E,B	T		**Yes**	No

2.3 Requirements on Trustworthy Double Auctions in the IoS

The mechanism this work is going to design and develop should focus on the trust relation between the set of buyers and the set of sellers. In addition, the coordination mechanism should be denoted by a Double Auction.

R 0 The mechanism to design must focus on the trusting relationship between the set of sellers and the set of buyers. Further, a Double Auction mechanism has to be used in order to fill the identified research gap.

In order to reach *H1* instead of *H2*, we have do find a configuration of the Double Auction protocol that shifts the B role (Beneficiary) from the Auctioneer to the SC in the pay-before case and respectively to the SP in the pay-later case (see table 2). If it is possible to design such a mechanism, the trustor can use reputation information in order to choose negotiation partners. For the trustee on the other hand, this information is not important as it can not be cheated during the interaction. We will hold this requirement as following:

R 1 The trustor must be able to use its information on the reliability of the matched partner or the partner to match in order to benefit from this information.

In order to design the mechanism as flexible as possible, it should be resistant against a change in the payment model. This means that a change of the payment model from pay-later to pay-before should have no effect on the system and vice versa. If the mechanism would be able to fulfil this requirement, a change of the payment model would not determine major changes in mechanism design. Further, the required flexibility regarding the trustor-trustee relationship will enable a pay-later or a pay-before or even an arbitrary combination of both models.

R 2 The protocol to be designed should work in a symmetric way in order to stay flexible to changes in the trustee-trustor relationship. This change can be determined by the change of the payment model.

The Auctioneer has to follow clear rules in order not to adulterate the economic outcome of the matching process. That is, the decision for matching a SC and a SP should be made by each side based on the reliability, which is indicated by the available trust and reputation information.

The decision should not be possible based on price information. More concrete, as soon as SP and SC are matched, each side should be able to refuse the partner based on its information on the former behaviour and not on the price.

R 3 SP and SC must be able to decide on the acceptance of an offer based on the former behaviour of the opponent and not based on the price of the matching (see for example [34])

2.4 Trusting Relationships between Auctioneer, Seller and Buyer

Within this subsection, we are going to consider approaches that regard the trusting relations between the Auctioneer, the set of sellers and the set of buyers. Figure 1 illustrates the three different emerging trusting relationships [30].

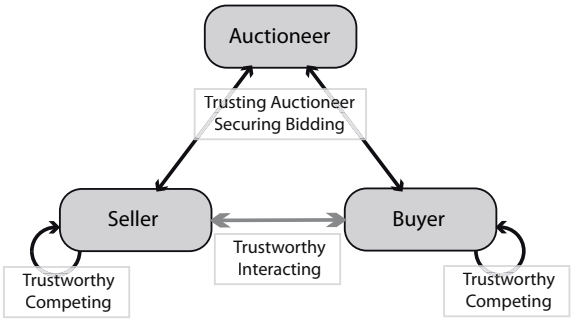


Figure 1: Classification of Competing Approaches, in analogy to [30, p. 145]

The first category regards the trusting relationship between the role of an Auctioneer and the role of bidders. Bidders can be divided into the set of sellers and the set of buyers. These groups are integrated to the group of bidders due to the fact that the trust relation does not differ between the set of sellers and the set of buyers.

Franklin and Reiter [14] mention in their work some of the most important challenges when designing electronic auctions with an Auctioneer instance: the Auctioneer can inform a collaborator regarding submitted bids, the closing time could be manipulated, such that interested bidders are not able to submit their intended bids, the Auctioneer could accept bids after the closing time, the Auctioneer could award the auction to another bidder than the winning bidder, the Auctioneer could collect money from the non-winning bidders and finally the winner could refuse to pay to the Auctioneer [14].

The same authors propose a mechanism for sealed bid auctions that ensures that bids are not revealed until the bidding period has been ended, the Auctioneer collects the money for the service from the winner. All other bidders are ensured that they do not have to pay and only the winner is able to consume the negotiated service [14]. Franklin and Reiter ensure this by the usage of cryptographic techniques.

Brandt [4] specifies this approach for Vickrey Auctions. In second price auctions the Auctioneer has an additional cheating opportunity. It could increase the second-highest bid in order to increase the effective price the winner has to pay. The mechanism of Brandt [4] considers especially the highest and the second price in Vickrey Auctions. Later, the same author proposes a mechanism in which the Auctioneer becomes obsolete as the bids are shared on all bidders [5]. This might solve the trusting problem between Auctioneer and bidders. Brandt [6] also proposes a mechanism for the (M+1)st price mechanism that is also used here in this paper. The author proposes a technique based on the El Gammal encryption. Following this mechanism, a trusted third party can substitute the trusting relationship between bidders and Auctioneer. A similar approach has been proposed by Baudron et al. [2]. An overview of cryptographic approaches without a trusted third party is given by Bogetoft et al. [3]. Wang [31] considers the problem of anonymity in Continuous Double Auctions and the traceability of false offers in this kind of Double Auctions.

The next category of trust relationships goes beyond the analysis of Turel and Yuan [30] (see figure 1). It is the trust relationship between participants within the groups of seller or buyers. Brandt and Weiss [7] name participants, which behave in a fashion that they reduce the profit of competitors, "antisocial" agents. These agents follow this goal beside the goal of maximizing the own profit. Following Brandt and Weiss, these agents "need to deviate from the dominant truth-telling strategy" [7, p. 335]. Among other researchers, Sandholm [26] adresses the problem of truth-bidding by the usage of the Clarke-Groves pricing mechanism, a generalized Vickrey auction. Zhou and Zheng generalize the truth-bidding for Double Auctions in their framework, called TRUST [34].

Finally, we are going to consider the category that represents the focus of this work: the trust relationship between the both bidder groups: seller (SP) and buyer (SC). This problem has been identified and examined by Braynov and Sandholm [8]. In a subsequent work the same authors address, beyond others, this problem [9], but use a single-sided Vickrey auction (in our terms: Auction) to solve this problem. Ramchurn [23] proposes a mechanism to ensure a trusted relation between sellers and buyers, but he also makes use of single-sided auctions. Tafreschi et al. [29] assume the same problem. But within their proposal to solve this problem they make use of a Fixed Price Auction protocol.

Summarizing, we can state that none of the existing approaches focuses the trusting relation between sellers and buyers within a Double Auction. They fail to meet requirement *R0*. We will now propose a design that allows to address the trusting problem within Double Auctions. Therefore, the Negotiation Protocol module has to be modified and the interface for agents to make decisions has to be designed.

3. DESIGN AND DEVELOPMENT

Following the Design Science process of Peffers et al. [21], the protocol is now designed. But before describing the results of this process in this paper, we will consider the underlying assumptions.

3.1 Assumptions and Simulation Environment

The approach assumes that participants might defect the system through not answering the service or payment requests not as expected. The Matcher, instead, is assumed to always act honestly. Keeping related work in mind, this assumption could be dropped in future. Further, no trusted third party for example to fulfil the payment process is available. Finally, the decision, if a trustee is playing honest or fraudulent, is modeled as a binary decision, that is the services can be clearly divided into honest and cheating services.

3.2 Protocol

In order to meet the requirements, the Double Auction protocol has to be re-designed. In this approach, the Matcher makes use of the (M+1)st price rule. The (M+1)st price rule sets the clearing price (the price buyers have to pay and suppliers earn) at the (M+1)st highest price of all bids [33]. For reasons of simplicity, the focus of the protocol will lie on the matched participants only: in case of the unmatched bidder, the Matcher sends a "Lost" message to the set of unmatched SPs and the set of unmatched SCs. For these participants the current Double Auction round is finished without success. In this case no trusting relationship occurs due to the failed attempt to find a suitable partner (to consume or to provide a service).

The protocol proposal is illustrated in figure 2 in terms of an UML sequence diagram. The process starts with an announcement of a new Auction round by the Matcher. Sending this broadcast message simplifies the service discovery process that is also a necessary part of the service life cycle. The process of how to find a service that promises the desired functionality is assumed to work. That means that the functionality can be exactly described with underlying service descriptions and the Matcher will only match with SPs and SCs that have equal service or demand descriptions. Within the following simulation the service is denoted by a textual string that represents the unique service and demand description.

The negotiation process starts with the aforementioned *Call-for-Bids* message that is sent by a broadcast message delivery to all participants. As soon as the SP has resources available to offer, it will answer this Call-for-Bids message with a *Sell* message. Symmetrically, the SC answers in case of an open demand the Call-for-Bids message with a *Buy* message. Both answer messages have to include the price for the proposal. For the SP the price denotes the lowest bid on which it will provide the service. For the SC on the other side, the price denotes the maximum price it is willing to pay for the service.

The Matcher follows the (M+1)st price matching algorithm [33] and stores the proposals in the corresponding proposal sets. In periodic matching proceedings, the proposals are matched regarding the (M+1)st price matching and price determination rule.

The matched services and demands instead, are notified by a *Ask-if-Opp-Is-Ok* message. This message includes the name of the matched opponent. The message does not include the (M+1)st price, such that no economic side effects can occur,

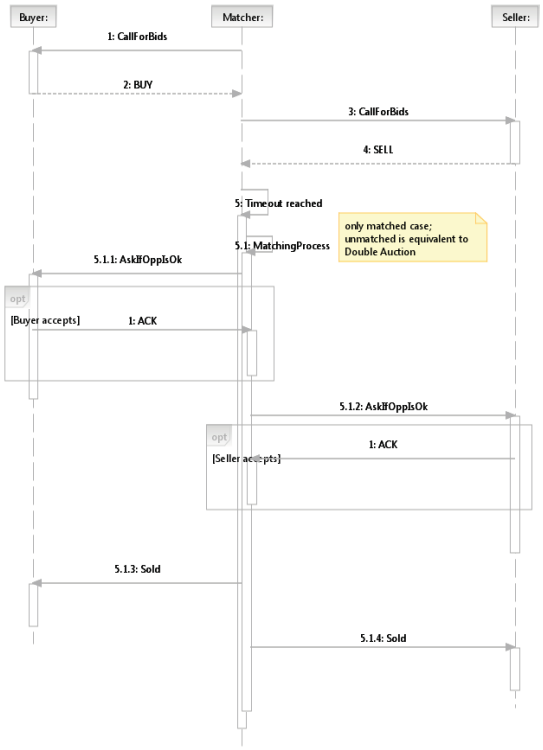


Figure 2: Modified Double Auction Protocol

i.e. refusing the matching due to the high or low price.

The Matcher stores the matched pairs in a new data structure. This data structure is able to store additional flags for each participant that denotes the fact if the corresponding agent has already confirmed the matching. The decision whether to confirm the matching or not, can be made by each agent on individual preferences: if the matched agent is sufficienty trustworthy, the matching is accepted, otherwise it is rejected. This decision point represents the interface to a potential usage of a trust and/or reputation mechanism to confirm the decision with more information, here on historic behaviour of the service or participant. The individual preference is modelled by an individual threshold that determines this decision.

As soon as the matched SP and the corresponding SC accepted the matching and announced this with an *ACK* message to the Matcher, the Matcher sends the final *Sold* message to both sides. This message includes the price on which offer and demand have been matched. At this point the agent can influence its future personal strategy regarding the negotiation itself and can increase or decrease its estimated market price.

4. DEMONSTRATION AND EVALUATION

4.1 Simulation Environment

In order to evaluate the mechanism later, we have to introduce a simulation environment, called SimIS [15], that follows the IoS vision. This system is able to model Internet-like networks where the nodes are hosting active services.

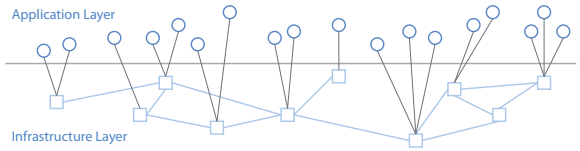


Figure 3: SimIS Architecture [15]

The messages follow the SOAP messages structure and the service interfaces follow real-world Web Services technology, like discussed in Lee and Winslett [17]. These interfaces conform to the widle-used interaction protocols in service economies.

4.1.1 Technological Base: Repast Toolkit

The SimIS[1] toolkit was implemented as an extension to the Recursive Porous Agent Simulation Toolkit [24], developed at the Argonne National Lab, Chicago. Repast is a free and open source agent-based modelling toolkit [18]. This foundation was chosen due to its comprehensive API, the very generic and easy to use set of data gathering and analysis functions as well as the support for network modelling (including respective programming libraries). Technically, the current version of SimIS is based on Repast Symphony and is completely implemented in the Java programming language.

4.1.2 SimIS Architecture

In order to map the abstract IoS architecture to our simulation model a two-tiered architecture for SimIS seems suitable. The overall system is thus divided into an Application Layer and an Infrastructure Layer. An overview of the overall architecture is illustrated in Figure 3.

The *Infrastructure Layer* models topological settings of the IoS. The basic idea is that all Application Layer Agents or Services are linked to a single Infrastructure Agent each, which is representing their server platform. This platform is therefore responsible for sending messages to other Application Layer Agents (including routing and communication patterns, such as broad- or multicast), and receiving messages from other Infrastructure Agents and passing them on to either other Infrastructure Agents (in case the agent represents only the next step on the message's route) or to one or more Application Layer Agents associated with it (in case these are the recipients) [15].

Within the *Application Layer* the actual services of the IoS vision are modelled. Basically the underlying Infrastructure Layer provides us with a high-enough flexibility for implementing any service logic in terms of Application Layer Agents communicating via the offered message objects and routing functionality. Each service (Application Layer Agent) is implemented as a plain Java class and can therefore exploit the full potential this programming language offers in addition to the libraries present within the SimIS toolkit.

4.2 Simulation Experiments

4.2.1 Simulation Scenario

[1]For more information see http://simis.sourceforge.net

As just mentioned, the topology will again be divided into an Infrastructure Layer consisting of nodes and edges between them, and an Application Layer. The network used for simulation experiments consists of 100 nodes that are connected not heavy-weighted and not long-tailed. The mean distance between the nodes is about 3.26 with an maximum distance between two nodes of six hops.

For the following simulation experiments, 200 SP agents and 200 SC agents will be deployed at the beginning of the simulation experiment. In order to introduce dynamics, participants are substituted by newcomers during the simulation experiments. The time range for the substitution process is set to a value that in average the complete population of agents is replaced once during one simulation experiment of 100,000 time ticks. Depending on the payment model 10% SP cheaters or 10% SC cheaters are deployed. This rate might fluctuate due to the dynamic character of the system.

For each simulation setting, the products that are negotiated are fixed by a certain functional attribute definition. This attribute combination is assumed to be defined by an underlying service description. The one and only attribute that is negotiated is denoted by the price. The negotiation protocol is determined by a (M+1)st price Double Auction.

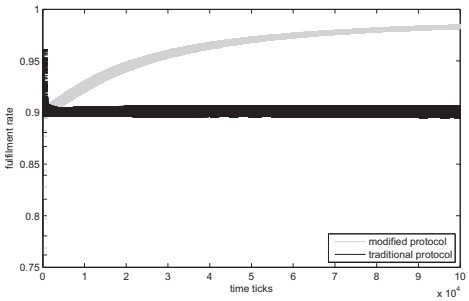
4.2.2 Metrics

As the simulation scenario has been defined, the metrics for the simulation experiments are introduced now. The fulfilment rate and the negotiation rate are plotted in dependence of the simulation time.

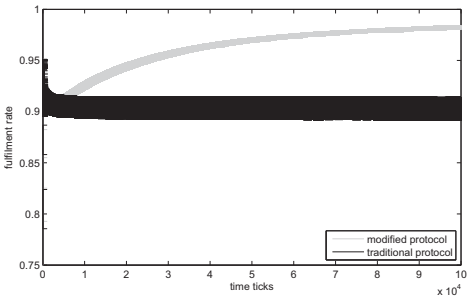
The concrete implementation of the fulfilment depends on the payment model of the simulation experiment. If the service has to be payed in advance, the fulfilment rate considers the service fulfilment: $fr = \sum_i^{|SCs|} \frac{services_i^{received}}{services_i^{paid}}$. For each participant the rate of successful services against failed services are noted. Without any reputation system, one would expect that the service fulfilment rate corresponds to the rate of cheaters in the system. If the service has to be payed after it has been fulfilled, the fulfilment rate considers the payment fulfilment. Then, for each SP the rate of successful payments against failed payments are plotted over time.

The negotiation rate is defined very close to the fulfilment rate. While the latter focuses on the service and payment fulfilment, the negotiation rate focuses on the rate of negotiation processes that have been finsihed successfully: $nr = \sum_i^{|SCs|} \frac{negotiation_i^{finished}}{negotiation_i^{started}}$. The outcome of the fulfilment afterwards is not relevant for this metric. In a system with a well-working reputation system the negotiation rate of defecting agents will fall down as soon as they are identified, and the other participants will not be willing to negotiate with them any more. If the trust and reputation system does not work properly, the rates might not differ at all between honest and cheating participants. The definition when a negotiation begins, depends on the negotiation protocol: in a Double Auction as soon as a bidder submits a bid to the Auctioneer is assumed as the negotiation start.

Finally a metric that covers indicators that base on the ne-

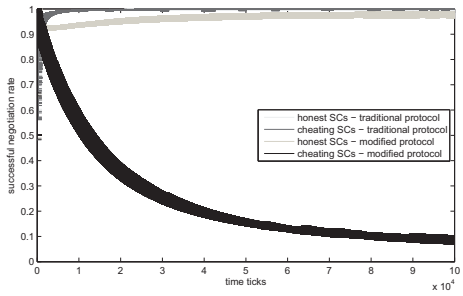


(a) Service Consumer negotiation rate (pay-later)

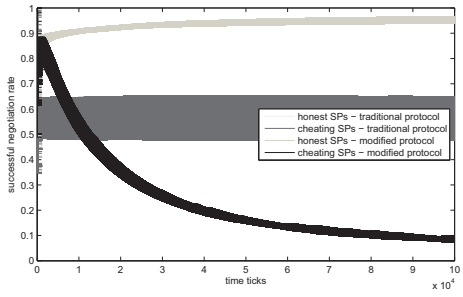


(b) Fulfilment Rate (pay-before)

Figure 4: Double Auction Fulfilment Rate



(a) Service Consumer negotiation rate (pay-later)



(b) Service Provider negotiation rate (pay-before)

Figure 5: Negotiation Rates in modified Double Auction

gotiation and trustworthiness will be introduced: fairness. While the negotiation and fulfilment rate will be plotted over time, this metric will analyse the agent population at the end of each experiment. When plotting the investments or revenues against the amount of services, fairness can be defined in a distribution close to the bisecting line (assuming similar valuations) [32]. The gradient of the bisecting line is determined by the valuation. To ensure the applicability of this metric in both trustor/trustee relationships, we will evaluate the deviation of a population by the root mean square deviation to the expected bisecting line. A deviation value of 0 for the whole polulation would denote a complete "fair" system, whereas a high deviation value denotes an "unfair" system. The deviation is calculated by the root mean square error that measures the differences of the measured values against the expected (fair) values.

4.2.3 Simulation Results

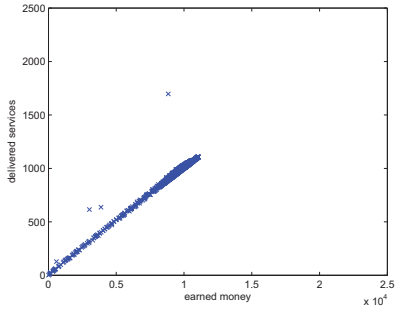
In the following simulations each single experiment is repeated for 30 times using different random valuations and makes use of the modified Double Auction Protocol. With 30 replications and uniformly distributed input parameters, the simulation experiments are expected to satisfy common statistical requirements. During the data analysis the 0.95 confidence interval of the time series are taken for further analysis. The simulation outcome is further compared to the case that uses the initial Double Auction protocol. In both cases, $AVALANCHE_{dec}$ is used as decentralized trust and reputation model for all participants (SPs and SCs).

Figure 4 illustrates the overall fulfilment rate of the scenario's simulation experiments. The figure denotes the payment fulfilment in the pay-later and the service fulfilment in the pay-before model. As we have again 10% of cheating agents, the fulfilment rate is expected at about 90%. As we have seen above, the reference case (initial Double Auction protocol) fulfils the expectation with a constant fulfilment rate of about 90% (dark area).

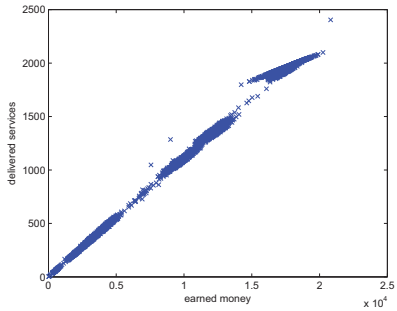
The grey areas illustrate the service or payment fulfilment rates when using $AVALANCHE_{dec}$, combined with the modified protocol. The service fulfilment rate increases significantly compared to the simulation outcome with the initial Double Auction protocol. Both payment models lead to analogous results.

A value of approximated 100% is unrealistic due to the following reasons: during the settlement phase of the simulation run the reputation system has to be filled with information. Within this settlement phase some interactions fail, such that the rate can not reach the 100% value. Further, during the simulation runs the implemented dynamics lead to a continuous arrival of unknown agents.

As we stated above, a working reputation system has to lead to a spreading of negotiation rates between honest and dishonest participants. Figure 5 illustrates the negotiation rates when using the initial or the modified Double Auction protocol. Subfigure 5a considers the pay-later case where the SCs are acting as trustees. Determined by the payment model, these agents are able to cheat the corresponding SPs. Following, with a well-working trust and reputation model the negotiation rate of cheating and honestly acting SCs

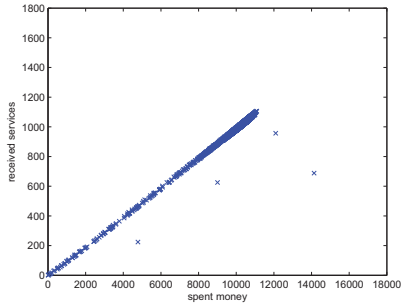


(a) modified protocol

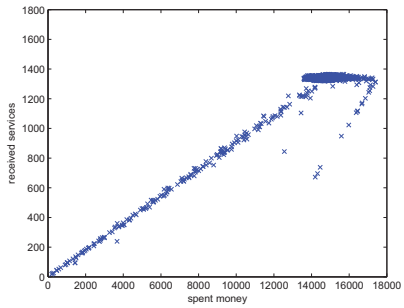


(b) initial protocol

Figure 6: Fairness between Service Providers in Double Auctions (pay-later)



(a) modified protocol



(b) initial protocol

Figure 7: Fairness between Complex Service Agents in Double Auctions (pay-before)

should spread.

The initial Double Auction protocol does not differentiate regarding the negotiation rate between honest and dishonest SCs. As soon as we implement the modified Double Auction protocol, the negotiation rate of dishonest SCs decreases over time and spreads compared to the rate of honest SCs (dark area vs. light grey area). This case (modified Double Auction, pay-later and $AVALANCHE_{dec}$) is now a well-working case regarding the trustworthiness.

The pay-before case is illustrated in figure 5b. Like in the pay-later case, an implementation of the modified protocol leads to a spreading of negotiation rates. Honest SPs, which represent the trustees now, are more often able to finish a negotiation process compared to cheating SPs (dark area vs. light grey area). As soon as the initial Double Auction protocol is used to coordinate the services, the negotiation rates of both SP groups do not differ at all. Demand and supply are not balanced in the scenario, such that the negotiation rates differ between SC and SP.

Figures 6 and 7 will now consider the fairness metric when coordinating the services in an IoS environment by the different Double Auction protocols.

Based on the findings before, figure 6a denotes a well-working combination of the modified Double Auction protocol with the pay-later payment model and $AVALANCHE_{dec}$ as reputation model. This combination leads to less negotiations with fraudsters involved. In consequence, a smaller amount of SPs are cheated. The root mean square error to the expected bisection (completely fair system) is denoted by 22.27.

Figure 6b illustrates the result of the same experiment, but with the initial Double Auction protocol. Compared to the simulation experiment with usage of the modified specification, we can immediately see that the reputation information of $AVALANCHE_{dec}$ does not influence the results. SPs receive less money in relation to the services they delivered. These cheated agents can be found above this imaginary bisection. The root mean squared error in this case is with 145.9 higher than in the same experiment using the extended protocol version.

Based on the findings before, figure 7a illustrates the second well-working combination of the modified Double Auction protocol: this time with the pay-later model. Here, the usage of $AVALANCHE_{dec}$ leads to less negotiations with fraudsters involved. In consequence, a smaller amount of SPs are cheated. The root mean square error to the desired bisection is denoted by 20.01. If we use the initial approach in the pay-before case, the simulation outcome decreases in its fairness metrics (see figure 7b). The simulation experiment leads to a root mean squared error of 149.1.

Additional sensitivity analyses show that the model is very robust against changes in the cheaters rate (up to about 40%) and regarding the amount of services (with more than

50 SPs and SCs). Running simulations for 100,000 ticks ensures that stable system states are reached. Neither increasing the simulation duration, nor changing the reputation model (for example by using the proposal by Abdul-Rahman [1]) affect the simulation outcomes.

Summarizing the results of the modified Double Auction experiments, we can state that both initial hypotheses regarding the Double-sided Auction can be confirmed: when using the original Double Auction protocol, the instantiation of $AVALANCHE_{dec}$ does not increase the trustworthiness of the system compared to the expected trustworthiness based on the cheating probability of trustees. This statement is unversally valid for both possible payment models and following both trustee/trustor relationships. As soon as we use the modified Double Auction protocol, which has been proposed in this paper, the impact of $AVALANCHE_{dec}$ renders the system more trustworthy and fairness between agents can be ensured.

If we consider this modification again in detail, we can notice that through the additional confirmation an additional decision point for SCs and SPs has been introduced. For these participants this additional decision point leads to the possibility to decide based on trust and reputation information. In terms of the introduced reputation roles, the trustors become Beneficiary of the reputation system, such that the reputation role overlapping determines a working system.

In order to close the section on the Demonstration and Evaluation of the modified protocol combined with the decentralized reputation model $AVALANCHE_{dec}$, we are going to review the deducted requirements.

R 0 *Trusting relationship between Sellers and Buyers:* The proposed mechanism adresses the trust relationship between the group of sellers (SPs) and buyers (SCs). This fact becomes obvious when changing the behaviour regarding the honesty of agents during the evaluation step. In both payment models, cheating trustees are detected and indirectly excluded. Further, the modified protocol is still denoted as Double Auction protocol, such as both groups, sellers and buyers, are bidding on services or demand within the same protocol.

R 1 *The trustor is able to use its reputation information:* With the modification of the protocol an additional decision point has been added. All participants can now decide if they accept the corresponding opponent. Through this additional possibility the agents are not confronted with an anonymous amount of agents. Instead they are already clearly allocated to a specific partner. Now they are able to use trust and reputation information to make the decision whether to confirm the matching or not. In terms of the reputational roles, SCs and SPs are now Beneficiaries of the reputation system.

R 2 *Mechanism that covers both payment models:* Such as for both sides, buyer and seller, this additional decision point has been introduced, the protocol is still a symmetric one. The fact that the protocol works well with both models could be shown within the evaluation.

R 3 *Decision on acceptance must not depend on price information:* This requirement can also be denoted as fulfilled. The "Ask-if-opponent-is-Ok" message does not include any price information. Instead it includes the name of the opponent, such that the agent is able to decide based for example on the opponent's past behaviour. The price information (in our simulation the (M+1)st price) is conveyed with the "Sold" message after both partners have confirmed the matching. This split into two steps avoids an influence of strategic economic behaviour at this point.

5. CONCLUSIONS AND FUTURE WORK

The paper identified trusting problems between autonomous services in the IoS vision. The IoS envisions a service ecosystem where services are traded without human interactions and even beyond company boundaries. The conflicting interests between participants lead to strategic behaviour of single actors.

In literature, trust and reputation models are proposed to set incentives to act honestly. But when using Double Auctions to match buyers and sellers, these trust and reputation models fail to close this "trusting gap". A theoretical investigation of the problem has led to the idea to introduce additional decision points. A modified Double Auction protocol fulfils the deducted requirements. Simulation experiments show that the usage of this modified protocol leads to increased trustworthiness for the participants. Concluding, the paper shows that it is possible to close the trusting gap in Double Auction Markets as soon as the protocol is modified as proposed. With a traditional Double Auction protocol, closing this trusting gap is not possible.

Future work should focus on the independence of the proposed mechanism from the used trust and reputation model. Even if first investigations have been positive, this question needs further investigations in future. Especially side-effects between Double Auction configurations and reputation model configurations should be investigated. Further, a detailed economic investigation of common Double Auction requirements, like truthful bidding is still to be done.

6. REFERENCES

[1] A. Abdul-Rahman. *A Framework for Decentralised Trust Reasoning*. PhD thesis, University of London, 2005.

[2] O. Baudron and J. Stern. Non-interactive private auctions. In P. Syverson, editor, *Financial Cryptography*, volume 2339 of *Lecture Notes in Computer Science*, pages 364–377. Springer Berlin / Heidelberg, 2002.

[3] P. Bogetoft, I. Damgård, T. Jakobsen, K. Nielsen, J. Pagter, and T. Toft. A practical implementation of secure auctions based on multiparty integer computation. In *Financial Cryptography and Data Security*, volume 4107 of *LNCS*, pages 142–147. Springer, 2006.

[4] F. Brandt. Cryptographic protocols for secure second-price auctions. In *COOPERATIVE INFORMATION AGENTS V*. Springer, 2001.

[5] F. Brandt. Secure and private auctions without

auctioneers. Technical Report FKI-245-02, Institut für Informatik, Technische Universität München, 2002.

[6] F. Brandt. How to obtain full privacy in auctions. *International Journal of Information Security*, 5(4):201–216, 2006.

[7] F. Brandt and G. WeiSS. Antisocial agents and vickrey auctions. In J.-J. C. Meyer and M. Tambe, editors, *Intelligent Agents VIII*, volume 2333 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 335–347. Springer, 2001.

[8] S. Braynov and T. Sandholm. Contracting with uncertain level of trust. *Computational Intelligence*, 18(4):501–514, 2002.

[9] S. Braynov and T. Sandholm. Auctions with untrustworthy bidders. In *Proceedings of the IEEE International Conference on E-Commerce, 2003*, pages 363–369, 2003.

[10] H. U. Buhl and R. Winter. Full virtualization - bise's contribution to a vision. *Business & Information Systems Engineering*, 1:1–4, 2009.

[11] R. Conte and M. Paolucci. *Reputation in arificial Societies*. Kluwer Academic Publisher, 2002.

[12] J. Eekels and N. F. M. Roozenburg. A methodological comparison of the structures of scientific research and engineering design: their similarities and differences. *Design Studies*, 12(4):197 – 203, 1991.

[13] T. Eymann, S. König, and R. Matros. A framework for trust and reputation in grid environments. *Journal of Grid Computing; Special Issue on Grid Economics*, 6(3):225–237, 2008.

[14] M. K. Franklin and M. K. Reiter. The design and implementation of a secure auction service. *IEEE Transactions on Software Engineering*, 22(5):302–312, 1996.

[15] S. König, S. Hudert, and T. Eymann. Socio-Economic Mechanisms to Coordinate the Internet of Services - the Simulation Environment SimIS. *Journal of Artificial Societies and Social Simulation (JASSS)*, 13(2), 2010.

[16] S. König, S. Hudert, T. Eymann, and M. Paolucci. Towards reputation enhanced electronic negotiations for service oriented computing. In *Proceedings of the CEC/EEE 2008, Washington, DC*, pages 285–292, 2008.

[17] A. Lee and M. Winslett. Towards standards-compliant trust negotiation for web services. In Y. Karabulut, J. Mitchell, P. Herrmann, and C. Jensen, editors, *Trust Management II*, volume 263 of *IFIP International Federation for Information Processing*, pages 311–326. Springer Boston, 2008.

[18] M. J. North, N. T. Collier, and J. R. Vos. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation*, 16:1–25, 2006.

[19] B. Padovan, S. Sackmann, T. Eymann, I. Pippow, and H. Eggs. A Prototype for an Agent-based Secure Electronic Marketplace Including Reputation Tracking Mechanisms. In *Proceedings of the Thirty-Fourth Annual Hawaii International Conference on System Sciences (HICSS 2001)*, 2001.

[20] B. Padovan, S. Sackmann, T. Eymann, I. Pippow, and H. Eggs. Automatisierte Reputationsverfolgung auf einem agentenbasierten elektronischen Marktplatz. In H.-U. Buhl, A. Huther, and B. Reitwiesner, editors, *Information Age Economy (Proc. 5. Internationale Tagung Wirtschaftsinformatik 2001, Augsburg, 18.-21. September)*, pages 517–530. Physica Verlag, Heidelberg, September 2001.

[21] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of Management Information Systems*, 24:45–77, 2007.

[22] C. Preist. Commodity trading using an agent-based iterated double auction. In *AGENTS '99: Proceedings of the third annual conference on Autonomous Agents*, pages 131–138, New York, NY, USA, 1999. ACM.

[23] S. D. Ramchurn. *Multi-Agent Negotiation using Trust and Persuasion*. PhD thesis, University of Southampton, 2004.

[24] Repast Development Group. Repast Home Page. Website. http://repast.sourceforge.net.

[25] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artif. Intell. Rev.*, 24(1):33–60, 2005.

[26] T. Sandholm. emediator: A next generation electronic commerce server. *Computational Intelligence*, 18(4):656–676, 2002.

[27] C. Sansores and J. Pavón. Agent-based simulation replication: A model driven architecture approach. In *Proceedings of the 4th Mexican international conference on artificial intelligence (MICAI), Monterrey, Mexico*, Lecture Notes on Computer Science, pages 244–253. Springer, 2005.

[28] W. Streitberger, S. Hudert, T. Eymann, B. Schnizler, F. Zini, and M. Catalano. On the simulation of grid market coordination approaches. *Journal of Grid Computing; Special Issue on Grid Economics*, 6(3):349–366, 2008.

[29] O. Tafreschi, D. Mähler, J. Fengel, M. Rebstock, and C. Eckert. A reputation system for electronic negotiations. *Computer Standards & Interfaces*, 30:351–360, 2008.

[30] O. Turel and Y. Yuan. You canŠt shake hands with clenched fists: potential effects of trust assessments on the adoption of e-negotiation services. *Group Decision and Negotiation*, 17:141–155, 2008.

[31] C. Wang and H. Leung. Anonymity and security in continuous double auctions for internet retails market. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.

[32] A. Wierzbicki. The case for fairness of trust management. *Electronic Notes in Theoretical Computer Science*, 197(2):73 – 89, 2008. Proceedings of the 3rd International Workshop on Security and Trust Management (STM 2007).

[33] P. R. Wurman, W. E. Walsh, and M. P. Wellman. Flexible double auctions for electronic commerce: theory and implementation. *Decision Support Systems*, 24:17–24, 1998.

[34] X. Zhou and H. Zheng. Trust: A general framework for truthful double spectrum auctions. In *IEEE INFOCOM*, pages 999–1007, 2009.

Kombinatorische Auktionen als Instrument zur Verhandlung von Verträgen mit interdependenten Eigenschaften

Fabian Lang
Professur für BWL, insbes. Wirtschaftsinformatik
Fakultät für Wirtschafts- und Sozialwissenschaften
Helmut-Schmidt-Universität Hamburg
Holstenhofweg 85, 22043 Hamburg
Tel. +49 40 6541 2321
fabian.lang@hsu-hamburg.de

Andreas Fink
Professur für BWL, insbes. Wirtschaftsinformatik
Fakultät für Wirtschafts- und Sozialwissenschaften
Helmut-Schmidt-Universität Hamburg
Holstenhofweg 85, 22043 Hamburg
Tel. +49 40 6541 2857
andreas.fink@hsu-hamburg.de

ZUSAMMENFASSUNG

Der technologische Fortschritt beim E-Commerce, bei der betrieblichen Planung und im Zusammenhang mit intelligenten Multi-Agenten-Systemen eröffnet eine Vielzahl an Möglichkeiten für automatisierte Verhandlungen durch Softwareagenten. Die Interdependenz verschiedener Verhandlungsgegenstände in der Form von Vertragseigenschaften kann dabei zu nicht-linearen Vertragsräumen führen, sodass Verträge mit hoher sozialer Wohlfahrt nicht erreicht werden. Als Lösungsansatz wurde ein kombinatorisches Auktionsverhandlungsprotokoll entwickelt, das Bestimmungsrechte zu Vertragseigenschaften versteigert. Das Verhandlungsprotokoll wurde simulativ getestet und mit weiteren Protokollen in Bezug auf Wohlfahrtsgüte, Rechenleistung und Anreizkompatibilität verglichen. Als Ergebnis erreichte das vorgeschlagene Auktionsverhandlungsprotokoll Verträge mit der höchsten Wohlfahrt innerhalb für die adressierten Anwendungen praxistauglicher Rechenzeit verbunden mit vollständiger Anreizkompatibilität.

Schlüsselwörter

automatisierte Verhandlungen, kombinatorische Auktionen, Multi-Agenten-Systeme, nicht-lineare Vertragsräume

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

1. Einleitung

Im Jahre 1899 trafen sich der belgische Chemiker Leo Baekeland und der Unternehmer George Eastman, um über den Kauf der neuesten Erfindung Baekelands, ein schnell entwickelbares Fotopapier, zu verhandeln. Baekeland plante, $ 50.000 als Kaufpreis zu verlangen, war aber bereit, Eastman bis $ 25.000 entgegen zu kommen. Eastman eröffnete zuerst die Verhandlung und kaufte das Patent für $ 1.000.000 [1]. Die Baekeland-Anekdote zeigt die Wichtigkeit der Verhandlungsgestaltung und die Bedeutung des Verhandlungsprotokolls für das spätere Verhandlungsergebnis. Da technologische Innovationen vielerlei Möglichkeiten zur Verhandlung geschaffen haben, haben Verhandlungen heute einen Stellenwert wie niemals zuvor. Im Bereich des E-Commerce gibt es Angebote, wie etwa das individuelle Zusammenstellen von Lebensmittelzutaten (z.B. Müsli – mymuesli.com) oder Name-Your-Own-Price-Auktionen (z.B. Reisen – priceline.com), die in einem klassischen Filialsystem aufgrund von hohen Transaktionskosten nahezu undenkbar wären. Doch nicht nur die Verhandlungsmöglichkeiten des Menschen wurden durch die Technologie erweitert, sondern es wurden auch die Grundlagen für automatisierte Verhandlungen zwischen intelligenten Softwareagenten geschaffen, die die Interessen eines menschlichen Prinzipals vertreten können. Beispiele hierfür sind agentenbasierte Konzepte zur Ablaufplanung von Fertigungsprozessen oder die Datenallokation in Rechenzentren [9,17]. Automatisierte Verhandlungen sind als zentrales Instrument zur Koordination von autonomen Agenten anzusehen [12]. Aktuelle Entwicklungen wie das Internet der Dinge oder Smart Houses tragen dazu bei, dass automatisierte Verhandlungen und intelligente Softwareagenten auch in Zukunft an Bedeutung gewinnen werden [14,15].

Verträge umfassen häufig mehrere festzulegende Vertragseigenschaften mit Abhängigkeit untereinander. So können in der Fertigungsplanung Abhängigkeiten aufgrund technisch bedingter Rüstaufwände oder der Relevanz der örtlichen Position einer Maschine innerhalb einer Fertigungskette entstehen. Diese Interdependenzen führen bei einer adäquaten Abbildung der Nutzenfunktionen zu nicht-linearen Vertragsräumen, die meist keine Monotonie der Nutzenfunktionen über den Verhandlungsverlauf aufzeigen, wodurch gängige

Verhandlungsformen nicht zwangsläufig zu Pareto-optimalen Einigungen führen [16].

Ein beispielhaftes Anwendungsszenario ist die Koordination von Flurfahrzeugen in einem von mehreren Unternehmen genutzten Lager. Ein von mehreren Parteien genutztes Lager kann Effizienzgewinne aufgrund einer höheren durchschnittlichen Auslastung realisieren. Unter der Annahme, dass die Flurfahrzeuge von verschiedenen Unternehmen gestellt werden, sind unterschiedliche Interessen zu berücksichtigen. Interdependenzen treten bspw. aufgrund der relativen Position auf; d.h., ein Flurfahrzeug, das eine Entität in der Nähe einer anderen einlagert, kann eine darauf folgende Auslagerung jener anderen Entität mit einer relativ hohen Effizienz durchführen. Eine ausschließliche Nutzung der Fahrzeuge für das eigene Unternehmen wäre damit im Allgemeinen ineffizient. Eine zentrale Steuereinheit kann den Interessen der verschiedenen Parteien nicht ausreichend gerecht werden (und höchstens als Mediator auftreten), insofern die Unternehmen ihre privaten Informationen nicht unbedingt preisgeben möchten. Die Flurfahrzeuge können jedoch als Agenten eingesetzt werden, ohne Informationen an andere Parteien komplett zu offenbaren, und mit Hilfe von Verträgen in der Form entsprechender Ablaufpläne die anstehenden Prozesse dezentral koordinieren.

Die zweckmäßige Gestaltung bilateraler und insbesondere auch multilateraler Verhandlungen ist bei nicht-linearen Vertragsräumen nicht trivial. Der Nutzengraph weist durch die Nicht-Linearität viele Hügel und Täler auf und es besteht die Gefahr, in einem lokalen Optimum zu verharren. Wie in Abbildung 1 illustriert wird, müssen Agenten unter Umständen zeitweise Verschlechterungen eingehen, um zu ihrem optimalen Vertrag im Verhandlungsprozess zu gelangen – unterstellt, dass nur schrittweise Veränderungen gemäß der gezeigten Vertragsanordnung möglich sind. So muss Agent 1 von seinem lokalen Maximum bei Vertrag 2 abweichen, um zu seinem bestmöglichen Vertrag (Vertrag 7) zu gelangen. Agent 2 erreicht seinen bestmöglichen Vertrag jedoch schon bei Vertrag 3 und wird nicht mehr bereit sein, von diesem abzuweichen. Sofern Agent 1 dies antizipiert oder sich zumindest des Risikos bewusst ist, wird er ebenfalls nicht mehr bereit sein, sein lokales Maximum bei Vertrag 2 zu verlassen, sodass weder Agent 1 noch Agent 2 noch die soziale Wohlfahrt (SW) ihr globales Optimum erreichen.

Dieses kontraproduktive, opportunistische Verhalten ist bei dem Design von Verhandlungsprotokollen zu berücksichtigen. Insbesondere bei multilateralen Verhandlungsproblemen erweist sich eine adäquate Einigung auf einen Vertrag durch iterative Verhandlung als zunehmend schwierig. Des Weiteren ist zu vermuten, dass die bekannten iterativen Protokolle bei einer sehr kleinen Anzahl zu verhandelnder Eigenschaften oder bei gegensätzlichen Bewertungen ebenfalls keine bestmögliche Einigung garantieren können. Daher versucht diese Arbeit, die daraus resultierende Forschungslücke zu schließen, indem ein neuartiges Verhandlungsprotokoll, das sich kombinatorischer Auktionen bedient, vorgeschlagen wird. Die Evaluation erfolgt durch eine simulative Untersuchung der oben genannten Aspekte.

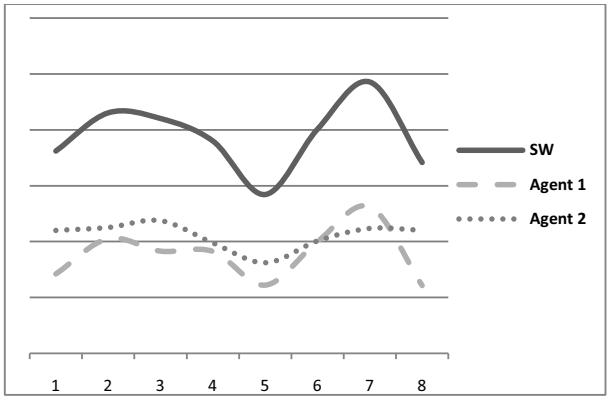


Abbildung 1. Nutzenwerte benachbarter Verträge über den Verhandlungsverlauf

Im Anschluss an die Einleitung wird in Kapitel 2 ein selektiver Überblick über die verwandte Literatur gegeben. In Kapitel 3 werden die Szenarien der Studie präsentiert. In Kapitel 4 werden das kombinatorische Auktionsprotokoll sowie zwei Referenzprotokolle vorgestellt. Kapitel 5 diskutiert die Ergebnisse der Simulation. Abschließend wird in Kapitel 6 ein Fazit gezogen und ein Ausblick auf die weitere Forschung gegeben.

2. Ausgewählte verwandte Literatur

Die Ausgestaltung von multidimensionalen Verträgen wurde bisher sowohl aus der Perspektive der Spieltheorie als auch im Gebiet der künstlichen Intelligenz untersucht [19].

Eine Einführung zur Forschung im Bereich der allgemeinen Multi-Agenten-Koordinationsprobleme gibt Conitzer [5] mit illustrativen Beispielen. Weiterführend geben Jennings et al. [12], Lomuscio et al. [21] und Ströbel und Weinhardt [33] einen Überblick zur automatisierten Verhandlung.

Die spieltheoretische Erforschung von Multi-Agenten-Verhandlungsproblemen findet bereits seit den 1940ern statt. Ein erstes Lösungskonzept erreichte Nash [24] mit der sogenannten *Nash-Lösung*, die auf vier Axiomen aufbaut, die von rationalen Individuen zu erfüllen sind: 1.) Unabhängigkeit von gleichwertigen Transformationen, 2.) Pareto-Optimalität, 3.) Unabhängigkeit von irrelevanten Alternativen, 4.) Symmetrie. Das ursprüngliche Nash-Verhandlungsproblem wurde durch ein *Ultimatumspiel* dargestellt, also eine Situation, bei der ein Gut oder Güterbündel als Spielgegenstand zwischen zwei oder mehreren Personen aufgrund eines einzigen Aufteilungsvorschlags geteilt werden müssen. Das Ultimatumspiel wird häufig anhand der Metapher eines zu teilenden Kuchens illustriert, bei dem bei Ablehnung eines Teilungsangebotes keiner der Spieler etwas erhält. Es muss folglich eine Kooperation bzw. eine Koalition zwischen den Spielern erzielt werden. Nashs ursprünglicher Beitrag befasst sich mit Nicht-Nullsummenspielen in der kooperativen Spieltheorie (Koalitionsspiele), was er jedoch bereits ein Jahr später auf die nicht-kooperative Spieltheorie und Nullsummenspiele erweiterte [25]. Einen wichtigen Beitrag zur kooperativen Spieltheorie lieferten auch Kalai und Smorodinsky [13], die das Nash'sche Axiom der irrelevanten Alternativen durch ein Monotonie-Axiom ersetzen. Rubinstein [29] entwickelte zur nicht-kooperativen Spieltheorie ein Verhandlungsmodell, bei dem in einem Zwei-Agenten-Ultimatumspiel eine Lösung durch abwechselnde,

alternierende Aufteilungsvorschläge erreicht wird. Sowohl das Rubinstein'sche als auch das Nash'sche Verhandlungsmodell nehmen vollständige Information über die Präferenzen der Vertragspendants an. In realen Szenarien ist jedoch ein Anreiz vorhanden, Informationen über eigene Präferenzen strategisch geheim zu halten oder sogar den Verhandlungspartner darüber zu täuschen. Es ist ebenso denkbar, dass die eigenen Präferenzen den Akteuren teilweise unbekannt sind. Die Spieltheorie kann deshalb nur stark eingeschränkt zur Lösung von angewandten Verhandlungsproblemen eingesetzt werden.

Aus dem Feld der künstlichen Intelligenz schlugen z.B. Klein et al. [16], Hattori et al. [10] und Lai et al. [19] Verhandlungsmechanismen vor, bei denen ein Mediator Vertragsvorschläge unterbreitet und somit die Suche im möglichen Vertragsraum nach Pareto-optimalen Lösungen unterstützt. So können bei Klein et al. [16] die Agenten über Annahme oder Ablehnung eines Vorschlages entscheiden. Verschiedene Agentenverhalten sowie deren Anreizkompatibilität wurden untersucht. Fink [8] entwickelte ebenfalls eine Strategie zur Anreizkompatibilität innerhalb dieses Verhandlungsrahmens. Homberger [11] verwendete und evaluierte das von Klein et al. [16] entwickelte und von Fink [8] erweiterte Protokoll auf der Basis eines Produktionskoordinationsproblems mit mehreren Produktionsanlagen. Aufbauend auf Klein et al. [16] konzipierten Hattori et al. [10] ein Verhandlungsprotokoll, das durch einen iterativen Bietprozess den Raum der möglichen Vereinbarung sukzessive verkleinert und den Vertrags- und somit den Suchraum einschränkt. Lai et al. [19] entwarfen einen Verhandlungsmechanismus, bei dem ein Mediator mehrere Basislösungen bestimmt, die den Vertragsraum und die Verhandlungsspannweite der Agenten auf der Suche nach Pareto-effizienten Verhandlungslösungen einschränken. Lai und Sycara [18] entwickelten des Weiteren ein Protokoll mit abwechselnden Angeboten, das auf dem Rubinstein'schen Verhandlungsmodell beruht.

Die Anwendung von Auktionen als Verhandlungsinstrument wird vielfach in der Literatur postuliert. So entwarfen Collins et al. [4] eine kombinatorische Auktionsarchitektur für Multi-Agenten-Verhandlungen zur marktlichen Koordination von Auftraggeber und -nehmer, z.B. für Bauprojekte. Bulow und Klemperer [3] verglichen den Erfolg von Auktionen mit dem von Verhandlungen am Beispiel einer Unternehmensakquisition und zeigten dabei, dass die Auktion in diesem Falle der Verhandlung überlegen ist.

In Bezug auf kombinatorische Auktionen geben Cramton et al. [6] eine gute allgemeine Einführung. Sandholm [31] entwickelte ein Auktionsdesign zur eingeschränkten Überwindung des *Winner Determination Problem*, einem $\mathcal{NP}$–schweren Problem, das bei der Bestimmung der Auktionsgewinner im Allgemeinen entsteht und welches zu dem gewichteten Mengenpackungsproblem äquivalent ist [30]. Hierbei werden Gebote auf Zweierkombinationen beschränkt und mit Hilfe eines Branch-on-Bid-Verfahren optimiert. Dieses Auktionsdesign kann zwar die $\mathcal{NP}$-Schwere nicht auflösen, aber eine effiziente Berechnung für bestimmte praxisrelevante Probleme gewährleisten. Lehmann et al. [20] präsentieren verschiedene Approximationsalgorithmen für das Winner Determination Problem (z.B. Greedy Winner Determination), die mit geringem Rechenaufwand adäquate Resultate erreichen.

3. Forschungsszenario

Vertragsverhandlungen können in sehr unterschiedlichen Ausgestaltungen und Formen stattfinden. So existieren sowohl Verhandlungen zwischen Unternehmer und Konsument (*B2C*) als auch Verhandlungen zwischen zwei Unternehmern (*B2B*), wobei einzelne Agenten (*One-to-One*) bilateral oder mit mehreren anderen Agenten (*One-to-Many*) sowie mehrere Agenten untereinander (*Many-to-Many*) verhandeln können [27]. Eine Zahlung oder ein Gütertausch kann dabei Bestandteil des Vertrages sein. Es existieren aber vielerlei Beispiele für Verhandlungen ohne Zahlung wie etwa beim Nierentausch oder bei Verhandlungen zur Planung von Maschinenbelegungen [4]. Die Bewertung der einzelnen Verträge kann ebenso unterschiedlich verlaufen. So wird zwischen privater Bewertung und gemeinsamem Wert unterschieden, was weitreichende Folgen für die Bereitstellung von Informationen hat. Bei der privaten Bewertung bilden die Agenten unabhängig von den anderen Agenten ihre private Präferenz und versuchen zudem, diese strategisch geheim zu halten. Im Gegenzug dazu kann ein gemeinsamer Wert vorliegen, der allerdings gegebenenfalls nicht offenliegt, so dass die Agenten diesen unterschiedlich abschätzen. Es existieren ebenso Mischformen, bei denen teilweise Einigkeit herrscht und teilweise unterschiedliche, private Bewertungen existieren [34].

Das hier betrachtete Szenario umfasst eine Many-to-Many-Verhandlung, bei der sich die verschiedenen *Akteure* $j \in \{\mathbb{N}|0 < j \leq J\}$ auf einen Vertrag mit N zweiwertigen Eigenschaften einigen müssen. Die daraus resultierenden möglichen Verträge $c = \{d_1, \dots, d_n, \dots, d_{N-1}, d_N\}$ innerhalb des Vertragsraumes C sind Bitvektoren; d.h., eine Vertragseigenschaft $n \in \{\mathbb{N}|0 < n \leq N\}$ kann die booleschen Werte $d_n \in \{0,1\}$ annehmen. Die einzelnen Nutzen der Vertragseigenschaften sind dabei nicht unabhängig, sondern *paarweise voneinander abhängig.*

Die Nutzen *einer* Vertragseigenschaft sind gleichverteilt im Intervall [-100, 100]. Der Nutzenzugewinn bei der Kombination *zweier* Eigenschaften wurde als normalverteilt mit $\chi \sim \mathcal{N}(\mu = 0{,}2; \sigma = 0{,}1)\,|\chi > 0$ angenommen. Es ergibt sich für einen Akteur *j* folgender Nutzen P_j für die Zweierkombination *n* und *n'*:

$$P_j(n,n')|\langle d_n = 1 \wedge d_{n'} = 1\rangle = [P_j(n,n')|\langle d_n = 1 \wedge d_{n'} = 0\rangle + P_j(n,n')|\langle d_n = 0 \wedge d_{n'} = 1\rangle] * [1 + \chi].$$

Wie hier sichtbar, umfasst das Szenario sowohl *Komplemente* als auch *Substitute*, d.h., der Nutzen einer Kombination aus zwei Eigenschaften kann größer bzw. geringer als die Summe der dazu gehörigen Einzelnutzen sein. Sofern die Summe der Nutzen zweier einzelner Vertragseigenschaften positiv ist, ist die Kombination noch wünschenswerter für den Akteur (Komplement). Bei den Substituten ist diese Beziehung genau umgekehrt. Ein Akteur j hat demnach für einen Vertrag c die Nutzenfunktion

$$U_j(c) = \sum_{p=1}^{N} \sum_{q=p}^{N} P_j(p,q)|\langle d_p = 1 \wedge d_q = 1\rangle * d_p * d_q,$$

während die Gesamtgüte der Verträge anhand der *sozialen Wohlfahrt SW* gemessen werden soll:

$$SW(c) = \sum_{j=1}^{J} U_j(c).$$

In der Literatur wird häufig die Pareto-Effizienz als Kriterium für einen optimalen Vertrag verwendet. Eine Allokation – oder hier

der Vertrag – ist Pareto-effizient, sobald ein Agent sich nicht mehr besser stellen kann, ohne einen anderen Agenten schlechter zu stellen. Die Maximierung der sozialen Wohlfahrt erfüllt, als Teilmenge der Pareto-effizienten Lösungen, das Pareto-Kriterium und ist sogar noch restriktiver als dieses, erfordert jedoch eine kardinale Nutzenquantifizierung.

Die im folgenden Kapitel präsentierten Verhandlungsprotokolle wurden anhand dreier konkreter Szenarioeinstellungen getestet:

Szenario 1: private Bewertung ($N = 50,\ J = 3$)

Im ersten Szenario werden ausschließlich private Bewertungen berücksichtigt. Die drei Akteure unterliegen keinen Externalitäten – die Bildung ihrer Präferenzen erfolgt also unabhängig von denen der anderen Agenten. Der zu verhandelnde Vertrag umfasst 50 binäre Eigenschaften. Dieses Szenario wird in der Literatur als Standardfall behandelt.

Szenario 2: gemischte Bewertung ($\tilde{N} = 5,\ J = 3$)

Im diesem Szenario findet eine teilweise gemeinsame Bewertung statt, sodass ein Teil des Vertrages ignoriert werden kann, da bei diesen Vertragseigenschaften sowohl Informationen über die Präferenzen der Akteure als auch Einigkeit über die Werte der Eigenschaften existieren. Bei einer gemeinsamen Bewertung und einer identischen Interessenrichtung haben die Akteure keinen Anreiz zur Geheimhaltung von Informationen. Ein Teil der Vertragseigenschaften wird allerdings in diesem Szenario weiterhin privat bewertet und muss verhandelt werden. Die Vertragsgröße weicht in diesem Szenario nicht von denen der anderen Szenarien ab ($N = 50$), jedoch resultiert – aufgrund des Konsenses der Akteure über 45 Eigenschaften – eine kleinere *zu verhandelnde* Vertragsgröße mit $\tilde{N} = 5$ unabhängig bewerteten Eigenschaften. Diesem Szenario liegt die Vermutung zu Grunde, dass verhandlungsbasierte Protokolle bei einer kleinen Anzahl an zu verhandelnden Eigenschaften eher in suboptimalen Zuständen verharren.

Szenario 3: gemischte – mit gemeinsamer, aber gegensätzlicher – Bewertung ($N = 50,\ J = 3$)

Im letzten Szenario sind die Nutzenfunktionen der drei Akteure teilweise abhängig und teilweise unabhängig. Agent 1 und Agent 2 haben eine gemeinsame, aber gegensätzliche Bewertung

$$U_1(c) = -U_2(c),$$

wohingegen Akteur 3 eine private Bewertung vornimmt. Dieses Szenario stellt demnach eine Situation mit zwei Konkurrenten mit gegensätzlichen Interessen und einer dritten Partei dar (also einen Extremfall eines Interessenkonfliktes). Interessenkonflikte treten z.B. bei Käufer- und Verkäuferbeziehungen oder bei zwei Anbieterkonkurrenten in Verhandlung mit einem Kunden auf. Solche Nullsummenspiele wurden in der Literatur für den bilateralen Fall gründlich untersucht, jedoch im Allgemeinen ohne die Hinzunahme einer dritten Partei. Verhandlungen erweisen sich in diesem Szenario als schwierig, da ein Konsens nicht möglich erscheint.

Evaluationskriterien

Um anwendungsorientierte Lösungen zu gewährleisten, werden an die Verhandlungsprotokolle folgende normative Anforderungen gestellt:

1. Die Protokolle sollten eine möglichst hohe soziale Wohlfahrt erreichen. Wie bereits erwähnt, entspricht die soziale Wohlfahrt dem Pareto-Kriterium und erreicht den besten Durchschnittsnutzen für die beteiligten Agenten.

2. Die Protokolle sollten eine möglichst geringe Rechenzeit aufweisen.
Eine Anwendungstechnologie muss eine effiziente, schnelle Lösung generieren, um einen praxistauglichen Einsatz zu gewährleisten.

3. Die Protokolle sollten das Kriterium der individuellen Rationalität erfüllen.
Die individuelle Rationalität besagt dabei, dass jeder Agent eigennützig handelt und seinen eigenen Nutzen, ohne Berücksichtigung der Nutzen anderer Akteure, zu maximieren versucht.

4. Verhandlungsprotokolle

Im Folgenden werden zwei auktions- und ein verhandlungsbasiertes Protokoll zur Vertragsgestaltung durch Softwareagenten präsentiert. Der Fokus liegt hier auf der auf dem Gebiet der Vertragsverhandlungen neu vorgestellten Methode der kombinatorischen Auktionen. Die weiteren Protokolle dienen dabei als zu vergleichende Referenzprotokolle.

4.1 Auktionsbasierte Protokolle

Innerhalb der auktionsbasierten Protokolle werden Rechte an der Bestimmung der Vertragseigenschaften von einem unabhängigen Auktionator an die Agenten versteigert. Der Auktionator handelt hierbei eigennützig, da der Auktionserlös maximiert wird. Der Auktionserlös ist für den Wert der sozialen Wohlfahrt irrelevant, da er z.B. durch Umverteilung wieder rückvergütet werden kann und somit den Wert der sozialen Wohlfahrt nicht beeinflusst. Erhält ein Agent ein Bestimmungsrecht über eine Eigenschaft, so wird der Agent, wenn er einen positiven Nutzen daraus erhält, sich für die Ausprägung $d_n = 1$ entscheiden, und, sofern er einen negativen Nutzen aus der Eigenschaft ziehen würde, sich für die Ausprägung $d_n = 0$ entscheiden. In Abbildung 2 wird die wechselseitige Beziehung der beteiligten Akteure der Auktion dargestellt.

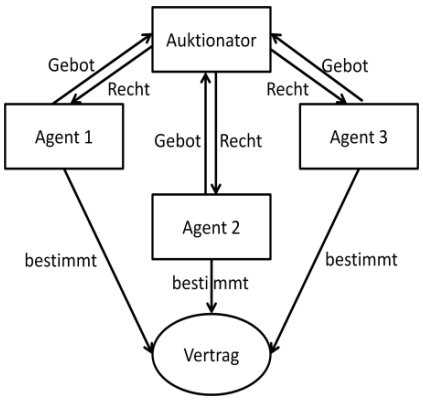


Abbildung 2. Agentenbeziehungen im auktionsbasierten Modell

4.1.1 Kombinatorische Auktionen (KA)

Kombinatorische Auktionen fußen auf der Annahme, dass bei einer Versteigerung verschiedener Güterkombinationen Interdependenzen innerhalb der Kombinationen existieren. Bspw. sind Nutzenwerte von Zeitfenstern auf Flughäfen nicht nur vom

Startfenster, sondern auch vom Landefenster am Zielort abhängig [28]. Ein weiteres Beispiel sind regional getrennte Frequenzspektrumauktionen; verschiedene Kombinationen von lizensierten Regionen – wie z.B. alle Bezirke eines Bundesstaates – sind von größerem Nutzen als dieselbe Anzahl an Bezirken über mehrere Bundesstaaten [2,32].

Da die Vertragseigenschaften im Szenario voneinander abhängig sind, entsteht eine Situation vergleichbar mit den oben genannten Güterkombinationen, wobei nun mit Hilfe von kombinatorischen Auktionen Verträge mit möglichst großer sozialer Wohlfahrt erreicht werden könnten. Ein Merkmal von Auktionen im Vergleich zu Verhandlungen ist ein Anreiz zur Informationspreisgabe. Während Agenten in Verhandlungen keinen ursprünglichen Anreiz haben, die Informationen über ihre Präferenzen, Reservationspreise etc. preiszugeben, wird bei der Auktion durch die Gebote der Agenten – mindestens aus der Perspektive des Auktionators – vielerlei Information preisgegeben.

Durch Implementierung sogenannter Vickrey-Clarke-Groves-Mechanismen (VCG-Mechanismen)[1] wird den Agenten zudem der Anreiz zum strategischen Bieten genommen, da die wahrheitsgemäße Angabe der Zahlungsbereitschaft im Falle der VCG-Mechanismen immer die optimale Strategie für ein Individuum ist und somit die dominante Strategie der Agenten darstellt [23].

Unter der Anwendung eines unparteiischen Auktionators wird das folgende kombinatorische Auktionsverhandlungsprotokoll (KA-P) vorgeschlagen (mathematische Formulierung nach [30]):

Das kombinatorische Auktionsverhandlungsprotokoll (KA-P)

Schritt 1. Die Agenten reichen ihre Gebote auf paarweise Kombinationen der Vertragseigenschaften sowie die allein stehenden Eigenschaften ein, woraus die Menge der maximalen Gebote (ein Gebot pro Kombination) entsteht $B = \{B_1, \dots, B_i, \dots, B_{I-1}, B_I\}$.

Ein Gebot B_i stellt dabei Tupel aus der Menge der enthaltenen Vertragseigenschaften S_i und dem Gebotspreis p_i dar: $B_i = (S_i, p_i)$.

Schritt 2. Der Auktionator maximiert den Ertrag der Auktion, indem er das Winner Determination Problem (WDP) löst:

$$\max_{x_i} \sum_{i=1}^{I} p_i * x_i \quad s.t. \sum_{i|n \in S_i}^{I} x_i \leq 1 \quad \forall\, n \in \{\mathbb{N} | 0 < n \leq N\}$$

Der Auktionator bestimmt die Vickrey-Preise, indem er das WDP für alle Akteure aus $\{\mathbb{N} \,|0 < j \leq J\} \setminus j$ erneut löst und somit die Opportunitätskosten für die Abwesenheit des Bieters *j* bestimmt.

Schritt 3. Der Auktionator vergibt als Ergebnis von Schritt 2 die Rechte an der Bestimmung der Eigenschaften an die jeweiligen Gewinner $winner = \{win_1, \dots, win_n, \dots, win_{N-1}, win_N\}, win_n \in \{\mathbb{N} \,|0 < j \leq J\}$.

Schritt 4. Die Agenten bestimmen die gewonnenen Eigenschaften des Vertrages und formen den abschließenden Vertrag wie folgt[2]:

$$d_n = \begin{cases} 1, & P_{j|j=win_n}(n,n) | \langle d_n = 1 \rangle \geq 0 \\ 0, & P_{j|j=win_n}(n,n) | \langle d_n = 1 \rangle < 0 \end{cases}$$

Die Agenten geben ihre Gebotspreise p_i entweder gemäß ihrem Nutzen aus der Vertragseigenschaft oder den Opportunitätskosten ab (d.h. dem Preis, den sie bereit sind, für die Vermeidung der Eigenschaft zu zahlen). Bspw. hat Agent 1 einen Nutzen für eine Eigenschaft von 15 Geldeinheiten (GE), während Agent 2 für dieselbe Eigenschaft einen Missnutzen von 20 GE hat. Agent 2 wäre bereit, bis zu 20 GE zu zahlen, damit diese Eigenschaft auf $d_n = 0$ gesetzt werden würde, und würde somit das Gebot von Agent 1, das 15 GE beträgt, überbieten. Der zu zahlende Betrag für Agent 2 würde im VCG-Verfahren allerdings nur 15 GE betragen, da dies das höchste Gebot ohne die Teilnahme von Agent 2 wäre. Agent 2 erhält also eine Vickrey-Zahlung in Höhe von 20 GE − 15 GE = 5 GE, da diese die Opportunitätskosten aus der Sicht des Auktionators für die Teilnahme darstellen.

4.1.2 Winner Takes It All (WTIA)

Das Winner-Takes-It-All-Protokoll (WTIA-P), als weiteres auktionsbasiertes Protokoll, soll einen Referenzwert für das KA-P darstellen. Das WTIA-P greift auf keine differenzierte Nutzenkombinatorik zurück, sondern bestimmt lediglich den Akteur mit der größten Zahlungsbereitschaft für den gesamten Vertrag.

Das Winner-Takes-It-All-Protokoll (WTIA-P)

Schritt 1. Die Agenten reichen ihre Gebote $GB = \{p_1, \dots, p_j, \dots, p_{J-1}, p_J\}$ auf den Gesamtvertrag ein. Die Zahlungsbereitschaft für den Gesamtvertrag bestimmt sich aus den Beträgen der Nutzen über alle Eigenschaften. Die Beträge resultieren daraus, dass bei negativem Nutzen die Opportunitätskosten zur Vermeidung einer Eigenschaft berücksichtigt werden.

Schritt 2. Der Auktionator maximiert den Ertrag der Auktion, indem das höchste Gebot ausgewählt wird:

$$\max \sum_{j=1}^{J} p_j * x_j \quad s.t. \sum_{j=1}^{J} x_j \leq 1$$

Der Auktionator bestimmt zudem den Vickrey-Preis, indem er das zweithöchste Gebot bestimmt und die Differenz zum Höchstgebot dem Auktionsgewinner gutschreibt (Vickrey-Zahlung).

Schritt 3. Der Auktionator vergibt als Ergebnis von Schritt 2 die Rechte an der Bestimmung der Vertragseigenschaften an den Gewinner $\widetilde{winner} \in \{\mathbb{N} \,|0 < j \leq J\}$.

Schritt 4. Der Gewinner bestimmt die gewonnenen Eigenschaften des Vertrages und formt somit den abschließenden Vertrag wie folgt:

---

[1] VCG-Mechanismen sind eine Verallgemeinerung der Zweitpreisauktion mit verdeckten Geboten (Vickrey-Auktion) für den Mehrgüterfall.

[2] Ein Risikonutzenkalkül für den theoretisch denkbaren Sonderfall, dass eine nicht selbst bestimmbare Eigenschaft eine größere Substitutionsstrafe als der Nutzen aus der Eigenschaft aufweist, wurde zur Vereinfachung außer Acht gelassen.

$$d_n = \begin{cases} 1, & P_{j|j=\widetilde{winner}}(n,n)|\langle d_n = 1\rangle \geq 0 \\ 0, & P_{j|j=\widetilde{winner}}(n,n)|\langle d_n = 1\rangle < 0 \end{cases}$$

4.2 Verhandlungsbasiertes Protokoll

Das verhandlungsbasierte Protokoll nutzt analog zum Auktionator einen unparteiischen Mediator – wie häufig in der Literatur vorgeschlagen (vgl. bspw. [8,16,19]). In diesem Fall unterbreitet der Mediator Vertragsvorschläge und die Agenten entscheiden über deren Annahme als aktueller Vertragskandidat. Abbildung 3 gibt einen Überblick über das Schema der Agentenbeziehungen.

In einem iterativen Prozess wird der Vertragsraum auf der Suche nach Pareto-effizienten Verträgen überprüft. Die Agenten bestimmen hierbei die Richtung der Suche des Mediators, indem sie durch die Annahme- bzw. Ablehnungsentscheidung Informationen über die relative Güte des Vertragskandidaten preisgeben. Der Vertragskandidat, der zuletzt die Zustimmung aller Agenten erhält, wird als finaler Vertrag c festgelegt.

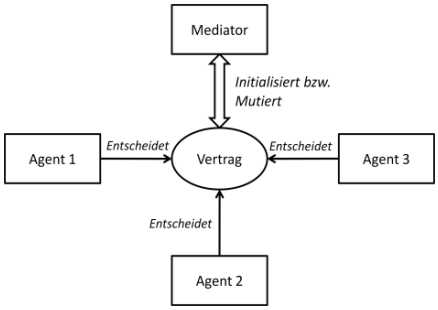


Abbildung 3. Agentenbeziehungen im verhandlungsbasierten Modell

Im Folgenden wird ein generisches Verhandlungsprotokoll basierend auf [16] vorgestellt:

Das generische verhandlungsbasierte Protokoll (GV-P)

Schritt 1. Der Mediator ermittelt zufällig einen Initialvertrag c_0', der als aktuell bester Vertrag c_0^* gesetzt wird[3]. Es gilt $t \leftarrow 0$.

Schritt 2. Der Mediator invertiert im aktuell besten Vertrag c_t^* ein Bit an einer zufällig ermittelten Stelle und erhält somit einen neuen Vertragskandidaten c_t' .

Schritt 3. Die Agenten entscheiden über die Zustimmung zum Vertragskandidaten. Eine Zustimmung durch einen Agenten hat den Wert $Z_j = 1$, während eine Ablehnung den Wert $Z_j = 0$ einnimmt.

Schritt 4. Der Mediator bestimmt den aktuell besten Vertrag:

$$c_{t+1}^* = \begin{cases} c_t', & \sum_{j=1}^{J} Z_j(c_t') = J \\ c_t^*, & \sum_{j=1}^{J} Z_j(c_t') \neq J \end{cases}$$

[3] Für c_t' und c_t^* gelten die gleichen Eigenschaften wie für c.

Sofern das Abbruchkriterium nicht erfüllt ist, beginnt der Mediator bei Schritt 2 mit $t \leftarrow t + 1$ erneut. Das Abbruchkriterium wird als eine bestimmten Anzahl an Wiederholungen $t \in \{\mathbb{N}|0 < t \leq T\} = T$ des Verhandlungsprozesses (Schritt 2 bis 4) festgelegt.

Schritt 5. Sofern das Abbruchkriterium erfüllt ist, legt der Mediator den finalen Vertrag $c = c_T^*$ fest.

Eine Eigenschaft des GV-P ist somit, dass sehr wenige Informationen für den Mediator zur Optimierung notwendig sind und daher auch nur eine geringe Informationspreisgabe von den Agenten verlangt werden muss.

Das GV-P wurde in der Simulation mit zwei verschiedenen Agententypen getestet. Der erste Agententyp agiert dabei opportunistisch, gemäß dem sogenannten Hill Climbing (siehe Abschnitt 4.2.1.), während der zweite Agententyp sich, gemäß dem probabilistischen Akzeptanzkriterium beim Simulated Annealing (siehe Abschnitt 4.2.2.), kooperativ verhält [9,16]. Eine nähere Diskussion zur Realitätsnähe der Verhaltensweisen sowie deren Anreizkompatibilität erfolgt in Kapitel 5.

4.2.1 Hill Climbing (HC)

Die Bezeichnung Hill Climbing (dt. Bergsteigen) bezieht sich auf eine Analogie zum sogenannten Bergsteigeralgorithmus. Der Bergsteigeralgorithmus ist eine iterative Optimierungsheuristik, die stets die beste Verbesserung in der momentanen Nachbarschaft sucht, bis keine weiteren Verbesserungen möglich sind und daher zumindest ein lokales Optimum gefunden ist [22]. Sinngemäß dazu stimmen per Annahme im GV-P, mit dem Agententyp Hill Climber (HC-P), die Agenten nur einem Vertrag zu, der ihnen einen größeren oder identischen Nutzen im Vergleich zum letzten Vertragskandidaten, der die Zustimmung aller Agenten erhalten hat, ermöglicht. Das Verhalten ist also gierig oder opportunistisch.

Die Zustimmungsfunktion in Schritt 3 des GV-P lautet somit für diesen Typus:

$$Z_j(c_t') = \begin{cases} 1, & U_j(c_t') \geq U_j(c_t^*) \\ 0, & U_j(c_t') < U_j(c_t^*) \end{cases}$$

4.2.2 Simulated Annealing (SA)

Die Bezeichnung Simulated Annealing (dt. simulierte Abkühlung) ist ebenfalls wie HC eine Analogie zu einer Metaheuristik. Die SA-Heuristik geht ähnlich wie die HC-Heuristik vor, jedoch ist beim SA eine Bereitschaft vorhanden, eine lokale partielle Verschlechterung einzugehen. Diese Bereitschaft ist abhängig von einer sogenannten Temperatur (τ), die über die Zeit bzw. Verhandlungsrunden gemäß einer Kühlfolge (*cooling schedule*) sinkt. Diese Vorgehensweise soll es ermöglichen, lokale Optima zu überwinden [22]. Beim GV-P mit dem Agententyp Simulated Annealer (SA-P) wurde den Agenten unterstellt, dass sie diese Bereitschaft zur Verschlechterung ebenfalls aufweisen. Somit können die Agenten schlechter stellenden Verträgen zustimmen, was den anderen Agenten gegebenenfalls eine Verbesserung ermöglichen kann (kooperatives Verhalten). Von diesem Verhalten wird eine verbesserte Wohlfahrt gegenüber dem HC-P erwartet, da der Verhandlungsprozesss – wie in der Einleitung gezeigt – sowohl individuelle als auch soziale lokale Optima aufweist.

Die Zustimmungsfunktion in Schritt 3 des GV-P lautet somit für diesen Typus

$$Z_j(c_t') = \begin{cases} 1, \ U_j(c_t') \geq U_j(c_t^*) \ \vee \ e^{-\left(\frac{U_j(c_t^*) - U_j(c_t')}{\tau}\right)} \geq zz_t \\ 0, \ U_j(c_t') < U_j(c_t^*) \ \wedge \ e^{-\left(\frac{U_j(c_t^*) - U_j(c_t')}{\tau}\right)} < zz_t \end{cases},$$

wobei zz_t eine gleichverteilte Zufallszahl aus dem Intervall von 0 bis 1 ist ($zz_t \sim \mathcal{U}(0,1)$).

5. Ergebnisse

Die verschiedenen Protokolle wurden anhand der in Kapitel 3 vorgestellten Kriterien – Leistung, Rechenzeit und Anreizkompatibilität – evaluiert. Hierzu wurde eine Simulationssoftware entwickelt, die die Protokolle implementiert. Zur Lösung des Winner Determination Problem im KA-P wurde eine Schnittstelle zu einem CPLEX-Solver verwendet, während das GV-P über T = 50.000 Verhandlungsrunden simuliert wurde. Während weitere Verhandlungsrunden eventuell bessere Ergebnisse hervorbringen könnten, liegt in diesem Bereich ein angemessener Kompromiss zwischen Rechenzeit und Leistung. Der Vertrag wird als bindend für alle Verhandlungsteilnehmer angenommen, sodass auch Nutzenniveaus im negativen Wertebereich nicht ausgeschlossen sind (aufgrund der Verteilung der Nutzenwerte im Intervall [-100,100]). Tabelle 1 zeigt die Ergebnisse der Simulationen nach 5.000 Simulationswiederholungen, deren Variationskoeffizienten c_v (Mittelwert-normierte Standardabweichung) in Abhängigkeit des wohlfahrtsmaximierenden Vertrages, der ebenfalls mit der Hilfe des CPLEX-Solver ermittelt wurde, sowie deren mittlerer Rang. Der Rang wurde mit allen Simulationsergebnissen als Bezugsgröße ermittelt, beginnend mit dem kleinsten Wert (Rang 1) und endend mit dem größten Wert (Rang 20.000). Aufgrund der Ordinalskala wurde der Median als Kennzahl für den mittleren Rang gewählt. Des Weiteren wurden die Ergebnisse paarweise in einem Wilcoxon-Vorzeichen-Rang-Test auf Gleichheit des Mittelwertes getestet und erreichten dabei ausschließlich sehr hohe Signifikanzniveaus (p-Wert < 1 ‰).

Tabelle 1. Ergebnisse der Protokolle in den drei Szenarien nach 5.000 Simulationswiederholungen in Abh. des optimalen, wohlfahrtsmaximierenden Vertrages

Szenario 1				
	KA	**SA**	**WTIA**	**HC**
Mittelwert	77,6%	70,4%	53,1%	39,7%
c_v	10,2%	16,8%	24,5%	44,8%
Rang (Median)	16103,5	13197,5	7026	3377,5

Szenario 2				
	KA	**SA**	**WTIA**	**HC**
Mittelwert	76,4%	29,8%	36,5%	-49,2%
c_v	82,2%	1088,3%	730,7%	-1622,6%
Rang (Median)	13921	9963,5	9963,5	5729,5

Szenario 3				
	KA	**SA**	**WTIA**	**HC**
Mittelwert	56,7%	38,8%	42,0%	-6,0%
c_v	24,5%	42,8%	120,0%	-441,7%
Rang (Median)	14330,5	104385	17214,5	3831

Das KA-P erweist sich anhand des Mittelwertes in allen drei Szenarien, sowohl bei privater als auch bei gemischter Bewertung sowie bei gegensätzlicher Bewertung, als leistungsstärkstes Protokoll. Das arithmetische Mittel liegt bei 77,6%, 76,4% bzw. 56,7% des optimalen Wertes. Gleichzeitig weisen die Ergebnisse das KA-P die geringste Streuung auf (c_v: 10,2%, 82,2% bzw. 24,5%). Das Benchmarkprotokoll WTIA-P erreicht allerdings in Szenario 3 anhand des mittleren Ranges ein besseres Ergebnis als das KA-P (17214,5 vs. 14330,5), im arithmetischen Mittel übertrifft das KA-P das WTIA-P (56,7% vs. 42,0%). Aufgrund dieser Diskrepanz zwischen Rang und Mittelwert wurde ein Wilcoxon-Mann-Whitney-Test (Wilcoxon-Rangsummentest) durchgeführt, der jedoch keine ausreichende Signifikanz für eine unterschiedliche Tendenz zeigen konnte. Wahrscheinlich wirken der hohe Variationskoeffizient ($c_v > 1$) und die große Anzahl an negativen statistischen Ausreißern hierbei verzerrend auf den Mittelwert des WTIA-P. Der Wilcoxon-Vorzeichen-Rang-Test, der im Gegensatz zum Wilcoxon-Rangsummentest nicht nur das Vorzeichen der Differenz, sondern auch deren Höhe in Betracht zieht, liefert beim Test der Alternativhypothese, dass der Mittelwert von KA größer als der Mittelwert von WTIA ist, jedoch ein sehr hohes Signifikanzniveau (p-Wert < 1 ‰).

Der Agententyp SA erreicht stets bessere Ergebnisse als der Agententyp HC, bei gleichzeitig geringerer Varianz. Das HC-P erreicht in Szenario 2 und 3 geradezu desaströse Ergebnisse mit negativer sozialer Wohlfahrt (−49,2% bzw. −6,0%) und erreicht einen Variationskoeffizienten von −441,7% bzw. −1622,6%, d.h., die Standardabweichung ist über viermal bzw. sechszehnmal größer als der Mittelwert. Das SA-P erreicht in Szenario 1 gute Ergebnisse (70,4%), ist jedoch in Szenario 2 und 3 beiden auktionsbasierten Protokollen – teilweise deutlich – unterlegen (29,8% bzw. 38,8%). Bei genauer Betrachtung ist das Abschneiden des GV-P nicht überraschend. Die gegensätzliche Bewertung in Szenario 3 führt dazu, dass beim HC-P kein Vertrag nach dem Initialvertrag des Mediators angenommen wird, da stets mindestens einer der Agenten den Vertrag ablehnt.

Beim SA-P werden diese Verträge nur aufgrund der Annealing-Funktion angenommen. Ein ähnliches Problem stellt sich in Szenario 2, in dem die kleine Größe des Vertragsvektors schneller zum Verharren in einem lokalen Maximum führt. Die Vertragsgröße hat nur einen sehr geringen Einfluss auf die Ergebnisse des KA-P, lediglich die Varianz steigt. Die gegensätzliche Bewertung stellt zwar im Vergleich ein größeres Hindernis dar, die Ausmaße fallen jedoch ebenfalls relativ gering aus. Die Abbildungen 4 bis 6 zeigen eine grafische Übersicht der Leistungsergebnisse in Prozent des optimalen Vertrages.

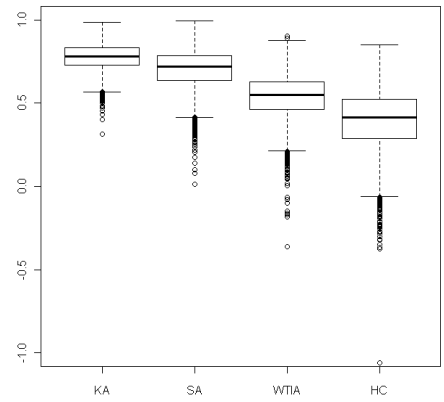


Abbildung 4. Boxplots der Ergebnisse – Szenario 1

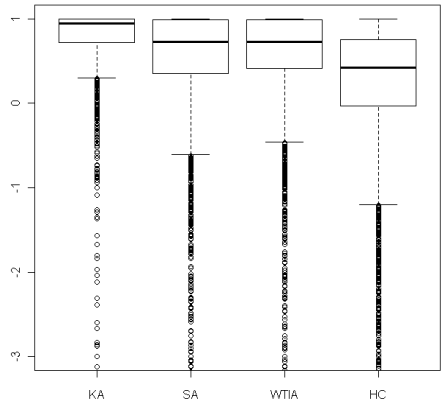


Abbildung 5. Boxplots der Ergebnisse – Szenario 2 (Ausschnitt)

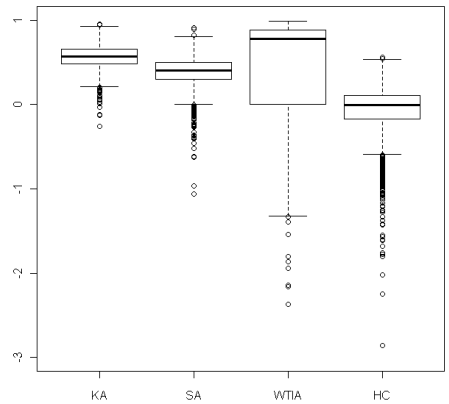


Abbildung 6. Boxplots der Ergebnisse – Szenario 3 (Ausschnitt)

Die Auswertung der Laufzeit zeigt ein heterogenes Bild. Hierzu wurden testweise 100 Simulationen mit verschiedenen Vertragsgrößen und Agentenzahlen für das Szenario 1 durchgeführt. Dabei wurde die durchschnittliche Rechenzeit gemessen. Tabelle 2 zeigt die Ergebnisse für das KA-P im Vergleich mit dem SA-P. Das HC-P erreicht vergleichbare Ergebnisse wie das SA-P (marginal schneller), während das WTIA-P im Laufzeittest äußerst schnell war und in keinem Fall eine bedeutende Laufzeit benötigte.

Tabelle 2. Rechenzeit des KA-P im Vergleich (SA-P = 100%)

	J=3	***J=6***	***J=9***
N=50	40%	46%	49%
N=100	25%	32%	37%
N=150	29%	36%	46%

Das KA-P (mit Hilfe von CPLEX zur Lösung des WDP) benötigte für die gewählten Parametrisierungen weniger als die Hälfte der Laufzeit des SA-P in der Simulation. Die Ermittlung der relevanten maximalen Gebote pro Eigenschaftskombination ($\frac{(N*N)+N}{2}$ Gebote) sowie die VCG-Mechanismen führen dazu, dass das KA-P bei einem Anstieg der Anzahl der Agenten eine längere relative Rechenzeit aufweist. Die Vergrößerung der Vertragslänge führt zuerst zu einer Abnahme der relativen Rechenzeit, anschließend jedoch zu einem Anstieg. Ein weiterer Anstieg der relativen Rechenzeit bei einer fortlaufenden Vergrößerung der Vertragslänge ist zu erwarten.

Bei der Laufzeitanalyse ist anzumerken, dass das GV-P ein iterativer Prozess mit skalierbarer Laufzeit darstellt. Die willkürlich gewählte Länge von 50.000 Verhandlungsrunden erwies sich im Experiment als angemessener Kompromiss zwischen Laufzeit und Leistung.

Beim KA-P sind insbesondere die VCG-Mechanismen sehr laufzeitintensiv, da diese das Winner Determination Problem mehrmals wiederholen und die Anzahl der Wiederholungen proportional zur Anzahl der Agenten steigt. Allerdings sind die VCG-Mechanismen unerlässlich für eine wahrheitsgetreue Gebotsabgabe.

Das KA-P wird somit bei einer großen Anzahl an Agenten schnell laufzeitintensiv und ist somit nicht für Anwendungsprobleme wie verteiltes Rechnen (außerhalb von kleinen Agentenzahlen wie in Forschungsnetzwerken) oder Routing geeignet. Wie bereits in der Zielsetzung in Kapitel 1 erläutert, soll das KA-P jedoch ein Protokoll für Anwendungen wie betriebliche Planung und Entscheidungsfindung, Smart Grids, Smart Houses oder das Internet der Dinge darstellen. Diese Problembereiche sind in der Regel durch eine überschaubare Anzahl an Agenten gekennzeichnet. Verhandlungen finden meist nur zwischen einer Teilmenge der Gesamtzahl an Agenten oder bilateral mit einem zentralen Steuerungsagenten statt. Sehr viel Rechenzeit musste für die Übermittlung der Probleminstanz an den CPLEX-Solver verwendet werden, während der Solver im Regelfall für das Lösen des Winner Determination Problem auch bei 200 Vertragseigenschaften weit weniger als eine Sekunde benötigte. Es existieren zudem, wie bereits in Kapitel 2 angesprochen, leistungsfähige Approximationsalgorithmen mit moderaten Laufzeitanforderungen, die eine übermäßige Laufzeitzunahme verhindern könnten.

Bezüglich der Anreizkompatibilität sind die auktionsbasierten Protokolle, KA-P und WTIA-P, aufgrund der VCG-Mechanismen individuell rational. VCG-Mechanismen, als Verallgemeinerung der Vickrey-Auktion (Zweitpreisauktion) für den Mehrgüterfall, führen aufgrund ihrer Eigenschaft als dominante Strategie zur wahrheitsgemäßen Offenlegung der Zahlungsbereitschaft und somit zu einer effizienten Allokation oder in diesem Fall zu einer effizienten Rechtevergabe [35].

Im Gegensatz dazu weist das SA-P keine Anreizkompatibilität auf. Bisher wurde angenommen, dass die Agenten entweder dem Typ HC oder dem Typ SA angehören. Sobald diese Festlegung jedoch aufgelöst wird und eine freie Wahlmöglichkeit gegeben wird, zeigt sich die fehlende Anreizkompatibilität. Tabelle 3 repräsentiert eine Auszahlungsmatrix bei einem Zwei-Agenten-Spiel. Es kann ein Agent dabei entweder auf denselben Agententyp ({HC;HC} bzw. {SA;SA}) oder einen Agenten anderen Typs ({HC;SA} bzw. {HC;SA}) treffen.

Das Nash-Gleichgewicht repräsentiert das strategische Gleichgewicht und führt zu einer individuell rationalen, stabilen Lösung, die somit anreizkompatibel ist. Das Nash-Gleichgewicht wird wie folgt definiert: „Ein Nash-Gleichgewicht ist eine Strategieprofil a^* mit der Eigenschaft, dass Spieler i sich nicht durch die Wahl einer Strategie verschieden von a_i^* verbessern kann, sofern Spieler j an der Strategie a_j^* festhält." ([26], S. 22, Originalzitat auf Englisch)

Das Spiel in Tabelle 3 stellt ein Gefangenendilemma dar. Das Nash-Gleichgewicht ist das Strategieprofil {HC;HC} und nicht das wohlfahrtsmaximierende Profil {SA;SA}. Des Weiteren ist das Nash-Gleichgewicht sogar das Strategieprofil mit der niedrigsten sozialen Wohlfahrt. HC kann man stellvertretend auch als opportunistisch und SA als kooperativ ansehen. Sofern der Konterpart sich kooperativ zeigt, ist eine Abweichung von der kooperativen Strategie zur opportunistischen lohnend (100% → 132%). Der andere Agent hat dann – und bereits vorher in {SA;SA} bei simultaner Wahl – ebenfalls einen Anreiz, zum Opportunismus zu wechseln (35% → 78%/74%). Im Strategieprofil {HC;HC} hat keiner der Agenten einen Anreiz, zur Kooperation (SA) zu wechseln, da er sich somit schlechter stellen würde (78%/74% → 35%).

Tabelle 3. Die beiden Agententypen im Vergleich (Nutzenwerte in Relation zum wohlfahrtsmaximierenden Vertrag – Mittelwerte über 250 Simulationen) [4]

		Agent 2	
		HC	SA
Agent 1	HC	72% 78% ; 74%	80% 132% ; 35%
	SA	80% 35% ; 132%	95% 100% ; 100%

Schema: [Soziale Wohlfahrt] [Agent 1] ; [Agent 2]

Zwar weist das SA-P folglich keine Anreizkompatibilität auf, es existieren jedoch in der Literatur verschiedene Ansätze, um dies zu beheben: So schlägt Klein et al. [16] einen „abkühlenden" Mediator vor, der jedoch signifikant schlechtere Ergebnisse hervorbringt als „abkühlende" Agenten. Ein weiterer Lösungsansatz wurde mit stärker oder schwächer gewichteten Zustimmungen bzw. Ablehnungen zu einem Vertrag vorgeschlagen, was allerdings wiederum einen Anreiz zu unwahrheitsgemäßen Verhalten seitens der Agenten bietet. Um dies zu beheben, wurde eine bestimmte Quote für jede der Gewichtungsmöglichkeiten gefordert. Ähnlich argumentiert auch Fink [7], der eine Mindestakzeptanzquote für die Agenten fordert, beruhend auf der Annahme, dass die Agenten daraufhin die Strategie SA in einer Form wählen, dass die geforderte Akzeptanzhäufigkeit genau erreicht wird. Abschließend lässt sich feststellen, dass alle Versuche, eine Anreizkompatibilität für den Agententyp SA durch Veränderung des Verhandlungsprotokolls herzustellen, im besten Fall zu Ergebnissen vergleichbar mit denen des SA-P führen können. Das SA-P stellt die obere Schranke dar, da das Protokoll sich nicht durch Mechanismen zur Anreizkompatibilität verbessern kann. Vielmehr ist von einer mindestens marginalen Verschlechterung, sowohl in Hinblick auf Leistung als auch auf Rechenzeit, aufgrund dieser Mechanismen auszugehen.

6. Fazit und Ausblick

Die Forschungsarbeit schlägt ein neuartiges kombinatorisches Auktionsverhandlungsprotokoll zu automatisierten Verhandlungen innerhalb bspw. der betrieblichen Planung, beim E-Commerce und im Zusammenhang mit intelligenten Multi-Agenten-Systemen vor und evaluiert dieses. Hierzu wurde ein Forschungsszenario mit drei konkreten Szenarioeinstellungen entworfen und das vorgeschlagene Protokoll sowie

[4] Die Tatsache, dass die soziale Wohlfahrt in diesem Zwei-Spieler-Spiel nicht dem Mittelwert der beiden Individualnutzen entspricht, und dass die Individualnutzen über 100% des wohlfahrtsmaximierenden Optimums erreichen, soll anhand eines Zahlenbeispiels erläutert werden: Agent 1 und 2 erhalten einen Nutzen in Höhe von 10 bzw. 5, wohingegen sie im Wohlfahrtoptimum 13 und 3 erhalten würden. Sie erhalten somit einen Nutzen in Abhängigkeit vom Optimum in Höhe von 76,9% bzw. 166,67%. Die Wohlfahrt der beiden Agenten beträgt allerdings nur 93,75%.

Benchmarkprotokolle auf Leistung, Rechenaufwand und Anreizkompatibilität untersucht. Aus dieser Untersuchung ergab sich, dass das vorgeschlagene Protokoll eine sehr gute soziale Wohlfahrt gewährleisten kann und zudem anreizkompatibel ist. Als Nachteil zeigt sich jedoch ab einer bestimmten Vertragsgröße respektive einer bestimmten Anzahl an Verhandlungsteilnehmern eine erhöhte Rechenzeit. Die Anwendungsmöglichkeiten können somit nicht um Massenumgebungen, charakterisiert durch eine große Anzahl an Akteuren (wie beim Routing), erweitert werden.

Ein kombinatorisches Auktionsrahmenwerk innerhalb von automatisierten Verhandlungen erweist sich als interessantes und leistungsstarkes Instrument und ermutigt zu zukünftiger Forschung. Daher werden sich künftige Arbeiten unter anderem auf die Implementierung von Heuristiken für das Winner Determination Problem erstrecken, um eine schnellere Rechenzeit zu erreichen. Des Weiteren soll eine anreizkompatible Variante des generischen Verhandlungsprotokolls mit einem Simulated-Annealing-Agententyp entwickelt und integriert werden. Zudem soll die Untersuchung auf weitere, nicht-binäre Vertragseigenschaften erweitert werden wie bspw. Vorgänger-Nachfolger-Beziehungen oder bestimmte Positionen in einer Abfolge (Fertigungsplanung). Schließlich sollen weitere Szenarioeinstellungen mit vier oder mehr Agenten, unterteilt in Gruppen mit teilweise ähnlichen Nutzenwerten, entworfen werden. Dies soll Aufschluss geben, inwiefern Koalitionsbildung einen Einfluss auf das Verhandlungsergebnis haben kann und wie das kombinatorische Auktionsprotokoll dadurch beeinflusst wird.

7. Literatur

[1] Asimov, I. 1982. *Asimov's Biographical Encyclopedia of Science and Technology*. 2. Auflage, Doubleday, Garden City, NY, USA.

[2] Brunner, C., Goeree, J.K., Holt, C.A. und Ledyard, J.O. 2009. An Experimental Test of Flexible Combinatorial Spectrum Auction Formats An Experimental Test of Flexible Combinatorial Spectrum Auction Formats. *Working Paper* No. 431, Zürich.

[3] Bulow, J. und Klemperer, P. 1996. Auctions Versus Negotiations. *The American Economic Review* 86, 1, 180-194.

[4] Collins, J., Gini, M. und Mobasher, B. 2002. Multi-Agent Negotiation using Combinatorial Auctions with Precedence Constraints. *Technical Report* 02-009, Minneapolis, MN, USA.

[5] Conitzer, V. 2010. Making Decisions Based on the Preferences of Multiple Agents. *Communications of the ACM* 53, 3, 84-94.

[6] Cramton, P., Shoham, Y. und Steinberg, R. 2007. An Overview of Combinatorial Auctions. *ACM SIGecom Exchanges* 7, 1.

[7] Fink, A. 2003. Automatisierte Koordinationsmechanismen zur Ausgestaltung komplexer Verträge im Rahmen mediatisierter Verhandlungsprozesse. In W. Uhr, W. Esswein, E. Schoop (Hrsg.), *Wirtschaftsinformatik 2003, Band I, Medien - Märkte - Mobilität*, Springer, Berlin/Heidelberg, 281-301.

[8] Fink, A. 2006. Supply Chain Coordination by Means of Automated Negotiations Between Autonomous Agents. In B. Chaib-draa, J. Müller (Hrsg.), *Multiagent based Supply Chain Management* (Studies in Computational Intelligence, Vol. 28), Springer, Berlin (2006), 351-372.

[9] Fink, A. 2007. Barwertorientierte Projektplanung mit mehreren Akteuren mittels eines verhandlungs basierten Koordinationsmechanismus. In A. Oberweis, C., Weinhardt, H. Gimpel, A. Koschmider, V. Pankratius, B. Schnizler (Hrsg.), *eOrganisation: Service-, Prozess-, Market-Engineering, Band 2*, Universitätsverlag Karlsruhe, 465-482.

[10] Hattori, H., Klein, M. und Ito, T. 2007. Using Iterative Narrowing to Enable Multi-Party Negotiations with Multiple Interdependent Issues. *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems*.

[11] Homberger, J. 2010. Decentralized Multi-level Uncapacitated Lot-sizing by Automated Negotiation. *4OR: A Quarterly Journal of Operations Research* 8, 2, 155-180.

[12] Jennings, N.R., Faratin, P., Lomuscio, A., Parsons, S., Sierra, C. und Wooldridge, M. 2001. Automated Negotiation: Prospects, Methods and Challenges. *Group Decision and Negotiation* 10, 2, 199-215.

[13] Kalai, E. and Smorodinsky, M. 1975. Other Solutions to Nash's Bargaining Problem. *Econometrica* 43, 3, 513-518.

[14] Karnouskos, S. und Holanda, T.N. 2009. Simulation of a Smart Grid City with Software Agents. *Third UKSim European Symposium on Computer Modeling and Simulation*, 424-429.

[15] Katasonov, A., Kaykova, O., Khriyenko, O., Nikitin, S. und Terziyan, V. 2008. Smart Semantic Middleware for the Internet of Things. *Proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics (ICINCO'08)*, 169-178.

[16] Klein, M., Faratin, P., Sayama, H. und Bar-Yam, Y. 2007. Negotiating Complex Contracts. *MIT Sloan Working Paper* No. 4196-01, Cambridge, MA.

[17] Kraus, S. 2001. Automated Negotiation and Decision Making in Multiagent Environments. In M. Luck, V. Marik, O. Stepankova and R. Trappl (Hrsg.), *Mutli-agents Systems and Applications*. Springer, New York, NY, USA, 150-172.

[18] Lai, G. und Sycara, K. 2008. A Generic Framework for Automated Multi-attribute Negotiation. *Group Decision and Negotiation* 18, 2, 169-187.

[19] Lai, G., Li, C. und Sycara, K. 2006. Efficient Multi-Attribute Negotiation with Incomplete Information. *Group Decision and Negotiation* 15, 5, 511-528.

[20] Lehmann, D., Müller, R. und Sandholm, T. 2006. The Winner Determination Problem. In P. Cramton, Y. Shoham and R. Steinberg (Hrsg.), *Combinatorial Auctions*. The MIT Press, Cambridge, MA, USA, 297-318.

[21] Lomuscio, A., Wooldridge, M. und Jennings, N.R. 2003. A classification scheme for negotiation in electronic commerce. *Group Decision and Negotiation* 12, 1, 31–56.

[22] Michalewicz, Z. und Fogel, D.B. 2004. *How to Solve It: Modern Heuristics*. 2. Auflage, Springer, Berlin/Heidelberg, Germany.

[23] Milgrom, P.R. 2004. *Putting auction theory to work*. 1. Auflage, Cambridge University Press, Cambridge, MA, USA.

[24] Nash, J.F. 1950. The Bargaining Problem. *Econometrica* 18, 2, 155-162.

[25] Nash, J.F. 1951. Non-Cooperative Games. *Annals of Mathematics* 54, 2, 286-295.

[26] Osborne, M.J. 2010. *An Introduction to Game Theory*. International Edition, Oxford University Press, New York, NY, USA.

[27] Pekec, A. und Rothkopf, M.H. 2003. Combinatorial Auction Design. *Management Science* 49, 11, 1485-1503.

[28] Rassenti, S.J., Smith, V.L. und Bulfin, R.L. 1982. A Combinatorial Auction Mechanism for Airport Time Slot Allocation. *The Bell Journal of Economics* 13, 2, 402-417.

[29] Rubinstein, A. 1982. Perfect Equilibrium in A Bargaining Model. *Econometrica* 50, 1, 97-110.

[30] Sandholm, T., Suri, S., Gilpin, A. und Levine, D. 2002. Winner Determination in Combinatorial Auction Generalizations. *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 69.

[31] Sandholm, T. 2006. Optimal Winner Determination Algorithms. In P. Cramton, Y. Shoham and R. Steinberg (Hrsg.), *Combinatorial Auctions*. The MIT Press, Cambridge, MA, USA, 337-368.

[32] Scheffel, T., Ziegler, G., Bichler, M. und Jacob, R. 2008. Selling Spectrum Licenses via Combinatorial Auctions: An Experimental Analysis of Bidding Strategies. *Working Paper* MS-0001-1922.65, München.

[33] Ströbel, M. und Weinhardt, C. 2003. The montreal taxonomy for electronic negotiations. *Group Decision and Negotiation* 12, 2, 143–164.

[34] Varian, H.R. 2010. *Intermediate Microeconomics: A Modern Approach*. 8. Internationale Auflage, W. W. Norton & Company, New York, NY, USA.

[35] de Vries, S., Schummer, J. und Vohra, R. 2007. On ascending Vickrey auctions for heterogeneous objects. *Journal of Economic Theory* 132, 1, 95-118.

User Acceptance of 'Smart Products': An Empirical Investigation

Peter Mayer
University of St. Gallen
Dufourstrasse 40a
CH-9000 St. Gallen
+41 71 224 7248
peter.mayer@unisg.ch

Dirk Volland
University of St. Gallen
Dufourstrasse 40a
CH-9000 St. Gallen
+41 71 224 7240
dirk.volland@unisg.ch

Frédéric Thiesse
University of Würzburg
Josef-Stangl-Platz 2
D-97070 Würzburg
+49 931 31 80242
frederic.thiesse@uni-wuerzburg.de

Elgar Fleisch
University of St. Gallen
Dufourstrasse 40a
CH-9000 St. Gallen
+41 71 224 7240
elgar.fleisch@unisg.ch

ABSTRACT

Smart Products pose a new class of IT artifacts based on sensors, ID-tags, haptic user interfaces, and other technologies usually subsumed under the notion of 'ubiquitous computing'. Such devices differ in many ways from traditional computers, e.g., with regard to their physical shape, computing power, and interaction paradigms. While a substantial body of literature already exists on underlying technological design challenges, only few researchers have attempted to quantitatively explore factors influencing user acceptance of Smart Products. Against this background, the present study is concerned with the use of Smart Products in a kitchen environment. Based on the Unified Theory of Acceptance and Use of Technology (UTAUT), we develop and empirically test a structural model of technology acceptance including five moderating factors. Our results indicate high overall acceptance of the proposed scenarios, corroborate the applicability of the UTAUT model for smart home environments, and confirm significant effects for two moderators.

Keywords

Smart Products, UTAUT, Pervasive computing/ubiquitous computing, Moderating Effect, Technology acceptance model

1. INTRODUCTION

'Ubiquitous Computing' [47], 'Pervasive Computing' [38], 'Things that think' [23], 'Ambient Intelligence' [1], 'Silent Commerce' [16] – a plethora of novel terms has evolved in recent years that propagate the coming of a new paradigm shift in information processing. Common to all these concepts is the shared vision of a future world of everyday physical objects equipped with digital logic, sensors, and networking capabilities [18]. Drivers behind the ongoing trend towards this vision are both miniaturization of microelectronic components and price decline as well as various new technologies reaching mass-market maturity, e.g., in the area of polymer electronics or wireless networks. On the one hand, these so-called 'Smart Products' allow manufacturing companies to differentiate themselves from their competitors by enriching physical items with digital functionality. On the other hand, the linkage of products with services in the Internet allows for the creation of novel product-service bundles that not only generate a continuous stream of additional revenues but also hold the potential to support new product development by providing companies with valuable information on their products' usage and to strengthen customer relationships.

Smart Products have become a fruitful research area on the interface of electrical engineering, computer science, and information systems. While a substantial body of literature on the associated design challenges (e.g., middleware architectures, multi-modal user interfaces, ad-hoc networking protocols) already exists, only few authors have so far conducted behaviorist research on the factors influencing user acceptance of this new class of IT artifacts. Smart Products differ in many ways from traditional computers, be it desktop PCs or mobile devices, which makes transferability of results from prior technology acceptance research seem questionable. For example, haptic user interfaces on the basis of gesture detection or acceleration sensors provide product owners with an entirely different user experience than the classical desktop environments known from today's graphical user interfaces. On the other hand, computing devices that merge with the physical world – and thus become 'invisible' to certain degree – might also lead to entirely different perceptions of IT than their classical counterparts and even evoke negative reactions to the point of fears from technology paternalism and ubiquitous surveillance [42].

It is against this background that the present study is concerned with the acceptance of Smart Products by end users. For this purpose, we consider the example of a 'smart kitchen' environment that encompasses a number of household appliances and associated digital services that are supposed to support their owner in everyday activities, such as preparing a meal. Based on the 'Unified Theory of Acceptance and Use of Technology' (UTAUT) proposed by Venkatesh et al. [46] and other prior research on technology acceptance, we develop and empirically test a structural model for the explanation and prediction of the users' intention to use a Smart Product. Our sample includes 166 responses to an online questionnaire covering five distinct application scenarios. This research contributes to the IS literature in two ways. First, we investigate the applicability of the UTAUT model to the domain of Smart Products and confirm its explanatory power for this new class of IT artifacts. Second, we extend the base model by five moderating factors and show that two of these play a significant role in varying acceptance behavior between different user groups. From a practical perspective, our

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

results indicate a generally positive perception of Smart Products by potential users and allow for drawing a number of managerial implications.

The remainder of the paper is organized as follows. We first provide an overview of the concept and the technologies underlying Smart Products. We continue with a review of related works on technology acceptance in general and smart products adoption in particular. Based on this review, we then develop our research model and formulate a set of hypotheses to be tested. Fourth, we describe our research methodology including survey design, data collection, and statistical analysis. The paper closes with a discussion of theoretical and managerial implications, limitations, and suggestions for further research.

2. Technological Background

Research on Smart Products is still scattered across different research streams covering aspects of technology and management. As a consequence, there is no unified definition of the term 'Smart Product' and different notions exist depending on the respective research perspective or application area. What can be said is that Smart Products denote an emerging class of products, which integrate different facets of Ubiquitous Computing technologies in order to provide a richer user experience particularly through connectivity to other products and proactive behavior. Smart Products possess capabilities to act jointly, complement each other, and thus establish a smart environment that goes beyond the isolated functionalities provided by conventional products. Motivated by various technological advances, a number of researchers have already considered this upcoming research issue in the past 20 years. An early proponent was Dhebar [15] who defines Smart Products as "physical products that have IT incorporated in them". In a similar way, Maass and Janzen [30] describe them as "hybrids of physical products and information products". However, as most electronic products today incorporate some kind of IT (e.g., microprocessors embedded in several household devices or cars), this definition is not sufficient for delimiting the scope of Smart Products. Allmendinger and Lombreglia [5] extend this conception by introducing the similar concept of "built-in product intelligence" as a combination of awareness and connectivity, which allows for creating a smart services portfolio around a product, particularly in an industrial setting. Examples for such services are remote maintenance, feature upgrades, or pay-per-use business models.

Smart Products are characterized by the fact that they make use of specific technologies and design principles mainly from the Ubiquitous Computing domain, in order to sense and communicate information about themselves, their condition, and the environmental context around them [18]. This real-time context awareness grants them the capability to act proactively with regard to internal state and context, adapt to different situations, interact with other Smart Products, and convey information across lifecycle boundaries. With respect to prior research on Ubiquitous Computing, Smart Products can be regarded as real-world manifestations of 'calm technologies' as formulated by Weiser [47], who envisioned "a physical world richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives and connected through a continuous network".

Aitenbichler et al. [4] summarize the idea of Smart Products by stating that "Smart Products are real-world objects, devices, or software services bundled with knowledge about themselves, others, and their embedding". This definition sets the focus on the knowledge dimension and emphasizes the capability of autonomous behavior depending on context information. Other contributions define Smart Products by listing a number of characteristic constituents. Rijsdijk and Hultink [35], for instance, postulate that seven dimensions determine the smartness of a product: autonomy, adaptability, reactivity, multi-functionality, the ability to cooperate, humanlike interaction, and personality. They also point out that the smartness of products is a broad continuum that is determined by the extent to which the seven dimensions are fulfilled, which leads to physical objects that "share the ability to collect, process, and produce information and can be described as 'thinking' for themselves".

From a technological perspective three aspects constitute Smart Products: First, network technologies such as Wi-Fi, Bluetooth, UMTS, and Auto-ID technologies such as RFID enable Smart Products to communicate with each other so that a smart environment can emerge. Second, sensors are required to capture the user context, which enables Smart Products to adapt to the user situation and act proactively and in a smart way. Third, sufficient computing power is required to execute smart behavior. To achieve smart behavior with limited computational resources, a number of research programs and initiatives have recently been started, which majorly focus on three aspects [4][6][40]: a first technology-oriented research stream explores the application of semantic modeling of context, product behavior, and interaction. A second stream is concerned with new techniques for superior human-computer interaction as not all smart products will be able to include conventional screen-based user interfaces. Third, researchers are exploring new middleware architectures that are tailored to the specific needs of Smart Products development. Such middleware is designed to connect the Smart Product to internal and external sensors and actuators, to establish communication to other Smart Products and back-end services, and to establish a programming platform that hides the details of a plethora of existing embedded technology stacks.

3. Related Work

Our study focuses on Smart Products in home environments. Strictly speaking, we investigate user acceptance towards a smart kitchen environment that consists of five functional scenarios. In this section we shortly review the literature on the theoretical foundations of our research as well as academic and industrial activities that relate to applications in the home appliances domain.

Research on user acceptance of information technology originates from different theoretical disciplines such as psychology, sociology, and information systems. Various alternative approaches have been proposed to analyze the acceptance and use of a new technology. The majority of technology acceptance models are based on the Theory of Reasoned Action (TRA) [17]. TRA posits that an individual behavioral intention towards a specific behavior can be considered as a proxy of the behavior itself [46]. The Technology Acceptance Model (TAM) [14] has become the most prevalent model for studying user acceptance in the field of information technology. TAM includes two major predictors of the dependent variable *Behavioral Intention*, which

TRA assumes to be closely linked to actual behavior: *Perceived Ease of Use* and *Perceived Usefulness*. More recently, the Unified Theory of Acceptance and Use of Technology (UTAUT) [46] has been proposed, which integrates TAM and the more advanced TAM2 with other technology acceptance research streams. UTAUT represents a parsimonious but still comprehensive framework to provide an understanding of factors that affect technology acceptance, and could be confirmed in a large number of research works (see [43] for a review).

Regarding empirical acceptance studies, there is only a relatively small number of prior studies investigating user acceptance of Ubiquitous Computing and related concepts. Garfield [21] presents results from a longitudinal, qualitative study of the acceptance of Tablet PCs based on interview data from four industries. Main findings include a list of factors that influence the predictors of *Behavioral Intention* in the UTAUT model as well as the identification of the technology's impact on work processes. Sheng et al. [41] studied interaction effects of personalization and context on intention to adopt. They conclude that increasing personalization raises privacy concerns, and the degree of this relationship is moderated by situational context.

Whereas these studies analyze various manifestations of the Smart Product concept, contributions on smart home environments in particular are rather scarce. Vastenburg [44] investigate in a simulated environment, to which degree consumers appreciate home automation applications. They conclude that, in general, consumers have a positive attitude towards home automation. Key success factors for home automation applications are *Ease of Use* and *Predictability*, the latter meaning that consumers understand and foresee the behavior of the system. After evaluating user acceptance of an intelligent thermostat control, Freudenthal and Mook [20] conclude that users carefully weigh benefits and drawbacks of new technologies. Major drawbacks are the difficulty to operate, the insufficient level of control, and privacy concerns, whereas usability is of utmost importance for user acceptance.

With regard to smart kitchen environments, previous studies focused only on a limited number of constituents. So far, research in the kitchen environment has mainly focused on nutrition [24], recipe planning [26], or communications [8]. Although having tested early prototypes with users, these studies are not based on the analysis of larger samples. The only exception we are aware of is a user acceptance study by Rothensee [37] concerning a simulated 'smart fridge', which offers various assistance functions (product information, automatic replenishment, recipe planner). The results indicate that *Perceived Usefulness* is the strongest predictor to *Behavioral Intention*, followed by emotional response to the product. The role of moderating factors (gender, technological competence, sense of presence in a simulation) could not be supported.

4. Research Model

In this section, we describe the research model underlying the study as depicted in Figure 1. Our research objective is to analyze the user acceptance of a 'smart kitchen' as an example of a Smart Product environment in the home appliance domain. The most obvious choice regarding the theoretical framework for a study like ours seems to be the classical TAM, which has been used as the foundation for several IT acceptance studies in recent years. For the present study however, TAM may have only limited ability to explain smart products acceptance because it neglects the social context in which a technology is being adopted. We consider the social context to be highly important, because smart kitchen appliances are targeting at the consumers' kitchens and homes. For this reason, we decided to construct and test a research model on the foundation of the more advanced UTAUT framework and its constructs as proposed by Venkatesh et al. [46].

Whereas UTAUT has served as the theoretical foundation to many analyses, particularly in industrial settings, it has not yet been applied specifically to smart environments in the domestic domain. Further, moderator variables proposed in the original model are not specifically targeted to the typically voluntary use of the investigated application in the private domain. While basic technology acceptance models have largely matured, the investigation of moderating effects to understand external factors that influence adoption decisions is still under-developed and needs to be further elaborated [13][43]. We intend to fill this two-fold research gap by applying the UTAUT model to the case of a smart kitchen environment and by introducing additional moderating variables to capture consumer traits and external factors that may influence adoption decisions.

The original UTAUT model posits that four independent variables determine an individual's intention to use a technology: *Performance Expectancy*, *Effort Expectancy*, *Social Influence*, and *Facilitating Conditions*. *Performance Expectancy* is defined as the degree to which an individual believes that using a particular technology will help him or her to attain performance gains. *Effort Expectancy* is defined as the degree of ease associated with the use of a particular technology. *Social Influence* is defined as the degree to which an individual perceives that important others believe he or she should use the new technology. *Facilitating Conditions* are defined as the degree to which an individual believes that an organizational and technical

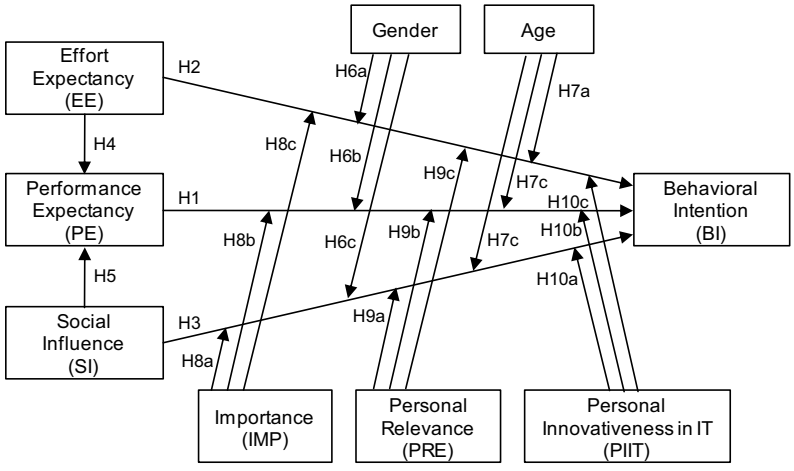


Figure 1: Research Model

infrastructure exists to support use of the new technology. *Gender*, *Age*, *Experience*, and *Voluntariness of Use* moderate the key relationships in this model.

To adjust the UTAUT model to our research setting, we made the following modifications to the original model. First, we eliminated the constructs *Use Behavior* and *Facilitating Conditions* because due to the lack of a working prototype, *Use Behavio*r cannot be observed. However, *Behavioral Intention* has shown to be a good predictor of actual behavior as posited by the TRA and could be confirmed in many studies [48]. Second, we added indirect relationships from *Effort Expectancy* and *Social Influence* on *Performance Expectancy* because this relationship was supported by the results from many prior technology acceptance studies [28][29][39]. Third, we eliminated two moderators from the original UTAUT model: *Voluntariness of Use* was eliminated, because the adoption of the proposed smart kitchen environment will, in contrast to workplace settings, always occur on a voluntary basis. *Experience* was eliminated, because in the original UTAUT study *Experience* was examined using a cross-sectional analysis from the time of the artifact's introduction to later stages of greater experience. Due to the early stage of development, and the unavailability of a commercial product, asking respondents at different points in time was not feasible. Fourth, we decided to introduce three additional moderating variables (*Importance*, *Personal Relevance,* and *Personal Innovativeness in IT*), which will be motivated below.

With regard to the direct and indirect relationships between the independent and the dependent variable, we therefore hypothesize the following:

H1: *Performance Expectancy* has a positive effect on *Behavioral Intention*.

H2: *Effort Expectancy* has a positive effect on *Behavioral Intention*.

H3: *Social Influence* has a positive effect on *Behavioral Intention*.

H4: *Effort Expectancy* has a positive effect on *Performance Expectancy*.

H5: *Social Influence* has a positive effect on *Performance Expectancy*.

Prior studies observed a high variability in the corresponding correlations, which suggests that moderator variables may exert a significant influence (e.g., [29][39][43]. Moderation occurs when the relationship between two variables depends on a third variable such as gender or age. As a consequence, the introduction of moderating factors can improve the often limited explanatory power and inconsistencies in existing technology acceptance studies. Therefore we introduce five moderating variables, which we regard as important in the proposed application setting.

First, we consider the differences in acceptance behavior between men and women [31][45]. Men have shown to be usually more pragmatic and task-oriented than women. Moreover, men usually feel more comfortable using new technologies. On the other hand, women compared to men have been found to have a higher awareness of other's feelings, and, in turn, are more influenced by others. Therefore, it seems likely that men are more driven by *Performance Expectancy*, whereas women are more driven by *Effort Expectancy* and *Social Influence*. Compared with *Gender*, *Age* has received less attention in the existing literature. Young users have been found to be more driven by *Performance Expectancy*, while older users are more driven by *Effort Expectancy* [32][46]. It has also been proposed that older users are more influenced by social factors, because affiliation increases with age and older people are more likely to conform to others' opinions [43]. In accordance with the original UTAUT model, we therefore hypothesize that *Gender* and *Age* play a moderating role in our research model.

H6a: For women the effect of *Effort Expectancy* on *Behavioral Intention* is higher than for men.

H6b: For men the effect of *Performance Expectancy* is higher than for women.

H6c: For women the effect of *Social Influence* on *Behavioral Intention* is higher than for men.

H7a: For older people the effect of *Effort Expectancy* is higher than for younger people.

H7b: For younger people the effect of *Performance Expectancy* is higher than for older people.

H7c: For older people the effect of *Social Influence* is higher than for younger people.

Beyond the logic of the original UTAUT model, we introduce additional hypotheses regarding the moderating influences of *Importance*, *Personal Relevance*, and *Personal Innovativeness in IT*. Prior work has investigated the role of involvement on consumer decisions [49]. Barki and Hartwick [7] investigated its role in the context of information systems development. They define involvement as "a subjective psychological state, reflecting the importance and personal relevance of an object or event". We argue that, following this definition and subsequent applications of the construct, involvement encompasses two different but important factors that influence technology adoption, namely *Importance* and *Personal Relevance*. In the context of our kitchen scenario, *Importance* denotes the extent of intrinsic desire or personal need for support throughout the preparation of a meal. In contrast to that, *Personal Relevance* denotes an individual's general dedication and interest in the application domain. The construct reflects to which extent cooking in general is relevant to an individual. As such it clearly differentiates from the *Importance* construct. For example, cooking can be very relevant for a person when he or she is often preparing food. At the same time, getting help in the kitchen may not be important for the same person because he or she is already very skilled. We therefore decided to split the originally proposed involvement construct into the two aspects *Importance* and *Personal Relevance* by introducing separate constructs.

One of the objectives behind the concept of a 'smart kitchen' is to help users to select and prepare healthier and more tasteful dishes. We theorize that the more a potential user feels that it is important for him to get support in the kitchen the more important becomes *Performance Expectancy* as a predictor, whereas the importance of *Effort Expectancy* and *Social Influence* will diminish.

H8a: The effect of *Effort Expectancy* decreases with higher *Importance*.

H8b: The effect of *Performance Expectancy* increases with higher *Importance*.

H8c: The effect of *Social Influence* decreases with higher *Importance*.

We further theorize that higher *Personal Relevance* increases the strength of the effect that *Performance Expectancy* exerts on *Behavioral Intention* because functional aspects will be more important than usability or social aspects. Consequently, the significance of *Effort Expectancy* and *Social Influence* should diminish.

H9a: The effect of *Effort Expectancy* decreases with higher *Personal Relevance*.

H9b: The effect of *Performance Expectancy* increases with higher *Personal Relevance*.

H9c: The effect of *Social Influence* decreases with higher *Personal Relevance*.

Finally, we add the construct *Personal Innovativeness in the domain of Information Technology* (*PIIT*) as a moderating factor to our model. Agarwal and Prahad [2] introduced this construct as a moderating variable into technology acceptance research. In the context of a novel technology that only few people are familiar with, it could be expected that innovativeness plays an important role in an individual's acceptance behavior. We therefore theorize that in the home domain, people with different levels of *Personal Innovativeness* show different adoption behavior.

H10a: The effect of *Effort Expectancy* decreases with higher *PIIT*.

H10b: The effect of *Performance Expectancy* increases with higher *PIIT*.

H10c: The effect of *Social Influence* decreases with higher *PIIT*.

Table 1: Validation of the measurement model

Construct	Item	Loading	Mean	SD	α	CR	AVE
Behavioral Intention (BI)	BI1	0.96	4.24	1.77	0.95	0.97	0.91
	BI2	0.95					
	BI3	0.94					
Effort Expectancy (EE)	EE1	0.90	5.01	1.47	0.82	0.89	0.74
	EE2	0.88					
	EE3	0.79					
Social Influence (SI)	SI1	0.78	3.69	1.74	0.82	0.89	0.74
	SI2	0.91					
	SI3	0.88					
Performance Expectancy (PE)	PE1	0.88	4.68	1.68	0.85	0.91	0.77
	PE2	0.88					
	PE3	0.88					
Importance (IMP)	IMP1	0.76	4.21	1.9	0.73	0.84	0.64
	IMP 2	0.88					
	IMP 3	0.76					
Personal Relevance (PRE)	PRE1	0.91	5.01	1.69	0.83	0.89	0.73
	PRE2	0.86					
	PRE3	0.79					
Personal Innovativeness in IT (PIIT)	PIIT1	0.76	5.40	1.69	0.77	0.87	0.77
	PIIT2	0.98					

5. Data Collection

5.1 Instrument Development

To test the research model and the associated hypotheses proposed above, we designed a questionnaire on the basis of existing scales from the technology acceptance literature (a list of questionnaire items can be requested from the authors). The measurement scales for the main constructs were operationalized by adopting items from [46] and adapting them to the specific context of our smart kitchen environment. For constructing measurement scales for *Importance* and *Personal Relevance*, we referred to [7] and [49]. *Personal Innovativeness in Information Technology* was operationalized using the scale developed by [2].

The focus of our study is on a complex smart kitchen environment, which incorporates different Smart Products that interact with each other and show context-aware behavior. It is constituted of the following functional blocks: A *Smart Kitchen Interaction Pad*, a Tablet-PC-like device, is the central user interface for the smart kitchen. It provides meal recommendations based on available ingredients and kitchen utensils as well as personal preferences. To guide users in their preparation process, textual and visual presentations provide step-by-step instructions that are synchronized with the actual preparation progress. Smart kitchen utensils can be parameterized according to recipe information, and they give feedback on ongoing activities and status information (e.g. temperature, weight, processing times). A recipe memorization function allows for recording preparation processes including sensorial information from the smart kitchen

tools. Once a recipe is chosen, the user can retrieve a shopping list either as a print-out or on a mobile phone. The shopping list considers which ingredients are already available in the household. Finally, the user can monitor his or her nutrition habits. Consumption in the smart kitchen is automatically recorded, and a mobile application enables users to track non-domestic consumption.

As the described smart kitchen environment is not yet physically available, we have taken a scenario-based approach. For each of the five functional blocks, we developed a detailed textual scenario description, which was complemented by a graphical illustration created by a professional graphics designer. For each scenario, interviewees were asked the same set of questions with minor adaptations to the specific context. All items were measured using a seven-point Likert scale. All constructs were formulated in a reflective mode. To further assure content validity, we followed a two-step process. First, each item was reviewed by three industry experts from a home equipment manufacturer and three academic experts in the area of Smart Products research. This resulted in a small number of changes to the wording and the overall structure of the questionnaire. The revised questionnaire was then circulated among the same group of experts and was then consistently rated as comprehensive and complete. In a pre-test, we then asked ten persons to fill in the questionnaire and provide us with feedback, which led to minor changes for reasons of clarity and comprehensiveness.

5.2 Sample and Descriptive Statistics

The data for the present study were gathered via an online survey, which was accessible for two months starting from September 2009. The participation was anonymous, voluntary, and there were no rewards for participation, which can be interpreted to mean that there should be no confounding effects from coercing subjects into participation or due to subjects that are just after some reward. The survey took about 25 minutes to complete.

600 people in different European countries were contacted by email, of which 175 completed the survey. The survey was designed in a way that participants had to answer all questions before they were able to submit the questionnaire. After an initial screening of the data, nine cases were removed from the sample, because of certain patterns that suggested unreliable responses (e.g., the same response category was checked for all questions). The resulting sample comprised 166 subjects corresponding to a final response rate of 28%. The proportion of gender is almost balanced with 46% of the respondents being female. 39% of respondents were younger than 30 years, 30% were between 31 and 40, 22% between 41 and 50, and 9% older than 51 years.

6. Data Analysis

6.1 Measurement Model

The questionnaire presented five partial scenarios, which were rated separately applying the same scales. This approach allowed for investigating a complex environment consisting of several different technological artifacts on a detailed level. To test whether the five scenarios had been rated in a consistent way, we applied t-tests on construct level to compare each scenario with each other. The results revealed that there were no significant differences at $p<0.05$ between construct means across all scenarios. Consequently, each scenario can be regarded as *pars pro toto* so that we could aggregate the five scenarios on item level for our further analysis instead of analyzing each scenario separately. As a consequence, we were able to use a questionnaire on a fine-grained functional level and at the same time investigate the smart kitchen environment as a whole.

We applied Partial Least Squares Path Modeling (PLS) as a Structural Equation Modeling (SEM) technique to test the research model. We favored PLS over first generation regression techniques because of its ability to model relationships among different constructs simultaneously and to handle measurement errors [10]. Furthermore, we favored PLS over the covariance-based SEM approach because under conditions of non-normality, moderate effect sizes, and smaller samples, the PLS approach appears preferable [22][34]. The data points of survey-collected data usually do not follow a multivariate normal distribution, which is an important precondition of the covariance-based approach but not for PLS [11]. In addition to that, we asked for the respondents' opinion regarding several different scenarios. Therefore, the observations in our study are not fully independent from each other, which is another assumption for the covariance-based approach. In contrast, independence of observations is not an assumption of PLS [11].

We employed the PLS implementation of Smart-PLS version 2.0M3 [36] with a 5000 sample bootstrapping technique for model assessment. All statistical tests were assessed with two-tailed t-tests. In a first step, we assessed the measurement model to ensure that good construct measures are represented in a valid structural model. Table 1 shows the results of our factor analysis. All item loadings are well above the threshold of 0.707, indicating that over half of the variance is captured by the latent construct [11][22]. No problematic cross-loadings could be observed. Further, Cronbach's α and composite reliability values as measures for internal consistency are well above the recommended value of 0.7 for each construct [33]. Convergent validity [12], which refers to the degree to which the items measuring the same construct agree, is examined by considering the average variance extracted (AVE). Table 1 shows that it is well above the recommended threshold of 0.5 for all constructs [19].

Discriminant validity, which refers to the degree to which measures of distinct concepts differ, was examined by comparing the correlations between the measurement items of distinct constructs with the squared root of the AVE by each construct. The squared root of the AVE for each construct was higher than its correlations with other constructs indicating satisfactory discriminant validity (Fornell-Larcker criterion [19]).

6.2 Structural Model

With sufficient evidence from reliability and validity measures, the next step was to test the hypothesized paths and the explanatory power of the model. The explanatory power is examined by inspecting the R^2 values (i.e., the explained variance) of the dependent variables. Chin finds that R^2 values of 0.67, 0.33, and 0.19 in PLS path models should be regarded as substantial, moderate, and weak, respectively [11]. Because PLS does not assume a particular distribution, re-sampling techniques such as bootstrapping have to be used to determine statistical significance of the path coefficients. The corresponding t-values indicate whether the hypothesis that the respective parameter estimates equal zero must be rejected.

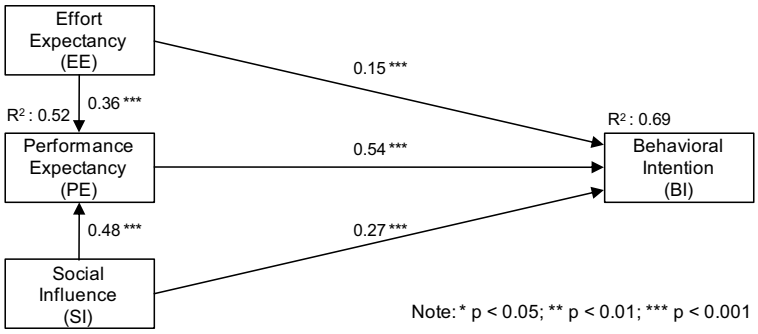


Figure 2: Results for the strutural model

For the basic model without moderators, Figure 2 shows that we obtained R^2 values of 0.69 for *Behavioral Intention* and 0.52 for *Performance Expectancy*. Moreover, the t-tests conducted on the relationships reveal that all relationships are significant, and the absolute path weights show that they are sufficiently substantial. Therefore we accept hypotheses H1, H2, H3, H4, H5. Consequently, the relations between independent and dependent variables as proposed by our modified UTAUT model can be confirmed.

For an examination of moderation effects, we need to distinguish between categorical variables such as *Gender* and latent variables such as *Personal Relevance*, which we measured on a Likert scale. As proposed in [27] and applied in [34], [45], we adopted multiple t-tests to examine the moderation effects of *Gender* and *Age*. The PLS t-test uses the standard errors obtained from bootstrapping to test for group equality of path coefficients. The following statistic, which is asymptotically t-distributed with *m+n-2* degrees of freedom, is calculated [27]:

$$t = \frac{\text{Path}_{\text{Sample 1}} - \text{Path}_{\text{Sample 2}}}{\left[\sqrt{\frac{(m-1)^2}{(m+n-2)} * \text{S.E.}^2_{\text{Sample 1}} + \frac{(n-1)^2}{(m+n-2)} * \text{S.E.}^2_{\text{Sample 2}}}\right] * \left[\sqrt{\frac{1}{m} + \frac{1}{n}}\right]}$$

In this formula, *m* and *n* denote the sample sizes of the two groups, $Path_{Sample1}$ and $Path_{Sample2}$ are the path coefficients for the path that is being compared, and $S.E.^2_{Sample1}$ and $S.E.^2_{Sample2}$ are the variances in each group for the paths that are compared. Our finding from this analysis is that only *Gender* has a moderating effect on the SI-BI relationship with p<0.01 (see Table 2). All other effects cannot be regarded as significant. Consequently, we accept hypothesis 6c, while we reject hypotheses 6a, 6b, 7a, 7b, and 7c.

To test *Importance*, *Personal Relevance*, and *Personal Innovativeness* for their moderating effects, we employed the product-indicator approach [10], which was specifically designed for continuous variables. With the product-indicator approach, a new interaction construct is created by using the products of the indicators of the moderating construct and the predictor construct. An F-value based on the effect size is calculated to decide via an F-test whether there is a significant moderator effect [3][9][25]. The F-value is calculated according to the following formula:

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)}$$

R_1 and R_2 are the explained variances before and after introducing the interaction term; k_1 and k_2 represent the number of predictors before and after introducing the interaction term; *N* is the sample size. *F* then follows an F-distribution with $df_1=(k_2-k_1)$ and $df_2=(N-k_2-1)$ degrees of freedom. An F-test reveals whether the explained variances are significantly different for the two models. Following the results presented in Table 3, we can conclude that only *Importance* has a significant moderating effect with regard to the explained variance, although this effect turns out rather week if we compare the explained variances with and without the interaction term. Further, Table 3 shows that the moderating effect of *Importance* is significant only for two of the three tested relationships for p<0.05, namely the PE-BI and the SI-BI relationship. The effect on the EE-BI is not sufficiently significant. Consequently, hypotheses H8a, H8b can be accepted, whereas hypotheses 8c, 9a, 9b, 9c, 10a, 10b, and 10c are rejected.

Table 2: Moderating effects of categorial variables

Moderator		R^2 BI	R^2 PE	Path Coefficients EE → BI	PE → BI	SI → BI
None		0.689	0.517	0.15 ***	0.54 ***	0.27 ***
Gender	Female	0.687	0.556	0.14 ***	0.51 ***	0.32 ***
	Male	0.693	0.480	0.16 ***	0.57 ***	0.21 ***
	T-Test	-	-	ns	ns	**
Age	≤ 40 y.	0.686	0.450	0.14 ***	0.54 ***	0.29 ***
	> 40 y.	0.700	0.647	0.17 ***	0.54 ***	0.22 ***
	T-Test	-	-	ns	ns	ns

Note: * p < 0.05; ** p < 0.01; *** p < 0.001

7. Discussion

7.1 Theoretical Implications

The objective of this study was (i) to test the applicability of the proposed modification of the UTAUT research model, which was adapted to the smart kitchen domain, and (ii) to analyze which variables exert a moderating role on the predictor relationship of *Effort Expectancy*, *Performance Expectancy*, and *Social Influence* on *Behavioral Intention*. For this purpose, we developed a modified version of the UTAUT model, extending it by *Gender*, *Age*, *Importance*, *Personal Relevance*, and *Personal Innovativeness in IT* as moderator variables. Furthermore, we added indirect effects from *Effort Expectancy* and *Social Influence* as proposed and affirmed in prior works.

Empirical analysis using PLS confirmed the applicability of the modified UTAUT model in a smart kitchen environment. *Performance Expectancy* has shown to have the strongest direct effect on *Behavioral Intention*. *Effort Expectancy* and *Social Influence* act as significant predictors, too, but at a weaker level. Moreover, our analysis has shown that *Gender* poses a significant moderator of the relationship between *Social Influence* and *Behavioral Intention*. For women, *Social Influence* seems to work as a stronger predictor than for men, which can be interpreted such that it is relatively more important to women that friends and family would appreciate adopting the proposed technology. In contrast to the original UTAUT model, our data do not support the assumption that *Gender* would exert a significant influence on the other relationships in our study. The same holds true for *Age*, which could not be confirmed to be a significant moderator. Regarding the moderators that were added to the original model, only *Importance* showed a significant effect on the relationships of *Performance Expectancy* and *Social Influence* on *Behavioral Intention*. With increasing *Importance*, *Performance Expectancy* has a relatively stronger effect on *Behavioral Intention*, whereas the effect size of *Social Influence* decreases. *Personal Relevance* and *Personal Innovativeness in IT* could not be affirmed as moderators.

All considered, the basic structural model could be confirmed, whereas only few moderating effects could be found. Whereas *Gender* could be approved as a moderator in many studies, *Importance* seems to have a significant, albeit so far underestimated moderating role. Against the background of our results, we encourage to further investigate *Importance* as a measure of intrinsic motivation to accept a novel technology in the private domain.

7.2 Practical Implications

Besides the aforementioned theoretical implications, our study also allows for drawing conclusions relevant to practice, particularly in the home appliance domain. Descriptive results indicate that the proposed smart kitchen environment was perceived positively across several population groups. In particular, there were no major differences between older vs. younger persons, innovative vs. non-innovative persons, persons with and without technology-related educational backgrounds. We are in favor of interpreting these results such that smart kitchen environments, and perhaps smart home environments in general, have the potential to leave their narrow market niches and become broadly adopted by the home appliance industry. Consequently, it may be the right time for managers in charge at the respective companies to develop innovative product portfolios that make use of the *Smart Product* concepts as described in this paper.

At the same time, more research effort should be focused on the question why home automation, although commercially available for more than a decade, does not gain more attraction. Reasons may be found in missing standards and consequently a lack of interoperability between different vendors, consumers that fear lock-in, long investment cycles for home appliances, or merely a price premium that is regarded to be inappropriate in relation to the additional value. Our study has shown that consumers basically have a positive attitude towards such technologies, so reasons for non-adoption decisions require further investigation.

Not least, we can learn from this empirical investigation that consumers regard functional capabilities of a smart kitchen environment as key to their adoption decision, whereas potential adopters are less concerned about usability issues. With regard to *Social Influence*, we have seen that it is quite important for consumers that friends and family appreciate the smart kitchen environment. As a consequence for market introduction, marketing measures should not only focus on technological capabilities. In addition, an image campaign seems to be appropriate such that potential adopters get the feeling that they improve their social image by using a smart kitchen environment at home.

8. Summary and Outlook

In this study, we investigated user acceptance towards a smart kitchen environment. Smart environments emerge from the interplay of individual smart products, a novel class of product/service bundles enabled by digital technologies that show complex behavior through context-awareness, communication among each other, processing power, and a paradigm shift away from PC-like interfaces towards tangible human-computer interaction. As such, smart environments can be regarded as a

Table 3: Moderating effects of continuous variable

Moderator		R^2 (BI)	Path Coefficients EE → BI	PE → BI	SI → BI
None		0.6887	0.15 ***	0.54 ***	0.27 ***
IMP	Direct Effect	0.6977	0.15 ***	0.52 ***	0.25 ***
	Interaction	-	0.03	0.08 *	-0.07 *
	F-Test	4.7923 *	-	-	-
PRE	Direct Effect	0.6934	0.14 ***	0.54 ***	0.27 ***
	Interaction	-	0.02	0.05	0.00
	F-Test	2.47	-	-	-
PIIT	Direct Effect	0.6915	0.14 ***	0.54 ***	0.27 ***
	Interaction	-	-0.02	0.00	0.06
	F-Test	1.46	-	-	-

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

concrete implementation of the ubiquitous computing paradigm for a specific domain. Our study has shown that potential adopters appreciate such novel approaches. In contrast to our initial expectation, usability was regarded as a minor issue whereas performance and social aspects turned out to be more important. We contributed to technology acceptance research by proposing an adapted version of the UTAUT model, and by demonstrating its practicability as an analytical tool in the smart home environment domain. Furthermore, we tested several constructs for their moderating effect and concluded that *Gender* and *Importance* play a significant role.

Even though every effort has been made to ensure the validity of our findings, the present study comes with limitations that point to opportunities for further research. First of all, while the size of our sample is sufficient for testing the constructed structural model, larger samples would be helpful to investigate simultaneously the differences in adoption behavior between geographic regions and additional demographic factors such as income, family status, etc. Second, although having achieved sufficient explanatory power, our results nevertheless leave room for additional factors not included in our research model that might influence adoption behavior. We therefore propose to discuss and empirically test the relevance of other constructs beyond the scope of the present study. Third, our investigation has been based on scenario descriptions, which limits the transferability to a commercial offering. As a consequence, the scenarios should next be implemented and tested in an experimental setting to increase the validity of our results.

9. References

[1] Aarts, E. et al. 2002. Ambient Intelligence. *The invisible future – the seamless integration of technology in everyday life*. P. Denning. McGraw-Hill. 138-144.

[2] Agarwal, R. and Prasad, J. 1998. A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*. 9, 2 (1998), 204-215.

[3] Aiken, L. and West, S. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Sage.

[4] Aitenbichler, E. et al. 2007. Engineering intuitive and self-explanatory smart products. *Proceedings of the 2007 ACM symposium on Applied computing* (New York, USA, 2007), 1632-1637.

[5] Allmendinger, G. and Lombreglia, R. 2005. Four strategies for the age of smart services. *Harvard Business Review*. 83, 10 (2005), 131-145.

[6] Baldauf, M. et al. 2007. A Survey on Context-aware Systems. *International Journal of Ad Hoc and Ubiquitous Computing*. 2, 4 (2007), 263-277.

[7] Barki, H. and Hartwick, J. 1994. Measuring user participation, user involvement, and user attitude. *MIS Quarterly*. 18, 1 (1994), 59-82.

[8] Bauer, J. et al. 2005. Fridgets: digital refrigerator magnets. *Proceedings of the Conference on Human Factors in Computing Systems* (Portland, USA, 2005), 2060-2064.

[9] Carte, T. and Russell, C. 2003. In pursuit of moderation: Nine common errors and their solutions. *MIS Quarterly*. 27, 3 (2003), 479-501.

[10] Chin, W.W. et al. 2003. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research*. 14, 2 (2003), 189-217.

[11] Chin, W.W. 1998. The partial least squares approach for structural equation modeling. Modern methods for business research. *Modern methods for business research. Methodology for business and management*. G. Marcoulides. Lawrence Erlbaum Associates Publishers. 295-336.

[12] Cook, M. and Campbell, D.T. 1979. *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin.

[13] Dabholkar, P. and Bagozzi, R. 2002. An attitudinal model of technology-based self-service: moderating effects of consumer traits and situational factors. *Journal of the Academy of Marketing Science*. 30, 3 (2002), 184-201.

[14] Davis, F.D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*. 13, 3 (1989), 319-340.

[15] Dhebar, A. 1996. Information technology and product policy: 'Smart' products. *European Management Journal*. 14, 5 (1996), 477-485.

[16] Ferguson, G. 2002. Have your objects call my objects. *Harvard Business Review*. 80, 6 (2002), 138–144.

[17] Fishbein, M. and Ajzen, I. 1975. *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley.

[18] Fleisch, E. and Thiesse, F. 2007. On the Management Implications of Ubiquitous Computing: An IS Perspective. *Proceedings of the European Conference on Information Systems (ECIS)* (St. Gallen, Switzerland, 2007).

[19] Fornell, C. and Larcker, D.F. 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*. 18, 1 (1981), 39-50.

[20] Freudenthal, A. and Mook, H. 2003. The evaluation of an innovative intelligent thermostat interface: universal usability and age differences. *Cognition, Technology & Work*. 5, 1 (2003), 55–66.

[21] Garfield, M. 2005. Acceptance of ubiquitous computing. *Information Systems Management*. 22, 4 (2005), 24-31.

[22] Gefen, D. et al. 2000. Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*. 4, 7 (2000), 1–77.

[23] Gershenfeld, N. 1999. *When Things Start to Think*. Henry Holt.

[24] Hanson-Smith, V. et al. 2006. NutriStat: tracking young child nutrition. *Proceedings of the Conference on Human Factors in Computing Systems* (Montreal, Canada, 2006), 1831-1836.

[25] Jaccard, J. et al. 1990. *Interaction Effects in Multiple Regression. University Paper series on Quantitative Applications in the Social Sciences, no. 72*. Sage.

[26] Ju, W. et al. 2001. CounterActive: an interactive cookbook for the kitchen counter. *Proceedings of the Conference on Human Factors in Computing Systems* (Seattle, USA, 2001), 269–270.

[27] Keil, M. et al. 2000. A cross-cultural study on escalation of commitment behavior in software projects. *MIS Quarterly*. 24, 2 (2000), 299–325.

[28] King, W. and He, J. 2006. A meta-analysis of the technology acceptance model. *Information & Management*. 43, 6 (2006), 740–755.

[29] Lee, Y. et al. 2003. The technology acceptance model: Past, present, and future. *Communications of the Association for Information Systems*. 12, 50 (2003), 752-780.

[30] Maass, W. and Janzen, S. 2007. Dynamic product interfaces: A key element for ambient shopping environments. *Proceedings of 20th Bled eConference* (Bled, Slovenia, 2007).

[31] Meyers-Levy, J. and Maheswaran, D. 1991. Exploring Differences in Males' and Females' Processing Strategies. *Journal of Consumer Research: An Interdisciplinary Quarterly*. 18, 1 (1991), 63-70.

[32] Morris, M. and Venkatesh, V. 2000. Age Differences in Technology Adoption Decisions: Implications for a Changing Work Force. *Personnel Psychology*. 53, 2 (2000), 375-403.

[33] Nunnally, J.C. and Bernstein, I.H. 1994. *Psychometric theory*. McGraw-Hill.

[34] Qreshi, I. and Compeau, D. 2009. Assessing between-group differences in information systems research: a comparison of covariance- and component-based SEM. *MIS Quarterly*. 33, 1 (2009), 197-214.

[35] Rijsdijk, S. and Hultink, E.J. 2009. How Today's Consumers Perceive Tomorrow's Smart Products. *Journal of Product Innovation Management*. 26, 1 (2009), 24-42.

[36] Ringle, C. et al. 2005. SmartPLS 2.0 (M3) Beta. *Hamburg: http://www. smartpls. de*. (2005).

[37] Rothensee, M. 2008. User Acceptance of the Intelligent Fridge: Empirical Results from a Simulation. *Lecture Notes in Computer Science (4952)*. Springer. 123-139.

[38] Satyanarayanan, M. 2001. Pervasive Computing: Vision and Challenges. *IEEE Personal Communications*. 8, 4 (2001), 10-17.

[39] Schepers, J. and Wetzels, M. 2007. A meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*. 44, 1 (2007), 90-103.

[40] Schmidt, D. 2002. Middleware for Real-time and Embedded Systems. *Communications of the ACM*. 45, 6 (2002), 43-48.

[41] Sheng, H. et al. 2008. An Experimental Study on Ubiquitous Commerce Adoption: Impact of Personalization and Privacy Concerns. *Journal of the Association for Information Systems*. 9, 6 (2008), 344-376.

[42] Spiekermann, S. and Pallas, F. 2005. Technology Paternalism – Wider Implications of RFID and Sensor Networks. *International Journal of Ethics of Science and Technology Assessment*. 4, 1 (2005), 6-18.

[43] Sun, H. and Zhang, P. 2006. The role of moderating factors in user technology acceptance. *International Journal of Human-Computer Studies*. 64, 2 (2006), 53-78.

[44] Vastenburg, M. et al. 2007. Measuring User Experiences of Prototypical Autonomous Products in a Simulated Home Environment. *Lecture Notes in Computer Science (4551)*. Springer-Verlag. 998-1007.

[45] Venkatesh, V. and Morris, M. 2000. Why don't men ever stop to ask for directions? Gender, social influence, and their tole in technology acceptance and usage behavior. *MIS Quarterly*. 24, 1 (2000), 115-140.

[46] Venkatesh, V. et al. 2003. User acceptance of information technology: Toward a unified view. *MIS Quarterly*. 27, 3 (2003), 425–478.

[47] Weiser, M. 1991. The Computer for the Twenty-First Century. *Scientific American*. 265, 3 (1991), 94-104.

[48] Yousafzai, S.Y. et al. 2007. Technology acceptance: a meta-analysis of the TAM: Part 1. *Journal of Modelling in Management*. 2, 3 (2007), 251-280.

[49] Zaichkowsky, J. 1985. User Measuring the involvement construct. *Journal of consumer research*. 12, 3 (1985), 341-352.

The Effects of Outcome Expectations on Monetary- and Non-Monetary Rewarded Product Recommendations in Open- and Invitation-only Social Networking Sites: An Empirical Comparison of Facebook and ASmallWorld

Christian M. Messerschmidt
Goethe-University Frankfurt
Grüneburgplatz 1
60323 Frankfurt
+49 (0)69 798 33882
messerschmidt@wiwi.uni-frankfurt.de

Deniz Yildirim
Goethe-University Frankfurt
Grüneburgplatz 1
60323 Frankfurt
+49 (0)69 798 33882
deniz.y@gmx.net

ABSTRACT
Social networking sites experience huge growth in their number of members. For marketing purposes they are very beneficial to spread Word-of-Mouth in terms of product recommendations. A closer view detects that social networking sites can be divided in open- (OSNs) and invitation-only social networking sites (ISNs). Their members may behave different in contributing knowledge in terms of product recommendation depending in which social networking site they are currently remaining. We therefore first analyze if the members are willing to recommend products for either monetary or non-monetary rewards in their preferred social networking site as well as if they consider these recommendations in their purchasing decision and connect this to a member's personal- and community-related outcome expectations. Second we compare the results between ISN- and OSN-members to conclude in which type of social networking site a product recommendation should be monetary rewarded or not.

Keywords
Social Networks, Online Communities, Web 2.0, Social Media, Outcome Expectations, Product Recommendation.

1. INTRODUCTION
Nielsen reported in their 2009 survey of global consumer trust in advertising, that 70% of all consumers trust consumer opinions posted online compared to 60% in 2007 [54] which is right behind the trust in the recommendations of friends (90%) on the second rank. Newspapers, tests from experts or any kind of advertising do not reach this high amount of trust. In academic research several studies emphasize that user generated content in terms of consumer product reviews significantly influence the consumers purchasing decisions. Active electronic communities are developing which provide a rich repertory of information about products and services [4]. Aral and Walker [3] found that active-personalized Word-of-Mouth-messages, although less frequently used, are more effective in encouraging the adoption of a product among the peers of the recommender compared to passive-broadcast Word-of-Mouth-messages. The analysis of how to enforce consumers to execute active-personalized product recommendations is a relevant question which has to be solved.

An application in the Internet which affords both the connection to friends as well as to unknown consumers and, thus, may boost the spread of active-personalized Word-of-Mouth-messages in terms of recommendations or opinions about products, are social networking sites. Currently these networks are subject to a strong trend of increasing members. According to Comscore, Facebook, the world's most popular and well-known social networking site, was the fourth largest site worldwide with 340 million unique users in July 2009, right after Google, Microsoft and Yahoo [41] and is the number one website in the US as of March 2010 according to current usage statistics [26]. Consumers spend on average almost six hours per month within a social networking site, while the search on Google captures only around two hours of the consumers.

Social networking sites can be generally classified into two categories: open social networks (OSNs) and invitation-only social networks (ISNs). OSNs have no entry restrictions, whereas ISNs are more exclusive as they require an invitation and are limited in their membership base.

Previous research has found that qualitative and rich knowledge contribution is essential for a successful development of online communities and, thus, for social networking sites, which are defined as subgroup of online communities. Chiu et al. [20] suggest that knowledge contribution is influenced by social capital and outcome expectations. Social capital is the network of relationships possessed by an individual or a social network and the set of resources embedded within it. Outcome expectations

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

represent the expectations that an individual has towards the possible outcome of knowledge contribution within a social network for him/herself or for the social network as a whole. As product recommendations and opinions posted online can also be seen as knowledge contribution, it is important to know from a marketing perspective how members can be motivated to recommend products or services they like to their social network connections and, thus, influence the purchasing decisions of other members. A key aspect in this case is, whether or not recommenders receive monetary or non-monetary rewards.

At first glance, one may think that a monetary reward should generally increase a member's willingness to recommend products on her/his social networking site. However, a monetary reward may primarily attract users with high personal outcome expectations (e.g. earning money for their knowledge-contribution) and may be rejected by those users who register on a social networking site because of community-related outcome expectations (e.g. helping others in the community via knowledge contribution). ISNs have entry restrictions and are therefore smaller, which increases the cohesion amongst their members since relationships are based on authentic connections and true information. Thus, in ISNs product recommendations shall be expressed regardless of a monetary reward. Helping other is hypothesized to be in the focus of the members. In contrast, OSN-members usually have weaker connections and therefore, the cohesion in the network may be also weaker. It could be expected that some OSN-members only recommend products to gain monetary rewards without consideration of whether their input helps the receiver of their recommendation. OSN-members may be less trusting in product recommendations of other members, because they cannot easily distinguish between valuable and invaluable products recommendations. This results in a decreasing influence on the purchasing decision, making a monetary rewarded product-recommendation-system less successful than initially believed. We assume, that the share of members with strong personal outcome expectations is higher in OSNs than in ISNs, which may result in a different acceptance of monetary rewarded product recommendations within each type of social networking sites.

The scenario outlined above shows, that the decision about the implementation of monetary or non-monetary rewards for recommendations is not trivial and sometimes not intuitive as different setups may lead to different outcomes depending on whether it is an open or invitation-only social network. The aim of this study is to analyze the differences in the impact of two outcome expectation aspects elements [20] on monetary and non-monetary rewarded recommendations between OSNs and ISNs and to identify which kind of reward will be successful in ISNs or OSNs.

This article first defines open and invitation-only social networks and discusses previous research concerning the motivations for product recommendations in social networking sites as well as outcome expectations. Chapter 3 defines the research model and sets up our hypotheses. Chapter 4 empirically measures and compares factors which influence knowledge contribution in ISNs and OSNs.

2. THEORETICAL BACKGROUND

2.1 Social networking sites

Social networking sites are web-based services where members can create personal profiles, connect with other members, share personal connections and establish or maintain relationships with others [14]. Social networks are usually organized around a specific subject or general demographic such as friends or business contacts. Social networking sites are one type of online communities (also called virtual communities) [50, 58], which also include markets and auction sites, electronic bulletin boards, list servers, blog sites, gaming communities and shared interest web sites [37].

Previous research has analyzed social networking sites from many different perspectives. Boyd and Ellison [14] focused on the history and development of social networking sites and gave a detailed overview of the existing sites. Hargittai [37] analyzed the usage of social networking sites based on demographic characteristics and social surroundings and found that gender, race, ethnicity and parental education have an influence on the usage of social networking sites. Lampe et al. [47] reported that the use and perception of Facebook sometimes changed over time, likely due to changes in an individual's social context or an introduction of major features to the site. Privacy on social networking sites was examined by Gross and Acquisti [35] who quantified social networking site users', especially Facebook member's willingness to share personal information and concluded that users were unconcerned about privacy implications at that time. Research on social networks has also shown that social networking sites can be extremely useful for marketers to generate positive Word-of-Mouth-communication, e.g. product recommendations and, thus, enhance brand loyalty and increase sales [5, 36].

Social networking sites can generally be classified into two categories: open social networks (OSNs) and invitation-only social networks (ISNs).

ISNs are private social networking sites or a type of so-called niche communities that require an invitation and are limited in their membership base. They target a selected audience by restricting access and, thus, are more exclusive [14]. Most of these social networking sites have their focus on the exclusive groups of successful (e.g. Decayenne.com, Internations.org), rich (e.g. Affluence.org) or beautiful people (e.g. Beautifulpeople.com) and are setting their selection on variables like income or social connections. ASmallWorld (ASW), the ISN, which is analyzed in this study, is an invitation-only online social network, founded in 2004, which aims to help confidentially connecting an existing international community of people with similar backgrounds, interests and perspectives online. Members are already directly or indirectly connected by three degrees of separation [7], which means that (almost) every person is connected to every other person through three contacts (or less). The aggregated source of valuable information, advice and help from trusted members enables individuals to manage their private, social and business lives [7]. In order to build a trusted community and a reliable source of information, the membership is only granted via invitation. Only 10 to 20% of the community is authorized to invite new members. These trusted and loyal members have to fulfill certain criteria in order to achieve invitation rights [7-9, 33]. Members participate in different discussions and offer

information, help and advice. Most popular topics are business opportunities (e.g. "I have a client with 4 billion dollars looking to invest" with over 49,060 views and 500 posts) or "top" and "best"-lists (e.g. "Best club in your city" with over 22,600 views and 500 posts). Individuals also ask for travel advice, or product to buy next [8]. Thus, the ASW community is important to marketers because it provides a trusted environment for luxury-brand advertisement. Manufacturers can increase the awareness of their brands by reaching an influential and sophisticated membership base with more than 520,000 members, whom can be described as opinion leaders.

OSNs are online social networking sites that have no entry restrictions. The first social networking site founded in 1997 was SixDegrees.com, where members were allowed to create profiles and connect to their friends [14]. Amongst other well-known social networking sites, e.g. Friendster, LinkedIn, Xing or Myspace, Facebook is currently the most successful OSN. Thus, Facebook will be analyzed in this study as representative for OSNs. Facebook is an open social networking site, launched in 2004, which helps to maintain and develop social relationships among friends, family and coworkers [28, 29]. Members of this social networking site are connected by six or less degrees of separation [30]. Millions of members share content such as web links, news stories, blog posts, notes, photo albums and also product recommendations on a daily basis, further establishing and broadening these social relationships [30].

A closer look at the comparison of ISNs and OSNs shows that distinct differences between each other exist. They especially differentiate themselves with the number of members. OSNs usually have at least a few million members, whereas in ISNs the membership is kept small with less than a million members worldwide. The aim of OSNs is to help maintaining or developing social relationships with friends, family and co-workers and to share all kind of happenings with each other. ISNs want to connect an existing community of likeminded people who share similar backgrounds, interests and perspectives and to manage their private, social and business lives. In OSNs, you will be able to find a wide range of your offline networks, which are sub-communities by themselves. Due to the openness of OSN to everybody, (almost) every person is connected to every other person through six contacts (or less). In ISNs you have to be invited from a trusted member who fulfills different requirements to join the social networking site. Therefore, members are connected with any other person via two or three other members.

2.2 Product recommendations and Word-of-mouth-communication

Product recommendations have a variety of sources. Senecal and Nantel [60] translated the typology of information sources stated by Andreasen [2] into computer-mediated environments. The four product recommendation sources are: 1) Personal source providing personalized information or 2) Non-personalized information, 3) Impersonal source providing personalized information or 4) Non-personalized information. They grouped online product recommendation sources into the following categories: 1) other consumers such as relatives, friends and acquaintances, 2) human experts such as salespersons and independent experts and 3) expert systems such as recommender systems.

Word-of-Mouth (WOM) is the communication and mutual exchange of positive, neutral and negative information about products and services between individuals. Product recommendations are a positive form of WOM [1]. Several studies proved that WOM significantly influences the aspects of the consumer behavior, which will be discussed in the following.

Previous research recognized the importance of WOM and found it to be more effective than e.g. printed advertisement, radio advertisement and personal selling [39, 44]. Katz and Lazarsfeld [44] conducted the earliest study on the influence of WOM, and found that it is especially effective on the purchase of household goods and food products. Herr et al. [39] studied the effect of WOM on product judgments by analyzing vividly presented information and found that WOM had a stronger influence on individuals due to its vividness when compared to printed information. Arndt's [6] approach to WOM was to identify the specific factors that influenced a consumer's decision and found that positive WOM increased the likelihood of purchase, whereas negative WOM decreased it. Brown and Reinigen [16] investigated the strength of ties between the communicator and the decision-maker. They found that consumers tend to choose more similar personal sources of information for a referral flow.

The expansion of the Internet in the last decade has made electronic Word-of-Mouth (eWOM), also called Word-of-Mouse an important source of consumers' product evaluations. Consumers gather product information from other consumers by reading comments or by posting their own experiences with a product [38]. Senecal and Nantel [60] investigated the influence of eWOM on consumers' product choices, taking into consideration the different effects of online recommendation sources, product and website types. They found that recommender systems are the most influential source, despite the fact that human experts possess more expertise and other consumers are more trustworthy. Vallerand [68] came to the same conclusion. Aral and Walker [3] investigated that passive-broadcast WOM-messages are in sum more successful than active-personalized WOM-messages, because they are expressed much more frequently. Regarded per message active-personalized gains more attention of the receiver of the message. Forman et al. [31] showed that the value of user-generated content in terms of product recommendations strongly depends on the available extent of information about the recommender. The mining of recommendations can also be helpful for manufacturers, who want to gain more insights into the valuation of subjective attributes of hedonic products, which has been a difficult task in the offline world [4]. Smith et al. [61] outlined the importance of peer recommenders to the consumer. They found that product recommendations influence consumers in their product choices, in the amount of search effort in the decision-making process and in the level of user interest in sponsored advertisement. Peer recommenders are generally preferred over editorial recommendations.

Depending on the common interest of an online community, members are willing to participate and provide product and service related information. Individuals who are committed to an online community are more likely to show a positive attitude and commitment to the products and brands favored by other online community members. Community members can act as objective sources of information that also create new uses and benefits from the brand [46].

Social networking sites are becoming an increasingly more important channel for eWOM because they enhance the ability of the consumer to share and provide information and advice about products and services. The main objectives of social networking sites are to share experiences and establish or maintain relationships with others [14]. Active and constant communication with friends and acquaintances through different channels such as forums, blogs, groups and instant messaging may strengthen the relationship within personal social networks [34]. The variety of online communication channels in social networking sites, give consumers many options to do eWOM behavior and share their product-related experiences or seek advice. Despite the huge impact of eWOM on purchasing decisions and the accessibility of consumer generated product recommendations, there is only little research on eWOM behavior and the influence on decision-making in social networking sites. Brown et al. [15] analyzed how eWOM impacts decision-making and attitude formation in the context of social networking sites and explained the role of tie strength, homophily and source credibility in the evaluation of marketing information.

2.3 Outcome expectations

Outcome expectations lead to an individual's behavior towards more favorable outcomes [21]. The consequences of knowledge contribution an individual is expecting for him/herself are defined as *Personal outcome expectations. Community-related outcome expectations* stand for the expected consequences of knowledge contribution for the whole online community [20]. Positive outcomes are seen as benefits and negative outcomes are seen as costs [43]. According to the social exchange theory, individuals try to maximize their benefits and minimize their costs [51]. When people perceive that their incentive to contribute knowledge exceeds costs, knowledge contribution becomes more likely [45]. Benefits that motivate behavior during social exchange can be classified as either intrinsic or extrinsic [12, 24, 66]. Extrinsic rewards can be monetary [12, 13]. Intrinsic rewards can be more subtle non-monetary benefits such as status or respect [12], enhanced reputation [67], improved sense of self-worth [13], increased access to useful information and expertise, additional social relationships [17], or the enjoyment in helping others [43, 67].

Prior research has found mixed results about the importance of rewards. Contrary to Bock et al. [13]'s hypotheses, extrinsic rewards resulted in significant but negative effect on knowledge contribution. Extrinsic rewards may even impede favorable behavior toward knowledge sharing. Whereas, Nahapiet and Ghoshal [53] argue that no external incentives are required in case of strong pro-sharing norms, Kankanhalli et al. [43] posit that rewards are an important motivation for knowledge contribution in case of weak pro-sharing norms. A number of studies found that insufficient extrinsic and intrinsic rewards in return for the cost of knowledge sharing accrued are a barrier to knowledge sharing [22, 23, 40].

While Nahapiet and Ghoshal [53] examined knowledge contribution from the network level, Wasko and Faraj [67] argued on an individual level. Former theory states that an individual contributes knowledge with (1) the expectation that her/his behavior creates value for the collective and (2) the anticipation that it will then create value for oneself in the future.

Helping behavior is a voluntary action to help others without any expectation of reciprocity [63, 64]. Information sharing and knowledge contribution are the two dimensions contained in helping behaviors. Prior research shows that knowledge contributors achieve satisfaction stemming from their intrinsic helping behavior [13, 49, 67]. Moreover, enjoyment of helping can significantly impact knowledge contribution [43] and encourages reciprocity between members [57, 58].

3. RESEARCH MODEL AND HYPOTHESES

In this section the research model is being developed. This study investigates the influence of outcome expectations on product recommendations as a special kind of knowledge contribution in social networks and compares the influence of outcome expectations between ISNs and OSNs (see Figure 1). In this study ISNs are represented by ASmallWorld (ASW) and OSNs are represented by Facebook (FB).

Chiu et al. [20] followed a similar approach to analyze knowledge sharing in online communities in terms of quantity and quality. Their model is based on Nahapiet and Ghoshal's [53] three dimensions of social capital and outcome expectations on a personal and community level. There are three notable differences between Chiu et al.'s [20] approach and this study: (1) we focus on the influence of two aspects of outcome expectations, (2) we examine knowledge contribution from a marketing perspective in terms of giving product recommendations for both monetary and non-monetary rewards as well as considering product recommendations (product-related knowledge contribution) and (3) we compare to different types of social networking sites, which are ISNs and OSNs.

Based on the literature review, our hypotheses for recommendations with monetary and non-monetary rewards and considering recommendations for ISN and OSN will be postulated in the following section.

Personal outcome expectations (POE) in this study correspond to knowledge sharing about products and services in social networking sites. Strong *Personal Outcome Expectations* mean that users associate the sharing of recommendations with additional benefits for themselves. Some FB-members may care more about themselves than ASW-members. Not every member of both ASW and FB may associate personal outcome expectations with knowledge sharing about products and services. But if this is the case, *Personal Outcome Expectations* may have a stronger influence on *Considering Recommendations (REC)* and giving recommendations for *Monetary Rewards (MR)* and giving recommendations for *Non-monetary Rewards (NMR)* for FB-members than for ASW-members.

Members, who expect that knowledge sharing about products and services will add value to them, should support knowledge sharing in form of *giving recommendations for non-monetary rewards*. The well being of one is more important to FB-members than amongst ASW-members. Hence, we hypothesize:

H1a. Personal Outcome Expectations have a stronger positive influence on Non-monetary Rewards amongst OSN-members compared to ISN-members.

Rewards are helpful to generate more knowledge sharing. FB-members with strong *personal outcome expectations* should also

have a stronger willingness to *give recommendations for monetary rewards* compared to ASW-members.

H1b. Personal Outcome Expectations have a stronger influence on Monetary Rewards amongst OSN-members when compared to ISN-members.

Members, who are convinced that knowledge sharing is adding value to them, will expect that recommendations are honest and qualitative, so that they can use them without doubting in their purchasing decisions. Due to the assumption that FB-members are more focused of their own well being in contrast to the well being of the whole community, we hypothesize:

H1c. Personal Outcome Expectations have a stronger influence on Considering Recommendations from other members amongst OSN-members when compared to ISN-members.

Community-related Outcome Expectations (COE) in this study correspond to knowledge sharing about products and services in social networking sites. A high value means that users associate the sharing of recommendations with additional benefit for the community. It can be assumed that ASW-members care more about the communities' well being than FB-members but not every member may associate that with knowledge sharing about products and services. Therefore, *Community-related Outcome Expectations* should have a stronger relationship with *Considering Recommendations, Monetary Rewards* and *Non-monetary Rewards* on ASW than for FB.

Members, who expect that knowledge sharing will add value to the community, should support knowledge sharing in form of *giving recommendations for non-monetary rewards*. Since ASW-members are expected to behave more community-orientated, we hypothesize:

H2a. Community-related Outcome Expectations have a stronger positive influence on Non-monetary Rewards amongst ISN-members compared to OSN-members.

Rewards are helpful to generate more knowledge sharing. ASW-members with strong community-related outcome expectations should also have a strong willingness to *give recommendations for monetary rewards*. This effect should also be stronger for ASW-members than for FB-members.

H2b. Community-related Outcome Expectations have a stronger positive influence on Monetary Rewards amongst ISN-members compared to OSN-members.

Members, who are convinced that knowledge sharing is adding value to the community, will expect that recommendations are honest and qualitative, so that they can use them without doubting in their purchasing decisions. The effect is assumed to be stronger amongst ASW-members than amongst FB-members.

H2c. Community-related Outcome Expectations have a stronger positive influence on Considering Recommendations from other members amongst ISN-members compared to OSN-members.

4. RESULTS AND DISCUSSION

4.1 Research methodology

An online survey was conducted from December 2009 until January 2010 in order to evaluate the hypotheses. All potential participants were contacted via the internal messaging system of the respective social networking site. The partial least squares (PLS) structural equation modeling approach was used to validate the construct measures (the measurement model) and test the hypothesized relationships (the structural model). PLS was chosen as the appropriate methodology because it has minimal demands about the normality of the data and the sample size relative to covariance-based approaches [18]. The conceptual model was tested with the software implementation SmartPLS [59].

At the beginning of the survey, participants were asked to which social networking site they belong to, FB, ASW or both, so that they were only asked questions about the social networking site that they are a member of. Afterwards an explanation about product-related knowledge contribution with examples of usage in the chosen social networking site followed. General questions about online- and social networking site usage followed and the items of the conceptual model were tested for that community. All items were customized for each social networking site, FB and ASW, which has to be answered along a seven-point Likert scale ranging from "completely disagree" to "completely agree". Questions for the demographic characterization concluded the survey. In December 2009, the survey was pretested with each 10 FB- and ASW-members.

The measurement items for *Personal outcome expectations* (6 items) and *Community-related Outcome Expectations* (4 items) were adapted from Chiu et al. [20] and modified to fit to the product recommendation context rather than solely on knowledge contribution.

Since this thesis wants to examine the influence of social capital on product recommendations, the dependent variables were self-developed items (following the procedure of Moore and Benbasat [52])to assess the following three aspects:

1. *Non-monetary Rewards (NMR)* (1 item) measures the enjoyment of a member to help other members via giving product recommendations.
2. *Monetary Rewards (MR)* (3 items) measures the intention to give product recommendations if commissions, coupons or miles / points can be earned.
3. *Considering Recommendations (REC)* (3 items) measures the degree a member will consider a product recommendation of other members in her/his purchasing decision.

4.2 Description of the sample

305 completed questionnaires were submitted with 131 participants stating that they were ASW-members and 174 that they were FB-members. This leads to a response rate of 20% for ASW- and 70% for FB-members. Demographic details of the respondents for both samples are shown in Table 1.

Among the participants both samples were balanced with slightly more male than female respondents. The ASW-member samples are slightly older than the FB sample with a mean age of 32 (median = 32, standard deviation =7.54) for ASW and a mean age of 30 (median = 29, standard deviation =7.88) for FB. One obvious characteristic of the respondents is that the large majority is highly educated. 92% of ASW-members and 79% of the FB-members have a Bachelor's degree or higher education. The demographics also show that ASW-members are more affluent than FB-members. 40% of the ASW-members and 37% of the FB-members refused to indicate their income. The other results showed that 39% of the ASW-members have a net monthly household income of € 5,000 and more which is also the median.

Regarding the income the biggest groups amongst the FB respondents € 1,001 - € 2,000 (17%) and more than € 5,000 (21%), with a median between € 3,001 and € 4,000. The high income of ASW respondents matches with previous internal member survey of ASW where the yearly median of the household income was $ 139,400.

Table 1: Demographic profile of the sample

Measure	Items	ASW Freq. (%)	FB Freq. (%)
Gender	Male	75 (57%)	96 (55%)
	Female	56 (43%)	78 (45%)
Age	<25	7 (5%)	28 (16%)
	26-35	85 (65%)	108 (62%)
	36-45	34 (26%)	30 (17%)
	>45	2 (2%)	3 (2%)
	n/a	3 (2%)	5 (3%)
Education	High school or below	2 (2%)	13 (7%)
	Apprenticeship	4 (3%)	14 (8%)
	Bachelor's Degree	42 (32%)	48 (28%)
	Master's Degree	74 (56%)	79 (45%)
	Doctoral Degree	5 (4%)	10 (6%)
	Other	4 (3%)	10 (6%)
Income	< € 1,000	3 (2%)	11 (6%)
	€ 1,001 to € 2,000	6 (5%)	29 (17%)
	€ 2,001 to € 3,000	7 (5%)	18 (10%)
	€ 3,001 to € 4,000	6 (5%)	12 (7%)
	€ 4,001 to € 5,000	5 (4%)	3 (2%)
	> € 5,001	51 (39%)	37 (21%)
	n/a	53 (40%)	64 (37%)
N		131	174

This survey also indicated that the respondents usually find out about new products and services that are relevant to themselves among others through a friend (ASW: 77%, FB: 80%), a website (ASW: 70%, FB: 74%), an online forum or a social networking site (ASW: 35%, FB: 29%). Only 23% of ASW-members stated that the recommendations on SNS do not influence their purchasing decisions, compared to 55% of FB-members. ASW-members mostly are influenced by recommendations for dining out (61%), hotels and airline tickets (58%) and events (51%). FB-members mostly get influenced by recommendations for events (39%).

4.3 Common method bias analysis

Common method bias is a potential problem for internal validity and usually the key source for measurement errors. Especially, self-reported data in surveys conducted with the same measurement context is possibly leading to errors [56]. Following the procedure recommended by Podsakoff et al. [56] and Liang et al. [48], a common method bias construct was integrated into the PLS research model with all the indicators used. The variances explained by the common method bias construct were computed relative to the variances explained by the substantive constructs. In our model the average variance explained by the substantive constructs is for ASW 0.612 and 0.681 for FB, while the average variance explained by the common method construct is 0.012 for ASW and 0.014 for FB. The method variance values are very low, which leads to the conclusion that the common method bias is not influencing the results of the research model.

4.4 Measurement model validation

All constructs in our model are measured in the reflective mode [42]. The quality of reflective constructs is determined by (1) convergent validity and (2) discriminant validity [10].

Convergent validity is assessed in two ways: (1) The indicator reliability and (2) the internal consistency. For the indicator reliability all indicators loaded significantly at least at the 0.01 level and all indicators met the suggested threshold of 0.707 [18]. Internal consistency is estimated by analyzing the composite reliability (CR), Cronbach's alpha and the average variance extracted (AVE) (see Table 2) [65]. All the values for CR and Cronbach's alpha exceeded the recommended threshold of 0.7 [55] and AVE the critical level of 0.5 [32]. A consolidated view indicated that the constructs fulfill all requirements for indicator reliability and internal consistency and therefore validate their convergent validity.

Table 2: Convergent validity of the reflective constructs

Construct (No. of Items)	CR	Cronbach's Alpha	AVE
ASW (FB)			
COE (4)	0.93 (0.95)	0.90 (0.94)	0.78 (0.84)
MR (3)	0.97 (0.99)	0.95 (0.94)	0.91 (0.88)
NMR (1)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
POE (6)	0.95 (0.96)	0.94 (0.95)	0.76 (0.81
REC (3)	0.91 (0.93)	0.85 (0.89)	0.77 (0.82)

Discriminant validity states to which degree a given construct differs from other constructs. It was analyzed by examining whether indicators load higher on their own constructs than on other constructs. Additionally, the square root of the AVE from the indicator should be higher than the correlations between constructs [62]. Furthermore, none of the correlations between a pair of constructs should be higher than the threshold of 0.9 [11]. Our model satisfies these criteria. Additionally the loadings of the indicators of the specific construct are always exceeding with this construct compared with others) [65], which also confirms discriminant validity.

4.5 Structural model validation

After assessing the measurement model, the explanatory power for each structural model was analyzed. The ASW-model explains 10.6% of the variance (R^2) in *Monetary Rewards*, 34.3% in *Non-monetary Rewards* and 44.1% in the latent variable *Considering Recommendations*. Whereas the FB-model explains 13.0% of the

variance (R^2) in *Monetary Rewards*, 38.5% in *Non-monetary Rewards* and 40.2% in the latent variable *Considering Recommendations*.

The theoretical model and hypothesized relationships were estimated using the bootstrapping procedure implemented in SmartPLS with 1,000 iterations. To examine the specific paths we assessed t-statistics for the calculated p-values based on two-tailed significance levels of 0.05. The results for ASW- and FB-members are summarized in Figure 1.

Four out of six paths for the ASW-model and five out of six paths for the FB-model exhibited a p-value less than 0.05 for bidirectional paths.

The ASW-model shows a positive and strongly significant influence of a members' *Personal Outcome Expectations* on all three dependent variables. The results display an insignificant path between members' *Community-related Outcome Expectations* and *Monetary Rewards*, while the path to *Non-monetary Rewards* (0.338; p<0.01) and *Considering Recommendations* (0.512; p<0.01) were both positive and high significant.

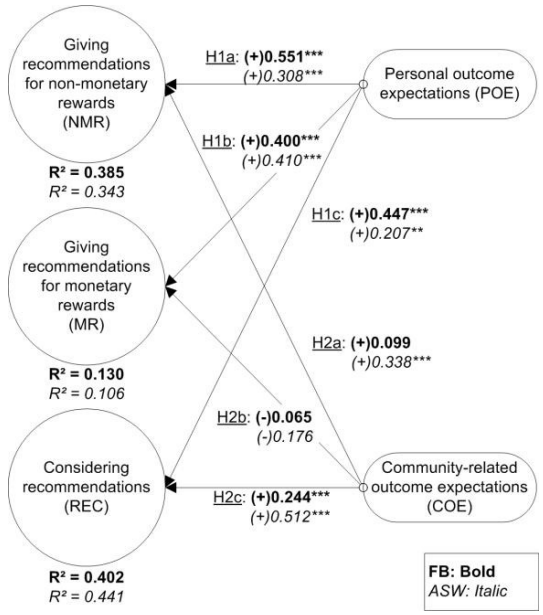


Figure 1: Path model with results for ASW and FB (*p < 0.01, **p < 0.05, p* < 0.1)**

The effect size f^2 indicates the importance of each influencing factor (see Table 3). All significant constructs in this model have at least a weak effect [18].

Table 3: Effect size f^2 of the outcome expectations

	f^2 REC		f^2 NMR		f^2 MR	
	ASW	**FB**	**ASW**	**FB**	**ASW**	**FB**
COE	0.27	0.06	0.10	0.01	0.02	0.01
POE	0.05	0.19	0.09	0.28	0.11	0.10

4.6 Comparison of path coefficients between OSNs and ISNs

The results of the structural model indicate that the influences of the independent constructs on the dependent latent variables differ between both models. To test whether there are the significant differences between the influencing constructs, a multi-group comparison is conducted by the PLS bootstrapping routine for each sub sample.

Chin et al. [19] argued that multi-group comparison with PLS is relatively naïve especially because of differences in path estimates for different sampled population, however, previous research applied multi-group comparison with PLS. Eckhardt et al. [27] analyzed the impact of social influence on IT adoption and non-adoption. Dibbern and Chin [25] evaluated a structural equation model and applied multi-group comparison for cultural differences in Germany and the USA. Based on these approaches, the hypotheses that there are different influencing factors for ASW- and FB-members to recommend products and services will be tested. 1,000 ß-coefficients for each sub-sample were generated with PLS bootstrapping and a t-test with the generated ß-coefficients was performed to test for significant differences and to verify each hypothesis. The results of the t-test (see table 4) show significant differences on a p<0.05 level between ISNs and OSNs for all constructs.

The t-test for mean equality for the *Personal Outcome Expectations* model indicates significant differences in the influence on the constructs *Considering Recommendations*, *Monetary Rewards* and *Non-monetary Rewards*. For the constructs *Non-monetary Rewards* and *Considering Recommendations*, the t-test indicated that *Personal Outcome Expectations* have a significantly higher influence amongst FB-members than it is observable amongst ASW-members, supporting hypotheses 1a and 1c. In contrast, for the construct *Monetary Rewards* the results showed a significantly stronger positive influence amongst ASW-members, which leads to the rejection of hypothesis 1b.

Table 4: T-test for mean equality

	Levene-test (F-value)	Mean Differences	
COE → MR	10.789***	-0.114***	ASW>FB
COE → NMR	30.157***	0.234***	ASW>FB
COE → REC	17.141***	0.256***	ASW>FB
POE → MR	0.527	0.125***	ASW>FB
POE → NMR	34.219***	-0.237***	FB>ASW
POE → REC	11.264***	-0.231***	FB>ASW

*** p<0.01; ** p<0.05; * p<0.1

The t-test for mean equality for the construct *Community-related Outcome Expectations* shows significant differences in its influence on all of the three dependent constructs. The effect of *Community-related Outcome Expectations* on *Non-monetary Rewards* and *Considering Recommendations* is significantly stronger for ASW-members than for FB-members, supporting hypotheses 2a and 2c. As already shown in the structural model *Community-related Outcome Expectations* has no significant effect on *Monetary Rewards* for both ASW- and FB-members.

Taking a closer look at the path coefficients, it is observable that the tendency of influence additionally is negative and not positive as stated in hypothesis 2b. The t-test of mean equality results in a significantly stronger negative influence for FB-members than for ASW-members, which also controverts hypothesis 2b. Summarizing these results hypothesis 2b has to be rejected.

4.7 Discussion

Our research question was to identify whether a product recommendation system with monetary or non-monetary rewards should be implemented in OSNs and ISNs, which will be evaluated in the following.

The members' *Personal Outcome Expectations* show a highly significant positive influence for FB and ASW on all three constructs *Non-monetary Reward, Monetary Rewards*, as well as *Considering Recommendations*. The result regarding the positive influence on *Considering Recommendations* is in line with the results of Aral and Walker [3] who found that active-personalized Word-of-Mouth-messages effectively increase the attention of the receiver. Thus, one can argue, that a higher attention also increases the consideration of a recommendation in the receiver's purchasing decision.

The mean comparison supports that *Personal Outcome Expectations* have a stronger influence on *Non-monetary Rewards* and *Considering Recommendations* for FB-members compared to ASW-members, but surprisingly not for *Monetary Rewards*, which is contrary to the assumption in our hypothesis. In that case, ASW shows a stronger influence. Thus, ASW-members with high *Personal Outcome Expectations* demand a monetary benefit even more than FB-members. This result can be explained on the basis of the findings of Forman et al. [31] who showed that more information about the recommender increases the value of the recommendation for its receiver. The members of ISNs are in closer connection to each other, which implies that they better know their contacts in the ISN. The value of a recommendation increases and, thus, the recommender is in the position to demand a monetary reward without the risk of losing reputation.

Community-related Outcome Expectations strongly influence *Non-monetary Rewards* for ASW-members but show an insignificant (positive) influence for FB-members. The mean comparison supported that the influence of *Community-related Outcome Expectations* on *Non-monetary Rewards* is significantly stronger for ASW-member than for FB-members. Hence, ASW-members seem to care more about the well-being of the community and enjoy adding value by giving recommendations. Contrary to our expectations, *Community-related Outcome Expectations* had a stronger influence on *Monetary Rewards* for FB-members compared to ASW-members and both paths showed a negative but insignificant influence. One plausible explanation is that members who care for the well-being of the community dislike when members may only recommend for monetary rewards. The effect is stronger for FB than for ASW. *Community-related Outcome Expectations* strongly influence *Considering Recommendations* for the individual paths of ASW and FB and also support the stronger influence of this construct for ASW-members compared to FB-members. Hence, ASW-members regard product recommendations more as valuable and qualitative community-outcome, which will be considered in their purchasing decisions, than FB-members.

In case of product recommendations which are contributed only because of monetary benefit, we suggest, based on our results, to implement a product recommendation system with non-monetary rewards in both OSNs and ISNs. In case of authentic product recommendations, where other members can rely on, a monetary rewarded product recommendation system can successfully be implemented in ISNs. Here, trust between members is distinctively existent and the strong significant positive influences of *Community-related Outcome Expectations* on *Considering Recommendations* and *Personal Outcome Expectations* on *Monetary Rewards* can be capitalized.

5. CONCLUSION

Social networking sites have become very popular for Internet users and give marketers the chance to target a variety of demographic profiles easy and cost efficiently. Furthermore, consumers show increased trust in opinions posted in online channels. Most of the social networking sites rely on knowledge contribution of their members. The challenge is to identify the factors that lead to knowledge contribution in form of product recommendations and the underlying process that enables to direct their advertising strategies to the consumers. Chiu et al. [20] proposed, as a future research to analyze the usefulness and sort of reward systems, intrinsic or extrinsic, which motivates individual's to share knowledge in online communities. We examined non-monetary rewards such as enjoyment of helping as an intrinsic reward and monetary rewards such as commissions, coupons, miles or points as extrinsic rewards to share recommendations in social networking sites. The results of this study imply that regarding the two facets of outcome expectations (*Personal Outcome Expectations* and *Community-related Outcome Expectations*) individuals will behave with a different impact to achieve desirable outcomes depending on the type of social networking site they are using.

Our results show that for FB-members, *Personal Outcome Expectations* have a significant higher positive effect on *Non-monetary Rewards* and *Considering Recommendations* than ASW-members. In contrast ASW-Members show significant stronger influences in the impact of *Community-related Outcome Expectations* on all dependent variables and additionally in the influence of *Personal Outcome Expectations* on *Monetary Rewards*. In contrast *Personal Outcome Expectations* amongst FB-Members are stronger connected to *Considering Recommendations* and *Non-monetary Rewards* compared to ASW-members. Thus, FB-Members seem to have a more salient focus on their own benefits than on the benefits for their community. Due to the significant higher path coefficients between *Community-related Outcome Expectations* and the dependent variables, ASW-members seem to care more for the community instead of their own benefits.

Future studies may examine which minimum amount must be offered within a monetary-rewarded recommendation system to motivate the consumer to recommend products. Since consumers can also be member in OSNs as well as ISNs at the same time, it should be examined whether these members show a different response behavior in terms of their evaluation of the constructs of our research model depending on the social networking site.

6. REFERENCES

[1] Anderson, E.W. 1998. Customer Satisfaction and Word of Mouth. *Journal of Service Research*, 1, 1, 5-17.

[2] Andreasen, A.R. 1968. Attitudes and Customer Behavior: A Decision Model. In, Kassarjian, H.H., and Robertson, T.S., (eds.), *Perspectives in Consumer Behavior*, Glenview, IL: Scott, Foresman and Company, pp. 498-510.

[3] Aral, S., and Walker, D. 2010. Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *NYU Stern Working Paper*, 1-44.

[4] Archak, N., Ghose, A., and Ipeirotis, P.G. 2009. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 55, 1, 47-57.

[5] Armstrong, A., and Hagel, J. 1996. The Real Value of Online Communities. *Harvard business review*, 74, 3, 134-142.

[6] Arndt, J. 1967. Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research (JMR)*, 4, 3, 291-295.

[7] ASmallWorld 2010. About. http://www.asmallworld.net/about, 2010/04/01.

[8] ASmallWorld 2010. Forums. http://www.asmallworld.net/forums, 2010/04/01.

[9] ASmallWorld 2010. Membership. http://www.asmallworld.net/membership, 2010/04/01.

[10] Bagozzi, R.P., and Yi, Y. 1988. On the Evaluation of Structural Equation Models. *Journal of the Academy of Marketing Science*, 16, 1, 74-94.

[11] Bagozzi, R.P., Yi, Y., and Phillips, L.W. 1991. Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, 36, 3, 421-458.

[12] Blau, P.M. 1964. *Exchange and Power in Social Life*. New York: Wiley, 1964.

[13] Bock, G.W., Zmud, R.W., Kim, Y.G., and Lee, J.N. 2005. Behavioral Intention Formation in Knowledge Sharing: Examining the Roles of Extrinsic Motivators, Social-Psychological Forces, and Organizational Climate. *MIS Quarterly*, 29, 1, 87-111.

[14] Boyd, D.M., and Ellison, N.B. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer Mediated Communication*, 13, 1, 210-221.

[15] Brown, J., Broderick, A.J., and Lee, N. 2007. Word of Mouth Communication within Online Communities: Conceptualizing the Online Social Network. *Journal of Interactive Marketing*, 21, 3, 2-20.

[16] Brown, J.J., and Reinigen, P.H. 1987. Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, 14, 4, 350-362.

[17] Butler, B.S., Sproull, L., Kiesler, S., and Kraut, R. 2002. Community Effort in Online Groups: Who does the Work and Why? In, Weisband, S., and Atwater, L., (eds.), *Leadership at a Distance*, Mahwah: Erlbaum.

[18] Chin, W.W. 1998. The Partial Least Squares Approach to Structural Equation Modeling. In, Marcoulides, G.A., (ed.), *Modern Methods for Business Research*, Mahwah: Lawrence Erlbaum, pp. 295-336.

[19] Chin, W.W., Marcolin, B.L., and Newsted, P.R. 2003. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. *Information Systems Research*, 14, 2, 189-217.

[20] Chiu, C.-M., Hsu, M.-H., and Wang, E.T.G. 2006. Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories. *Decision Support Systems*, 42, 3, 1872-1888.

[21] Compeau, D.R., and Higgens, C.A. 1995. Computer Self-efficiency: Development of a Measure and Initial Test. *MIS Quarterly*, 19, 2, 189-211.

[22] Constant, D., Kiesler, S., and Sproull, L. 1994. What's mine is Ours, or Is It? A Study of Attitudes about Information Sharing. *Information Systems Research*, 5, 4, 400-422.

[23] Constant, D., Kiesler, S., and Sproull, L. 1996. The Kindness of Strangers. *Organization Science*, 7, 2, 119-135.

[24] Deci, E.L., and Ryan, R.M. 1980. The Empirical Exploration of Intrinsic Motivational Processes. In, Berkowitz, L., (ed.), *Advances in Experimental Social Psychology*, New York: Academic Press, pp. 39-80.

[25] Dibbern, J., and Chin, W.W. 2005. Multi-Group Comparison Testing a PLS Model on Sourcing of Application Software Services across Germany and the USA Using a Permutation Based Algorithm. In, Bliemel, F., Eggert, A., Fassot, G., and Henseler, J., (eds.), *Handbuch PLS Pfadmodellierung*, Stuttgart: Schaeffer-Poeschel.

[26] Dougherty, H. 2010. Facebook Reaches Top Ranking in US. http://weblogs.hitwise.com/heather-dougherty/2010/03/facebook_reaches_top_ranking_i.html, 2010/08/23.

[27] Eckhardt, A., Laumer, S., and Weitzel, T. 2009. Who Influences Whom? Analyzing Workplace Referents' Social Influence on IT Adoption and Non-Adoption. *Journal of Information Technology*, 24, 1, 11-24.

[28] Facebook 2010. Factsheet. http://www.facebook.com/press/info.php?factsheet, 2010/08/23.

[29] Facebook 2010. Product. http://www.facebook.com/press/product.php, 2010/08/23.

[30] Facebook 2010. Statistics. http://www.facebook.com/press/info.php?execbios#!/press/info.php?statistics, 2010/08/23.

[31] Forman, C., Ghose, A., and Wiesenfeld, B. 2008. Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets. *Information Systems Research*, 19, 3, 291-313.

[32] Fornell, C., and Larcker, D.F. 1981. Structural Equation Models With Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, 18, 3, 382-388.

[33] Frank, R. 2007. MySpace for Millionaires. *The Wall Street Journal*, W2.

[34] Goldsmith, R.E., and Horowitz, D. 2006. Measuring Motivations for Online Opinion Seeking. *Journal of Interactive Advertising*, 6, 2, 1-16.

[35] Gross, R., and Acquisti, A. 2005. Information Revelation and Privacy in Online Social Networks. *Workshop on Privacy in the Electronic Society*, Alexandria, VA: ACM Press.

[36] Hagel, J., and Armstrong, A.G. 1997. *Net Gain: Expanding markets Through Virtual Communities*. Boston, MA, 1997.

[37] Hargittai, E. 2007. Whose Space? Differences Among Users and Non-Users of Social Network Sites. *Journal of Computer-Mediated Communication*, 13, 1, 276-297.

[38] Hennig-Thurau, T., Gwinner, K.P., Walsh, G., and Gremler, D.D. 2004. Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing*, 18, 1, 38-52.

[39] Herr, P.M., Kardes, F.R., and Kim, J. 1991. Effects of Word of Mouth and Product Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective. *Journal of Consumer Research*, 17, 4, 454-462.

[40] Huber, G.P. 2001. Transfer of Knowledge in Knowledge Management Systems: Unexplored Issues and Suggested Studies. *European Journal of Information Systems*, 10, 72-79.

[41] Internet World Business 2009. Die größten Webseiten im Netz., http://www.internetworld.de/Nachrichten/Zahlen-Studien/die-groessten-Webseiten-im-Netz-20539.html, 2010/08/23.

[42] Jarvis, C.B., Mackenzie, S.B., and Podsakoff, P.M. 2003. A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, 30, 2, 199-218.

[43] Kankanhalli, A., Tan, B.C.Y., and Wei, K.K. 2005. Contributing Knowledge to Electronic Knowledge Repositories: An Empirical Investigation. *MIS Quarterly*, 29, 1, 113-143.

[44] Katz, E., and Lazarsfeld, P.F. 1955. Personal Influence: The Part Played by People in the Flow of Mass Communications. New York: Free Press.

[45] Kelley, H.H., and Thibaut, J.W. 1978. *Interpersonal Realtions: A Theory of Interdependence*. New York: John Wiley, 1978.

[46] Kim, J.W., Choi, J., Qualls, W., and Han, K. 2008. It Takes a Marketplace Community to Raise Brand Commitment: The Role of Online Communities. *Journal of Marketing Management*, 24, 3-4, 409-431.

[47] Lampe, C., Ellison, N.B., and Steinfield, C. 2008. Changes in Use and Perception of Facebook. *ACM Conference on Computer-Supported Cooperative Work (CSCW)*, San Diego, CA: ACM Press, pp. 721-730.

[48] Liang, H., Saraf, N., Hu, Q., and Xue, Y. 2007. Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. *MIS Quarterly*, 31, 1, 59-87.

[49] Ma, M., and Agarwal, R. 2007. Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities. *Information Systems Research*, 18, 1, 42-67.

[50] Miller, K.D., Fabian, F., and Lin, S.-J. 2009. Strategies for Online Communities. *Strategic Management Journal*, 30, 305-322.

[51] Molm, L.D. 1997. *Coercive Power in Social Exchange*. New York: Cambridge University Press, 1997.

[52] Moore, G.C., and Benbasat, I. 1991. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*, 2, 3, 192-222.

[53] Nahapiet, J., and Ghoshal, S. 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review*, 23, 2, 242-266.

[54] Nielsen 2009. Global Online Consumer Survey. http://blog.nielsen.com/nielsenwire/global/online-advertising-has-work-to-do-to-elicit-consumer-trust/, 2010/08/23.

[55] Nunally, J.C., and Bernstein, I.H. 1978. Psychometric Theory. New York: McGraw-Hill.

[56] Podsakoff, P.M., MacKenzie, S.B., Jeong-Yeon, L., and Podsakoff, N.P. 2003. Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88, 5, 879-903.

[57] Raymond, E.S. 2003. The Hacker Milieu as Gift Culture. http://futurepositive.synearth.net/gift-economy/, 2010/08/06.

[58] Rheingold, H. 1993. *The Virtual Community: Homesteading on the Electronic Frontier*. Reading MA: Addison-Wesley, 1993.

[59] Ringle, C.M., Wende, S., and Will, A. 2006. SmartPLS. University of Hamburg.

[60] Senecal, S., and Nantel, J. 2004. The Influence of Online Product Recommendations on Consumers' Online Choices. *Journal of Retailing*, 80, 2, 159-169.

[61] Smith, D., Menon, S., and Sivakumar, K. 2005. Online Peer and Editorial Recommendations, Trust, and Choice in Virtual Markets. *Journal of Interactive Marketing*, 19, 3, 15-37.

[62] Son, J.-Y., and Benbasat, I. 2007. Organizational Buyers' Adoption and Use of B2B Electronic Marketplaces: Efficiency- and Legitimacy-Oriented Perspectives. *Journal of Management Information Systems*, 24, 1, 55-99.

[63] Taylor, G.S., Brown, D., Nelson, L., Longton, J., Gassman, T., Cohen, J., Swartz, J., Horner, R., Sugai, G., and Hall, S. 1997. School-wide Behavioral Support: Starting the Year Off Right. *Journal of Behavioral Education*, 7, 1, 99-112.

[64] Taylor, N.J. 2007. Public grid computing participation: An exploratory study of determinants. *Information & Management*, 44, 1, 12-21.

[65] Teo, H.H., Wei, K.K., and Benbasat, I. 2003. Predicting Intention to Adopt Interorganizational Linkages: An Institutional Perspective. *MIS Quarterly*, 27, 1, 19-50.

[66] Vallerand, R.J. 1997. Toward a Hierarchical Model of Intrinsic and Extrinsic Motivation. *Advances in Experimental Social Psychology*, 29, 271-360.

[67] Wasko, M.M., and Faraj, S. 2005. Why should I share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly*, 29, 1, 35-57.

[68] Zhang, J.Q., Craciun, G., and Shin, D. 2010. When does Electronic Word-of-Mouth Matter? A Study of Consumer Product Reviews. *Journal of Business Research*, 63, 12, 1336-1341.

Revealing Business Relationships – Eavesdropping Cross-organizational Collaboration in the Internet of Services

André Miede
Multimedia Communications Lab (KOM) – TU Darmstadt
Rundeturmstraße 10, 64283 Darmstadt, Germany
miede@KOM.tu-darmstadt.de

Gökhan Şimşek
Multimedia Communications Lab (KOM) – TU Darmstadt
Rundeturmstraße 10, 64283 Darmstadt, Germany
simsek@KOM.tu-darmstadt.de

Stefan Schulte
Multimedia Communications Lab (KOM) – TU Darmstadt
Rundeturmstraße 10, 64283 Darmstadt, Germany
schulte@KOM.tu-darmstadt.de

Daniel F. Abawi
Hochschule für Technik und Wirtschaft des Saarlandes
Waldhausweg 14, 66123 Saarbrücken, Germany
abawi@htw-saarland.de

Julian Eckert
Multimedia Communications Lab (KOM) – TU Darmstadt
Rundeturmstraße 10, 64283 Darmstadt, Germany
eckert@KOM.tu-darmstadt.de

Ralf Steinmetz
Multimedia Communications Lab (KOM) – TU Darmstadt
Rundeturmstraße 10, 64283 Darmstadt, Germany
steinmetz@KOM.tu-darmstadt.de

ABSTRACT

The Internet of Services is envisioned as a global Service-oriented Architecture enabling collaboration across organizational boundaries. However, by monitoring communication endpoints, attackers can create detailed profiles of service consumers and providers even if typical security mechanisms such as message encryption are used. In a business context, this traffic analysis threatens the relationship anonymity of the participants and can reveal sensitive information about an organization's underlying business processes or a service provider's client base. In this paper, we discuss the simulation-based evaluation of different attack scenarios regarding the identification of the service compositions an organization uses. Thus, we offer insights regarding the limits of anonymity for cross-organizational collaboration in the Internet of Services.

Keywords

Security, Anonymity, Internet of Services, Service-oriented Architectures, Cross-organizational Collaboration

1. INTRODUCTION

Modern global economies have become fast-paced and highly competitive, thus, requiring organizations to adapt both quickly and continuously to changing circumstances and requirements. An important factor to achieve this goal is the underlying enterprise Information Technology (IT), which has to integrate both internal and external systems.

The paradigm of *Service-oriented Architectures* (SOAs) [22] offers technological and organizational means in order to improve the alignment between the functional and the IT side, i. e., by enabling *service-based, cross-organizational workflows*. In the last years, *Web services* have become both a mature and successful technology for implementing the SOA paradigm.

For the near future, the *Internet of Services* is envisioned as a global SOA further facilitating cross-organizational collaboration [4,26]. The Internet of Services provides the foundation for complex business value networks by supporting the composition and aggregation of existing services to value-added services, i. e., using market places as intermediaries between service consumers and providers. Furthermore, it is a business model using the Internet as a medium for the retrieval, combination, and utilization of interoperable services. For example, market places could build compositions using services from different providers and offer these compositions as best practices for recurring process needs to service consumers.

In order to enable such service-based, cross-organizational collaboration, the security of the communication channels used, exchanged messages, and participating systems is a necessity. Regarding the security of Web service technology, substantial advancements have been achieved in the last years as discussed in the standard literature on Web service security [3,12,25]. However, several technology-independent and service-specific attacks on SOA have been identified recently, especially in the Internet of Services context [17,18].

One of these attacks aims at identifying the existence of relationships between collaborating organizations: By ob-

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

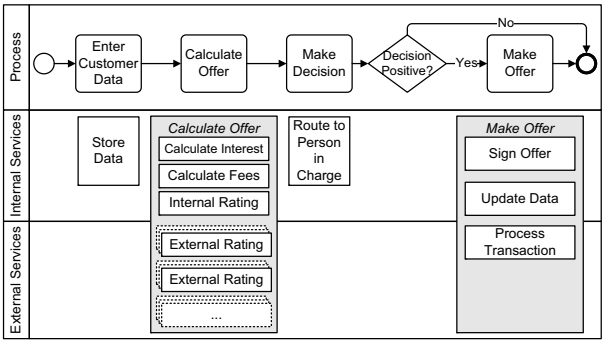


Figure 1: A fragment of a generic credit application process and possible mappings to both internal and external services.

serving the communication between the participants in the Internet of Services, attackers can create detailed profiles of service consumers, providers, and also of market places. Because only the message exchange endpoints have to be monitored, the use of encryption or other standard mechanisms is no protection against this kind of attack, which is in general communication networks also known as "traffic analysis" [24]. In addition, due to its passive nature, it is hard to detect and depending on the monitoring means used this attack may not even be illegal. However, the obtained information reveals important organizational details, e. g., consumers exploring new business opportunities, the anticipation of mergers and acquisitions, or providers changing their business models.

The security goal that is threatened by this attack is referred to as "relationship anonymity" in the standard literature on anonymity research [23]. This means that an adversary cannot sufficiently distinguish whether the sender and recipient of a particular message are related or not. It is important to understand that this kind of anonymity does not apply to the sender and recipient of the message, i. e., they know each other. It refers only to third parties, i. e., parties that are neither sender nor recipient of the message.

A simple, but tangible example from the financial services domain is a generic credit application process, i. e., where credit ratings for customers are retrieved from an external rating agency. More about such an example is shown in Figure 1: The bank works on credit applications from its customers, e. g., first entering and storing the customer's data in its systems (using an internal access service "Store Data"). A possible next step would be calculating the concrete credit offer, which is a service composition "Calculate Offer" consisting of both internal and external services, e. g., for external credit history ratings. After that, another service would route the resulting information to a human decider "Route to Person in Charge" for triggering the next steps. Finally, if this person's decision is positive, the offer would be made. This results in another service composition "Make Offer", again consisting of both internal and external services, e. g., for notifying the customer about the decision, processing the payout of the credit, etc.

However, just by monitoring the message exchange between the bank and the rating agency, an attacker can gain information about how many people apply for credit, when peak times are, when the bank works on the applications, and so on. If complete or parts of service compositions can be monitored, e. g., if information about successful credit applications is transferred to mailing and payout services via Web services, attackers could also infer a percentage of how many applications are granted or denied. This is very sensitive information rather easily available for attackers and it is not protected by the common and currently used security technology, i. e., for Web services.

A comprehensive overview of mechanisms and systems in order to achieve different types of anonymity in communication networks is given, e. g., by Edman and Yener [8]. However, even if such standard anonymity mechanisms are deployed and used correctly, attacks mounted at the edges of such networks and aiming at typical long-term business relationships are very likely to be successful. Thus, the goal of the paper at hand is the following: We investigate how an adaptation of a typical anonymity attack with respect to service compositions threatens the relationship anonymity of service consumers and providers in the Internet of Services. This is done by measuring the attacker's success using metrics from the field of Information Retrieval while varying key system parameters such as the number of service providers, the composition complexity, or the number of observed collaborations.

The rest of the paper is structured as follows: Section 2 and 3 outline the analysis and design of our evaluation, i. e., how it was set-up and why we made certain design decisions. Subsequently, Section 4 analyzes and discusses selected results. In order to place our contributions within the body of existing research, Section 5 discusses the most relevant related work in this area. Section 6 sums up the findings and closes with a brief outlook on future work.

2. ANALYSIS

In this section, the foundation of our research is presented, i. e., the underlying assumptions, the research question, and reasons for the selected means to answer this question.

2.1 Attack Selection

First of all, how does a typical attack on anonymity and anonymity systems look like and where has it to be mounted? Assuming that basic anonymity systems are being used by the communicating participants, there are two basic choices for an attack:

1. *Attack the anonymity network itself*: The attacker tries to follow the trail of a message along the nodes of an anonymity system, e. g., as described by Guan et al. [9] (cf. Figure 2). However, this is very difficult because of the (usually) high number of participating nodes and the used security mechanisms. Thus, many nodes would have to be compromised in order to cover the whole route. In addition, a single missing node on the route makes this kind of attack even more difficult, because messages are hard to correlate between the nodes.

2. *Attack the anonymity network edges, i. e., incoming and*

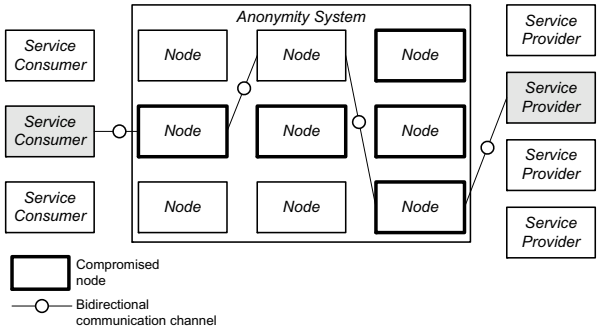


Figure 2: Attack the anonymity system itself.

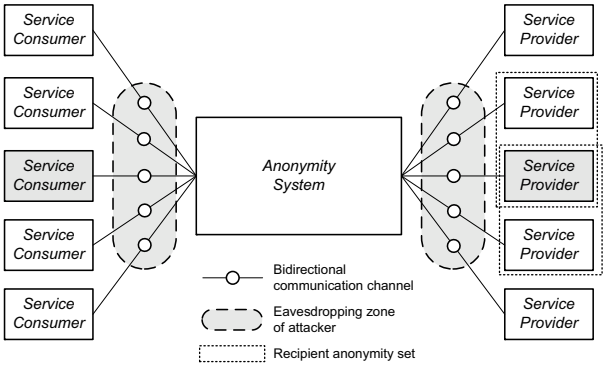


Figure 3: Attack the edges of the anonymity system.

outgoing messages (cf. Figure 3): This kind of attack focuses on the communication relationships of the participants and not on the anonymity network itself. These attacks are called "Intersection/Disclosure Attacks" and were introduced by Kesdogan et al. [13]. Their name is based on the intersection of the anonymity sets of senders and recipients for each communication round. However, this intersection was proven to be equal to a known NP-complete problem, thus, a statistical variant was developed by Danezis [5], reducing the required resources for the attack while still achieving good results.

Unlike the standard disclosure attacks, the so-called "*Statistical Disclosure Attack*" does not provide the attacker with definite information about the communication relationships but with a probability of each potential relationship. Basically, this requires the attacker to observe a large amount of interactions from which he can calculate the respective relationship probabilities. Selected details of these calculations are discussed as part of our simulation model in Section 3.2.

We chose this attack type for our evaluation, because it is a particular threat for strategic, long-term relationships, i. e., relationships that are custom in the field of service-based cross-organizational collaboration. In addition, the attack is basically independent from the used anonymity system, thus, based on certain assumptions that will be outlined below, it is a threat for most deployed anonymity systems.

For this paper, the Statistical Disclosure Attack is adapted for the Internet of Services scenario, i. e., attackers aim to identify the service compositions that organizations use for executing their processes. More details on these adaptations are given in Section 3.

2.2 Research Question

As outlined above, we assume for our research that organizations use external services (and compositions thereof) for executing their processes. Furthermore, we assume that basic countermeasures against traffic analysis are in place, thus, non-trivial attacks are needed because an attacker cannot just intercept any message in order to retrieve sender and recipient information from its header.

From this and the selected attack type follows the research question we try to answer in this paper: *"How does an adaptation of the Statistical Disclosure Attack regarding service compositions threaten the relationship anonymity of service consumers and providers in the Internet of Services?"*

For answering this question, we varied key system parameters such as the number of service providers, the composition complexity, or the number of observed collaborations and measured the attacker's success using metrics from the field of Information Retrieval. We chose a simulation-based approach because it will be an initial indicator for how well business relationships in the Internet of Services are protected against dedicated attacks. The reasons for using a simulation instead of, e. g., a testbed, are the following:

An important reason is noise reduction, i. e., dealing with specific technology (and its integration challenges) for implementing a cross-organizational testbed distracts from the attacker's focus, which is on revealing relationships between organizations. Furthermore, the attacks need to be performed in a completely controlled environment. Our focus is not on providing a real implementation of such attacks but on developing a model that gives us information about how dangerous they are. In addition, although enabling technologies such as Web services exist, a truly global SOA such as the Internet of Services is not yet available and, thus, cannot serve as a foundation for investigating dedicated attack scenarios.

3. SIMULATION DESIGN AND SETUP

This section discusses the underlying design decisions of our simulation model, i. e., the general assumptions, an overview of the model, brief implementation information, and the different evaluated configurations are presented.

3.1 General Simulation Assumptions

For our simulation model, we assume the following regarding the different entities: The system uses end-to-end encryption that cannot be broken in time. Furthermore, it delivers messages to recipients in batches, e. g., using a so-called "Threshold Mix" [5, 13].

The attacker is passive and static, i. e., the attacker observes only and does not adapt his attack behavior. In addition, he can observe messages leaving and entering the network (not necessarily *all* messages) and can guess when a message

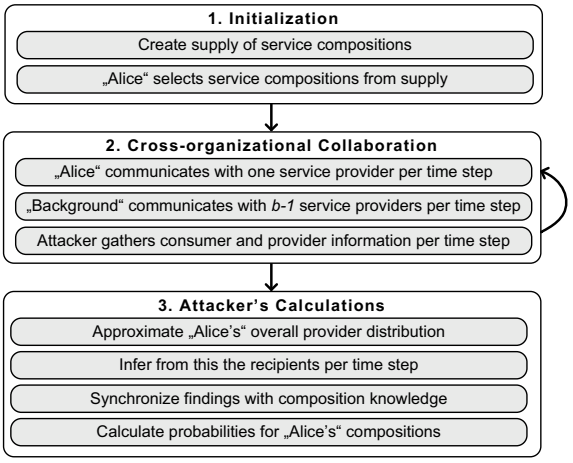


Figure 4: Schematic flow of the simulation model.

entering is likely to leave. Furthermore, the attacker knows the anonymity system's parameters (e.g., batchsize) and the market place's offerings, i. e., what compositions are available and what providers they consist of. Also, the participants have a consistent communication behavior, i. e., they have strategical, long-term communication relationships and do not change their service providers frequently.

These are typical and well-proven assumptions in the field of anonymity research similar to those in the related work, e. g., assuming nearly worst-case scenarios from the anonymity system's point of view. Based on this foundation, the model is described in the next section.

3.2 Simulation Model

The attack is modeled as a stochastic model of a time-step simulation. For an overview, the basic flow of the simulation is shown in Figure 4. In addition, these steps are described in the following in more detail:

1. *Initialization*: The first step of the initialization is the generation of the overall supply of *service compositions*. Here, a service composition consists of at least one service, i. e., the ID of the respective *service provider*. The assignment of service providers to compositions is done randomly, in our model based either on a *uniform* popularity distribution of the providers or a *Zipfian* one [31]. A Zipfian distribution means that the relative probability of the i-th most popular service provider to be used is proportional to $1/i^{\alpha}$, leading to a more realistic selection probability of service providers. Breslau et al. showed that the requests of Web pages follow a Zipfian distribution with an exponent α of about 0.75 [2]. This finding is adopted for our simulation because it provides a realistic estimation of service offerings on the Internet. Zipfian distributions were used before in the area of anonymity research, e. g., by Shmatikov and Wang [27]. From the generated service compositions, the organization under observation, here called Alice Corp. ("Alice") selects a certain number for executing her (business) processes.

2. *Cross-organizational Collaboration*: With the service offerings and the compositions used by Alice determined, the cross-organizational collaboration starts. At each time-step of the simulation, Alice contacts one of the service providers that are part of her used service compositions. As in real collaboration scenarios, Alice is not the only one communicating with service providers. Thus, there is also the so-called "background", i. e., other service consumers communicate with different service providers as well at each time-step. This background fills the remaining slots of the anonymity system's batch of size b. The recipients of the background are denoted by the vector $\vec{u}$ and distributed according to the general provider distribution, i. e., either uniform or Zipfian as described above. The attacker is assumed to know or approximate this distribution for his calculations. At each time-step i, the attacker intercepts the batch of messages $\vec{o_i}$ or a fraction thereof, depending on the attacker's spread.

3. *Attacker's Calculations*: At regular intervals, e. g., time-step t, the attacker performs calculations for identifying Alice's service providers in general and the corresponding service compositions in particular. The core of the calculations is based on the formal model of the classic Statistical Disclosure Attack [5]. Thus, as proven by Danezis, the attacker approximates Alice's recipients ($\vec{v}$) after t time-steps based on the observed output of the anonymity system ($\sum_{i=1}^{t} \vec{o_i}$), the batchsize b, and the known/approximated background distribution ($\vec{u}$):

$$\vec{v} \approx b\frac{\sum_{i=1}^{t} \vec{o_i}}{t} - (b-1)\vec{u} \tag{1}$$

Using vector $\vec{v}$ as approximated above and the stored observed vectors $\vec{o}$, the attacker calculates each vector $\vec{r_k}$ by multiplying each element of $\vec{o_k}$ (observed in round k) with the respective element of $\vec{v}$, afterwards normalizing the results using their dot product ($|\vec{v} \cdot \vec{o_k}|$):

$$\vec{r_k} = \frac{\vec{v} \cdot \vec{o_k}}{|\vec{v} \cdot \vec{o_k}|} \tag{2}$$

This then contains the probabilities about the service providers Alice communicated with in time-step k, i. e., the higher the resulting value of an element in $\vec{r_k}$, the more likely this service provider was used by Alice in round k.

The attacker then uses the maximum probability of each $\vec{r_k}$, i. e., the service provider Alice most likely communicated with in time-step k. This knowledge is then combined with the knowledge about the available service compositions, e. g., retrieved from the market places in the Internet of Services. Thus, the probability of each composition containing the most likely provider of time-step k is increased. Because the attacker does this iteratively for all observed time-steps, he builds an internal model of the service compositions Alice is using, assigning a probability to each possible composition.

The validation of the model is done as suggested by North and Macal [20]. It is based on the identified requirements, the plausibility of the assumptions, and the general development process, because the Internet of Services is not yet available for a comparison validation. These necessary aspects were discussed above and found to be valid for our model.

Using this specification as a foundation, a brief overview of

the model's implementation is given in the next section.

3.3 Model Implementation

The simulation model is implemented using *Repast Simphony*, an agent-based modeling toolkit[1]. Repast has the advantage of providing a frame for the general simulation, such as methods that are executed at each time-step of the simulation, a graphical user interface for configuring simulation parameters, and built-in functionality for tracing and logging simulation results. Furthermore, Repast models can be implemented using the Java programming language, thus, there is no need to learn yet another special modeling language.

A particular implementation aspect is the generation of random numbers based on the Zipfian (or Zeta) distribution in order to achieve more realistic results than using a basic uniform distribution for randomly selecting service providers. We calculated numbers of this distribution based on the following procedure, where $F(x)$ can be any cumulative distribution function [14]:

$$F(x) = \Pr(X \leq x), \quad y = F(x) \iff x = F^{-1}(y) \qquad (3)$$

Thus, a random number X of distribution $F(x)$ can be generated by using $X = F^{-1}(U)$, where U is uniformly distributed. In our case, we used the *Apache Commons Mathematics Library* version 2.1[2] for calculating the inverse cumulative probability $F^{-1}(U)$ of the Zipfian distribution, extending it regarding much faster random number generation as required for our simulation runs.

For verification purposes, test cases with pre-calculated results of the attack are compared to the (non-stochastic) results of simple attack runs of our model. These test cases can be used, e. g., for verifying the results of the simulation model after changes to the underlying algorithms have been made.

The next section describes how the model and its implementation can be configured in order to reflect different attackers and attack scenarios.

3.4 Configuration

For the evaluation runs of our simulation model, it can be configured in a variety of ways, modeling different attack scenarios and attacker capabilities. The used configuration parameters are described in the following:

Service compositions are generated based on the maximum number of services per composition (mSC) and the total number of (different) service compositions (C). From these, Alice selects randomly a number of used compositions (aC).

Service providers influence the simulation by their overall number (N) and their popularity distribution, which can be either uniform or Zipfian. Furthermore, the Zipfian distribution is detailed by it skewness. Based on Breslau et al.'s seminal work on Zipfian distributions in the Web as shown above, we chose a skewness of $\alpha = 0.75$ for our simulations [2].

The *anonymity system* is characterized by the batchsize (b), i. e., the number of messages leaving the system per time-step.

The *attacker's capabilities* are modeled by the parameter spread (S), which denotes the percentage of how many outgoing messages the attacker can intercept.

For the attack to have any chance of success, the following relationship must hold, as shown by Danezis [5]:

$$m < \frac{N}{b-1} \qquad (4)$$

However, the parameter m, i. e., the number of Alice's recipients, is no longer directly available in our model, because it is partially based on random variables. It can be approximated before-hand by $aC \times mSC$, which serves as an upper bound for m. At run-time, i. e., after the initialization phase, m can be determined exactly by counting the number of *distinct* service providers in all service compositions used by Alice.

As a preparation for the simulation runs, we performed a number of calibration runs in order to determine the most important parameters to be observed. The main distinction is the provider popularity, modeled by a uniform or Zipfian distribution. These are then evaluated regarding the impact of the overall number of service providers, the maximum number of services per composition, the number of compositions used by Alice, and the attacker's spread, i. e., his access to outgoing messages. Based on these configuration decisions, the next section describes the performed simulations and discusses selected results.

4. OUTPUT ANALYSIS AND DISCUSSION

This section discusses the used evaluation metrics and selected results of the performed simulations. Due to space-constraints, some results are omitted here, e. g., the impact of the number of available compositions (C).

For each single configuration, e. g., each different value for N, 100 simulation runs were performed in order to achieve a suitable level of confidence for assessing the results [11].

4.1 Evaluation Metrics

In order to evaluate the attacker's performance, i. e., his success regarding the identification of Alice's service compositions, we use well-proven metrics from the field of Information Retrieval [15]. These metrics were chosen, because we consider the problem of retrieving a set of "relevant" documents from a larger set of documents to be very similar to the attacker's goal of identifying certain compositions from the overall supply of service compositions. In addition, our scenario has the advantage of specifying definitely, what "relevant" documents, i. e., compositions, are: The ones used by Alice. Figure 5 shows the respective sets of our scenario in order to apply typical Information Retrieval metrics for our evaluation. Alice's compositions (relevant) are denoted with A, the attacker's identifications (retrieved) with B.

[1] `http://repast.sourceforge.net`, last access on January 3, 2011.
[2] `http://commons.apache.org/math/`, last access on January 3, 2011.

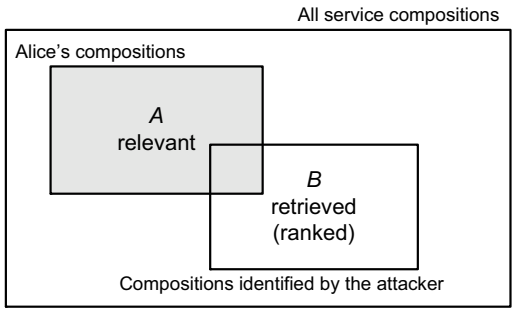


Figure 5: Retrieved/relevant sets of our attack scenario for applying Information Retrieval metrics.

Thus, the metrics mean in our scenario the following: *Precision* denotes the fraction of the identified compositions that are actually used by Alice, i. e., $\frac{|A \cap B|}{|B|}$. *Recall* denotes the fraction of how many of Alice's compositions could be identified by the attacker, i. e., $\frac{|A \cap B|}{|A|}$.[3]

As the attacker assembles a ranked list of Alice's compositions, i. e., sorted by their respective probability, this can be considered for the evaluation as well: *Mean Average Precision* (MAP) considers the position of Alice's compositions in the attacker's list of identified compositions. The more of Alice's compositions are at the top of the list, thus, having a high probability, the higher the MAP. However, even if more of Alice's compositions are identified correctly, MAP will decrease if these are ranked lower. More details on these metrics can be found in the works by Manning et al. or Moffat and Zobel [15, 19].

4.2 Impact of the Number of Service Providers

The results for different numbers of service providers, i. e., precision, recall, and MAP, are shown in Figure 6. In all these figures, the three metrics on the left are based on a uniform provider distribution while the three on the right are based on a Zipfian one. Furthermore, the y-axis uses a logarithmic scale in order to facilitate the comparison between the two distributions.

The measurements were taken after a rather short amount of interactions, i. e., 1,000 collaborations between Alice and her service providers. The reason behind this is to investigate how variations of certain system parameters influence the attacker's results.

Uniform: Recall is basically not affected by the number of service providers, as all of Alice's compositions are identified. However, opposed to what one might expect, precision *rises* if the number of different service providers increases. The reason for this might be, that Alice's and the background's interactions are distributed over a larger number of possibilities, thus, they stand out more prominently. MAP differs significantly from precision and has a very high and about constant value over the observed N. This means, that Alice's compositions are always at the top of the list of compositions assembled by the attacker. In general, the attacker

[3]However, a recall of 1.0 can be achieved easily by identifying *all* available service compositions as Alice's.

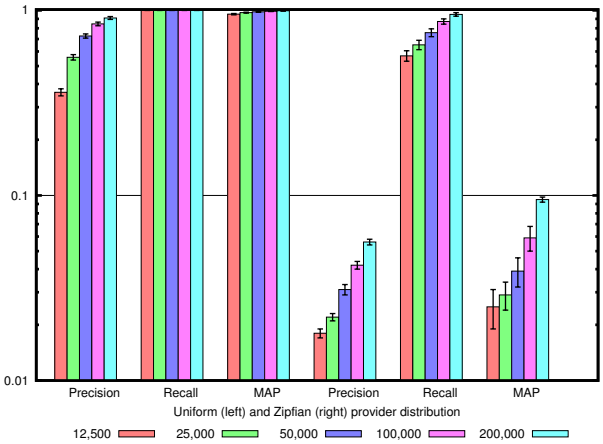


Figure 6: Varying the number of service providers N, each after 1,000 time-steps, 95% confidence intervals ($b = 125$, $C = 1{,}000$, $aC = 10$, $mSC = 8$, $S = 100\%$).

is very successful in this type of scenario, being mostly independent from the overall number of service providers in the system.

Zipfian: This scenario's results are completely different for the attacker. While recall is rather good, both precision and MAP are insufficient for the attacker, achieving not even 10%. This is due to the clustering around prominent service providers, which makes it for the attacker hard to distinguish between Alice and the background. However, as seen above for the uniform distribution, the increase in providers works for the attacker, leading to more providers on the "long tail" of the Zipfian curve, thus, making profiling easier. For example, an increase by factor 4 of the service providers leads to a doubling of MAP.

4.3 Impact of the Number of Compositions Used by Alice

The results for different numbers of Alice's service compositions are shown in Figure 7.

Uniform: The variety, i. e., number, of Alice's compositions has a strong impact on the metrics. While the recall for 5 to 50 compositions is still very good, it declines heavily with 100 and 200 compositions used. Precision is not affected as heavily, but MAP declines even more than recall, leaving the attacker with a reasonable amount of found compositions, but which are very late in his ranked list (thus, not of much use for him). Therefore, as could be expected, Alice using more different service compositions is more difficult for the attacker.

Zipfian: As above, the attacker's results against a Zipfian provider distribution are much worse than for the uniform one. The general trend is similar, but the degradation is more graceful than for N, e. g., recall decreases with the increase of Alice's compositions. However, precision is rather unaffected by this increase, but for high numbers of used compositions, it even increases as well (with low and about

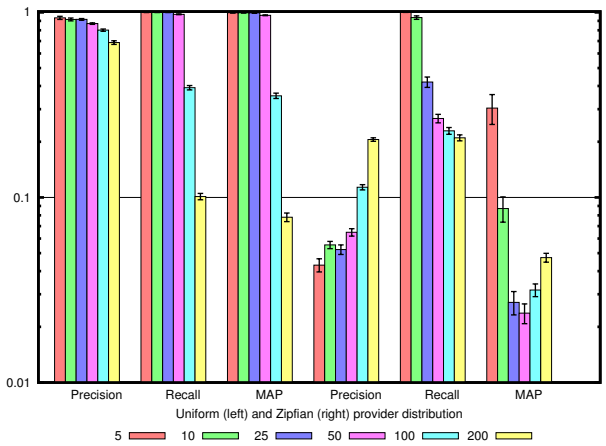


Figure 7: Varying the number of compositions used by Alice aC, each after 1,000 time-steps, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, mSC = 8, S = 100%).

constant values of MAP, though). This trend is due to the rising chances of the attacker identifying compositions because there are more of them.

4.4 Impact of the Maximum Number of Services per Composition

The results for different values for the maximum of services per composition are shown in Figure 8.

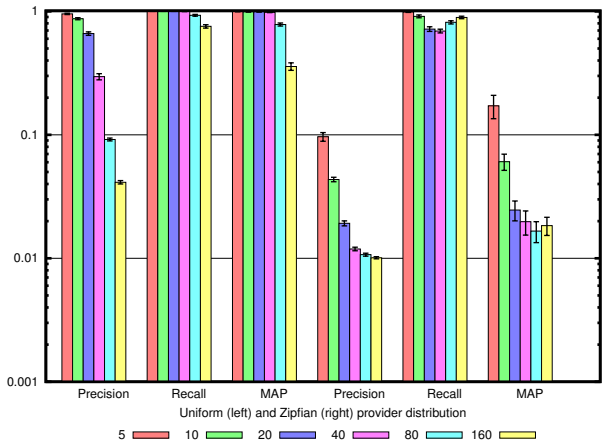


Figure 8: Varying the number of services per composition mSC, each after 1,000 time-steps, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, aC = 10, S = 100%).

Uniform: This parameter has a stronger impact than the number of Alice's compositions. The results are affected significantly, especially for complex compositions, i. e., high mSC between 80 and 160, which make the identification more difficult. However, recall is not affected as much, it is at most down to about 75%. The impact on MAP is not so heavy, at least less than on precision and not as much as for high aC values. In total, this is still a success for the attacker.

Zipfian: The trend is here similar as for the uniform distribution, but the results are by far worse for the attacker. Although recall is rather high, both precision and MAP decrease significantly with a decreasing number of services per composition due to the same reasons as above. In total, this is to be regarded as insufficient for the attacker.

4.5 Impact of the Attacker's Spread

The results for different spread values of the attacker are shown in Figure 9.

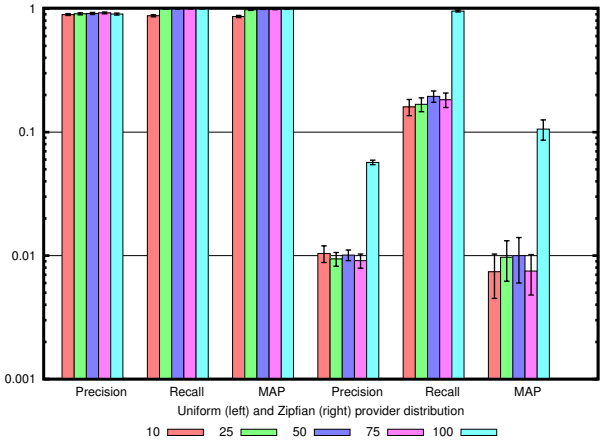


Figure 9: Varying the attacker's spread S, each after 1,000 time-steps, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, aC = 10, mSC = 8).

Uniform: Interestingly, the spread of the attacker does not seem to influence the attacker's results very much. Except for 10% access to the anonymity system's messages, the attacker achieves very good results that are in accordance with the N = 200,000 runs as discussed above. A possible explanation is, that the spread only affects the time required by the attacker for achieving certain results, e. g., a 10% spread at 1,000 time-steps could lead to similar results as a 100% spread at about 100 time-steps. This aspect will be investigated further in our future work.

Zipfian: This has a devastating effect on the attacker's results. Precision and MAP fall in general below 1% (with the exception of 100% spread that is similar to the obtained results above for N = 200,000). Recall is better, but does not reach 20%, which is insufficient for the attacker. The presumption of the attacker needing more time if less messages can be observed should be investigated further as well in this scenario.

4.6 Impact of the Number of Time-Steps

The above evaluations considered a short, fixed amount of time in order to determine the impact of different system parameters. In addition, it is also beneficial to investigate how the metrics evolve over time, i. e., where the limits of anonymity in cross-organizational collaboration or of

the attacker could be. For this, we used the following selected scenarios from above with an underlying Zipfian service provider distribution.

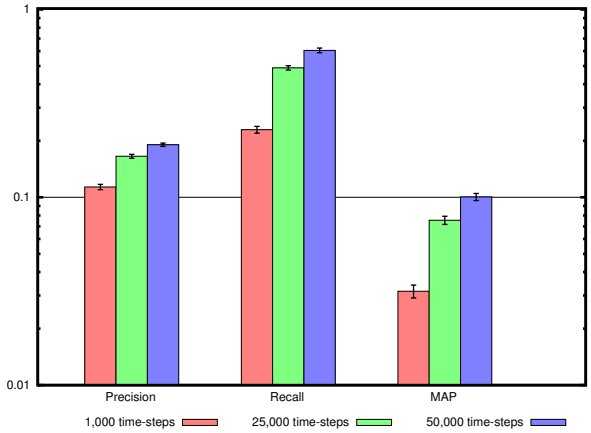


Figure 10: Evolution of attacker's results with $aC = 100$ and Zipfian distribution, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, mSC = 8, $S = 100\%$).

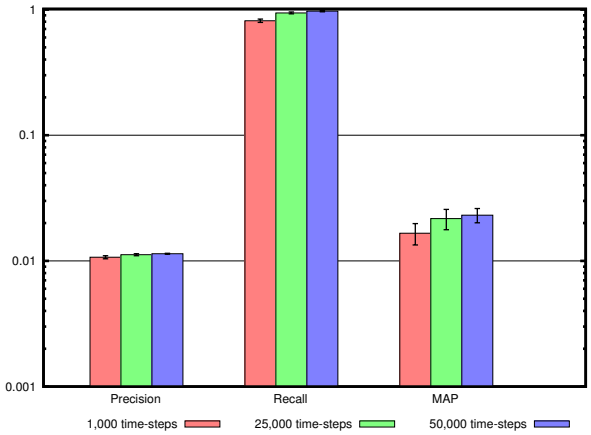


Figure 11: Evolution of attacker's results with $mSC = 80$ and Zipfian distribution, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, aC = 10, $S = 100\%$).

High number of Alice's compositions ($ac = 100$): As shown in Figure 10, more time, i. e., more observations, gives the attacker a slight advantage. While the additional knowledge regarding MAP from 1,000 to 25,000 time-steps is high, additional 25,000 observations do not contribute much, reaching in total still only about 10%. Precision and recall do not improve much as well, so that a significant improvement after even more observations is unlikely.

High composition complexity ($mSC = 80$): Observing a scenario with a high maximum number of services per composition over a longer time does not improve the attacker's results as depicted in Figure 11. Precision and MAP remain

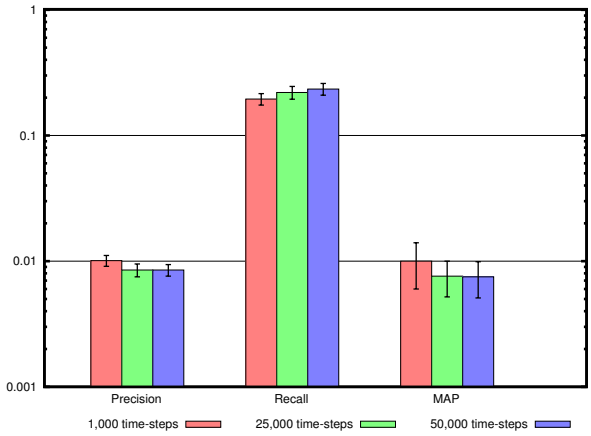


Figure 12: Evolution of attacker's results with $S = 50\%$ and Zipfian distribution, 95% confidence intervals (N = 200,000, b = 125, C = 1,000, aC = 10, $mSC = 8$).

about constant at their very low values between 1 and 2%. However, the very high recall can be slightly improved from 1,000 to 25,000 time-steps, but not much after that.

Medium spread ($S = 50\%$): Regarding the attacker's bad results for irregular access to the anonymity system's messages, additional time does not help as shown in Figure 12. All observed metrics remain nearly constant at their low values, precision and MAP at most reaching 1%. As pointed out above, significant gains cannot be expected even after more time-steps.

In general, making more observations is only one approach for the attacker and not a very good one, i. e., it is most likely only used as a last resort. Other approaches, e. g., improving the internal calculations, the general model, etc. are more likely to threaten the overall anonymity in the Internet of Services. Such possibilities will be discussed below as future work.

A summary and further discussion of the overall results is given in Section 6.

5. RELATED WORK

Regarding specific attacks on anonymity, this paper focuses on attacks on the boundaries of anonymity systems, i. e., the Statistical Disclosure Attack [5] from the general class of Intersection/Disclosure Attacks [13]. Within this attack class, only simple sender-recipient-relationships were investigated so far.

However, the concept of service compositions in the Internet of Services introduces additional complexity, i. e., with respect to the relationships between service consumers, the compositions they use, and the networks of service providers that constitute these compositions. Thus, in order to gain knowledge about an organization's processes, for example, by identifying the service compositions it uses, an attacker

has to confirm the relationships between service consumers and providers, inferring from this knowledge the service compositions this organization is most likely to use.

As a starting point for our investigations of this new scenario, we adapted the basic variant of the Statistical Disclosure Attack. Therefore, other variants and extensions were not considered so far, e. g., the use of Mix networks or different batching algorithms as described by Mathewson and Dingledine [16] or utilizing graph theory in order to relax specific user behavior assumptions of the attack model as introduced by Troncoso et al. [28].

Furthermore, the attack on relationship anonymity investigated in this paper must not be mixed up with the extensive research on Web service privacy, e. g., [10, 29, 30]. Web service privacy deals with the content of the exchanged messages, e. g., users' personal data, and how this information is further processed and possibly shared. It is an important aspect of the overall security goal "confidentiality", not of anonymity [1, 7].

On the other hand, the important aspect of anonymous communication between the different organizational participants of an SOA, i. e., with respect to third parties in order to conceal important business relationships has not been addressed so far. Further aspects of anonymity, i. e., the issue of anonymous Web service provision as well as consumption is addressed, e. g., by Papastergiou et al. [21]. However, it is questionable whether this is a desirable functionality for cross-organizational collaboration where it is important that both service consumer and provider know and trust each other, i. e., for legal reasons such as compliance or audit.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated the impact of attacks that aim at revealing business relationships of collaborating organizations. These attacks are of a particular danger in the field of cross-organizational service-based collaboration, because attackers can create detailed profiles of service consumers, providers, and also of market places by monitoring communication endpoints. Thus, sensitive information about the underlying business processes of the communicating organizations can be inferred easily.

Sophisticated countermeasures exist for achieving the required type of anonymity, so that an attacker cannot sufficiently distinguish whether the sender and recipient of a particular message are related or not. However, even if such standard anonymity mechanisms are deployed and used correctly, attacks mounted at the edges of such networks and aiming at typical long-term business relationships are very likely to be successful.

Therefore, this paper investigated the following research question: "How does an adaptation of the Statistical Disclosure Attack regarding service compositions threaten the relationship anonymity of service consumers and providers in the Internet of Services?"

In order to answer this question, the well-known "Statistical Disclosure Attack" was extended regarding service compositions in the Internet of Services scenario. This extension was then evaluated using a simulation model of different attacker models and attack scenarios, which was implemented with the Repast Simphony toolkit.

While the results based on a uniform provider distribution look promising for the attacker's success, such a distribution cannot necessarily be expected in the real world, i. e., the future Internet of Services. The used Zipfian distribution, whose skew parameter is inspired by the access distribution of Web pages, has the strongest impact on the attacker's results. This leads to a clear defeat of the attacker for the observed parameters, even if more observations are made.

In addition, if the observed organization ("Alice") uses many different compositions or mainly ones with a high composition complexity, this makes the attacker's defeat even clearer. As an organization cannot just increase its process complexity for improved security, this could be achieved by extending the concept of "dummy traffic" with respect to using "dummy compositions" or "dummy services" therein, e. g., obfuscating real compositions with additional (irrelevant) services.

Further impact can be achieved by increasing the number of offered compositions or the batchsize of the anonymity system. However, increasing the batchsize is likely to have serious side-effects, e. g., regarding important Quality of Service parameters such as the response time of service requests.

These findings might suggest anonymity is not that much in danger in an Internet of Services with a suitable provider distribution. However, the attacker can also improve his chances of success by including further information such as the providers' replies into his internal calculations. In addition, outside knowledge can be used as well, e. g., a bank is more likely to collaborate with other financial service providers than with providers from the logistics, pharmaceutical, or automotive sector. These aspects will be addressed in our future work, because the threat of revealing sensitive process information remains. Thus, our next steps will be to evaluate the existing model with other configuration parameters, e. g., even more observations, and to extend the attacker model regarding industry sector information and replies from service providers, e. g., continuing the work by Danezis et al. [6].

Acknowledgments

This work is partially supported by E-Finance Lab e. V., Frankfurt am Main, Germany (`www.efinancelab.de`) and BearingPoint Management and Technology Consultants. Furthermore, the authors would like thank Mr. Enkh Amgalan Ganbaatar for his support regarding the simulation runs.

7. REFERENCES

[1] R. J. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley, 2nd edition, 2008.

[2] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of the Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 1999)*, pages 126–134. IEEE, 1999.

[3] Bundesamt für Sicherheit in der Informationstechnik. SOA-Security-Kompendium: Sicherheit in Service-orientierten Architekturen, 2009. Version 2.0. `https://www.bsi.bund.de/SOA`. Last access on January 3, 2011.

[4] J. Cardoso, K. Voigt, and M. Winkler. Service Engineering for The Internet of Services. In *Enterprise Information Systems*, pages 15–27. Springer, 2008.

[5] G. Danezis. Statistical Disclosure Attacks: Traffic Confirmation in Open Environments. In *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC 2003)*, pages 421–426. Kluwer, 2003.

[6] G. Danezis, C. Díaz, and C. Troncoso. Two-sided Statistical Disclosure Attack. In *Proceedings of the 7th International Conference on Privacy Enhancing Technologies (PET 2007)*, pages 30–44. Springer, 2007.

[7] C. Eckert. *IT-Sicherheit: Konzepte – Verfahren – Protokolle*. Oldenbourg, 5th edition, 2007.

[8] M. Edman and B. Yener. On Anonymity in an Electronic Society: A Survey of Anonymous Communication Systems. *ACM Computing Surveys (CSUR)*, 42(1):1–35, 2009.

[9] Y. Guan, X. Fu, R. Bettati, and W. Zhao. A Quantitative Analysis of Anonymous Communications. *IEEE Transactions on Reliability*, 53(1):103–115, 2004.

[10] P. C. K. Hung, E. Ferrari, and B. Carminati. Towards Standardized Web Services Privacy Technologies. In *Proceedings of the IEEE International Conference on Web Services (ICWS 2004)*, pages 174–181. IEEE Computer Society, 2004.

[11] R. Jain. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. Wiley, 1991.

[12] R. Kanneganti and P. Chodavarapu. *SOA Security*. Manning Publications, 2008.

[13] D. Kesdogan, D. Agrawal, and S. Penz. Limits of Anonymity in Open Environments. In *Revised Papers from the 5th International Workshop on Information Hiding (IH 2002)*, pages 53–69. Springer, 2003.

[14] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley Professional, 3rd edition, 1997.

[15] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[16] N. Mathewson and R. Dingledine. Practical Traffic Analysis: Extending and Resisting Statistical Disclosure. In *Proceedings of Privacy Enhancing Technologies Workshop (PET 2004)*, pages 17–34. Springer, 2004.

[17] A. Miede, T. Ackermann, N. Repp, D. F. Abawi, R. Steinmetz, and P. Buxmann. Attacks on the Internet of Services – The Security Impact of Cross-organizational Service-based Collaboration. In *Proceedings of Multikonferenz Wirtschaftsinformatik (MKWI 2010)*, pages 2151–2162. Universitätsverlag Göttingen, 2010.

[18] A. Miede, N. Nedyalkov, D. Schuller, N. Repp, and R. Steinmetz. Cross-organizational Security – The Service-oriented Difference. In *Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops*, pages 72–81. Springer, 2010.

[19] A. Moffat and J. Zobel. Rank-biased Precision for Measurement of Retrieval Effectiveness. *ACM Transactions on Information Systems (TOIS)*, 27(1):1–27, 2008.

[20] M. J. North and C. M. Macal. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press, 2007.

[21] S. Papastergiou, G. Valvis, and D. Polemi. A Holistic Anonymity Framework for Web Services. In *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments (PETRA 2008)*, pages 1–8. ACM, 2008.

[22] M. P. Papazoglou. Service-oriented Computing: Concepts, Characteristics and Directions. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering (WISE 2003)*, pages 3–12, 2003.

[23] A. Pfitzmann and M. Hansen. A Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management, Aug. 2010. v0.34. `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml`. Last access on January 3, 2011.

[24] J.-F. Raymond. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In *International Workshop on Designing Privacy-enhancing Technologies*, pages 10–29. Springer, 2001.

[25] J. Rosenberg and D. Remy. *Securing Web Services with WS-Security: Demystifying WS-Security, WS-Policy, SAML, XML Signature, and XML Encryption*. Sams Publishing, 2004.

[26] C. Schroth. The Internet of Services: Global Industrialization of Information Intensive Services. In *Proceedings of the Second IEEE International Conference on Digital Information Management (ICDIM 2007)*, pages 635–642, 2007.

[27] V. Shmatikov and M.-H. Wang. Measuring Relationship Anonymity in Mix Networks. In *Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society (WPES 2006)*, pages 59–62. ACM, 2006.

[28] C. Troncoso, B. Gierlichs, B. Preneel, and I. Verbauwhede. Perfect Matching Disclosure Attacks. In *Proceedings of the 8th International Symposium on Privacy Enhancing Technologies (PETS 2008)*, pages 2–23. Springer, 2008.

[29] W. Xu, V. N. Venkatakrishnan, R. Sekar, and I. V. Ramakrishnan. A Framework for Building Privacy-Conscious Composite Web Services. In *Proceedings of the IEEE International Conference on Web Services (ICWS 2006)*, pages 655–662. IEEE Computer Society, 2006.

[30] G. Yee and L. Korba. Privacy Policy Compliance for Web Services. In *Proceedings of the IEEE International Conference on Web Services (ICWS 2004)*, pages 158–165. IEEE Computer Society, 2004.

[31] G. K. Zipf. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Routledge & Sons, 1999.

Assisted Social Identity Management

Enhancing Privacy in the Social Web

Michael Netter
Department of Information Systems
University of Regensburg
Regensburg, Germany
michael.netter@wiwi.uni-regensburg.de

Moritz Riesner
Department of Information Systems
University of Regensburg
Regensburg, Germany
moritz.riesner@wiwi.uni-regensburg.de

Günther Pernul
Department of Information Systems
University of Regensburg
Regensburg, Germany
guenther.pernul@wiwi.uni-regensburg.de

ABSTRACT

The rise of the social web shifts personal identity management to the online world. As a result, personal information is persistently available to all of a user's contacts without distinguishing between different contexts such as *Work* and *Leisure*. Personal information being available to audiences outside the intended context violates contextual integrity and poses a threat to users' privacy. We argue that a formal description and a conceptualization of the problem scope is required to systematically address current challenges of personal identity management in online social settings. Based thereupon, we propose assisted social identity management to support the user in finding segregated audiences among his contacts as a first step to advance from the current situation and sketch further improvements. We evaluate our approach using real-world data, demonstrating the feasibility of our proposal.

Keywords

Assisted Social Identity Management, Audience Segregation, Contextual Integrity, Social Web, Privacy

1. INTRODUCTION

Besides the established services of the WWW – distribution of information and commercial use – the social web, consisting of communication and interaction between users enabled through easily usable applications, has gained importance rapidly [19]. Today it offers an infrastructure for communication, information and self-expression. Even established services on the Internet, such as news portals incorporate related features, thus adding to the pervasiveness of the social web. A current example is the "Like-Button" introduced by the 500-million user social networking platform Facebook that can be embedded into regular websites to connect their content with user-generated comments and recommendations. In enterprise settings, social applications generate new opportunities such as tapping into new knowledge pools originating from employees connected by social networks and assessing potential for innovation through better customer integration into the value chain [2].

This rise of relevance and quantity of social web services has also been accompanied by concerns regarding privacy and thus raises organizational and legislative implications. Politicians criticize possible data protection issues, for example with services offered by Google and Facebook, and demand stricter regulation. Similarly, users of the social web voice demand for more protection of privacy as well.

While this common criticism usually targets enterprises and their handling of customer data, there is also a need for research regarding the sharing of personal data between users of the social web. In fact, as the social web is used to shift personal identity management (IdM) to the online world, the differences to the offline world pose new challenges. In current implementations of the social web, personal information is available persistently, digitally and thus unchanged. Contrary, in the real world, personal information is transient and and availability is limited to certain settings. Existing notions of privacy, usually involving confidentiality [18] and access control [20][5], cannot be applied without modification, as users of the social web have the desire to selectively share personal information depending on situation and recipient. However this demand for personal information on the one hand being publicly available, but on the other hand not for the public [16] poses a challenge to users of the social web. This trade-off between privacy and comfortable disclosure of personal information is referred to as "privacy paradox" [13], calling for contextually aware information sharing. Unlike the real-world, where different contexts such as coworkers and a group of friends are automatically separated by mere physical and spatial separation, providers of services for the social web do not distinguish between different contexts, thus often exposing the complete digital identity of a user to her contacts.

Identity management is commonly used in organizations and enterprises to administrate individuals and control their access to resources. On the contrary, social IdM is a user-centric concept to cope with the challenges of presenting different facets of the self to different audiences and to keep those views consistent. In this work, we introduce ways to

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

improve social IdM by distinguishing between different contexts. This enables context-aware segregation of audiences of the online identity which in turn can be realized by the possibility to present different values for the same attribute to different audiences. The proposal is accompanied by a formal presentation of the problem scope and by a prototype application aiding the user in managing her online contexts.

The remainder of the paper is structured as follows: In the next Section, we give an overview of related work in the field of social IdM and audience segregation. In Section 3, we show that contextually segregated audiences can be realized through directed attribute presentation and introduce the problem scope in a formal fashion. After conceptualizing the problem scope we propose several approaches to implement assisted social IdM that support the user in recognizing and organizing online contexts in Section 4. The prototypical implementation of two of these approaches and a subsequent evaluation are presented in the following Section 5. Finally, in Section 6 we summarize our findings.

2. RELATED WORK

The increasing relevance of the online digital identity has been recognized by a plethora of research. Nowadays, personal data resides in a growing number of databases, while information on the Internet becomes easier to find and harder to hide [11]. The users' digital identity is defined as their attributes and their values [4], while a subset of them constitutes a partial identity as described in [14] and [8]. Cameron sees the increasing number of personal digital identities as a problem, as users lose track of their disclosed data [3]. While presenting different attribute values to different audiences is related to the concept of partial identities and these notions are related to our work, these works are mostly concerned with interaction with service providers, while we focus on selected disclosure of personal data to contacts in the social web.

This difference can also be seen when looking at the state of end-user IdM systems (IMS), which focus mostly on account management issues such as Single-Sign-On. The need for IMS offering "context-dependent role and pseudonym management" has been postulated [1], however, since the identity of a contact in the social web is usually disclosed, we see a shift from pseudonym management to context-based attribute management.

The notion of privacy as contextual integrity was first introduced by Nissenbaum [12] to cope with the challenges of information technologies processing personal data. She argues that contextual integrity is a measure of how closely gathering and dissemination of personal information conforms with the intended context for this information. Privacy is breached if that personal information is available outside this context. This requires the user to be free from constraints on the construction of his identity as stated in the concept of privacy as practice that is introduced in [7].

To facilitate contextual integrity, audience segregation is a valuable concept. Originally developed by Goffman [6], it states that each individual performs multiple and possibly conflicting roles in everyday life, and it needs to segregate the audiences for each role, in a way that people from one audience cannot witness a role performance, that is intended for another audience and thereby keep a consistent self-image. In [17] this concept is applied to social networks, emphasizing the increased complexity of audience segregation compared to the everyday life.

A few Facebook applications share similarities with the prototypical implementation presented in Section 5.1, which creates segregated audiences by grouping the user's contacts based on their mutual friends. These applications differ from our work by aiming at a visually appealing presentation of a user's network of contacts while this work focuses on finding disjoint segregated audiences in the context of assisted social identity management. Friendwheel[1] allocates all contacts on a circle based on their proximity to each other, which is derived from mutual friend connections. Unlike our approach however, it doesn't create disjunct groups of contacts. TouchGraph[2] and Social Graph[3] do create disjunct contact groups. Yet, for all of these implementations, neither the algorithm, nor the exact input parameters to compute the friend groups are available. Also, these implementations are specifically built for the Facebook platform while we aim for an approach that is agnostic of a particular social network by solely employing contact relations for finding segregated audiences.

The social web is increasingly being recognized by EU research projects. Padgets[4] acknowledges the relevance of the social web and harnesses its community knowledge to get input for policy making while being committed to preserve participants' privacy. PrimeLife[5] [15][10], a European research project, employs the concept of audience segregation to develop an advanced social network which allows to define different social contexts and assign different audiences. This work differs from ours as we propose means to support the user in finding segregated audiences within her contacts while PrimeLife focuses on how to assign contacts to different social contexts and thus here audiences must be defined manually.

3. SEGREGATED AUDIENCES IN SOCIAL IDENTITY MANAGEMENT

In this section, we explain the need for audience segregation and directed attribute presentation as a means to implement it in the social web, followed by a formal definition of the problem scope.

This work is agnostic to how the online identity is hosted and how the communication between contacts is organized and implemented. While possibly enabled through a single social networking website, other ways to implement social IdM are also possible, such as distributed solutions. Also, we only consider other members as a possible audience while privacy issues regarding entities such as the operator of a social net-

[1]http://thomas-fletcher.com/friendwheel/
[2]http://www.touchgraph.com/TGFacebookBrowser.html
[3]http://www.mihswat.com/labs/app/facebook-social-graph/
[4]Policy Gadgets Mashing Underlying Group Knowledge in Web 2.0 Media - http://www.padgets.eu
[5]Privacy and Identity Management in Europe for Life - http://www.primelife.eu

working site or other service providers are out of scope of this work. Furthermore, we do not consider possible inferences through publicly available information in connection with shared personal data through linkability.

The research is carried out from the point of view of a person ("user") interacting with other individuals through an online social setting. Those other individuals are commonly viewed as "contacts" or even called "friends". To construct an attacker model, one has to see them as privacy-attacking adversaries that may be able to discover information that was meant to be kept undisclosed to them. We define the attacker as "honest but curious", acting within the rules set by the services providing the social web, thus only accessing legitimately available information and not exploiting security weaknesses of the site or, for example, weak passwords. Therefore, the technical protection of the confidentiality of certain attribute values is out of scope of this work, while we focus on the assignment of attribute value visibility to contacts depending on contexts.

3.1 The Need for Segregated Audiences

Successful personal IdM in the social web requires means to organize one's identity depending on social contexts and to act accordingly. The user needs to be supported in recognizing and distinguishing between different social contexts online similarly to offline contexts such as work, school, family and friends. Like with offline contexts, which each have a different appropriate and accepted behavior, one must be able to choose how to present one's identity depending on online contexts as well. As one performs multiple and possibly conflicting roles in various contexts, the need for audience segregation occurs, meaning that different audiences are kept from witnessing role performances that were meant for other audiences [6].

An identity in the social web consists of several attributes and their associated values, while individuals being able to view certain attribute values are referred to as an audience. In the online social web, *directed attribute presentation* is needed to realize audience segregation. In other words, one must be able to present different values for the same attribute to different audiences and to hide certain attribute values from other audiences, thus keeping presented partial identities consistent. One example that has been brought up is a teacher that feels the need to hide certain spare time activities and friends from her student contacts on a social networking site after irritations about her hobby occurred [15].

As in real life, audiences need to be disjoint with each contact only seeing the predetermined attribute values and thereby preserving the integrity of the user's partial identity. In case of overlapping audiences, i.e. contacts spanning several segregated audiences, these contacts are assigned to a newly created audience. Enforcing this policy raises the problem of two of the users's contacts from different segregated audiences possibly exchanging information on the user without his knowledge and thereby violating the contextual integrity, which is unsolved even in the real world.

A context, describing each instance in which certain attribute values and a particular audience come together [15],

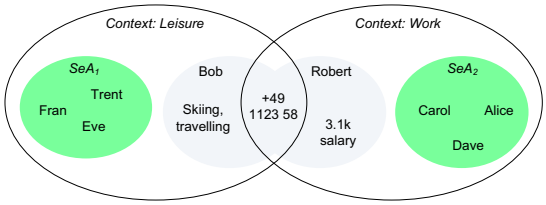


Figure 1: Two Sample Contexts of User Bob

can be seen as a means to connect sets of attribute values and their corresponding audiences. Thus, users must be supported – assisted or automatically – in allocating both online contacts and appropriate attribute-value pairs, thereby forming a specific context. Figure 1 shows two sample contexts, one for work and one for spare time activities. Some attributes, such as the salary, are only shown in one context, some, such as the first name have different values in each context while others are shown in both contexts.

Presently however, directed attribute presentation is not or only in a limited way implemented in the field of the social web. We argue that such controls are essential to empower users for context-based identity presentation. Currently, users wishing to customize their digital appearance to different audiences need to resort to workarounds such as creating multiple accounts at the same or distinct social platforms. By doing so, they are showing different sides of their identity to different audiences and thereby create distinct contexts. The option to hide certain attributes from a subset of one's contacts is already available on some sites, but does not reach far enough.

There are several reasons for the current lack of more fine-grained customization possibilities, for example providers of social platforms may not have seen a need to implement such features due to a lack of customer interest and business value. Also, the complexity of such controls poses a challenge both for developers and users. We improve the situation by showing further approaches to segregate audiences and to enable directed attribute presentation in Section 4.

3.2 A Formal Problem Scope Description

Followingly, we express the notions related to contexts, audiences and directed attribute presentation formally and thus more precisely to ensure a clear presentation of the problem scope and to support future work. Firstly, for the user, there is a set of attributes

$$A = \{a_1, a_2, ..., a_n\}.$$

For each of these attributes, the user may enter a set of values

$$AV_m = \{av_{m1}, av_{m2}, ...av_{mk}, \emptyset\} \forall m = 1...n.$$

Always included is the empty value representing that the attribute is not shown. Note that in current services the user is restricted to only one attribute value and the empty value. Further, there is a set of contacts

$$C = \{c_1, c_2, ..., c_p\}$$

denoting user's contacts in the social web.

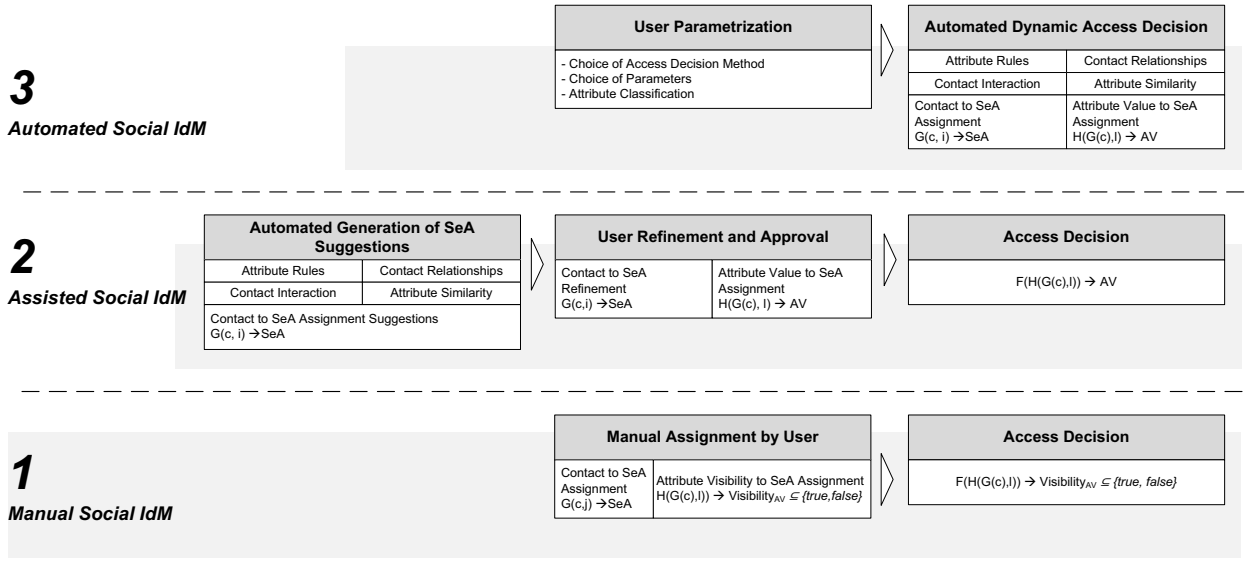


Figure 2: Conceptualization of Current and Proposed State of Social Identity Management

For the selected, directed and context-based release of attribute values, the following question must be answered: Which value av_x of the attribute a_x is presented to contact c at the time of access? To answer this question, one also needs to consider on which information, denoted as i, the decision should be based, which can by summarized by a function

$$f(a, c, i) \Rightarrow av$$

f can be seen as a *view* on an attribute and may be invoked each time a contact attempts to access an attribute of the user. Similarly, we define a view V on all of the user's attributes as the set of attribute values

$$F(c, i) \Rightarrow V \subseteq AV_q \forall q = 1..n.$$

It returns the set of attribute values a particular contact may see, thus representing a partial identity of the user.

Before elaborating further on the possible ways to implement f, we need to consider the previously mentioned audience segregation. We consider a segregated Audience (SeA) as a subset of the user's contacts C, therefore

$$SeA_i = \{c_1, ..., c_r\} \subseteq C.$$

As the purpose of SeAs is to prevent the presentation of inconsistent partial identities or single attribute values, all contacts that form a SeA have to be presented with the same attribute values leading to

$$f(a_x, c_y, i_y) = f(a_x, c_z, i_z) \forall x \subseteq 1..n; c_y, c_z \subseteq SeA.$$

In other words, f has to return the same attribute values for all contacts in the same SeA. In order to prevent contacts from seeing different and therefore conflicting attribute values, one contact can only be part of one SeA, thus all SeAs are disjunct subsets of C. To resolve the case of overlapping SeAs, they may be split up so that the intersection forms a new SeA.

We define a context as a set of contacts forming a SeA and a corresponding consistent set of attribute values that may be exposed to these contacts, as illustrated in Figure 1. Thus, F, returning an attribute value set for a given contact can be used to establish the connection between contacts and attributes that is necessary to define a context. To reduce complexity, the problem of defining F can be split into two tasks represented by Functions G and H, namely assigning contacts to contexts and the corresponding SeA ($G(c, j) \Rightarrow SeA$) on the one hand and assigning attribute values to audiences ($H(SeA, l) \Rightarrow V$) on the other hand with j and l being additional information on which the decision is based on.

One straightforward way to implement G and H are manual assignments conducted by the user, thus resulting in j and l being static lists. However, with the complexity arising from the ever-increasing number of contacts, the abundance of attributes and the added possibility of using multiple values for each attribute, this task is prone to become tedious at best. Further, the dynamic nature of the social web would not be considered. Contacts previously not classified, changes of contact's attributes and the nature of the relationship with them are partly out of control of the user and would require reconsidering the attribute values released to them. Also, notions such as SeAs, directed attribute release, contexts and views may not be intuitive to the average user. These challenges require an advancement from the current state of the art both conceptually and as seen in current social web implementations and are addressed in the following sections.

4. APPROACHES TO DETERMINE SEGREGATED AUDIENCES

In this section, we propose a conceptualization of social IdM, showing that currently users have to manually deal with the complexity imposed by the social web. We advance the situation by proposing assisted social IdM and illustrate several approaches to support the user.

4.1 Current and Proposed State of Social IdM

Today's social network landscape consists of many service providers offering platforms for different contexts such as leisure (e.g. Facebook) and work (e.g. LinkedIn). However, the possibilities to cope with multiple contexts within the same platform are limited, as available possibilities to enforce audience segregation and ways to exert directed attribute presentation leave much to be desired. This is shown in Figure 2, in which the current state is illustrated at the bottom level as *Manual Social Identity Management*. While it is already possible for the user to assign contacts to groups and set attribute visibility based on group membership, all of these steps have to be performed manually by the user before the attribute in question is accessed by a contact. It is currently not possible to show different attribute values of the same attribute to different contacts. Additionally, current provider implementations lack default groups with settings that provide only limited access to personal information. A further shortcoming is the lack of transparency for the user about who has access to which disclosed information stemming from large contact group sizes and inappropriate tools to visualize audiences. Thus it is difficult to act appropriately for a given context and thereby maintain contextual integrity. As mentioned in Section 3.1, the common approach to cope with this limited status quo is to employ workarounds.

We propose improving the current state by managing segregated audiences through directed attribute release, e.g. displaying different attribute values to contacts depending on the current context. As the transition to the use of different attribute values increases the complexity of social IdM even more, the need for *Assisted Social Identity Management* arises, e.g. the user needs to be supported in creating and managing groups of contacts that correspond to SeAs and their respective contexts. As illustrated in the middle level of Figure 2, this assistance encompasses automatically generating suggestions for SeAs, namely allocations of the user's contacts to groups. Thus, the initial burden of creating groups from possibly hundreds of contacts is lifted. The user may then refine these suggestions manually, approve them and assign attribute values to form a context. In Section 4.2, we present and discuss four new approaches to generate SeA suggestions.

While assisted social IdM eases the initial allocation of existing contacts into groups, there is still room for improvement: Firstly, any suggested group allocation needs to be verified manually and refined if necessary, as errors would possibly lead to the unintended release of personal information. Also, after user approval, no further information will be considered for the access definition. Yet, dynamic incorporation of new information would be desirable, for example to prevent a coworker that has left the company from viewing work-related attributes. Thus, we propose *Automated Social Identity Management* as a next step to improve social IdM, as shown on the top level in Figure 2. It advances from assisted social IdM through the following properties:

1. **User parametrization of contact allocation:** Rather than having to decide or approve each contact's group individually, the user sets parameters for automatic contact allocation.
2. **Contact allocation after user interaction:** Unlike as in assisted social IdM, user interaction occurs before contacts are assigned to groups, thus allowing for dynamic contact-to-audience allocation and incorporation of the most recent available information. However, this poses high requirements on the algorithm's reliability.

Besides their adequacy to produce suggestions for SeA, in Section 4.2, the aforementioned approaches are examined regarding their use in a possible automated social IdM scenario.

4.2 Approaches to Segregate Audiences

In the following we develop and discuss more advanced approaches to support the user in creating and managing segregated audiences in order to maintain contextual integrity. The goal of all approaches is to create disjunct groups of contacts, or, applying the formal notation introduced in Section 3.2, to define parameter j of function $G(c, j) \Rightarrow SeA$. Table 1 provides an overview of our proposed approaches.

Manual group assignment. This static approach requires the user to manually cluster her contacts in a reasonable manner and thereby create segregated audiences. Using manual assignment, j simply represents the user's knowledge on each contact, i.e. she has contextual information about the contact and is therefore able to assign it to an appropriate segregated audience.

Theoretically, this approach allows for a fine-grained assignment of all contacts. However, with the increasing number of contacts, the user has to deal with the complexity of overlapping audiences which need to be split up, in order to ensure consistency of the presented attribute values. Furthermore, due to the static nature of this approach the user has to reconsider the classification each time a new contact is added. In summary, manual assignment is a powerful concept to create and manage segregated audiences, however due to its high complexity it needs to be combined with other approaches that aim at supporting the user. This approach is implemented in many of today's social networks and is listed for reference purposes.

Attribute rules. Employing attribute rules makes use of a contact's attributes to assign her to a segregated audience. In more detail, the user predefines a set of rules that constitute the input j of function G. A simple rule might be $R : \{av_{lastname} = Doe\} \Rightarrow SeA_{Family}$ to assign each contact having *Doe* as the attribute value for the attribute *lastname*

Table 1: Proposed Audience Segregation Approaches

No.	Type	Description	Required input data j	Categorize new user at time of access	Combinable with
1	Manual	Manual group creation and user assignment	Manual SeAs definition	no	2, 3, 4, 5
2	Attribute Rules	Dynamic assignment based on predefined attribute rules	Set of Rules	yes	1
3	Contact Interaction	Analyze past user behavior and interaction	Quantity and quality of past contact interaction	yes	1, 5
4	Attribute similarity	Assign new users according to similarity to predefined and populated classes	Manual SeAs definition, contact's attributes	yes	1, 5
5	Contact Relationships	Determine distinct contact groups	All contacts' contact lists	yes	1, 3, 4

to the segregated audience *Family*. More sophisticated rules are conceivable to get finer-grained segregated audiences. A set of predefined rules for reoccurring problems could be provided to the user. As it is possible to create contact groups using this approach can be employed in assisted social IdM to make SeA suggestions based on the user's preferences.

Once a rule set is defined thoroughly, this approach can operate without further user interaction, hiding all the complexity from the user and allowing for automated social IdM. New contacts can be classified automatically and segregated audiences can by created dynamically at access time. However, the assignment precision of a contact and thereby the effectiveness highly depends on the quality of the rule set. As the assignment decision is based on the contact's attributes, without further assurance about those attributes (e.g. trust authority), this approach is vulnerable to a malicious contact that alters his attributes in order to be assigned to a different segregated audience and thereby seeing different attributes and attribute values.

Contact Interaction. This approach relies on the user's preexisting interaction with her contacts and the fact that both quality and quantity of interaction depends on the contact's context. In other words, we propose to analyze both the frequency and the content of all previous communication that is available (e.g. private messages, wall postings, etc.). As an example, a user might frequently communicate with her best friend using everyday language while she might have written only few formal messages to her colleague.

In a more formal way, qualitative and quantitative communication patterns constitute the input j of function $G(c, j) \Rightarrow SeA$. As a result, function G delivers a set of segregated audiences that can be further refined by the user. New contacts without any prior interaction must be assigned to a default group requiring the user to finally assign them.

Besides the use of this approach for SeA suggestions in assisted social IdM, applying it to automated social IdM is also feasible. For instance, contacts without interactions in a certain time period could be automatically put into a group with less access privileges.

Attribute similarity. This approach is based on the assumption that all people in a given context are similar to some extent or share a common attribute (e.g. the same affiliation). We propose to make use of this similarity between a user's contacts to create segregated audiences.

Two variations of employing attribute similarity can be conceived: First, in order to find meaningful segregated audiences among all contacts, unsupervised learning techniques, such as clustering algorithms can be applied. Without any user interaction, all contacts are assigned to subsets in a way that all contacts in one subset share similarities. The approach is useful to suggest a set of segregated audiences which the user can adapt and refine and is thus use in assisted social IdM. For example this approach can be used to discover the segregated audience $SeA_{FellowStudents}$ since all associated contacts have the same affiliation and are of similar age.

Second, supervised learning techniques can be used to classify new contacts based on preexisting groups. In more detail, after the user has defined segregated audiences (e.g. using one of the other approaches depicted in Table 1), a classifier is determined and used to suggest the correct segregated audience for any new contact.

For the first variant of the approach the input j of function G is a vector of attributes that are suitable for similarity measurement. The second variant additionally requires the predefined segregated audiences as input value. Machine learning techniques are a powerful concept to cope with the high complexity of finding segregated audiences among a user's contacts [21]. However as the algorithms lack contextual knowledge the assignment precision is low and therefore the user has to finally decide upon the classification. Similar to the rule-based approach, the decision is made upon a contact's attribute and is therefore vulnerable to a malicious contact that changes her attributes in order to be classified with another context. Thus, the use in automated social IdM is questionable.

Contact Relationships. The rationale of this approach is that information about a contact's context is embedded in the relations this contact has with other contacts. We propose to employ this inherent property of common social net-

Table 2: Contact Groups from the Test Scenario

No.	Name	Description	Group Size
1	Hometown	Contacts from the student's hometown	3
2	High school year	Contacts from a student exchange year spent in the United States	8
3	Study abroad year	Mostly international contacts from a study abroad year during college	25
4	USA 2	Contacts in the US unrelated to group "High School Year"	8
5	Summer school	Contacts from an international summer school attended during PhD studies	18
6	University	Contacts known through the University	77
7	Other	Other contacts not related to groups above	6
Sum			145

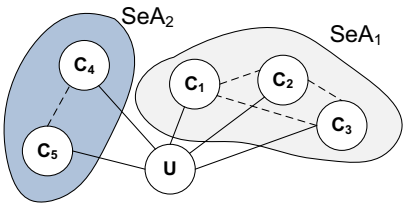


Figure 3: Relationships Between Contacts

works to arrive at segregated audiences. It can be seen as a special case of the attribute similarity approach with the contacts of a contact as the only attributes being considered.

To be more specific, both the user and a specific contact have relations to other contacts as depicted in Figure 3. It can be seen that the user's contacts c_1, c_2, and c_3 have a relation to each other and thereby form $SeA_1 = \{c_1, c_2, c_3\}$. Analogously the contacts c_4, c_5 constitute $SeA_2 = \{c_4, c_5\}$.

We assume that contacts that have relationships among each other also communicate with one another and possibly about the user and her attribute values. This approach has the benefit that it makes communication between members of different SeA and therefore the discovery of different attribute values by the same contact less likely. However, this cannot be prevented completely for two reasons: First, not all relationships in real life are also represented online in the social web, therefore there may always be links between contacts that cannot be discovered. Secondly, only allowing SeA allocations in which there are no links between members of different SeAs would be the strictest form of this approach. It would possibly lead to only a few or one big SeA, as there are usually relations between contacts from different contexts such as a user's friend knowing a user's colleague from work (note that the two SeAs in Figure 3 satisfy the requirement). In order to find reasonable segregated audiences it is necessary to allow linkage between the contacts of two SeAs.

The approach will gain importance even more once the current service provider landscape with its distinctive social networks for different contexts (e.g. Facebook for *Leisure*- and LinkedIn for *Professional*-activities) converges to one large or several interconnected networks for all contexts.

To arrive at segregated audiences, this approach uses information about relations between users and contacts as input parameter j. The approach can be employed to discover segregated audiences within the user's contacts and enable assisted social IdM, which is demonstrated in Section 5. It is also conceivable to employ such an approach of automated social IdM by classifying new contacts based on their relationships to contacts in existing groups.

5. IMPLEMENTING ASSISTED GROUP ASSIGNMENT

In this section we evaluate the effectiveness of *Contact Relationships* as one of our proposed approaches to discover segregated audiences among a user's contacts. We first present our algorithm that implements the function G as introduced in Section 3.2. Subsequently, we describe the test scenario based on a test person's real-world data and employ this dataset to evaluate our results using a prototype.

5.1 Algorithmic Foundations and Prototypical Implementation

The test person's friend list and their connections to each other were gathered using an application connecting to the social networking site Facebook. With *FQL (Facebook Query Language*[6]*)*, the site allows applications to access some of the information visible to their users through a number of available tables. We queried the table *connection* to get a list of the user's contacts and the table *friend* to obtain the links between those contacts.

We further implemented a prototype to derive segregated audiences within the social networking site dataset and evaluated the effectiveness of our approach. The prototype is available for download and testing on our website[7]. For evaluation purposes it allows for manual creation of segregated audiences (cf. Manual approach in Table 1) that serve as a test set. The tool contains an algorithm implementing the *Contact Relationship* approach to find segregated audiences among a user's contacts and a matching algorithm to evaluate the result against the predefined test set (see Figure

[6]http://developers.facebook.com/docs/reference/fql/
[7]http://www-ifsresearch.wiwi.uni-regensburg.de/paper/wi/

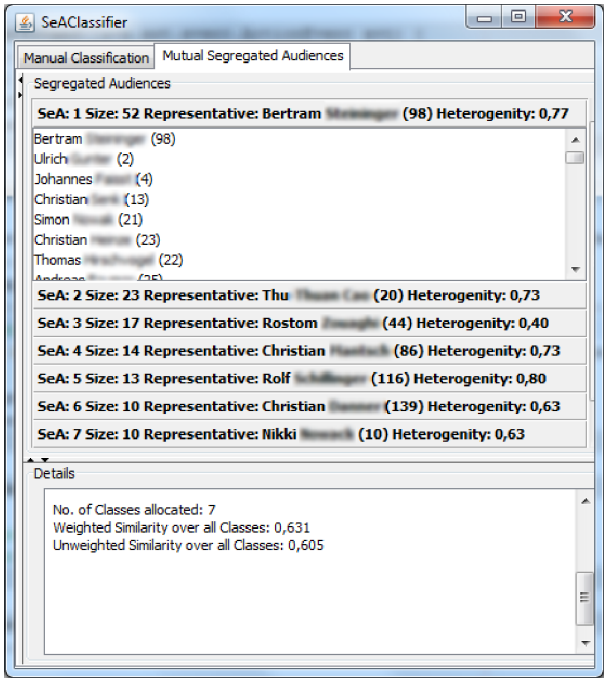


Figure 4: Our Prototype (Screenshot)

4). Furthermore we included a simulator to rapidly determine the optimal initialization parameters for the *Contact Relationship* algorithm.

The prototype groups the test person's contacts into classes, using the relationships among them as a criterion, based on the assumption that contact groups with a high number of relationships among them also belong together semantically and thus form a possible context. A "relationship" is a binary property derived from the "is friend" property that popular social networking sites employ.

For clustering, we employed a heuristic that uses the Jaccard index [9] – a frequently used distance metric – to calculate the similarity between two contacts. First, it calculates a matrix containing similarities between any combination of two contacts based on the overlap of their contact sets employing the Jaccard similarity coefficient. Thus, two contacts with a high number of mutual contacts receive a high similarity value. Next, representatives for the first two classes are selected, with the first one being the contact with the highest number of similarity values that exceed a predefined threshold and the second one being the contact that has the lowest similarity to the first representative. Further class representatives are chosen based on the maximal distance to all existing representatives as a criterion, until the distance of the next possible representative falls below the previously chosen threshold. Next, all remaining contacts are assigned to classes based on their similarity to the class representatives. The threshold greatly influences the outcome of the algorithm, as the number of created classes is directly related to it. In Section 5.3, a detailed discussion on the optimal threshold determination is provided.

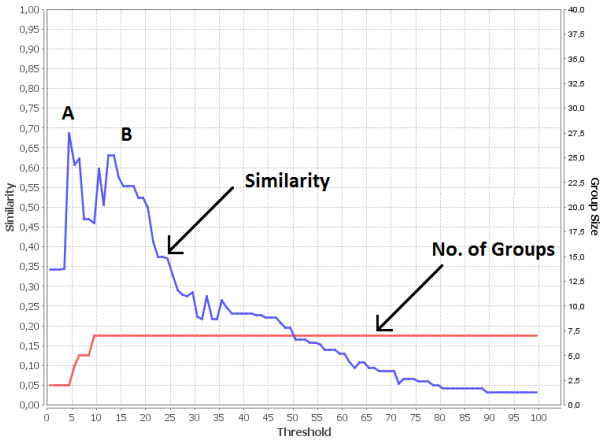


Figure 5: Threshold Determination

In terms of machine learning, this approach is a type of unsupervised classification, as it requires no prior training of the algorithm by the user. Note that the existence of a connection between two contacts does not weigh more towards their similarity value than any other mutual contact they have. Also, only the existence of connections and thus mutual friends influence the outcome, not their absence. This matches the nature of the data set, in which instances where two contacts have no connection far outweigh the number of connections.

5.2 Test Scenario

For evaluation, we obtained actual user data containing 145 contacts and their relationships from a PhD-student's social network account and created SeAs automatically using the algorithm. Then, we compared this output to a manual assignment that the test person was asked to create.

The test person was asked to classify his contacts into semantically meaningful groups similar to real-life contexts to the best of his knowledge, leading to the allocations presented in Table 2. Groups such as the international *summer school* and *study abroad year* containing contacts from various countries and with differing affiliations show that a classification algorithm cannot be realized trivially by only using these attributes for classification.

5.3 Evaluation

We employed a cluster to classes evaluation methodology [21], i.e. the test person was asked to create a set of optimal segregated audiences, as shown in Table 2. The manual classification was then compared to the classes our algorithm created. In more detail, starting with the largest manually created group the most similar automatically created group was chosen for evaluation and a similarity value was determined based on member overlap. If the algorithm created more groups than defined in the manual classification, the smallest manually created group was compared to a backup-group, that was created of all members of the residual automated groups.

As mentioned before, the quality of the results highly de-

Table 3: Classification Results

No.	SeA_{man}	SeA_{man} Members	SeA_{aut} Members	Sim.
1	University	77	52	0.633
2	Study abroad year	25	23	0.846
3	Summer school	18	17	0.842
4	High school year	8	8	1.000
5	USA 2	8	10	0.800
6	Other	6	14	0.111
7	Hometown	3	23	0.000
Weighted similarity				0.631

pends on the similarity threshold the algorithm uses to create classes of segregated audiences. To determine the optimal threshold we employed a simulator implemented in our prototype, calculating the average cluster to classes similarity for each threshold between 0 and 100 percent. The results for are depicted in Figure 5. As explained in the previous paragraph, if the algorithm finds more classes than in the test set, a backup class is created. This is shown by the red line, which is capped at the number of user-defined classes. The optimal threshold is not at peak A, as the algorithm only creates two classes with that threshold, whereas at peak B the number of found classes equals the number of classes in the test set and thus peak B denotes the optimal threshold.

Using a threshold of 13.5 percent similarity, Table 3 depicts the classification results for our algorithm and the cluster to classes evaluation against the predefined set (cf. Table 2). The similarity of each pair of classes (manual and automated) is weighted with its average number of members resulting in an overall class similarity of 63.1 percent. This means the automatically created classes overlapped the manually created classes by 63.1 percent in average with the average being weighted using the class sizes.

As Table 3 shows, the result quality highly depends on the contact group. For example, for the classes *Study abroad year* and *Summer school* there is a high similarity between manually and automatically created classes. This can by explained by the high interconnection between the group members and their distance to other groups. As the algorithm highly depends on interconnections between group members, the results are less optimal for other contact groups like *Hometown*. As the members in this group do not know each other, this explains the algorithm's inability to find an appropriate segregated audience for that group.

The results clearly demonstrate the feasibility of automated finding of segregated audiences within a set of contacts which can be further refined by the user and thereby implement assisted social IdM.

A further evaluation comparing our approach with the related Facebook applications mentioned in Section 2 would be desirable. Yet, due to the different nature of the approaches, one could not rule out misleading findings. Friendwheel does not produce any contact groups, thus providing no results that could be compared. For the other two approaches, due to the limited available documentation, one cannot assess if they consider any other input parameters besides the contact's relationships among each other. Also, we consider a clear understanding of the employed algorithms as a precondition before comparing outputs in a meaningful way.

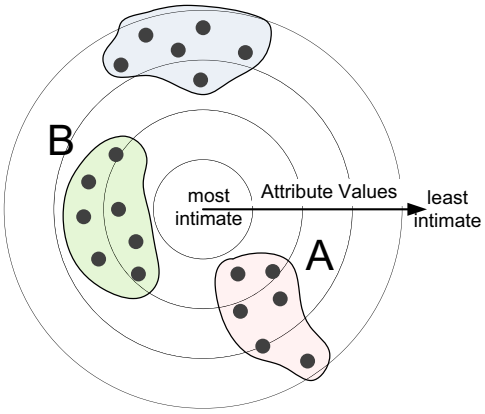


Figure 6: Audience Distribution

If such a comparative evaluation were carried out, it would have to be performed on exactly the same input data, ideally by adding the other application to the test user's Facebook-profile. The cluster to classes evaluation could then be applied to compare the resulting classes to the manually created user groups, yielding a similarity value that would be easy to compare.

5.4 Ideas for improvement

While our implementation has shown the feasibility of assisted social IdM, there are various angles on which both the algorithm and the general idea of assisted IdM can be improved.

The presented algorithm produces promising and usable results, yet there are various other clustering approaches [21] with a number of possible settings and parameters which allow for further research on their suitability in this problem setting. Furthermore, using other input attributes in addition to the contact's mutual friends opens up future work opportunities to further optimize our approach. Still, groups such as *Summer School* (cf. Table 2) show that members of valid segregated audiences may have little or nothing in common besides mutual friends.

The focus of the demonstrator lays on the assignment of contacts to groups, corresponding to function G described in Section 3.2. For a full specification of contexts in social IdM however, an assignment of attributes to contexts is also necessary. The attribute-to-context assignment could happen after the assignment of users to contexts, however there could be interdependencies. For example, adding a certain attribute visibility to a context may require reconsidering the assigned contacts and splitting the group, as only some of the corresponding contacts were meant to see the new at-

tribute. To provide fully assisted social IdM, suggestions for attribute allocations should be presented as well.

The algorithm is capable of classifying contacts into semantically associated groups based on their mutual contacts. However, while this classification is valid and corresponds to the user's classification, as demonstrated by similarity values up to 100%, the user might still prefer to share different information with members of one group. For example, there may be both loose acquaintances met in a student club, as well as members that have grown to be deep friends. This is illustrated in Figure 6, in which the circles represent classes of attribute values, with the most intimate values, such as sexual preferences, located in the center and the least intimate values located in the outer circle. Contacts, represented by dots are placed on those circles corresponding to the user's willingness to share personal information with them. As one can see, groups like group A cause problems for audience segregation, as they cover multiple attribute classes. Groups similar to group B are more suitable, as they only cover a small range of attribute classes.

6. CONCLUSIONS

The pervasiveness of the social web poses many challenges for future research, especially in the field of privacy. Unlike the real-world, where personal information is ephemeral, in the online-world, this information is almost infinitely available while new information is added constantly increasing the existing complexity of managing different identities consistently. This permanency of information poses a great challenge for personal social IdM, since we are no longer free in constructing our identities because contradictory information may be available online. While the social web provides the platform for social IdM to everyone, regardless their technical expertise, privacy controls to raise awareness of implications using the social web and support the users in constructing and managing different identities are still in their infancy.

In this paper we propose assisted social IdM as a means to advance from the current state. In more detail, we provide a conceptualization of current and future social IdM and its formal foundations. Building on that, we propose several approaches to segregate audiences, which is a necessary step to disclose personal information within its belonging context. Our proposal is backed by a prototypical implementation and an evaluation that prove the applicability of the approach.

Acknowledgements

The research leading to these results is partly funded by the European Union within the Padgets project under grant agreement no. 248920.

7. REFERENCES

[1] M. Bauer, M. Meints, and M. Hansen. Structured overview on prototypes and concepts of identity management systems. http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp3-del3.1.overview _on_IMS.final.pdf, September 2005.

[2] H. Buhl and G. Müller. Der "gläserne Bürger" im Web 2.0 - Herausforderungen des "virtuellen Striptease". *Wirtschaftsinformatik*, 4:193–197, 2010.

[3] K. Cameron. The laws of identity. http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf, March 2005.

[4] S. Claubeta, D. Kesdogan, and T. Kölsch. Privacy enhancing identity management: protection against re-identification and profiling. In *Proc. of the 2005 Workshop on Digital Identity Management*, 2005.

[5] European Parliament. EU-Directive 95/46/EC. *Official Journal of the European Communities*, 1995.

[6] E. Goffman. *The Presentation of Self in Everyday Life*. Anchor, 1959.

[7] S. Gürses and B. Berendt. The Social Web and Privacy: Practice, Reciprocity and Conflicts in Social Networks. In *Privacy-Aware Knowledge Discovery: Novel Applications and New Techiques*. Ferrari, Elena and Bonchi, Francesco, 2009.

[8] M. Hansen and M. Meints. Digitale Identitäten - Überblick und aktuelle Trends. *Datenschutz und Datensicherheit - DuD*, 30:543–547, 2006.

[9] P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901.

[10] R. Leenes. Context Is Everything Sociality and Privacy in Online Social Network Sites. In *Privacy and Identity Management for Life*. Springer, 2010.

[11] B. Malin. k-Unlinkability: A privacy protection model for distributed data. *Data & Knowledge Engineering*, 64:294–311, 2008.

[12] H. F. Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79:119–157, 2004.

[13] I. Oomen and R. Leenes. Privacy risk perceptions and privacy protection strategies. In *Policies and Research in Identity Management*. Springer, 2008.

[14] A. Pfitzmann and M. Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, 2010.

[15] PrimeLife. D1.2.1 - Privacy Enabled Communities. http://www.primelife.eu/images/stories/deliverables/d1.2.1-10.04.23-privacy_enabled_communities-public.pdf, 2010.

[16] C. Shirky. Here comes everybody: The power of organizing without organizations. 2008.

[17] Z. Tufekci. Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites. *Bulletin of Science, Technology & Society*, 11:544–564, 2008.

[18] S. D. Warren and L. D. Brandeis. The right to privacy. *Harward Law Review*, 4:193–220, 1890.

[19] E. Weiser. *The Functions of Internet Use and Their Social, Psychological, and Interpersonal Consequences*. PhD thesis, Texas Tech University, 2001.

[20] A. Westin. *Privacy and freedom*. The Bodley Head, 1967.

[21] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan

Kaufmann, 2005.

Anwendungsszenarien als Werkzeug zur (V)Ermittlung des Nutzens von Corporate Social Software

Alexander Richter* Susanne Mörl+, Matthias Trier#, Michael Koch*

* Forschungsgruppe Kooperationssysteme München
Universität der Bundeswehr München
+49(0)89-6004-4406
{a.richter ; michael.koch}@unibw.de

+ Siemens IT Solutions & Services
SIS GO FCS PMC
81739 München
+49(0)89-636-29193
susanne.moerl@siemens.com

University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
+31(0)20-525-5561
trier@uva.nl

ABSTRACT
Im vorliegenden Beitrag stellen wir Anwendungsszenarien als Grundlage zur Ermittlung des Nutzens von Corporate Social Software zur Diskussion. Zu diesem Zweck betrachten wir die Aneignung eines Social Networking Services im Kontext einer Fallstudie, in der wir die auf der Plattform stattfindende Kommunikation mit Hilfe einer Genreanalyse untersuchen. So entwickeln wir ein Verständnis für das „Phänomen Corporate Social Networking“, dessen Anwendungspotentiale, vor dem Hintergrund nur weniger bestehender Studien, bisher nur beschränkt untersucht wurden. Unsere Analyse zeigt, dass die Plattform vor allem zum Teilen von Informationen und zum Diskutieren genutzt wird, aber auch für die Suche nach Unterstützung und zur Vermarktung. Entlang dieser vier Kategorien von Anwendungsszenarien aus Nutzersicht diskutieren wir den Nutzen des Dienstes im vorliegenden Unternehmenskontext und demonstrieren dazu eine Multi-Level-Nutzenanalyse, die neben dem individuellen Nutzen auch den strukturellen Nutzen auf Netzwerkebene berücksichtigt.

Keywords
CSCW, Social Software, Collaboration.

1. MOTIVATION
Die internationale Markt- und Wettbewerbssituation führte in den letzten Jahren zu organisatorischen Veränderungen in vielen Unternehmen und unter anderem zu einem gestiegenen Bedarf an Kollaboration und Informationstransfer zwischen geographisch getrennt arbeitenden Experten. Dieser Bedarf stellt hohe Anforderungen an die innerbetriebliche Expertensuche und insbesondere deren IT-Unterstützung (z.B. [1]).

Im privaten Nutzungskontext haben sich in den vergangenen Jahren Web-Plattformen etabliert, welche die nötigen Funktionen besitzen um soziale Netzwerke aufzubauen, zu pflegen und somit die informelle Kommunikation zu fördern. Allerdings haftet diesen (und auch vergleichbaren kollaborativen Diensten wie Wikis und Weblogs) trotz vielfach positiv belegter Erfahrungen in den Unternehmen (z.B. [2], [15], [19]) noch ein negatives Image bei der Einführung an, da sie mit überwiegend privater Nutzung und Zeitverschwendung in Verbindung gebracht werden. Es bietet sich auf den ersten Blick an, dieser skeptischen Haltung in Unternehmen durch eine Aufwands- und Nutzenkalkulation, beispielsweise die Messung des Return of Investment (ROI), zu begegnen. Dieser und vergleichbare quantitative Ansätze beziehen sich jedoch meist auf Informationstechnologien (IT), die einen (a priori) festgelegten, messbaren Geschäftszweck verfolgen. Da Social Software jedoch eine Unterstützungsfunktion im Unternehmen übernimmt und keinen klar abgegrenzten Geschäftszweck verfolgt (z.B. [17]), lassen die Ansätze sich schwer übertragen. Zudem ist die Ermittlung des Nutzens derart nutzungsoffener kollaborativer Dienste vor der Einführung kaum möglich, da der Nutzen erst nach der Aneignung durch die Anwender nachgewiesen werden kann. Die Dienste müssen hierfür zu einem Bestandteil der Gruppenpraktiken und organisationalen Prozesse werden (z.B. [31]). Durch diese Einbettung in Interaktionsstrukturen und –prozesse kann weiterhin ein Nutzen auf Gruppen- bzw. Netzwerkebene identifiziert werden (vgl. auch [10]). Insgesamt erscheint das Potential des Einsatzes von Corporate Social Software jedoch noch nicht ausreichend erforscht, um den Nutzen der Dienste zu belegen (vgl. [24]).

Ziel dieses Beitrags ist es daher, zur Ermittlung des Nutzens von Corporate Social Software, die Analyse von Anwendungsszenarien als geeignete Perspektive vorzuschlagen und anhand eines empirischen Beispiels aufzuzeigen. Neben der theoriegeleiteten Entwicklung des an Anwendungsszenarien orientierten Konzepts zur Nutzenermittlung, stellt dabei die Identifikation der geläufigen Anwendungsszenarien über eine Genreanalyse den wesentlichen Beitrag dieses Artikels dar. In diesem Kontext positionieren wir weiterhin unseren Ansatz als ein Multi-Level Konzept, welches, über die Messung der individuellen Nutzung durch die Mitarbeiter hinausgehend, auch die wichtigen interpersonellen Entwicklungen im Netzwerk als eine strukturelle Ebene der Nutzenmessung einbezieht. Diese theoriegeleitete Integration der Netzwerkebene zur Nutzenermittlung von Social Software ist jedoch Teil eines laufenden umfassenderen Forschungsprojekts und wird in diesem Beitrag aufgrund der Platzbeschränkung nur in den Grundzügen kurz vorgestellt.

Wir beginnen unseren Beitrag mit einer Einführung zu Social Networking (Services) (Kap. 2) und einer Diskussion bisheriger Ansätze zur Erfolgsmessung dieser und vergleichbarer Dienste (Kap. 3). Nach Einführungen in Untersuchungsfeld (Kap. 4) und -methodik (Kap. 5) zeigen wir exemplarisch anhand eines international eingesetzten Social Networking Service fünfzehn in der vorliegenden Studie identifizierte Anwendungsszenarien des Einsatzes der Dienste auf (Kap. 6) und diskutieren diese vor dem Hintergrund der Nutzenanalyse. Wir beenden den Beitrag mit dem Ausblick auf geplante und weitere mögliche Arbeiten (Kap. 8).

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

2. SOCIAL NETWORKING

Plattformen zum Social Networking erfreuten sich in den vergangenen Jahren zunehmender Beliebtheit unter den Internet-Nutzern. Gleichzeitig zog das Thema auch zunehmendes Interesse aus der Wissenschaft auf sich. Die meisten existierenden Forschungsarbeiten im Themenfeld konzentrieren sich auf die überwiegend zur Privatnutzung vorgesehenen Internet-Plattformen und es finden sich nur wenige Studien zum Einsatz der Dienste innerhalb der Unternehmensintranets. Dabei stellen die Ergebnisse letztgenannter Studien den Ausgangspunkt dafür dar, sich im weiteren Verlauf des Beitrags tiefergehender mit dem Einsatz von Corporate Social Networking Services zu beschäftigen.

2.1 Social Networking im Internet

Web-Plattformen, die Internetnutzer dabei unterstützen soziale Netzwerke aufzubauen, zu pflegen und somit die Kommunikation mit anderen Nutzern aufrecht zu erhalten, werden als Social Networking Sites/Services (SNS) bezeichnet. Laut Boyd und Ellison [6] sind SNS "web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system". Koch et al. [20] definieren SNS als „Anwendungssysteme, die ihren Nutzern Funktionalitäten zum Identitätsmanagement (d.h. zur Darstellung der eigenen Person i.d.R. in Form eines Profils) zur Verfügung stellen und darüber hinaus die Vernetzung mit anderen Nutzern (und so die Verwaltung eigener Kontakte) ermöglichen". Um beide Definitionen zusammen zu bringen, sind die wichtigsten Funktionen eines SNS (1) die Darstellung des eigenen Selbst in Form eines Profils und (2) die Fähigkeit, andere Nutzer als Kontakte hinzuzufügen. Gleichzeitig stellen diese beiden Kernfunktionen auch die wichtigsten Motivationen für deren private Nutzung im Web dar. So wurde in mehreren Studien nachgewiesen, dass es den Anwendern wichtiger ist, bestehende Freundschaften zu pflegen als neue Leute (über Such- oder Empfehlungsfunktionen) kennen zu lernen (z.B. [23], [38]). Eine Anzahl weiterer Studien beschäftigt sich damit, wie die Plattformen genutzt werden, um sich selbst darzustellen (z.B. [7], [36]).

Neben den beiden Kernfunktionen lassen sich vier weitere Funktionen von SNS identifizieren [29]: Expertensuche, Unterstützung von Kontext- und Netzwerkawareness sowie Unterstützung eines gemeinsamen Austauschs. Erstere Funktion wird insbesondere auf Web-Plattformen mit geschäftlichem Fokus genutzt, unter anderem, weil diese als Tor zu neuen Arbeitsplätzen oder geschäftsrelevanten Kontakten gesehen. Diese werden von ihren Nutzern weniger spielerisch und kommunikations-zentriert wahrgenommen [34]. Skeels und Grudin [33] untersuchten in diesem Zusammenhang die geschäftliche Nutzung von zwei öffentlichen SNS (Facebook und LinkedIn) durch Microsoft-Mitarbeiter. Das Ergebnis zeigt eine intensive Nutzung sowohl im privaten als auch im geschäftlichen Umfeld mit komplexen Nutzungsmustern abhängig von dem Softwaresystem und dem Alter der Nutzer. Obwohl die Ergebnisse der Studie Aufschluss über die Nutzungsmöglichkeiten solch offener Web-Plattformen im Unternehmenskontext geben, muss festgestellt werden, dass sich diese von der Nutzung interner SNS unterscheidet, z.B. weil alle Informationen außerhalb der Firewall und somit außerhalb des Einflusses der Firma bleiben.

2.2 Social Networking im Intranet

Als Vorreiter von Forschungsarbeiten zum Einsatz von Corporate Social Networking Services können mehrere Studien einer Gruppe um Dimicco und Geier gesehen werden, die die Nutzung eines SNS namens "Beehive" bei IBM untersuchten. Dieses wurde im Jahr 2007 als Prototyp aufgesetzt, um die Vernetzung von IBM Mitarbeitern zu fördern. Der Fokus des Prototypen lag nicht darauf, die Zusammenarbeit in wissensintensiven Projekten zu unterstützen, sondern vielmehr Leute zu ermutigen, sich in einem weniger geschäftlichen Kontext zu vernetzen. In diesem Kontext wurden drei Motive für die Nutzung identifiziert: (1) Nutzer wollten sich auf einer persönlichen Ebene mit ihren Kollegen vernetzen, (2) ihre Karriere im Unternehmen vorantreiben und (3) ihre Projekte vermarkten [14]. Des Weiteren wurden in der Studie die auf der Plattform ausgetauschten Inhalte und genutzten Funktionen analysiert [13] sowie Faktoren, die auf die Nähe der (generellen und persönlichen) Beziehungen der Nutzer hinweisen [40].

In einem Vergleich dreier weiterer Unternehmens-SNS von SAP, Accenture und IBM zeigen Richter und Riemer [30] wie die Dienste zur Zusammenarbeit genutzt werden (können), beispielsweise um Experten zu identifizieren, einen persönlichen Kontext aufzubauen und existierende professionelle Beziehungen aufrecht zu erhalten. Darüber hinaus wird von den Autoren die Tatsache diskutiert, dass es sich bei SNS um nutzungsoffene Dienste handelt und zur besseren Aneignung der Dienste vorgeschlagen, im Rahmen eines „Ko-Evolutions-Prozesses" neue Funktionen in bestehende Systeme zu integrieren.

Schließlich untersuchte Brzozowski [8] die Nutzung einer Plattform namens „WaterCooler" bei Hewlett Packard. Dabei kam er zum Schluss, dass der Dienst seine Nutzer besonders dabei unterstützt sich besser zu vernetzen und ein Gewahrsein für die Aktivitäten anderer Geschäftsbereiche zu entwickeln.

Den wenigen dokumentierten Anwendungsfällen von Corporate Social Networking ist gemein, dass sie sich zwar die Frage nach Aneignung und Nutzung der SNS stellen, aber nicht nach dem Nutzen einer stärkeren Vernetzung der Mitarbeiter für die Organisation. Obwohl wir vor dem Hintergrund dieses beschränkten Forschungskanons glauben, mit unserer Studie auch zum Verständnis dieses Phänomens beitragen zu können, möchten wir einen Schritt weitergehen und aufbauend auf den Anwendungsszenarien den Nutzen der Dienste für Anwender und Unternehmen diskutieren. Im folgenden Abschnitt gehen wir daher zunächst einmal auf den Status quo der Erfolgsmessung von Informationssystemen im Allgemeinen und von Social Software bzw. SNS im Konkreten ein.

3. NUTZEN VON SOCIAL SOFTWARE

Nicht erst seit Carrs Artikel „IT Doesn't Matter" [9], in dem argumentiert wird, dass es historisch gesehen keine Korrelation zwischen Investitionen in IT und Produktivität gegeben hat, wird kontrovers diskutiert, ob und wie IT zur Verbesserung der Produktivität beiträgt. Zweifellos ist die Messung des Erfolges (oder der Wirksamkeit) von Informationssystemen von entscheidender Bedeutung für die Planung von Investitionen in die Unternehmens-IT.

Im Folgenden betrachten wir zunächst die individuelle Nutzung. Dann charakterisieren wir die Netzwerkstruktur als eine zusätzliche und bisher wenig betrachtete Ebene der Erfolgsmessung. Als Kombination beider theoretischer Dimensionen demonstrieren wir dann das Potential einer Multi-Level-Analyse.

3.1 Individuelle Nutzung

In einem weithin bekannten Modell schlagen Delone und McLean [12] eine Taxonomie als Rahmenwerk für die Konzeption und Operationalisierung des Erfolges von Informationssystemen vor. Als Ergebnis ihrer Meta-Analyse identifizieren die Autoren insb. die wesentliche Rolle der Faktoren Nutzung (Häufigkeit, Volumen) und Nutzerzufriedenheit (bzgl. Angebot auf Informationsnachfrage), welche im Ergebnis individuelle Vorteile (Recall, Information Awareness, Effektivität der Entscheidungsfindung) und organisatorische Vorteile (Kostenreduktion, Umsatzsteigerungen) erzeugen. Das Modell wurde auch für die Erfolgsmessung von Social Software angepasst (z.B. von Räth et al. [28]). Letztgenannte weisen jedoch selbst darauf hin, dass ein solchermaßen angepasstes Modell lediglich einige Anhaltspunkte zu möglichen Variablen der Erfolgsmessung liefern kann, nicht aber ein Hilfsmittel ist, um zu tatsächlichen Größen zu kommen.

Melville et al. [25] haben anhand einer Literaturrecherche ein Modell zum geschäftlichen Beitrag von IT-Systemen abgeleitet. Sie argumentieren, dass der Effekt eines IT-Systems anhand der Veränderungen der Prozesseffizienz gemessen werden kann. Aufbauend auf dem Modell von Melville et al. schlagen Kettles und David [18] vor, die durch Social Software veränderte Prozess-Performanz mit Hilfe von Nutzer-Workshops zu ermitteln. Andere von Praxis und Wissenschaft vorgeschlagene Ansatzpunkte zur Erfolgsmessung von Social Software sind beispielsweise die Quantifizierung des Kommunikationsverhaltens (weniger E-Mails, weniger CC-E-Mails, weniger For-Your-Interest-Mails, …) oder des Grades der Beteiligung an einzelnen Dokumenten und Entscheidungen sowie die Messung der allgemeinen Mitarbeiterzufriedenheit (z.B. [19]).

Den o.g. und vergleichbaren Modellen (wie beispielsweise weiteren bekannten Modellen von Gable et al. [16] sowie Shang und Seddon [32]) ist gemeinsam, dass der Nutzen über Fragebögen (oder Workshops) ausgehend von durch die Nutzer wahrgenommenen Veränderungen ermittelt wird. Dies ist die Erkenntnis von Urbach et al. [37], die in einer Metanalyse mehrdimensionale Ansätze zur Erfolgsmessung von Informationssystemen untersuchen und einen Stand der aktuellen Forschung auf diesem Gebiet geben. Wie einführend bereits erläutert, ist das problematische bei Erhebungen anhand von Nutzer-Fragebögen (oder vergleichbaren Methoden), dass aufgrund der Nutzungsoffenheit von Social Software (oder vergleichbaren Diensten) der Nutzungskontext, nach dem gefragt wird, klar sein muss. Allgemeine Fragen zum Erfolg bringen ansonsten genauso allgemeine Ergebnisse zu Tage.

Eine andere Herangehensweise verfolgen Muller et al. [24] für den Einsatz von Social Software, indem sie den Return on Contribution (ROC) betrachten. Auch mit dem ROC wird der Nutzen des Systems den entstandenen Kosten gegenüber gestellt. Allerdings stellt er die Menschen, die von dem System profitieren denjenigen Anwendern gegenüber, die aktiv in dem System beitragen. Somit ist der ROC auch in Situationen einsetzbar, in denen eine direkte monetäre Bewertung nicht möglich ist. Mit dem ROC kann jedoch wiederum nur nachgewiesen werden, in welchem Ausmaß das System Nutzen für die Anwender bringt. Ob und vor allem welcher Nutzen dadurch für die Organisation generiert wird, bleibt weiter unbekannt.

Aus diesem Grund argumentieren Cooper et al. im Juni 2010 [10], dass die Definition einer Metrik für den Einsatz von Plattformen zum Social Networking davon abhängig ist, welche Anwendungsszenarien unterstützt werden sollen. Sie schlagen die folgenden drei Schritte zur Erstellung eines Rahmenwerks zur Erfolgsmessung vor: 1. Die Identifikation eines klaren Anwendungsfalls und dessen Ziele. 2. Anhand dieser Ziele kann festgelegt werden welche Art von Metrik zum Einsatz kommt und in welchem Detailgrad. 3. Davon abhängig können die entsprechenden Instrumente zur Messung der festgelegten Metrik ausgewählt werden.

Zur Messung der Zielerreichung eines Anwendungsfalls im zweiten Schritt unterscheiden Cooper et al. drei Ebenen von möglichen Metriken, die mit den abhängigen Variablen „Nutzung“ „Individueller/Organisatorischer Nutzen“ des o.g. Delone/McLean-Modells korrespondieren:

1. Vitality („Nutzung“): Untersucht werden grundlegende Kennzahlen für die Benutzeraktivität, wie die Zahl der geschriebenen Weblog-Einträge oder die Aufrufe eines Weblogs. Diese geben einen kurzfristigen Überblick über die Aneignung eines Dienstes auf der Messung des individuellen Verhaltens.

2. Capability („Individueller Nutzen“): Diese Metriken sollen dazu beitragen den Wert der eingesetzten Dienste für die Organisation zu beurteilen. Es wird vorgeschlagen zu evaluieren, wie viel schneller und effizienter Mitarbeiter Expertenwissen innerhalb der Organisation ausfindig machen können, um ein Problem zu lösen. Die Messung könnte beispielsweise erfolgen, indem die Häufigkeit der Mitarbeiterinteraktion auf der Plattform oder die Zunahme der Vernetzung der Mitarbeiter untereinander untersucht wird.

3. Business Value („Organisatorischer Nutzen“): Diese Metriken orientieren sich zur Messung des langfristigen Erfolgs an „Key Process Indicators“ (KPIs) oder an „Key Value Indicators“ (KVIs), die stark abhängig von der Geschäftsstrategie eines Unternehmens sind. Als Beispiel wird ein Call Center genannt, in dem mehr Anrufe bearbeitet oder Kundenprobleme schneller gelöst werden können.

Mit dem vorliegenden Beitrag gehen wir den von Cooper et al. vorgeschlagenen Weg. Wir identifizieren Anwendungsszenarien und legen damit die Grundlage für weitere Arbeiten, die (aufbauend auf den ermittelten Anwendungsszenarien und der damit verbundenen Zielsetzung) Überlegungen zu Metriken und Instrumenten anstellen (möchten/werden). Dabei ist ein Anwendungsszenario durch die Nutzung des Dienstes im Rahmen einer konkreten Praktik und durch ein damit zusammenhängendes Ziel definiert.

Im Kontext der Messung von „Capability“ schlagen Cooper et al. [10] unter anderem die Messung der Vernetzung der Mitarbeiter vor und verweisen somit auf die wichtige strukturelle Ebene der Mitarbeiterinterkation als einen relevanten Aspekt der Nutzenmessung. Im folgenden Abschnitt wird genauer hergeleitet, wie diese Netzwerkebene als Teil eines Multi-Level Konzepts zur Nutzenmessung fungieren kann.

3.2 Netzwerkebene

Social Software zeichnet sich dadurch aus, dass die Nutzer im Zeitablauf über ihre Interaktion Netzwerke bilden. Das Netzwerk mit seinen speziellen strukturellen Eigenschaften (z.B. Effizienz von Suche, Informationsaustausch, Themenintegration) kann als schwer zu imitierende organisatorische Ressource interpretiert werden [41]. Daher bietet es sich an, die interpersonellen strukturellen Effekte der Softwarenutzung auf Netzwerkebene in die Konzeption der Nutzenermittlung miteinzubeziehen. Als theoretische Basis können hierbei die Nutzungsprozesse durch Individuen und Gruppen herangezogen werden. Sie wurden für die Bewer-

tung des Nutzens eines eingeführten Systems mehrfach als herausragende Einflussgröße für den Unternehmenserfolg einer Systemeinführung festgestellt [35].

Nutzung ist der wesentliche moderierende Faktor zwischen Technik und Geschäftserfolg. Der theoretische Zusammenhang zwischen technischen Systemeigenschaften und Nutzung wird dabei über sozialpsychologische Ansätze mit Faktoren wie z.B. wahrgenommene Nützlichkeit erklärt [11]. Vor kurzem wurde dieses Basismodell um weitere Faktoren ergänzt, welche sich in einer interaktiven Mehrbenutzerumgebung (z.B. eine Community) als sehr einflussreich gezeigt haben. Bagozzi et al. [3] belegen empirisch die Rolle von sozialer Identität (Gruppenidentifikation) für die erfolgreiche Nutzung von Online Communities. Analog diskutieren Bouman et al. [5] die Rolle von „Sociality" als wichtiges Designelement von sozialer Software. Hierunter fallen Aspekte wie soziale Interaktivität. Eine weitere Perspektive, welche die Rolle von individuellen Aktivitäten der technologischen Aneignung und Nutzerinteraktion bei gruppenorientierten Systemen in den Mittelpunkt stellt, ist der institutionstheoretische Ansatz sowie die darauf aufbauende „Adaptive Structuration Theory" nach Poole and DeSanctis [27]. Diese Theorie motiviert eine Betrachtung der strukturbildendenden Effekte der technischen Aneignung innerhalb einer sozialen Struktur.

Beide Perspektiven, die Adaptive Structuration Theory und die neuen sozialpsychologischen Modelle zur Systemnutzung legen nahe, soziale Interaktionsprozesse und ihre Resultate auch als Einflussfaktor für Effizienz- und Qualitätssteigerungen heranzuziehen. Neben der Messung individueller Aktivität wird daher im vorliegenden Beitrag das Potential der Betrachtung der strukturellen Ebene i.S.d. Vernetzung zwischen den Nutzern aufgezeigt. Zur Operationalisierung entsprechender Messgrößen in den o.g. Domänen Capability und Business Value, kann die Methode der Sozialen Netzwerkanalyse (SNA) eingesetzt werden.

Cooper et al. [10] schlagen in ihren Messkonzept für Social Software beispielsweise die Dichte des Netzwerks als Indikator für größeren Informationsdurchsatz und effektivere Suchprozesse vor. Weitere Faktoren können die Zentralisierung oder einfach die durchschnittliche Anzahl an Kontakten sein. Durch den Einbezug einer SNA wird eine nachvollziehbare Brücke von individuellem und organisatorischem Nutzen geschlagen: Die Aktivitäten Einzelner verdichten sich zu einer messbaren strukturellen Eigenschaft der informellen Organisation. Basierend auf diesem theoretischen Multi-Level Ansatz liegt der Schwerpunkt des weiteren Beitrags auf der Identifikation der wesentlichen Anwendungsfälle in einer Fallstudie des Unternehmenseinsatzes. Danach demonstrieren wir die Analyse des Nutzens (bzw. der Erfolgsmessung der Plattform) auf individueller und struktureller Ebene.

4. TECHNOWEB 2.0

Bereits seit über zehn Jahren sammelt die Siemens AG mit verschiedenen Tools Erfahrungen im Bereich Wissensvernetzung. So wurde im Jahr 1998 in einem Software-Entwicklungs-Bereich mit 4000 Mitarbeitern bei Siemens Österreich das sog. TechnoWeb 1.0 eingeführt. Dieses wurde in den Jahren 2007 und 2008 auf den gesamten Bereich Siemens IT Solutions und Services und damit auf rund 10.000 potenzielle Nutzer weltweit ausgedehnt.

Im Juni 2009 wurde vom Zentralvorstand der Siemens AG entschieden, diese Plattform auf eine neue Technologie, Liferay, zu stellen und konzernweit auszurollen, weil man das Potential des Dienstes für das ganze Unternehmen überprüfen wollte. Die neue Plattform sollte eine heterogene Gruppe von über 80.000 potenziellen Nutzern dabei unterstützen, sich gegenseitig zu vernetzen und auszutauschen. Ziel der Einführung auf Konzernebene war und ist es, Synergien, v.a. auf technologischer Ebene, zwischen den verschiedenen Sektoren (Healthcare, Industry, Energy) sowie den sektorübergreifenden Einheiten zu realisieren. Es ist keine Sprache vorgegeben, der größte Teil der Kommunikation findet jedoch auf Englisch statt.

Das zentrale Element von TechnoWeb 2.0 sind Netzwerke, in denen Mitarbeiter neue Ansprechpartner zu bestimmten Themen finden und Erfahrungen austauschen können. Grundsätzlich kann jeder Mitarbeiter sich allen Netzwerken seines Interessengebietes anschließen oder neue Netzwerke gründen, wenn es zu diesem Thema noch keines gibt. Die Netzwerke selbst enthalten neben einer Beschreibung ihres Zieles eine Liste aller Netzwerk-Mitglieder und deren Aktivitäten im Netzwerk. Die Mitglieder eines Netzwerks haben die Möglichkeit, Nachrichten an alle anderen Mitglieder zu schreiben und können auch mit Hilfe von RSS-Feeds über interne oder externe Neuigkeiten zu dem jeweiligen Thema informiert bleiben. Netzwerke können sich in Partner-Netzwerken organisieren und vernetzen. Die Plattform greift darüber hinaus auf viele der in Abschnitt 2.1 genannten typischen Funktionen eines SNS zurück: Jeder Nutzer hat die Möglichkeit ein persönliches Profil anzulegen, in dem er sich selbst beschreiben kann. Zusätzlich werden hier vom System die Netzwerke angezeigt, in denen er Mitglied ist, Tags, die der Nutzer für Netzwerke oder Beiträge vergeben hat und sämtliche Aktivitäten des Nutzers im System (falls vom Nutzer freigegeben). Der Nutzer vernetzt sich im TechnoWeb 2.0 nicht direkt mit anderen Personen, wie es bei SNS üblich ist, sondern er folgt diesen indirekt, indem er sich wie oben dargestellt Netzwerken zu bestimmten Themen anschließt.

In den Netzwerken können die Nutzer, vergleichbar einem Diskussionsforum, Nachrichten (im Folgenden der Kürze halber auch News genannt) austauschen. Zusätzlich bietet TechnoWeb die Möglichkeit, über sog. „Urgent Requests" zu kommunizieren. Hat ein Nutzer ein spezifisches Problem und findet kein entsprechendes Netzwerk dazu oder erhält keine befriedigende Antwort aus dem entsprechenden Netzwerk, kann er sich damit an eine sehr große Zahl von TechnoWeb-Nutzern wenden. Der Urgent Request wird dazu vom Fragenden einer oder mehreren von neun Kategorien zugeordnet und wird an alle Nutzer, die diese Kategorie abonniert haben, per Mail versandt. Damit ist der Urgent Request ein sehr mächtiges Instrument, um schnell geschäftsrelevante Probleme mit dem versammelten Wissen von Siemens zu lösen. Nach einer Siemens-internen Auswertung stellten die TechnoWeb-Nutzer zum Zeitpunkt der Studie ca. 1-2 Urgent Requests am Tag die durchschnittlich 8 Antworten erhielten. Dabei variiert die Anzahl der Antworten stark abhängig vom Detailgrad der Fragestellung.

TechnoWeb 2.0 ging im Oktober 2009 in einen sechsmonatigen Pilotbetrieb. Ziel der Pilotphase war es unter anderem herauszufinden, welche Potentiale das TechnoWeb 2.0 für einzelne Siemens-Bereiche und für den Konzern als Ganzes bringt. Des Weiteren sollte in der Pilotphase eine kritische Masse an Anwendern und Themen erreicht werden, um den konzernweiten Rollout zu unterstützen. Daher sollten schon so viele Netzwerke zu den verschiedensten Themen vorhanden sein, dass es attraktiv ist sich anzuschließen. In der Pilotphase wurde die Existenz des TechnoWeb 2.0 ausschließlich in den Piloteinheiten beworben (und begleitet). Das System selbst war aber von Anfang an für alle

Siemens-Mitarbeiter frei zugänglich, so dass am Ende dieser Phase Mitarbeiter aus allen Sektoren vertreten waren.

5. METHODIK

Wie einführend erläutert, ist es das Ziel dieses Beitrags Anwendungsszenarien als Grundlage zur Ermittlung des Nutzens von Corporate Social Software zu identifizieren. Als Untersuchungsfeld wurde mit Siemens ein Unternehmen ausgewählt, bei dem wir aufgrund seiner Größe und der Heterogenität des Geschäftes eine Vielfalt an Anwendungsszenarien vermuteten. Dabei wurde die Nutzung der Plattform mit Hilfe einer Forschungsmethode der qualitativ empirischen Sozialforschung, der Genre-Analyse, untersucht. Das diesen Methoden zugrundeliegende interpretative Paradigma verweist auf eine soziale Konstruktion der Realität, die nicht objektiv verstanden oder nicht direkt erfahren werden kann, sondern vom Forscher über seine Sinne interpretiert wird. Zur empirischen Analyse einer Situation wird es folglich als unbedingt notwendig erachtet, den Kontext zu berücksichtigen (vgl. z.B. [22]). In diesem Fall liegt das qualitative Vorgehen darin begründet, dass damit insbesondere die Kontextabhängigkeit sozialen Handelns betont wird, die zur Identifikation und dem Verständnis der Anwendungsszenarien, also eines fallbezogenen Handelns, notwendig ist. Ziel des Vorgehens ist es somit, möglichst viele Aspekte des Untersuchungsgegenstands zu erfassen und insbesondere auch die Beziehungen von Merkmalen des Untersuchungsgegenstands untereinander und im Kontext zu berücksichtigen und nicht kausale Verallgemeinerungen daraus abzuleiten (vgl. [39]).

In der vorliegenden Studie werden daher die Anwendungsszenarien einer Plattform zum Social Networking in einer Fallstudie identifiziert und somit die Möglichkeit ergriffen die unmittelbare Nutzung im Kontext zu untersuchen, wie von Orlikowski [26] vorgeschlagen. Einer der Autoren war aktiv in die Einführung der Plattform involviert. Aufgrund mehrerer über Monate verteilte Gespräche und dreier Workshops mit Treibern und Nutzern bekamen auch die weiteren Autoren Einblicke in die Nutzung von TechnoWeb 2.0. Für unsere Analyse wählten wir die 10 Netzwerke mit den meisten Mitgliedern (zwischen 123 und 437 Personen) aus. Zusätzlich untersuchten wir 40 zufällig ausgewählte Urgent Requests. Davon stammt jeweils die Hälfte aus der Pilotphase und aus dem Zeitraum nach der Pilotphase.

Zur Identifikation der Anwendungsszenarien untersuchten wir im Rahmen einer Kommunikationsgenreanalyse wiederkehrende Kommunikationsmuster innerhalb der Netzwerke, die sich als Reaktion auf wiederkehrende Kommunikationssituationen bildeten (vgl. dazu [3]). Auf diese Weise bekamen wir einen Überblick über die Kommunikationspraktiken der Mitglieder der Netzwerke (vgl. dazu [42]) und über die Art und Weise wie diese kommunizierten [4]. Die Genres dienten also zum Verständnis der Kommunikationspraktiken einer sozialen Gruppe, indem diese anhand der Dimensionen Inhalt, Form und Zweck gekennzeichnet werden [21]. Im vorliegenden Fall wurde jeder Post im Kontext der stattfinden Kommunikation sowohl nach dem Kommunikationszweck („Welches Ziel wird mit diesem Post verfolgt?“, z.B. die Suche nach einem Diskussionspartner) als auch nach der Form („Urgent Request“ oder „Nachricht in einem Netzwerk“) kodiert, wobei einem Post mehrere Genres zugeordnet werden können.

Zur Kodierung wurden die ausgewählten „Urgent Requests“ und „Nachrichten“ in die Software Atlas.ti 6.0 geladen und von zweien der Autoren gelesen und kodiert. In einem zirkulären Prozess wurden die jeweils kodierten Teile verglichen und von den Autoren diskutiert. Wenn ein neuer Genrekandidat auftauchte, der nicht einem der bereits bestehenden Genres zuordenbar war, wurden auch die bereits kodierten Daten unter Berücksichtigung des neuen Genres kodiert. Dies geschah solange bis alle Daten kodiert waren und kein neuer Kandidat mehr auftauchte. So wurden beispielsweise die Nachrichten “I would like to inform you that there is a webinar organized this Friday with the following description …” und „It might be interesting for you following our planned idea contest on sustainability which is going to run in the intranet from April 19 on.“ zunächst als *Einladung* kodiert. Im weiteren Verlauf der Analyse stellt sich jedoch heraus, dass die zweite Einladung dem primären Ziel der Vermarktung einer internen Initiative diente. Somit wurde diese Nachricht umkodiert und die bereits kodierten Daten wurden nach ähnlichen Beispielen durchsucht.

Insgesamt wurden 640 einzelne Genrecodes vergeben, davon 331 für die untersuchten Urgent Requests und 309 für die Beiträge in den Netzwerken. In einigen Fällen wurden auch mehrere Genres für dieselbe Nachricht vergeben, wie z.B. für „An interesting article about Carbon Mgt Software in the economist: ... We're starting to get enquiries about this topic and we're not real in the position to deal with them. If anyone has any experience here, then please get in contact with me.“ Hier wurde der erste Teil der Nachricht mit dem Hinweis auf den Artikel dem Genre „Weitergabe von Links zu Artikeln, ...“ kodiert, während der restliche Teil dem Genre „Suche nach Diskussionspartner“ zugeordnet wurde. Auf diese Weise wurden insgesamt fünfzehn verschiedene Einzelgenres identifiziert, die für die Nutzung des Dienstes im Rahmen einer konkreten Praktik und ein damit zusammenhängendes Ziel stehen. Diese Praktiken werden wir im Folgenden als Anwendungsszenarien erläutern (vgl. dazu Abschnitt 3.1).

6. ANWENDUNGSSZENARIEN

Die insgesamt fünfzehn identifizierten Anwendungsszenarien lassen sich in vier Kategorien einteilen: „Teilen von Informationen“, „Diskutieren“, „Suche nach Unterstützung“ und „Vermarkten“. Tabelle 1 enthält eine vollständige Aufstellung der identifizierten Anwendungsszenarien und ihre Zuordnung zu den Kategorien. Darin sind auch die Häufigkeit der einzelnen Anwendungsszenarien in den Netzwerk-Nachrichten und der Urgent Requests aufgeschlüsselt (zu deren Unterscheidung vgl. Abschnitt 4).

6.1 Teilen von Informationen

Das Teilen von Informationen ist sowohl von der Häufigkeit als auch von der Vielfalt der Anwendungsszenarien führend. Allgemein kann unterschieden werden, ob Informationen als Antwort auf eine Suche nach Unterstützung geteilt werden (häufig als Antwort auf einen Urgent Request) oder ob die Kommunikation mit dem Teilen einer Information beginnt. Beide Arten grenzen sich aber klar von der Diskussion ab, da hier nur eine Information weitergegeben wird ohne, dass darauf eine Antwort erwartet bzw. nachgefragt wird.

Neuigkeiten oder Statusmeldungen werden beispielsweise von Zentralabteilungen publiziert, um diese unternehmensweit zu verbreiten. So gibt es ein Netzwerk „Quality Management @ SIS“, das von drei Mitarbeitern aus der zentralen Qualitätsabteilung gepflegt wird. Diese stellen regelmäßig Informationen zur Verfügung, wenn sich qualitätsrelevante Regelungen verändert haben, z.B. “Dear colleagues, please mention that the Quality Gate checklist PM-20-40-70 is updated.” Statusmeldungen können aber auch Antworten eines Mitarbeiters enthalten, der nach dem aktuellen Status eines Dokumentes oder Projektes gefragt

wurde. Eine spezielle Art des Teilens sind *Einladungen* zu Konferenzen oder Messen, die von Mitarbeitern gepostet werden, um Kollegen mit denselben Interessen auf diese Veranstaltungen hinzuweisen. Dies erfolgt meistens als Initiative eines Mitarbeiters ohne dass dem eine konkrete Frage voraus geht. Zudem werden auch *Links* gepostet, die auf Whitepapers, Bücher oder Blogeinträge verweisen, beispielsweise „The Cloud Computing White Paper is published in...".

Hinweise auf Personen erfolgten in der Stichprobe stets als Antwort auf eine Frage, in der nach einer Lösung oder einem Ansprechpartner gesucht wird. Hier kennt die antwortende Person selbst zwar nicht die Lösung, meint aber jemanden zu kennen, der weiterhelfen kann und schlägt dem Suchenden vor, diese Person zu kontaktieren. Die *Weitergabe von Erfahrungswissen* erfolgt auch meist als Antwort auf eine Frage, indem ein Kollege berichtet, wie ein ähnliches Problem bereits in seiner Abteilung gelöst wurde.

Tabelle 1: Überblick über alle 15 Anwendungsszenarien

Anwendungsszenario	Alle	UR	News
Teilen von Informationen	**49%**	**56%**	**44%**
Neuigkeiten/ Statusmeldungen	20%	19%	20%
Einladungen	4%	0%	7%
Weitergabe von Links zu Artikeln, ...	11%	9%	13%
Hinweise zu einem Ansprechpartner	5%	11%	0%
Weitergabe von Erfahrungen	10%	17%	4%
Diskutieren	**28%**	**20%**	**34%**
Suche nach Diskussionspartner	3%	0%	5%
Nachfrage	4%	4%	4%
Meinungsäußerung	20%	16%	23%
Weiterentwickeln einer Idee	1%	0%	1%
Suchen nach Unterstützung	**9%**	**17%**	**3%**
Lösung eines technischen Problems	3%	7%	0%
Entscheidungs-Vorbereitung	2%	3%	1%
Suche nach einer Referenz	1%	1%	1%
Identifikation eines Ansprechpartners	3%	6%	1%
Vermarkten	**6%**	**2%**	**9%**
Zentrale Initiativen	4%	0%	7%
Interne Dienstleistungen, neue Technologien oder Methoden	2%	2%	2%

6.2 Diskutieren

In Diskussionen tauschen sich Mitarbeiter üblicherweise innerhalb eines Netzwerkes aus. Die Diskussionen beginnen beispielsweise, wenn ein Mitarbeiter einen *Diskussionspartner sucht*, um ein Konzept, das er entworfen hat zu diskutieren. („Hi there, as you know, we plan to carry out a jam on disruptive ICT trends – do you think this makes sense? We have identified 6 initial topics ..."). Dabei kommt es regelmäßig vor, dass der Antwortende *nachfragt*, um das Konzept besser zu verstehen und auf eine Erklärung hofft. Eine Diskussion kann sich aber auch aus einer Frage nach einer technischen Lösung entwickeln, in dem verschiedene *Meinungen geäußert* bzw. kontrovers diskutiert werden. ("Thank you for an interesting mindmap and for sharing your experience with us. From my perspective I'm missing the topics…")

Wie sich in der Analyse zeigte, kommt es nicht häufig, aber regelmäßig vor, dass die Mitarbeiter das TechnoWeb dazu nutzen, um in einem frühen Stadium eine *Idee zur Diskussion zu stellen*. Im Netzwerk: „iPhone@Siemens" hatte z.B. ein Mitarbeiter die Idee für eine neu iPhone app und wollte wissen, wie Kollegen diese Idee beurteilen. („…With the amount of people having iphones, would it not be good to have an app ... the application ideas are many, but I have not seen any, so I thought it would be a good idea to throw some idea into the pool. Comments welcomed."). In der Folge brachten verschiedene Kollegen ihre Sichtweisen und Erfahrungen ein, um die Idee weiterzuentwickeln.

6.3 Suchen nach Unterstützung

Aus Gesprächen mit den Nutzern wissen wir, dass es häufig vorkommt, dass diese bei Kunden vor Ort arbeiten und kurzfristig Lösungen für komplexe technische Probleme suchen. Unter Umständen gibt es weltweit noch andere Teams, die dieses Problem kennen und aus Erfahrung berichten können wie der Fehler bei ihnen behoben wurde. Hierfür beschreibt der Mitarbeiter sein Problem möglichst genau in TechnoWeb und Kollegen, die bereits ähnliche Probleme gelöst haben, geben ihm Tipps. Es ließen sich vier verschiedene Arten von Suchen nach Unterstützung entsprechend der Ziele der Suche unterscheiden:

So suchte z.B. ein griechischer Mitarbeiter, der im Bereich IT Solutions und Services tätig ist eine *technische Lösung*: „Dear colleagues, I am facing a problem in the customer's PROD database (ORACLE 10g). Cost based optimizer (CBO) creates an inefficient execution plan that has a result vey slow queries. CBO wrongly uses the bitmap INDEX TTME_DATETYPE. This problem is not reproduced in our test environment. Any help is highly appreciated." Aufgrund vier hilfreicher Hinweise von Kollegen weltweit konnte der Mitarbeiter das Problem beim Kunden noch innerhalb des gleichen Tages lösen. Ein anderes Szenario ist die Suche nach *Unterstützung bei der Vorbereitung einer Entscheidung*. Mitarbeiter fragen z.B. ihre Kollegen, welchen Hersteller sie für einen bestimmten Anwendungsfall einsetzen würden. („Has anybody of you already made a comparision between these suppliers for workplace design?")

Schnelle Hilfe wurde in der Stichprobe nicht nur bei technischen Problemen und bestehenden Projekten angefragt, sondern auch um an Ausschreibungen teilzunehmen und einen neuen Kunden von der eigenen Kompetenz zu überzeugen. Hierfür wurden *Referenzen auf einen Spezialgebiet* abgefragt, die dem Vertrieb nicht bekannt sind wie in folgenden Beispiel: „Dear all, do we have references/customer success stories for Fleet Management solutions provided/ developed by Siemens or SIS?" Auf diese Anfrage bekam der Mitarbeiter dreizehn Antworten aus unterschiedlichen Sektoren. („CT T MSO hat im Bereich Flottenmanagement ausgewiesene Kompetenzen und auch schon einiges gemacht.", „We developed application „Safari" which is used in Germany to administrate the whole lifecycle of company cars for Siemens Germany ...", „... is the IT lead for the recently started fleet Management Tool for Energy Fossil services..."). Die *richtigen Ansprechpartner* werden zum Einen gesucht, um als Experten Lösungen auf Probleme zu finden, die noch zu diffus sind, um das Problem selbst in TechnoWeb zu beschreiben oder Experten für einen längerfristigen Projekteinsatz.

6.4 Vermarkten

Entgegen unseren Vermutungen (und anders als von Dimicco et al. [14] beschrieben) wird das TechnoWeb nur in sehr begrenztem Umfang zur Vermarktung von zentralen Initiativen, Dienstleistungen und Technologien genutzt. Ein Anwendungsfall, in dem Informationen vermarktet werden, ist die Erhöhung des Bekanntheitsgrades von *zentralen Initiativen.* So wird z.B. im Netzwerk Sustainability für einen anstehenden Ideenwettbewerb Werbung gemacht: „It might be interesting for you following our planned idea contest on sustainability which is going to run in the intranet from April 19 on."

Zum Teil bieten diese Zentralabteilungen auch Dienstleistungen rund um ein Thema wie Datensicherheit an. Hier wird TechnoWeb zur *Vermarktung ihrer Dienstleistungen* genutzt um potenzielle Kunden kennenzulernen und durch die Diskussion mit den Anwendern Vertrauen zu schaffen. Wie z.B. im Netzwerk Secure Software Development, das von der Zentralabteilung CERT gegründet wurde und zum Zeitpunkt der Studie siemensweit 99 Mitglieder hatte. „ ... In this network which is mainly driven by the Siemens Computer Emergency Response Team (CERT), we want to answer those important security questions." Im Vergleich zur Vermarktung von zentralen Initiativen, die das Ziel haben ihren Bekanntheitsgrad zu erhöhen, steht beim Vermarkten von Dienstleistungen/Services die Anbahnung eines Geschäftes im Vordergrund. Nicht nur Dienstleistungen von Zentralabteilungen nutzen TechnoWeb zur Vermarktung sondern auch Abteilungen in Sektoren, die neue Technologien entwickelt haben und dafür Partnerschaften suchen, um diese Technologien auf den Markt zu bringen: „I am looking for an expert within Siemens, who is interested in a new insulating material."

6.5 Meta-Kommunikation

Zusätzlich zu den o.g. fünfzehn Genres konnte noch ein weiteres Genre identifiziert werden, das jedoch nicht bei den Anwendungsszenarien berücksichtigt und als Meta- Kommunikation bezeichnet wurde. Dieses Genre trat in 8% aller Posts auf, wobei sich Urgent Requests (4%) und Beiträge in den Netzwerken (12%) erheblich unterscheiden. Mit Meta-Kommunikation wurden Beiträge kategorisiert, die mit den eigentlichen Inhalten nichts zu tun haben, z.B. wenn ein Mitarbeiter einen Kollegen darauf hinweist, dass der gepostete Link nicht funktioniert. Ein anderes Beispiel sind Nachrichten, in denen Nutzer sich für die Hilfe der Kollegen bedanken. TechnoWeb verfügt über eine eigene Funktion, die sich „thank you" nennt und als Antwort auf einen Post gewählt werden kann. Diese Antwort wird dann besonders hervorgehoben. Aufgrund dieses technischen Unterschiedes wurde das Bedanken nicht oben als Anwendungsszenario genannt, sondern als Meta-Kommunikation gesehen.

6.6 Interpretation

Zusammenfassend lässt sich festhalten, dass die Kategorie „Teilen" fast die Hälfte aller analysierten Genres ausmacht. Diese Erkenntnis ließ sich nach einer quantitativen Auswertung damit begründen, dass eine Suchanfrage durchschnittlich acht Antworten erhält, die sich häufig der Kategorie „Teilen" zuordnen lassen. Der zweithäufigste Grund TechnoWeb zu nutzen ist das Diskutieren. Dabei ist zu berücksichtigen, dass eine Diskussion aus einer längeren Verkettung von Antworten bestehen kann, die jeweils den Anteil des Genres erhöht. Die kleinste Kategorie stellt das Vermarkten dar. Praktiken dieser Kategorie wurden in der Stichprobe nur von einer eingeschränkten Zielgruppe genutzt, die hauptsächlich in Zentralabteilungen angesiedelt sind oder als interner Dienstleister fungieren.

Die unterschiedliche Verteilung von Genres zwischen in den Netzwerken erstellten Nachrichten und den Urgent Requests ist mit der Funktion an sich verbunden. Rein technisch funktionieren beide gleich: Eine Nachricht / ein Urgent Request wird geschrieben und jeder der das entsprechende Netzwerk oder die entsprechende Urgent Request-Kategorie abonniert hat, wird darüber per Mail benachrichtigt. Bei der Einführung der Plattform wurde der Sinn der beiden Funktionen klar kommuniziert: Urgent Requests sollen nur bei sehr dringenden Anfragen gestellt werden oder als Eskalation, falls die Frage im entsprechenden Netzwerk nicht zur Zufriedenheit beantwortet wurde. Grund hierfür ist, dass die Hilfsbereitschaft von Kollegen nicht ausgenutzt werden sollte, in dem sie mit Anfragen überflutet werden. Netzwerk News dienen dagegen weniger dringlichen Fragen und der allgemeinen Diskussion. In den Ergebnissen spiegelt sich dies in der Tatsache wider, dass nur in 4 % der News nach Unterstützung gesucht wird, während es bei den Urgent Requests 15 % sind. Die Kategorie Diskutieren weist den gegenläufigen Effekt auf und kommt nur in 17 % der Urgent Requests vor, dafür aber in 40% der News.

Der Nutzen von TechnoWeb ist nicht immer aus der Kommunikation der Plattform selbst erkennbar. So kann der Kommunikation in TechnoWeb nicht immer entnommen werden, was der übergeordnete Nutzen der Problemlösung ist. Desweiteren können zwar Anfragen und deren Antworten TechnoWeb entnommen werden, falls aber keine Antwort enthalten ist, die darauf schließen lässt, dass das Problem erfolgreich gelöst wurde, z.B. ein Danke-Kommentar, kann der Nutzen von TechnoWeb nur durch eine zusätzliche Befragung der Betroffenen ermittelt werden. Gerade bei der Vermarktung ist der Nutzen von TechnoWeb für einen höheren Bekanntheitsgrad nicht direkt ersichtlich und kann nur über zusätzliche Messgrößen ermittelt werden (Bekanntheitsgrad vor und nach Nutzung von TechnoWeb). Auch Mitarbeiter, die sich für ein gemeinsames Thema interessieren, tauschen sich üblicherweise innerhalb eines Netzwerkes aus. Manchmal wird der initiale Kontakt über ein Netzwerk hergestellt und der eigentliche Austausch findet dann über andere Medien wie MS Livemeeting, Telefon oder E-Mail statt.

7. ERMITTLUNG DES NUTZENS

Aufbauend auf den Ergebnissen der Genreanalyse diskutieren wir im Folgenden den Nutzen des TechnoWeb 2.0 für das Unternehmen und demonstrieren zudem Potential einer Multi-Level-Verknüpfung mit der strukturellen Ebene über eine Netzwerkanalyse. Anschließend schlagen wir aufbauend auf Cooper et al. [10] und am Beispiel der Kategorie "Suche nach Unterstützung" ausgewählte Metriken vor, die auf den Anwendungsszenarien aufbauen und zur Ermittlung des Nutzens beitragen können.

7.1 Erkennbarer Nutzen

Im Folgenden wird der Nutzen der vier Kategorien von Anwendungsszenarien für das Unternehmen herausgestellt, die in Kapitel sechs identifiziert wurden. Durch das Konkretisieren des Potentials der Plattform in Form von Anwendungsszenarien und deren Veranschaulichung mit realen Beispielen wird der Nutzen des Social Networking Systems für die Unternehmensleitung klar erkennbar.

7.1.1 Teilen von Informationen

Der Hauptnutzen des Teilens von Informationen ist offensichtlich: Informationen können aufgrund der Aufteilung in zahlreiche

Netzwerke ausschließlich interessierten Mitarbeitern (kanalisiert) zur Verfügung gestellt werden. Andere Mitarbeiter bleiben von möglichen „cc-Massenmails“ verschont und sparen sich damit auch Zeit, die sie mit dem aussortieren irrelevanter Nachrichten verbringen würden. Ein weiterer Nutzen dieser Nutzungskategorie ist, dass Mitarbeiter sich mit Fragen auch direkt an bestimmte Mitglieder des Netzwerkes wenden können, die aufgrund ihrer geteilten Informationen als Experten angesehen werden können. Zusätzlich sind diese Informationen in TechnoWeb durch eine Suche für jedermann leicht auffindbar. So finden sich Kollegen aus verschiedenen Ländern und Sektoren, die sich mit ähnlichen Themen beschäftigen im Tagesgeschäft aber keine Berührungspunkte haben und sich vor TechnoWeb gar nicht gekannt haben. Zudem werden Informationen häufig auf ganz konkrete Suchanfragen geteilt. Somit ist die Motivation des Antwortenden viel höher sein Wissen als „just in time reuse“ zur Verfügung zu stellen als „lessons learned“ aufzuschreiben von denen er nicht weiß ob sie jemals benötigt werden. So wird Doppelarbeit vermieden, indem bestehende Best Practices genutzt werden anstatt neue zu erfinden.

7.1.2 Diskutieren

Im Unterschied zu bestehenden Diensten bietet TechnoWeb als Diskussionsplattform einen Mehrwert. So ist im Vergleich zu E-Mail die gesamte Kommunikation transparent. Jeder Mitarbeiter hat die Möglichkeit bestehende Diskussionen zu verfolgen, oder auch seine eigene Meinung zu äußern, auch wenn er gar nicht direkt angesprochen wurde. Zudem gehen die Funktionalitäten des Technoweb auch über die von typischen im World Wide Web genutzten Diskussionsplattformen hinaus, beispielsweise durch die Integration von RSS-Notifikation oder die Vernetzungsmöglichkeit verschiedener Netzwerke. Zusammenfassend lässt sich sagen, dass der Dienst es den Mitarbeitern leicht macht über geschäftsrelevante Themen zu diskutieren. Das Unternehmen profitiert von diesem Austausch, weil sich Best Practices schneller verbreiten und damit die Qualität der Leistungserbringung verbessert wird. Das Zusammenführen von im Unternehmen vorhandener Expertise fördert auch dass Innovationen zukünftig schneller auf den Markt gebracht werden können.

7.1.3 Suchen nach Unterstützung

Ein weiterer Beitrag von TechnoWeb besteht darin den Mitarbeiter, der ein Problem hat, mit einem erfahrenen Kollegen zusammenzubringen. Für diesen Fall wird vom Personalwesen ein Expertenfinder zur Verfügung gestellt, der mit der Expertise der einzelnen Mitarbeiter gepflegt ist. Manchmal wird allerdings Spezialwissen benötigt, das in keiner Datenbank mehr gepflegt wird wie z.B. Programmiersprachen, die heute nicht mehr eingesetzt werden oder sehr aktuelles Wissen zu einem neuen Trendthema, das noch nicht erfasst ist. Hierbei kann TechnoWeb helfen die richtige Person zu finden, da hier die entsprechenden Netzwerke mit ihren Mitgliedern dynamisch auffindbar sind. Somit kann das Unternehmen die Expertise des ganzen Konzerns ausschöpfen, um ein bestehendes Problem bei einem Kunden schneller zu lösen oder ein Projekt in kürzerer Zeit zu Ende zu bringen. Dadurch, dass auf Erfahrungswissen (in Form des richtigen Ansprechpartners) oder Informationen zurückgegriffen werden kann spart das Unternehmen Geld und erhöht gleichzeitig die Kundenzufriedenheit. Durch das schnelle Ansprechen von vielen Personen kann auch die Meinung einer breiten Basis in Entscheidungen einbezogen werden. Zudem wird die Akzeptanz der Entscheidungen erhöht, wenn Experten aktiv in die Diskussionen und Entscheidungen eingebunden werden.

7.1.4 Vermarkten

Die Vermarktung von zentralen Initiativen oder internen Dienstleistungen auf der Plattform trägt dazu bei, dass Initiativen schneller einen höheren Bekanntheitsgrad erreichen. So können viele interessierte Mitarbeiter gleichzeitig angesprochen werden und deswegen einem Netzwerk beitreten. Somit erfolgt die Kommunikation viel zielgerichteter. Anbieter von internen Dienstleistungen können sich in TechnoWeb schnell ein Bild davon machen was ihre „Kunden“ bewegt und darauf mit einem entsprechenden Leistungsangebot reagieren. Auch Technologien, die in einem Sektor entwickelt wurden können in anderen Sektoren von Nutzen sein und mit TechnoWeb bekanntgemacht werden, was die Nutzung von Synergien im Unternehmen erhöht.

7.2 Nutzen auf struktureller Ebene

Wie weiter oben theoretisch begründet, erzeugt die Nutzeraktivität auch auf der strukturellen Ebene des Mitarbeiternetzwerks Nutzenpotentiale. Als Indikator für die Vernetzung zwischen den Mitarbeitern lässt sich beispielsweise die Netzstruktur der Themengruppen als Koaffiliationsnetzwerk untersuchen. Die entsprechende Visualisierung der Interaktivität ist in Abbildung 1 exemplarisch dargestellt. Die Linkstärke ist hierbei die Anzahl der Mitarbeiter, welche in zwei Themengruppen mitarbeitet und diese somit näher zusammenbringt.

Im abgebildeten Beispiel aus TechnoWeb sind 58 Netzwerke dargestellt, die durch je mindestens 10 User verknüpft werden. In diesem Core-Netzwerk haben sich im Zeitablauf 74 solcher starken Themenbezüge entwickelt. Zwei Themen sind dabei sehr hierarchisch verbunden, mit zwei zentralen Domänen und vielen peripheren untereinander wenig-vernetzten Unterthemen. Die dynamische Verfolgung dieser Struktur zeigt auf, inwiefern die Themenverknüpfungen stärker werden (durch mehr Mitglieder verknüpft) und damit der Wissenstransfer wahrscheinlicher wird. Dicht vernetzte Themen legen weiterhin eine Basis für erhöhte Innovationsfähigkeit.

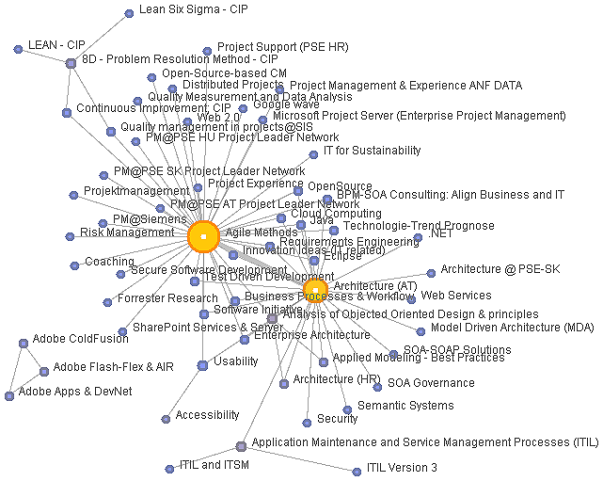


Abbildung 1: Vernetzung von Themengruppen. Knotengröße: Anzahl verknüpfter Themen, Bezeichnung: Thema.

Ein weiteres noch zu untersuchendes Netzwerk zur Messung struktureller Effekte ist das Ko-Mitgliedschaftsnetzwerk, in dem ein Link repräsentiert, dass die zwei Akteure gemeinsam an einem oder mehreren Themennetzwerken teilnehmen und somit einen thematischen Kontext teilen. Je stärker der Link desto breiter ist

der verbindende Kontext und desto wahrscheinlicher sind Vertrauen oder Wissenstransfer. Es soll also mit diesem Ansatz gezeigt werden, dass die Wahrscheinlichkeit, dass die Mitarbeiter Informationen miteinander teilen sich durch die stärkere Verbindung erhöht. Zudem zeigen dichte Cluster erhöhte semantische und strukturelle Ähnlichkeit zwischen Akteuren an. Es ist wahrscheinlicher, dass Akteure länderübergreifend Expertise besser identifizieren können, da sie über die Interessen anderer informiert sind.

7.3 Abgeleitete Metriken

In Anlehnung an Cooper et al. [10] können die identifizierten Anwendungsszenarien als Grundlage zur Definition von Metriken für den Regelbetrieb dienen. Denn eine Kausalität zwischen dem Nutzen der Plattform und der tatsächlichen Nutzung kann nur über konkrete Anwendungsbeispiele, wie sie hier untersucht wurden, angenommen werden.

Auch wenn es nicht das Ziel dieses Beitrags ist Metriken zu entwickeln, möchten wir am Beispiel der Kategorie "Suche nach Unterstützung“ ausgewählte Metriken vorschlagen, die zur Ermittlung des Nutzens beitragen können. Dabei greifen wir auf die von Cooper et al. vorgeschlagene Unterteilung zurück und unterscheiden darüber hinaus, gemäß unserer eingangs erläuterten Multi-Level Analyse Metriken des individuellen und strukturellen Nutzens:

1. Vitality: Wie erklärt dient die Messung der Benutzeraktivität dazu einen kurzfristigen Überblick über die Aneignung eines Dienstes zu erhalten. Um diese zu erheben bieten sich als Metriken des individuellen Nutzens beispielsweise die Messung der Anzahl der Urgent Requests oder News an.

2. Capability: Um den Wert der eingesetzten Dienste (innerhalb der o.g. Anwendungsszenario- Kategorie) für die Organisation zu beurteilen können, bieten sich die Anzahl der Antworten auf News und Urgent Requests oder auch die Anzahl der "Thank you"-Kommentare an. Auf struktureller Ebene können die strukturelle Ähnlichkeit im Ko-Mitgliedschaftsnetzwerk (gleiche Kontakte) oder die Anzahl der Kontakte (Degreezentralität) als Maß für den Aufbau von Austauschbeziehungen und für den zeitlichen Aufwand und das Potential der Ansprechpartnersuche dienen.

3. Business Value: Auf struktureller Ebene bieten sich Netzwerkeigenschaften wie die Netzwerkdichte, die Gleichmäßigkeit der Verteilung der Kontaktanzahl über die Akteure (statt starker Zentralisierung und Bottlenecks), die durchschnittliche Relationsstärke und die durchschnittliche Closeness-Zentralität (durchschnittliche Pfadlänge eines Akteurs zu allen anderen Kontakten) zur Messung der Informationsbeschaffungszeit und der Antwortzeit auf Technik- oder Kundenfragen an. Um den tatsächlichen Nutzen für die Organisation beurteilen zu können, genügt es dabei nicht auf Ebene der Kategorien Metriken festzulegen, sondern diese müssen für jedes Anwendungsszenario definiert werden. Zum Beispiel, dass die *Lösung eines technischen Problems* schneller gefunden wurde.

8. FAZIT UND AUSBLICK

Im vorliegenden Beitrag haben wir Anwendungsszenarien als Werkzeug zur Ermittlung des Nutzens von Corporate Social Software vorgeschlagen. Wie die Genreanalyse der vorliegenden Fallstudie zeigte, wird die untersuchte Plattform vor allem zum Teilen von Informationen und zum Diskutieren genutzt wird, aber auch für die Suche nach Unterstützung und zur Vermarktung. Entlang dieser vier Kategorien von Anwendungsszenarien aus Nutzersicht haben wir den Nutzen des Dienstes für das Unternehmen diskutiert. Wir haben weiterhin demonstriert, dass die Netzwerkanalyse eine sinnvolle Ergänzung darstellt, um interpersonelle strukturelle Nutzeneffekte zu erfassen und so einen Multi-Level-Ansatz zur Nutzenmessung zu entwickeln.

Für die vorliegenden Ergebnisse muss festgehalten werden, dass nur ein Ausschnitt der Kommunikation auf der Plattform untersucht wurde (10 von zuletzt 760 Netzwerken). Im weiteren Verlauf der Studie soll diese Stichprobe deswegen noch ausgeweitet werden. Zudem haben die Ergebnisse nur im Kontext der vorliegenden Fallstudie Gültigkeit. Aus diesem Grund ist es wünschenswert, noch weitere Fälle des Einsatzes von vergleichbaren Diensten in anderen Unternehmen mit ähnlicher Methodik zu untersuchen und zu vergleichen.

9. LITERATUR

[1] Ackerman, M. S., Pipek, V., und Wulf, V. 2003. *Beyond Knowledge Management: Sharing Expertise*. MIT Press, Cambridge.

[2] Back, A., Gronau, N., und Tochtermann, K. 2009. *Web 2.0 in der Unternehmenspraxis.* Oldenbourg, München.

[3] Bagozzi, R.P., Dholakia, U.M. und Mookerjee, A. 2006. Individual and Group Bases of Social Influence in Online Environments. *Media Psychology,* 8, 8, 95–126.

[4] Berkenkotter C, Huckin T.N. 1995. *Genre knowledge in disciplinary communication: cognition, culture, power*. Hillsdale, New Jersey.

[5] Bouman, W., de Bruin, B., Hoogenboom, T., Huizing, A., Jansen, R. und Schoondorp, M. 2007. The Realm of Sociality: Notes on the Design of Social Software, In *Proceedings International Conference on Information Systems*, Montreal.

[6] Boyd, D. und Ellison, N. 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 1, article 11.

[7] Boyd, D. und Heer, J. 2006. Profiles as Conversation: Networked Identity Performance on Friendster, In *Proceedings Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Kauai.

[8] Brzozowski, M. J. 2009. WaterCooler: exploring an organization through enterprise social media. In *Proceedings GROUP 2009,* Sanibel Island.

[9] Carr, N. 2003. IT Doesn’t Matter. *Harvard Business Review*, Mai 2003, 41–49.

[10] Cooper, C., Martin, M., und Kiernan, T. 2010. *Measuring the value of social software*. IBM Whitepaper, Cambridge.

[11] Davis, F.D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 319-340.

[12] DeLone, W.H., und McLean, E.R. 1992. Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research,* 3 1, 60-95.

[13] DiMicco, J. M., Geyer, W., Dugan, C., Brownholtz, B., und Millen, D.R. 2009. People Sensemaking and Relationship Building on an Enterprise Social Networking. In *Proceedings Hawaii International Conference on System Sciences*, Waikoloa, IEEE Computer Society Press.

[14] DiMicco, J. M., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., und Muller, M. 2008. Motivations for Social Networking at Work. In *Proceedings Conference on Computer Supported Cooperative Work*, San Diego, ACM Press.

[15] Günther, O., Krasnova, H., Riehle, D., und Schönberg, V. 2009. Modelling Micro-Blogging Adoption in the Enterprise. In *Proceedings Americas Conference on Information Systems*, San Francisco.

[16] Gable G. G., Sedera, D., und Chan, T. 2008. Re-conceptualizing Information System Success: The IS-Impact Measurement Model. *Journal of the Association for Information Systems*, 9, 7, 377-408.

[17] Howlett, D. 2007. *ROI is so Business 1.0: not.* http://blogs.zdnet.com/Howlett/?p=183, 1 Oct 2007

[18] Kettles, D., und Smith David, J. 2008. The Business Value of Social Network Technologies: A Framework for Identifying Opportunities for Business Value and an Emerging Research Program. In *Proceedings Americas Conference on Information Systems*, Toronto.

[19] Koch, M., and Richter, A. 2009. *Enterprise 2.0 - Planung Einführung und erfolgreicher Einsatz von Social Software im Unternehmen*. Oldenbourg, München.

[20] Koch, M., Richter, A., und Schlosser, A. 2007. Produkte zum IT-gestützten Social Networking in Unternehmen. *Wirtschaftsinformatik*, 49, 6, 448-455.

[21] Kwasnik, B. H., und Crowston, K. 2005. Introduction to special issue: Genres of digital documents. *Information Technology & People*. 18, 2, 76-88.

[22] Lamnek, S. 1993. *Qualitative Sozialforschung. Band I: Methodologie*. 2. Aufl., Deutscher Studien Verlag, Weinheim.

[23] Lampe, C., Ellison, N., und Steinfeld, C. 2006. A face(book) in the crowd: social searching vs. social browsing. In *Proceedings Conference on Computer Supported Cooperative Work*, Banff, ACM Press.

[24] Muller, J.M., Freyne, J., Dugan, C., Millen, D.R., Thom-Santelli, J. 2009. Return on contribution (ROC): A Metric for Enterprise Social Software. In *Proceedings European Conference on Computer Supported Cooperative Work*, Wien.

[25] Melville, N. 2004. Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly* 28, 2.

[26] Orlikowski, W. J. 2000. Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11 4. 404-428.

[27] Poole, M.S., und DeSanctis, G. 1992. Microlevel structuration in computer-supported group decision making. *Human Communication Research*, 19, 1, 5-49.

[28] Räth, P., Smolnik, S., Urbach, N., und Zimmer, C. 2009. Towards Assessing the Success of Social Software in Corporate Environments. In *Proceedings Americas Conference on Information Systems*. San Francisco.

[29] Richter, A., und Koch M. 2008. Functions of Social Networking Services. In *Proceedings International Conference on the Design of Cooperative Systems*, Carry-le-Rouet.

[30] Richter, A., und Riemer, K. 2009. Corporate Social Networking Sites – Modes of Use and Appropriation through Co-Evolution. In *Proceedings Australasian Conference on Information Systems*.

[31] Riemer, K., Frößler, F., und Klein, S. 2007. Real Time Communication – Modes of Use in Distributed Teams. In *Proceedings European Conference on Information Systems*. St. Gallen.

[32] Shang, S., und Seddon, P. B. 2002. Benefits of Enterprise Systems. *Information Systems Journal*, 12, S. 271–299.

[33] Skeels, M. M. und Grudin, J. 2009. When Social Networks Cross Boundaries: A case study of workplace use of Facebook and LinkedIn. In *Proceedings GROUP 2009*, Sanibel Island.

[34] Sledgianowski, D., und Kulviwat, S. 2008. Social Network Sites: Antecedents of Adoption and Usage. In *Proceedings Americas Conference on Information Systems*, Toronto.

[35] Todd, P. und Benbasat, I. 2000. Evaluating the Impact of DSS, Cognitive Effort, and Incentives on Strategy Selection', *Information Systems Research*, 10, 4, 356-374.

[36] Tufekci, Z. 2008. Grooming, Gossip, Facebook and Myspace: What can we learn about these sites from those who won't assimilate? *Information, Communication, and Society*, 11 4, 544-564.

[37] Urbach., N., Smolnik, S., und Riempp, G. 2009. Der Stand der Forschung zur Erfolgsmessung von Informationssystemen – Eine Analyse vorhandener mehrdimensionaler Ansätze. *Wirtschaftsinformatik* 51, 4, 363–375

[38] Utz, S. 2010. Show me your friends and I will tell you what type of person you are: how own profile, number of friends, and type of friends influence impression formation on social network sites. *Journal of Computer-Mediated Communication*, 15, 1, 314-335.

[39] Wilson, T.P. 1982. Qualitative oder quantitative Methoden in der Sozialforschung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. 487-508.

[40] Wu, A., DiMicco, J.M., und Millen, D. R. 2010. Detecting Professional versus Personal Closeness Using an Enterprise Social Network Site. In *Proceedings Conference on Human factors in computing systems*, Atlanta, ACM Press.

[41] Wade, M. R., and Hulland, J. 2004. Review: The Resource-Based View And Information Systems Research: Review, Extension, And Suggestions For Future Research. *MIS Quarterly*, 28, 1, 107-142.

[42] Yates, Y., Orlikowski, W. J. und Okamura, K. 1999. Explicit and Implicit Structuring of Genres: Electronic Communication in a Japanese R&D Organization. *Organization Science*. 10, 1, 83-103

Exploration & Promotion: Einführungsstrategien von Corporate Social Software

Alexander Richter
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
D-85577 Neubiberg
+49 089 6004 4406
a.richter@unibw.de

Alexander Stocker
Institut DIGITAL, JOANNEUM RESEARCH
Steyrergasse 17-19
A-8010 Graz
+43 316 876-2611
alexander.stocker@joanneum.at

ZUSAMMENFASSUNG

Im Kontext des Einsatzes von Social Software zur Unterstützung der innerbetrieblichen Zusammenarbeit wird regelmäßig über die „richtige" Einführungsstrategie diskutiert und es werden die beiden Paradigmen Top-Down (vom Management getrieben) und Bottom-Up (von den Mitarbeitern getrieben) gegenüber gestellt. Der vorliegende Beitrag zeigt anhand einer vergleichenden Analyse von 21 Fallstudien, dass die Diskussion um o.g. Dichotomie nicht zielführend ist. Die Ergebnisse der Untersuchung lassen vielmehr erkennen, dass die betrachteten Unternehmen zwei durchaus miteinander vereinbare Strategien anwandten: Die Art der Nutzung blieb im Rahmen eines partizipativen Vorgehens zunächst den Nutzern überlassen und die Anwendungsszenarien wurden nach und nach identifiziert („Exploration") oder/und die Plattformen wurden im Unternehmen mit Unterstützung des Managements koordiniert vermarktet und deren gezielte Nutzung geschult („Promotion").

Keywords

CSCW, Social Software, Collaboration, Enterprise 2.0, Exploring, Promoting, Wikis, Weblogs, Microblogging, Social Networking.

1. MOTIVATION

Aufgrund des großen Erfolgs von „Web 2.0"-Plattformen wie Wikipedia oder Twitter und der darauffolgenden umfangreichen Berichterstattung aus der Presse, haben viele Organisationen in den letzten Jahren damit begonnen, vergleichbare Dienste auch prototypisch „hinter der Firewall" zu nutzen. Die Einführung dieser Dienste zur Unterstützung der Zusammenarbeit in Unternehmen, stellt die Beteiligten dabei vor spezifische Herausforderungen, wie z.B. die Einbeziehung von Organisationsstrukturen und -prozessen. Diese gehen über die Anforderungen an Web-Plattformen hinaus, welche primär durch informelle Strukturen gekennzeichnet sind [23] und müssen bei der sozio-technischen Gestaltung der Dienste berücksichtigt werden. Das erzeugt in Unternehmen ein Spannungsfeld zwischen der mit dem Web 2.0 implizit verbundenen Selbstorganisation mit der durch Organisationsstrukturen oktroyierten Fremdorganisation der Mitarbeiter.

Während das Potential der Dienste wie Weblogs, Wikis oder Social Networking Services (SNS) im Unternehmenskontext zunächst überwiegend prototypisch untersucht und durch die Wissenschaft in explorativen Studien eine Vielzahl einzelner Gestaltungsparameter aufgezeigt wurde (z.B. [7], [14], [21]), hat es inzwischen den Anschein, dass diese neuen Dienste zunehmend den Weg in die Arbeitsprozesse finden und ernster genommen werden [37]. Vor diesem Hintergrund wird in Wissenschaft und Praxis vermehrt diskutiert, welche Auswirkungen der Einführungsprozess auf den tatsächlichen Erfolg der Dienste für Wissensarbeiter und Organisation hat. Insbesondere wurde zuletzt mehrfach die Frage aufgebracht, ob es gegen die Natur der Dienste wäre diese „Top-Down" (durch das Management gesteuert und koordiniert) einzuführen (z.B. [4]). Diese und vergleichbare Fragen sind -aufgrund der Tatsache, dass der Einsatz von Corporate Social Software in mehrfacher Hinsicht einen Paradigmenwandel mit sich bringt - mit den Erkenntnissen der Erforschung (traditioneller) Groupware nicht vollständig zu erklären [26]. Aus diesem Grund ist es notwendig über den umfangreichen Erfahrungsschatz zur Einführung von Groupware in Unternehmen hinaus Ansätze zu finden, um mehr über diese neuartigen Phänomene zu lernen.

Hier setzt die vorliegende vergleichende Analyse von 21 von den Autoren dieses Beitrags in den vergangenen vier Jahren erhobenen Fallstudien an. In der Untersuchung wurden unter anderem das von den Unternehmen bei der Einführung der neuen Dienste gewählte Vorgehen (Potenzial, Art der Nutzung, Treiber) und dessen Auswirkungen auf die Zusammenarbeit analysiert. Dabei stellte sich heraus, dass in den Unternehmen die Frage nach Top-Down und Bottom-Up schnell in den Hintergrund trat, oder sich oftmals gar nicht erst stellte. Vielmehr kristallisierten sich zwei (durchaus miteinander vereinbare) Vorgehensweisen zur Einführung als praktikabel bzw. praktiziert heraus: Die kontinuierliche Ermittlung möglicher Anwendungsszenarien für derart nutzungsoffene Dienste im Rahmen eines partizipativen Vorgehens (im Folgenden „Exploration" genannt) und die mit Unterstützung des Managements koordiniert vermarktete und gezielt geschulte Nutzung mit Fokus auf bestimmte Nutzungspotentiale („Promotion").

Es folgt ein Überblick über bestehende Arbeiten, die sich mit der Einführung von Corporate Social Software beschäftigen (Kap. 2). Dann wird das methodische Vorgehen erläutert (Kap. 3) und als (Teil-)Ergebnis des Forschungsprojektes werden die gerade bezeichneten Vorgehensweisen, Exploration und Promotion, anhand von besonders aussagekräftigen Beispielen aus den erhobenen Fallstudien veranschaulicht (Kap. 4). Im darauffolgenden Kapitel (Kap. 5) werden diese Beispiele diskutiert sowie Implikationen für Wissenschaft und Praxis aufgezeigt. Zusammenfassung und Ausblick schließen diesen Beitrag (Kap. 6).

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

2. ARBEITEN ZUR EINFÜHRUNG VON CORPORATE SOCIAL SOFTWARE

Bereits seit mehreren Jahren finden verschiedene IT-Werkzeuge, welche sich im Web 2.0 durch eine stetig ansteigende Nutzung privater Art auszeichnen, auch in der Unternehmenspraxis zunehmend Anwendung ([1], [5], [30]). Obwohl das Potential der Werkzeuge noch nicht vollständig erforscht ist [37], wird diesen bereits vielfach bescheinigt, Informationsmanagement, Identitäts- und Netzwerkmanagement sowie Interaktion und Kommunikation in einem Unternehmen verbessern zu können [26]. Der Einsatz von Corporate Social Software bringt einen Wandel auf mehreren Ebenen mit sich. Dieser Wandel ist nicht nur in den Werkzeugen selbst zu sehen, sondern in den Paradigmen, die hinter dem Web 2.0 stehen und die durch die Werkzeuge ermöglicht bzw. umgesetzt werden: Beispielsweise werden die Benutzer einer Plattform als Produzenten von Informationen und Inhalten betrachtet, die die Inhalte freiwillig und selbstorganisiert erstellen („Participation"), (z.B. [30]). Damit zusammenhängend ist Social Software durch eine wesentlich stärkere Orientierung an den Bedürfnissen der Nutzer charakterisiert (oftmals als „me-centricity" bezeichnet; siehe z.B. [26]).

Inzwischen gibt es einen umfangreichen Bestand an Forschungsarbeiten darüber, wie Social Software im Allgemeinen in einer Unternehmensumgebung eingesetzt werden kann. In mehreren Studien wurde der Einsatz der einzelnen Dienste wie z.B. Weblogs (z.B. [10], [22]), Microblogs (z.B. [31], [42]), Wikis (z.B. [6], [52]) und Social Networking Services (SNS; z.B. [7], [39]) innerhalb der Grenzen eines Unternehmensintranets untersucht. In überwiegend explorativen Studien wurde dabei bereits eine Vielzahl einzelner Aspekte, wie etwa die Frage nach der Art und Anzahl von Beiträgen [50], dem Verhältnis zwischen lesender und schreibender Nutzung [3], der Qualität der nutzergenerierten Beiträge [19], der Motivation der Nutzer [7], dem Nutzen für Individuum und Organisation [50] oder den bei der Nutzung empfundenen Hindernissen oder Regeln untersucht [26].

Nicht vollständig geklärt ist jedoch die Frage, welche Einführungsstrategien von Unternehmen verfolgt werden, die Social Software für sich nutzen möchten. Diesbezüglich wurde in bisherigen Arbeiten beispielsweise untersucht, welche Rolle die Unternehmenskultur [3] oder die Existenz verschiedener Nutzergruppen spielen (können) [10]. Weiterhin aufgebracht wurde die Diskussion, wie viel Freiwilligkeit möglich bzw. wie viel Kontrolle nötig ist (z.B. [4]). Im Spannungsfeld Freiwilligkeit/Kontrolle wird auch regelmäßig darüber diskutiert, ob Corporate Social Software Top-Down oder Bottom-Up eingeführt werden sollte. In dieser Diskussion wurde bereits argumentiert, dass sich nicht die Frage stellt, ob die Einführung vom Management angeordnet oder von den Mitarbeitern gewünscht und vorangetrieben wird, sondern wie das Management eine Bottom-Up-Einführung zielführend unterstützen kann [26].

Die in diesem Zusammenhang betonte Notwendigkeit des Commitments von Seiten des Managements (das heißt, dass die Dienste durch die Unternehmensführung konsistente, überzeugende und verlässliche Unterstützung erhalten) wurde bereits für ähnliche Systemgattungen aus dem Bereich Groupware untersucht und für notwendig befunden (vgl. z.B. [16], [24]). Dabei sollte das Management aufgrund seiner Vorbildfunktion deutlich machen, dass es vom Sinn der Plattform überzeugt ist und diese selbst nutzt (z.B. [8]). Zu vergleichbaren Ergebnissen kommen auch bisherige Forschungsarbeiten zum IT-gestützten Wissensmanagement. In dieser Domäne wurden bereits zahlreiche Rahmenwerke zur Einführung der Dienste vorgestellt und Einführungsstrategien diskutiert (vgl. z.B. [17]). Gerade beim IT-gestützten Wissensmanagement sind es häufig Top-Down-Einführungsstrategien, welche sich durch die Manifestation von Management-Rollen wie beispielsweise dem Chief Knowledge Officer (CKO) darstellen. Klassische Wissensmanagement-Ansätze vernachlässigen dabei sehr oft die Perspektive der Wissensarbeiter [17] und dementsprechend auch die Faktoren, welche Wissensarbeiter zur Teilung von Wissen motivieren. Gerade diese „Grenze" will Wissensmanagement durch die Konvergenz mit Web 2.0 und Social Software überwinden [33].

Ein wesentlicher Unterschied zwischen (traditioneller) Groupware und Social Software wird darin gesehen, dass Social Software oftmals „Bottom-Up" eingeführt wird (z.B. [4]). Sie wird also, als logische Folge der o.g. „me-centricity", von den Mitarbeitern gewünscht und vorangetrieben, oftmals ohne, dass das Management bereits vom Dienst überzeugt wurde, aber gleichzeitig auch ohne umfangreiche Vorschriften und Genehmigungsprozesse (z.B. [26]). Dies hat den Vorteil, dass sich die Software nicht so sehr nach bestehenden Prozessen und sozialen Strukturen richten muss (wie eine umfangreich „von oben" oktroyierte Plattform), sondern auch neue Wege gehen kann.

Diesem Bottom-Up-Ansatz werden in der Diskussion um Corporate Social Software häufig verschiedene Einwände entgegengebracht. Beispielsweise wird argumentiert, dass eine derart eingeführte Software sich nur schwer an den Zielen des Unternehmens orientieren könne. Zudem ergeben sich Fragestellungen nach dem Zusammenspiel verschiedener Einflussfaktoren wie z.B. der Unternehmenskultur, die bei der Einführung betrachtet werden müssen (vgl. z.B. [37]). Zudem werde Bottom-Up häufig als Ausrede benutzt, um entweder kein Budget in die Einführung von Social Software zu investieren, oder nach der Umsetzung die Nutzer für Fehler verantwortlich zu machen, da sie die Einführung nicht ausreichend unterstützt haben, obwohl ihnen genügend Freiraum dafür gegeben wurde [28]. Deswegen, so wird gerne interpretiert, müsse die Einführung einer langfristig orientierten Kollaborationsplattform, wie auch bei anderen Diensten, Top-Down erfolgen.

Aus dieser öffentlichen und interdisziplinären Diskussion heraus, wurden in der vorliegenden Studie das in Unternehmen gewählte Vorgehen und dessen Auswirkungen systematisch untersucht. Ziel war es festzustellen, welchen Einfluss die o.g. Paradigmen des Web 2.0 einerseits und bestehende Strukturen in den Unternehmen andererseits auf die Einführungs-Vorgehensweisen hatten. Bevor die Forschungsergebnisse der vergleichenden Fallstudienanalyse präsentiert und daraus Erkenntnisse für Theorie und Praxis abgeleitet werden, folgt im nächsten Abschnitt eine Erläuterung des methodischen Vorgehens.

3. METHODIK

3.1 Vergleichende Fallstudienanalyse

Wie bereits zuvor ausgeführt, wurde der Einsatz von Corporate Social Software in der Vergangenheit bereits anhand einer Anzahl verschiedener, explorativer Fallstudien untersucht. Der Einsatz der Fallstudienforschung als Hilfsmittel zur Erklärung des Phänomens Corporate Social Software ist darin begründet, dass sich

aufgrund der Vielzahl und Komplexität verschiedener Geschäftsprozesse und Arbeitspraktiken Informationsinfrastrukturen zunehmend schwieriger modellieren lassen [32]. Aus diesem Grund ist eine Sammlung von Regeln (z.B. in Form von praxistauglichen Leitfäden) allein nicht ausreichend. Vielmehr sollte ein Anwendungssystem-Verantwortlicher sich Anregungen von Anwendungserfahrungen anderer Unternehmen einholen können. Fallstudien stellen in diesem Zusammenhang eine gute Möglichkeit dar, Gestaltungs- und Aneignungsformen von Informationstechnologien in ihrem Kontext zu beschreiben [53].

Für die vorliegende Analyse wurde von den Autoren auf insgesamt 21 Fallstudien von Unternehmen in Deutschland, Österreich und der Schweiz zurückgegriffen. Die zugrundeliegenden Daten wurden im Rahmen zweier Promotionsprojekte und fünf weiterer Forschungsprojekte zwischen April 2007 und Juli 2010 selbst erhoben, es handelt sich also um eine Studie auf Basis von Primärdaten. Schon während der Promotionsprojekte haben sich die beiden Autoren über die eigenen Vorarbeiten hinaus intensiv ausgetauscht, um später eine gemeinsame Vergleichsbasis zu schaffen.

Pro Fallstudie standen jeweils sehr unterschiedliche Datenquellen zur Verfügung, die in die Analyse eingingen:

- Für jede Fallstudie, d.h. in jedem Unternehmen wurde mindestens ein mehrstündiges Interview mit Mitarbeitern geführt, die für Einführungsprozess und/oder Nutzerbegleitung verantwortlich waren. In zehn Fällen waren es zwei oder mehr Interviews.
- Zusätzlich wurden über alle Unternehmen verteilt zusätzlich über 30 weitere Interviews und mehrere Workshops mit Nutzern geführt. In sechs Fällen wurden quantitative Online-Befragungen mit Mitarbeitern als Nutzer der neuen Dienste geführt. Insgesamt wurden so rund 200 Nutzer befragt.
- Bis auf fünf Ausnahmen konnten die Autoren sich zudem vor Ort einen Überblick über die neuen Dienste und deren Nutzung verschaffen. In vier Fällen standen auch Nutzungsdaten zur Verfügung, die qualitativ ausgewertet (codiert) wurden.
- Darüber hinaus wurden meistens interne und in jedem Fall zur Verfügung stehende externe Dokumente ausgewertet. In zwei Fällen wurde zudem auf zusätzliche Informationen, die entlang eines vergleichbaren umfassenden Rasters durch das Enterprise-2.0-Fallstudiennetzwerk (e20cases.org) dokumentiert wurden, zurückgegriffen.

Zusammenfassend lässt sich sagen, dass die Triangulation mehrerer Datenquellen und mehrerer Verfahren zur Datenerhebung dabei wesentlich zur Sicherstellung einer größtmöglichen Objektivität und Konstruktvalidität (vgl. z.B. [9], [51], [54]) der aus den Fallstudien gewonnenen Erkenntnisse beiträgt.

Eine wesentliche Herausforderung bei der vergleichenden Analyse mehrerer Fallstudien stellen die oftmals unterschiedlichen, durch die jeweiligen Autoren gewählten Zielsetzungen der verschiedenen Untersuchungen dar [54]. Im vorliegenden Fall war es von Anfang an das Ziel der Autoren, alle Fallstudien im Rahmen der Datenerhebung möglichst strukturiert zu beschreiben und es so zu ermöglichen, Fallstudien in einzelnen Punkten strukturiert miteinander vergleichen zu können. Dabei wurde auf gängige Strukturierungshilfen für Fallstudien zurückgegriffen ([45], [46], [47]). Die strukturierte Dokumentation der Fallstudien orientierte sich dabei an den folgenden Punkten: (1) Unternehmen, (2) Problemstellung, (3) „alte“ Vorgehensweise, (4) Einführung der neuen Lösung, (5) Änderungen, (6) Zielerreichung und „Lessons Learned“.

Für jeden analysierten Fall wurde ein detaillierter Fallstudienreport auf Basis des Rasters erstellt. Auf diese Weise ließen sich eine Vielzahl von Parametern (z.B. Ausgangssituation, Betrieb, Nutzung, etc.) detailliert vergleichen. Vor allem die zahlreichen Bemerkungen, Aussagen und Feststellungen von Verantwortlichen und Nutzern im Zuge der Befragungen zum Einführungsprozess und zu den Zielen waren für die in diesem Beitrag vorgestellte Studie äußerst wertvoll und trugen zur internen Validität (vgl. [54]) bei.

So waren in den Interviews beispielsweise folgende Fragen an Entscheider und Nutzer herangetragen worden:

- Was war der Leidensdruck für die Einführung eines neuen Dienstes? Wer hat diesen Leidensdruck erstmalig festgestellt? (Ausgangssituation)
- Was war das Ziel des einzuführenden Dienstes? Welches Potenzial wurde wahrgenommen? (Potenzial)
- Welche Prozesse bzw. Aktivitäten sollte dieser neue Dienst unterstützen? (Nutzung)
- Nach welchen Kriterien wurde der neue Dienst ausgewählt? Wie wurden Nutzeranforderungen erhoben? (Auswahl)
- Wer sollten die Nutzer des neuen Dienstes sein: Projektteams, Abteilungen oder das gesamte Unternehmen? (Umfang)
- Wer hat den neuen Dienst ausgewählt und festgelegt, in welchem Kontext dieser verwendet wird? (Treiber)
- Welche Aktivitäten wurden hinsichtlich Projektmarketing, Bekanntmachung und Akzeptanzsteigerung gewählt und was war der Inhalt? (Akzeptanz)
- Wurde eine einführende Schulung bezüglich Ziele und Nutzung des neuen Dienstes durchgeführt? (Training)

Auch die zur Verfügung stehenden Dokumente wurden jeweils nach diesen Inhalten und Fragestellungen (Ausgangssituation, Potenzial, Nutzung, Auswahl, Umfang, Treiber, Akzeptanz, Training, …) durchsucht, sodass sich ein zufriedenstellendes Gesamtbild über die tatsächliche Art der Einführung ergab.

Aufbauend auf den zur Verfügung stehenden und die Forschungsfrage adressierenden Daten wurden alle Fälle von den Autoren unabhängig voneinander analysiert und auf gemeinsame Muster untersucht. Dabei wurden die konkreten Situationen der realen Welt in den Fallstudien als Basis zur induktiven Theoriebildung benutzt, wobei jeder Fall eine eigenständige analytische Einheit bildete [12]. Bei der Theoriebildung folgten wir der in der Grounded Theory definierten Methode des konstanten Vergleichs [13]. Gemäß dieser Methode wurden alle Fälle systematisch miteinander verglichen, um ähnliche Rahmenbedingungen und Vorgehensweisen zu identifizieren. Die Grundidee besteht darin, durch die Feststellung von Gemeinsamkeiten und Unterschieden zwischen verschiedenen Phänomenen im Datenmaterial ein konzeptuelles Unterscheidungsvermögen zu gewinnen. Hinter der Analy-

se von Gleich- und Andersartigkeit der untersuchten Vorkommnisse soll ein Sinn entdeckt und abstrahierend mit Bezug auf die Forschungsfrage in ein Konzept gegossen werden. Als Ergebnis dieser durchgeführten Analyse stellten sich schließlich zwei unterschiedliche Vorgehensweisen für die Einführung von Corporate Social Software, sowie deren Kombination heraus, auf die in Abschnitt 4 näher eingegangen wird, Exploration und Promotion.

3.2 Qualität der Stichprobe

Die als Fallstudienteilnehmer gewählten Unternehmen zeichnen sich durch eine Reihe von Gemeinsamkeiten aus. In allen Unternehmen existieren namentlich bekannte Verantwortliche für die Durchführung der Einführungsprojekte. Alle Unternehmen sind im deutschsprachigen Raum angesiedelt und weisen somit kulturelle Ähnlichkeiten auf. In allen Fällen konnten auch trotz der Novität und Komplexität von Social Software bereits hinreichend Erfahrungen mit den eingesetzten Werkzeugen Wiki, Weblog, Microblog und Social Networking Services gesammelt werden.

Die 21 erhobenen Fallstudien werden im Anhang in einer Tabelle skizziert, wobei jeweils die Aspekte „konkrete Umsetzung der Technologie“, „Ursprung/Konzept“ und „Art der Einführung“ herausgestellt werden. Auf eine umfangreichere Darstellung aller 21 Fälle muss aufgrund des beschränkten Platzes in diesem Beitrag leider verzichtet werden.

Der größte Teil der Fallstudien (18) wurde bereits durch mindestens einen Autor des vorliegenden Beitrags auf wissenschaftlichen Konferenzen vorgestellt oder in Fachzeitschriften bzw. Fachbüchern publiziert (vgl. Tabelle 1). Aufgrund dieses bereits stattgefundenen Peer-Review-Prozesses, wurde, die wissenschaftliche Qualität der in diesem Beitrag analysierten Fallstudien von externen Experten bereits positiv bewertet, was sich positiv auf die externe Validität (vgl. [54]) der mit Hilfe des verwendeten Datenmaterials gewonnenen Erkenntnisse auswirkt. So wird trotz unterschiedlicher Zielsetzungen der einzelnen Publikationen aufgrund der umfangreichen Dokumentation auch die Nachvollziehbarkeit der Fälle für den Leser sichergestellt (vgl. [47]).

Tabelle 1: Referenzen zu den Fallstudien (18)

Untersuchte(s) Unternehmen	Referenz
Accenture, IBM, SAP	[39]
Bosch	[52]
Communardo	[42]
Deutscher Skiverband	[40]
Infineon Technologies, IVM, Raiffeisen	[49]
Fraport, Maschinenfabrik Andritz, Microsoft, Pentos Spirit Link, Synaxon, TAO	[50]
Siemens Building Technologies	[31]
Siemens	[38]

4. ERGEBNISSE

Auf Basis der bereits eingeführten Diskussion zum Spannungsfeld zwischen Bottom-Up und Top-Down wurde in der vorliegenden Studie der Frage nachgegangen, auf welche Weise die Dienste in den betrachteten Unternehmen eingeführt wurden und welche Implikationen sich daraus ergeben. Dabei wurden in den betrachteten Fallstudien zwei durchaus vereinbare Vorgehensweisen identifiziert: Exploration und Promotion.

In diesem Abschnitt werden die beiden Vorgehensweisen zunächst kurz allgemein erläutert und dann anhand mehrerer besonders aussagekräftiger Beispiele veranschaulicht. Dabei werden sowohl Fälle, in denen nur Exploration oder Promotion verfolgt wurden, als auch Fälle, in denen beide Ansätze zum Tragen kamen, vorgestellt.

Für die Gesamtheit aller **21 Fälle** ergibt sich folgendes Bild: Während in **fünf Fällen „Exploration“** als dominierende Einführungsstrategie durch die Autoren identifiziert werden konnte, findet sich ebenfalls in **fünf Fällen „Promotion“** vor. In **elf Fällen** wurden **beide Einführungsstrategien kombiniert** angewandt.

Nachfolgend wird die Einführung von Corporate Social Software exemplarisch anhand von jeweils drei fokussierten Fallbeispielen pro Einführungsstrategie, also insgesamt neun Fällen (Exploration, Promotion, Kombination beider Strategien) kurz erläutert. Die nachfolgenden Fallbeispiele sollten eine gewählte Einführungsstrategie möglichst klar und deutlich widerspiegeln und wurden daher genau aus diesem Grunde durch die Autoren zur Präsentation herangezogen. Für die Definition der beiden Vorgehensweisen Exploration und Promotion werden die Aspekte **Potenzial des Dienstes**, **Art der Nutzung des Dienstes** und **Treiber des Dienstes** analysiert

4.1 Exploration

In rund einem Viertel der untersuchten Fallstudien wurde Exploration als dominierende Einführungsstrategie verwendet.

Unter Exploration wird in diesem Beitrag die kontinuierliche Ermittlung möglicher Anwendungsszenarien für neue nutzungsoffene Dienste mittels eines partizipativen Vorgehens verstanden.

Die Vorgehensweise der Exploration war in den Fallstudien auf unterschiedliche Art und Weise ausgeprägt. Im Laufe der Untersuchung kristallisierten sich folgende Merkmale als charakteristisch für dieses Vorgehen heraus:

- Das *Potential* des neuen Dienstes war nicht (oder zumindest nicht vollständig) klar und sollte während der Nutzung erstmals oder noch weitergehend erschlossen werden.
- Die *Art der Nutzung* des neuen Dienstes wurde nicht, oder nur zu einem kleinen Teil, vorgegeben.
- Die *Treiber* des Dienstes waren von dessen Nutzen überzeugt, aber es fehlte ein klarer „Business-Case“ mit definierter Zielsetzung und festgelegten Einsatzszenarien.

Als erstes Fallbeispiel ist die bereits im Jahr 1998 in einem Software-Entwicklungs-Bereich der *Siemens AG* erfolgte Einführung der Plattform TechnoWeb zu nennen. Die Nutzung wurde in den Jahren 2007 und 2008 zunächst auf den gesamten Bereich Siemens IT Solutions und Services und damit auf ca. 20.000 potentielle Nutzer weltweit ausgedehnt. Im Juni 2009 wurde vom Zentralvorstand der Siemens AG entschieden, die Plattform auf eine neue Technologie, Liferay, zu stellen und um mehrere Funktionalitäten, vergleichbar einem Social Networking Service, zu erweitern. Ziel war es, das TechnoWeb in dieser neuen Gestalt konzernweit auszurollen, um das Potential des Dienstes für das ganze Unternehmen zu überprüfen. Das „neue“ TechnoWeb sollte eine

heterogene Gruppe von ca. 100.000 potenziellen Nutzern dabei unterstützen, sich gegenseitig zu vernetzen und auszutauschen. In der Pilotphase sollte herausgefunden werden, *„welchen Beitrag das TechnoWeb zum Geschäftserfolg einzelner Siemens-Bereiche und für den Konzern als Ganzes leisten kann"*. Aus diesem Grund sollten mögliche Szenarien identifiziert werden, in denen die Anwendung des Dienstes nutzbringend ist und die Art und Weise der Nutzung wurde nicht vorgegeben. Nutzer hatten die Möglichkeit, mit dem neuen Dienst zu experimentieren, um erfolgreiche Szenarien zu explorieren. Um in einem weiteren Schritt Management und Mitarbeiter vom Nutzen der Plattform zu überzeugen sowie eine kritische Masse an Anwendern und Themen zu erreichen, wurde das Ziel verfolgt *„bereits identifizierte Szenarien als Erfolgsgeschichten zu kommunizieren, um mehr Transparenz zum Potential der Dienste zu schaffen"*. Gleichzeitig wurde aber darauf hingewiesen, dass es jedem Nutzer weiterhin ungenommen bleibt, das Potential von TechnoWeb für sich zu entdecken.

Ein weiteres Beispiel stellt die Nutzung des Microblogging-Dienstes Yammer.com durch inzwischen über 17.000 Mitarbeiter des international agierenden Beratungsunternehmens *Capgemini* dar. Yammer ist eine im World Wide Web erreichbare Plattform, die speziell für Unternehmen, Organisationen und sonstige geschlossene Gruppen angeboten wird. Aufgrund eines sehr einfachen Registrierungsverfahrens, kann jeder Mitarbeiter eines Unternehmens durch die Eingabe der eigenen (Firmen-)E-Mailadresse Mitglied eines geschlossenen Netzwerks auf Yammer werden, das sich einzig und allein durch die Domäne der E-Mail-Adresse definiert. So können nur Mitarbeiter desselben Unternehmens in das Netzwerk gelangen. Im März 2009 beschlossen mehrere Capgemini-Berater ein firmeneigenes Yammer-Netzwerk zu gründen, in welchem sich nach und nach immer mehr Kollegen in über 100 Themengruppen organisierten und austauschten. Nach einigen Monaten schlossen sich auch mehrere hochrangige Manager des Unternehmens, unter anderem der Chief Technology Officer (CTO), der Nutzergruppe an, um wie die anderen Nutzer die Potentiale des Dienstes für sich zu entdecken. Stellvertretend für viele andere Nutzer stellte der CTO fest, dass er *„vom Nutzen für das Unternehmen überzeugt [ist], und weiter daran interessiert [wäre], neue Anwendungsszenarien zu identifizieren"*.

Bei *Fraport*, einem Großunternehmen in der Flugbranche mit international rund 70.000 Beschäftigten, wurde der Wunsch nach einem unternehmensweiten Wiki im Rahmen einer firmeninternen Mitarbeiter-Befragung offenbar. In einem im Jahr 2006 initiierten Projekt begann zunächst eine für zwei Jahre angesetzte experimentelle Erprobung eines *„Wiki als Web-2.0-Werkzeug im Wissensmanagement"*. Eine Kerngruppe bestehend aus fünf Mitarbeitern aus Personalabteilung, Unternehmenskommunikation, Wissensmanagement und IT begann mit der inhaltlichen Betreuung und Erstbefüllung des Wiki sowie mit der Umsetzung des Projekts. Nach und nach meldeten sich die ersten Autoren, wobei ein Teil von ihnen, vorwiegend Themenexperten, persönlich durch die Kerngruppe angesprochen wurde. Zur internen Vermarktung der Lösung wurden zahlreiche Flyer inklusive eines Vorworts des Arbeitsdirektors online und offline im Unternehmen mit folgender Aussage verteilt: *„Inhalt des Wiki darf jegliches Wissen sein, welches für die Unternehmen und seine Mitarbeiter wichtig ist. Mitarbeiter bei Fraport sind daher frei, Inhalte im Wiki nach bestem Wissen und Gewissen zu erstellen sowie zu verändern."* „Von oben" wurden weder einzelne gewünschte Themenbereiche vorgeschlagen, noch Prozesse in denen das Wiki zum Einsatz kommen soll. Das Wiki entwickelt sich im Unternehmen evolutionär, getrieben durch die Initiative der Beitragenden.

4.2 Promotion

Ebenfalls in rund einem Viertel der Fallstudien wurde Promotion als dominierende Einführungsstrategie verwendet.

Unter Promotion wird in diesem Beitrag die (mit Unterstützung des Managements) koordiniert vermarktete und gezielt geschulte Nutzung der neuen Dienste mit Fokus auf bestimmte Nutzungspotentiale verstanden.

Dabei war die Vorgehensweise der Promotion in den Fallstudien insbesondere durch folgende Merkmale charakterisiert:

- Das *Potential* des neuen Dienstes war den Treibern größtenteils bereits vor der Einführung dieses Dienstes bekannt.
- Die möglichen Arten der *Nutzung* wurden im Rahmen der Einführung (mit Unterstützung des Managements) koordiniert und kommuniziert.
- Die *Treiber* der Einführung hegten bereits eine klare Erwartung an den neuen Dienst und erkannten eine klare Zielsetzung sowie einen bestimmten Nutzen.

Als Beispiel für eine Einführung bei welcher der Ansatz der Promotion verfolgt wurde, lässt sich *Infineon Technologies Austria* nennen, die österreichische Tochter eines weltweit operierenden Konzerns in der Halbleiterindustrie. In diesem Fall bestand das klar definierte Ziel des einzuführenden Wiki in der Schaffung einer elektronischen Wissensbasis, um das *„Kerngeschäft des Unternehmens durch effizienteren Tool- und Methodensupport zu unterstützen"*. Der Bedarf nach einer neuen Lösung für den Transfer von internem Supportwissen zwischen Supportmitarbeitern wurde durch den verantwortlichen Manager des „Design Application Engineering" Teams festgestellt. Dezentral in Projektgruppen operierende Mitarbeiter benötigten eine zentralisierte Lösung, um wiederkehrende Fragen und Probleme effektiv zu bearbeiten. Im Hinblick auf den Adressatenkreis war das durch den Manager einführte Wiki in erster Linie dazu gedacht, Supportmitarbeiter in ihrem Kerngeschäft zu unterstützten. Verbunden mit internen Marketingmaßnahmen, wie beispielsweise der Präsentation des Wiki inklusive Zielen und erwartetem Nutzen in anderen Abteilungen, wird sukzessive daran gearbeitet, auch die eigentlichen Adressaten des Supports, Techniker und Chipentwickler, zur eigenständigen Nutzung des Support-Wiki zu überzeugen.

Bei der *Maschinenfabrik Andritz*, einem österreichischen Unternehmen aus der Branche Maschinen- und Anlagenbau, wurde ein passendes Werkzeug gesucht, um die Abwicklung eines Projekts zur Einführung eines neuen Customer-Relationship-Management-Systems zu optimieren. Der Nutzen des durch die beiden Projektleiter eingeführten Wiki war allen Mitarbeitern sehr schnell klar. Geht ein Projekt über eine gewisse Komplexität hinaus, wird eine geeignete technologische Lösung für die Unterstützung benötigt. Das Wiki erfüllte alle an eine solche Lösung gestellten Anforderungen ausgezeichnet und stellte sich als einfach und rasch implementierbar heraus. Alle Projektmitarbeiter verstanden diesen Umstand schnell und nutzten das Wiki, auf die von den Projektleitern beabsichtigte und an die Projektmitarbeiter kommunizierte Art und Weise.

Bei der *Pentos*, einem deutschen IT-Systemdienstleister, wurde durch den Vorstand festgestellt, dass der Wissenstransfer zwischen den Mitarbeitern nicht optimal war. So wurden Mitarbeiter-Weblogs auf der Basis von Lotus-Notes implementiert und durch den Vorstand in Abstimmung mit der Personalabteilung eingeführt. Das Ziel der neuen Lösung bestand darin, die interne Mitarbeiter-Kommunikation effektiver zu gestalten, als über klassische Medien wie Telefon und E-Mail möglich und die Mitarbeiter durch das Schreiben von Beiträgen zu mehr Selbstreflexion über ihre Tätigkeiten zu ermuntern. Mitarbeiter sollen in ihren „Wochenberichten" einmal pro Woche festhalten, was sich in ihrem Arbeitsbereich ereignet hat und diese Informationen dann allen Kollegen zur Verfügung stellen. Die Informationen in den Mitarbeiter-Weblogs sollen Mitarbeiter-Ziele und -Tätigkeiten sowie den Bezug dieser Tätigkeiten zum Unternehmenserfolg transparent machen. Sowohl die Nutzungsart, als auch der erzielte Nutzen war bei Pentos also von Anfang an durch den Vorstand klar definiert und kommuniziert.

4.3 Exploration und Promotion

Wie sich in den untersuchten Fallstudien zeigte, sind beide Vorgehensweisen auch miteinander vereinbar, ohne dass eine die andere dominiert. In elf Unternehmen, also in mehr als der Hälfte der untersuchten Fälle, konnten beide Vorgehensweisen in Kombination festgestellt werden, teilweise auch nachgelagert (zuerst Exploration, dann Promotion). Drei Beispiele für diese Mischform werden im Folgenden angeführt.

Zur Verbesserung der unternehmensinternen, internationalen Zusammenarbeit und zur Unterstützung des unternehmensweiten, informellen Expertenaustauschs führte Ende 2005 der zentrale IT-Bereich von *Bosch* zusammen mit einer Entwicklungsabteilung des Geschäftsbereichs Diesel Systems eine mehrmonatige Studie zur unternehmensweiten Wiki-Einführung durch. Das Ziel der Wiki-Studie bestand darin, eine einheitliche Wiki-Plattform auszuwählen, diese im Testbetrieb zu untersuchen und schließlich Anfang 2007 im Produktivbetrieb jedem Mitarbeiter der Bosch-Gruppe zur Verfügung zu stellen. Schon vorher waren vereinzelt Wiki-Server im Einsatz - in der Regel genutzt innerhalb einer Abteilung und auf unterschiedlichen Plattformen. Aufgrund dieser vorhergegangenen „*Explorationsphase*" waren bereits verschiedene Wiki-Nutzungsszenarien identifiziert worden, die bei der *Promotion* des Wikis in den weltweiten Abteilungen hilfreich waren. Dazu gehörte die Nutzung zum Bug Tracking, d.h. zum Verfolgen von Software-Fehlern, als Glossar, als Artikelsammlung, als Dokumentationsplattform, zum Experten-Debriefing und zum interkulturellen Austausch.

Die 180 Mitarbeiter der *Communardo GmbH* arbeiten in wissensintensiven Softwareentwicklungs- und Beratungsprojekten und in Form verschiedener Teams zusammen. Bedingt durch das schnelle Wachstum entstanden Schwierigkeiten bei der Weitergabe von Informationen und Ideen zwischen Organisationseinheiten und zwischen Projekten. So wurde im Frühjahr 2008 von einem Mitarbeiter vorgeschlagen das Potential von Twitter oder einen Twitter-ähnlichen Instrument zur Erleichterung der Kommunikation im Team zu untersuchen. Diese Idee wurde von der Geschäftsführung unterstützt und mitgetragen. Schnell entschied man sich gegen einen öffentlichen Microblogging-Dienst und entwickelte eine eigene Lösung. Jedoch war unklar, welche Nutzungsszenarien die Plattform insbesondere unterstützen würde. Also wurde beschlossen, die Potentiale der Plattform nach und nach für sich zu entdecken (*Exploration*). Die Geschäftsführer waren von den Vorteilen der Nutzung überzeugt und begannen „Erfolgsgeschichten" regelmäßig an die Belegschaft zu kommunizieren (*Promotion*). Bereits Mitte 2009 hatte der Dienst so nicht nur einen festen Platz in der IT-Landschaft des Unternehmens erobert, sondern es waren eine Vielzahl von Nutzungsszenarien bekannt, in denen der Dienst genutzt werden sollte (*Promotion*).

Bei *Spirit Link*, einem deutschen IT-Systemhaus für die Healthcare Industrie, wurde durch die Geschäftsleitung ein Leidensdruck in Folge einer Überverwendung von klassischen Medien wie E-Mail festgestellt. Die Effizienz einer über Kanäle im Gegensatz zu Portalen gesteuerten Informationsverteilung wurde im Unternehmen als sehr gering eingestuft. Das Ziel der durch die Geschäftsleitung initiierten Themen-Weblogs bestand darin, das Kerngeschäft des Unternehmens durch eine optimierte Verteilung innovativer Inhalte im Unternehmen zu unterstützen (*Promotion*). Inhalte und inhaltliche Strukturen werden jedoch weitestgehend von den themenverantwortlichen Mitarbeitern in Selbstorganisation erstellt (*Exploration*), welche die von ihnen verantworteten Weblogs im Unternehmen auch eigenständig vermarkten. Obwohl Nutzen und primäre Nutzung der Themen-Weblogs durch die Geschäftsleitung mittels der Zuweisung von Themen und Themenverantwortlichen vorgegeben wird (*Promotion*), obliegen inhaltliche Arbeiten und Ausrichtung der Weblogs und das Erstellen von Beiträgen den Mitarbeitern selbst (*Exploration*).

5. DISKUSSION

Obwohl die beiden Vorgehensweisen aus Sicht der Autoren klar beschrieben und voneinander abgegrenzt wurden, wird aus den neun fokussiert erläuterten Fallbeispielen klar, dass jede Einführung von unterschiedlichen Faktoren beeinflusst wird und daher nur im Kontext betrachtet werden kann. So können, wie aufgezeigt, beispielsweise bereits bestehende Vorgängerlösungen Einfluss auf das Vorgehen bei der Einführung eines neuen Dienstes ausüben, ebenso wie verschiedene Zielsetzungen sowie Anforderungen von Stakeholdern an die Dienste und noch eine Vielzahl weiterer Einflüsse. Aus diesem Grund fordert eine saubere Trennung der Vorgehensweisen und anschließende Einordnung der einzelnen Fälle eine gründliche Auseinandersetzung mit jedem unterschiedlichen Einführungsprozess (vgl. dazu z.B. [24]). Dabei waren mehrere bestehende Denkansätze zur Einführung von Groupware aus dem Forschungsgebiet Computer Supported Collaborative Work (CSCW) hilfreich, die in diesem Abschnitt mit den in der Studie gewonnenen Erkenntnissen in Beziehung gesetzt werden. Gleichzeitig wird im Folgenden auch auf die bestehenden Unterschiede von (traditioneller) Groupware und Social Software hingewiesen.

5.1 Exploration

Zusammenfassend lässt sich sagen, dass das Ziel der Exploration in den untersuchten Fällen darin bestand, einen neuen Dienst als Hilfsmittel in bestehende (soziale) Prozesse und Mitarbeiter-Aktivitäten zu integrieren und diese damit zu unterstützen. Gleichzeitig sollten aber auch weitere noch nicht identifizierte Anwendungsszenarien mithilfe des Dienstes unterstützt werden. Es ging es den Verantwortlichen bei prototypischen Umsetzungen oftmals gar nicht darum, einen sogenannten „viralen Effekt" (z.B. [2]) zu erzielen und damit eine starke und schnell skalierende Nutzerbasis zu gewinnen. Vielmehr hing die Entscheidung davon ab, ob das Potential eines Dienstes bereits vollkommen klar war,

oder ob Management bzw. Mitarbeiter nach und nach weitere bisher unbekannte Nutzungsmöglichkeiten identifizierten.

Die dieser Fragestellung zugrundeliegende Nutzungsoffenheit kann dabei als eine Kerncharakteristik von Corporate Social Software und als Unterschied zu (traditioneller) Groupware gesehen werden [39]. Das heißt, das Potential der Dienste zeigt sich erst nach der Aneignung durch die Anwender. Die Dienste müssen hierfür erst zu einem Bestandteil der Gruppenpraktiken und organisationalen Prozesse werden (vgl. [41]).

Wie in den vorliegenden Fallstudien deutlich wurde, eigneten die Nutzer sich die Dienste im Rahmen ihrer Arbeitspraktiken an. Bei dieser Aneignung wurden sie in der Regel unterstützt – nicht indem man ihnen die Nutzung strikt vorschrieb, sondern indem man ihnen lediglich eine moderne und zeitgemäße Infrastruktur für ihre Arbeitspraktiken zur Verfügung stellte. Ein klarer „Business-Case“ mit definierter Zielsetzung, die Festlegung der für die Nutzung beteiligten Mitarbeiter und eine detaillierte Beschreibung des erwarteten Nutzens für Mitarbeiter und Unternehmen fehlten. Erst im Laufe der Zeit konnte sich dieser aufgrund der langsam aufkeimenden Nutzung der neuen IT-Werkzeuge kontinuierlich herausbilden. Ein solches Vorgehen wird in der Literatur auch „Infrastructuring“ genannt (vgl. [35]). Dabei ist eine Informationsinfrastruktur als “a shared, evolving, heterogeneous installed base of IT capabilities among a set of user communities based on open and/or standardized interfaces” zu verstehen ([18], p. 208).

Im o.g. Vorgehen kommt auch die Relevanz der partizipativen Einführung von Software zum Tragen, die bereits seit Jahren im Bereich CSCW betont wird (z.B. [16], [24]). Denn neben den zunehmend komplexeren Beziehungen in solchen Systemen, ändern sich die Arbeitssysteme häufig durch die Technikeinführung und erst nach der Einführung erster Versionen werden die Möglichkeiten der Nutzung richtig klar. Aufgrund von immer kürzer werdenden Innovationszyklen ändern sich technologische Systeme viel rascher, als die Arbeitspraktiken der Mitarbeiter. Organisationale Änderungen benötigen aufgrund der höheren Komplexität sozialer Systeme mehr Zeit. Jedoch kann eine Technologie ihr volles Potenzial nur dann ausspielen, wenn sich beide Systeme aneinander anpassen.

Dabei zeigt sich auch die hohe Bedeutung einer sozio-technischen Systemgestaltung, d.h., dass bei der Entwicklung und Einführung sowohl das soziale als auch das technische Teilsystem einer Organisation gleichmäßig beachtet werden müssen [34]. Nur wenn sich beide Teilsysteme regelmäßig im Rahmen einer Ko-Evolution einander annähern, ist der Nutzen des Dienstes sichergestellt [39]. Sehr oft wird dem technischen System jedoch eine weit höhere Aufmerksamkeit geschenkt und ein nicht adaptiertes soziales System konterkariert den Nutzen einer neuen Technologie.

5.2 Promotion

Das Ziel der Promotion besteht darin, möglichst rasch eine kritische Masse [29] an echten Nutzern zu erreichen, was gerade für Social Software eine wichtige Rolle spielt (z.B. [39]). Hier kann das Management als Vorbild dienen, d.h. die Nutzer motivieren und zur Nutzung der neuen Dienste aufzurufen. Es sollte aber gleichzeitig nicht diktieren, wie Nutzer damit zu arbeiten haben. Zudem sollten keine Hierarchien abgebildet werden, wo sie nicht notwendig sind und sichergestellt werden, dass die richtigen Anreize zur Nutzung bestehen.

Weiterhin ist die Vorgehensweise der Einführung dadurch charakterisiert, dass ein „Business-Case“ bereits vor der Einführung vorhanden ist und meist durch das Management definiert wurde. Entsprechende IT-Werkzeuge wurden in mehreren Fällen auch bereits vor dem Hintergrund dieses speziellen Business-Cases ausgewählt. Dieser Business-Case wird auch im Rahmen von Begleitmaßnahmen wie beispielsweise Akzeptanzmaßnahmen, Präsentationen oder Schulungen laufend an die Mitarbeiter kommuniziert.

Gleichzeitig zeigte sich in mehreren Fällen, dass die Phasen des klassischen Veränderungsprozesses (beispielsweise Sensibilisierung, Vision entwickeln und kommunizieren, Steuerung der Veränderung, etc., vgl. z.B. [25]) auch für Corporate Social Software durchaus sinnvoll anwendbar sind. Interessant erscheint in diesem Zusammenhang ebenfalls, dass allgemein auch die Vertreter nachhaltig gemeinter Veränderungsprozesse dafür plädieren, alle Stakeholder frühestmöglich einzubeziehen (z.B. [43]). Gerade an diesem letztgenannten Punkt zeigt sich, dass die beiden Vorgehensweisen nicht diametral gegenüber stehen müssen, sondern sich durchaus ergänzen können bzw. sich teilweise ähnliche Schritte anbieten.

5.3 Limitationen

Die vorgestellten Forschungsergebnisse müssen im Lichte der folgenden Limitationen der Analyse betrachtet werden. Wie an der eben geführten Diskussion ersichtlich, ist die Abgrenzung der beiden Vorgehensweisen Exploration und Promotion stark vom Kontext abhängig. Als Grund hierfür ist die bereits angesprochene Vielzahl unterschiedlicher Faktoren, welche die Entscheidung für eine oder beiden Vorgehensweisen beeinflussen können, zu sehen. Es wurden Unternehmen unterschiedlicher Größe und mit sich unterscheidenden Organisationskulturen untersucht – um nur zwei Unterschiede zu nennen.

Während statistische Generalisierbarkeit durch die durchgeführte Forschung ohnehin nicht beabsichtigt wird, so ließe sich durch die Erweiterung der Cross-Case-Analysis die analytische Generalisierbarkeit der gewonnenen Ergebnisse erhöhen (vgl. dazu z.B. [54]). Dabei lässt sich die in diesem Beitrag entwickelte Theorie („Exploration“, „Promotion“) als Schablone benutzen und mit den empirischen Ergebnissen weiterer Fallstudien vergleichen. Unterstützen zwei oder mehrere Fälle die Theorie, kann von Replikation gesprochen werden [54].

Des Weiteren ist anzumerken, dass aufgrund der vorliegenden einheitlichen Struktur aller Fallstudien, der eingegrenzten Forschungsfrage und der Vertrautheit der Autoren mit den einzelnen Fällen auf eine zusätzliche Codierung der Fallstudien verzichtet wurde. Dabei wird nicht in Frage gestellt, dass für die weiteren Auswertungen eine Codierung (wie z.B. in [44] verfolgt) einen zusätzlichen Nutzen bringen kann.

6. ZUSAMMENFASSUNG UND AUSBLICK

Im vorliegenden Beitrag wurden Teilergebnisse aus einer vergleichenden Analyse von 21 Fallstudien zum Einsatz von Corporate Social Software vorgestellt und diskutiert. Durch die systematische Betrachtung der Vorgehensweisen bei der Einführung stellte sich heraus, dass in den Unternehmen zwei unterschiedliche, jedoch nicht unbedingt widersprüchliche, Vorgehensweisen angewandt wurden: „Exploration“ (in fünf Fällen), „Promotion“ (in fünf Fällen), sowie eine Kombination beider Vorgehensweisen (in elf Fällen).

Eine Flankierung der Einführung von Corporate Social Software von „links“ und „rechts“ erscheint dabei zielführender als die nicht enden wollende Diskussion darüber, ob Unternehmen Corporate Social Software „von oben“ oder „von unten“ einführen sollten. Die in diesem Beitrag vorgestellten Forschungsergebnisse räumen auch mit dem Vorurteil auf, dass neue IT-Werkzeuge in Unternehmen immer Top-Down eingeführt werden müssen, um erfolgreich zu sein. Neue Dienste können auch nutzenstiftend wirken, wenn sie nicht Top-Down, sondern explorativ eingeführt werden. Den Autoren liegt nach bestem Wissen und Gewissen noch keine wissenschaftliche Arbeit vor, welche diese von Praktikern schon länger gemutmaßte Erkenntnis anhand einer großen Anzahl an Fallstudien bestätigt.

Die vorgestellte Studie hat noch weitere Erkenntnisse zu Tage gebracht, die jedoch nicht im Fokus des vorliegenden Beitrags stehen. Beispielsweise ist den Fallstudien zu entnehmen, dass der Einsatz von Social Software zunehmend weniger prototypisch abläuft, den Weg in die Arbeitsprozesse findet und mit anderen Diensten integriert wird (oder werden soll). Auf Basis der erhobenen Daten lässt sich vermuten, dass Werkzeuge wie Wikis und Weblog im Kontext von Unternehmen bereits eine gewisse Reife besitzen, sodass sich Promotion als dominierende Strategie herauskristallisiert, während Microblogging als neuer Dienst verstärkt mittels Exploration in Unternehmen durch die Mitarbeiter „erforscht“ wird. Eine Einführung neuer Dienste, bei denen noch größtenteils Unklarheit über Szenarien und Nutzungsmuster herrscht, tendiert eher in Richtung Exploration, während klassische Anwendungen, für die bereits reichlich Verständnis zu Nutzung und Nutzen vorliegt, von „Promotion“ stark profitieren.

7. LITERATUR

[1] Back, A., Gronau, N., und Tochtermann, K. 2009. *Web 2.0 in der Unternehmenspraxis*. Oldenbourg, München.

[2] Barabasi, A.L. 2003. *Linked: How Everything Is Connected to Everything Else and What It Means*. Plume, London.

[3] Böhringer M, Richter, A., und Koch, M. 2009. Awareness 2.0 - Ein Anwenderbeispiel von Microblogging im Unternehmen, *Information, Wissenschaft & Praxis*, 60, 4, 275-279.

[4] Buhse W., und Stamer S. 2008. *Enterprise 2.0 - Die Kunst, loszulassen.* Rhombos, Berlin.

[5] Bughin, J., und J. Manyika 2007. *How businesses are using Web 2.0: A McKinsey Global Survey.* McKinsey Research.

[6] Danis, C., und Singer, D. 2008. A Wiki Instance in the Enterprise: Opportunities, Concerns and Reality. In *Proceedings of the 11th Conference on Computer Supported Cooperative Work*, San Diego, ACM Press.

[7] DiMicco, J.M., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., und Muller, M. 2008 Motivations for Social Networking at Work. In *Proceedings of the 11th Conference on Computer Supported Cooperative Work*, San Diego, ACM Press.

[8] Disterer, G. 2000. Individuelle und soziale Barrieren beim Aufbau von Wissenssammlungen. *Wirtschaftsinformatik*, 42 6, 539-546.

[9] Dubé L. und Paré G. 2003. Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly* 27, 4, 597-635.

[10] Ebersbach, A., und Glaser, M. 2009. Wiki als zentrales Suchportal. Das Beispiel bluepedia. Information, Wissenschaft & Praxis, 60, 4, 197-201.

[11] Efimova, L,. und Grudin, J. 2007. Crossing boundaries: A case study of employee blogging. In *Proceedings of the 40th Hawaii International Conference on System Sciences*, Waikoloa, IEEE Computer Society Press.

[12] Eisenhardt K. M. 1989. Building Theories from Case Study Research. *The Academy of Management Review*, 14, 4, 532-550

[13] Glaser B.G., und Strauss, A.L. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Publishing Company, Chicago

[14] Grace, T. 2009. Wikis as a Knowledge Management Tool. *Journal of Knowledge Management*, 13, 4, 64-74.

[15] Gratton, L. und Erickson, T.J. 2007. 8 Ways to Build Collaborative Teams. *Harvard Business Review* (Nov. 2007), 100-109.

[16] Gross, T, Koch, M. 2007. *Computer-Supported Cooperative Work*, Oldenbourg, München.

[17] Han, B.M., und Anantatmula, V. 2006. Knowledge Management in IT Organizations From Employee's Perspective. In *Proceedings 39th Hawaii International Conference on System Sciences*

[18] Hanseth, O., und Lytinnen, K. 2004. Theorizing about the Design of Information Infrastructures: Design Kernel Theories and Principles, *Sprouts: Working Papers on Information Systems*, 4, 12.

[19] Happel, H., und Treitz, M. 2008. Proliferation in Enterprise Wikis. In: *Proceedings International Conference on the Design of Cooperative Systems*, Carry-le-Rouet.

[20] Holsapple, C.W., und Joshi, K.D. 1999. Description and Analysis of Existing Knowledge Management Frameworks. *In Proceedings of the 32th Hawaii International Conference on System Sciences*.

[21] Holtzblatt, L., L. Damianos. und Weiss, D. 2010. Factors Impeding Wiki Use in the Enterprise: A Case Study. In *Proceeding of the 28th annual SIGCHI conference on Human factors in computing systems*, Atlanta, ACM Press.

[22] Ip, K.F.R., und Wagner, C. 2008. Weblogging: A study of social computing and its impact on organizations. *Decision Support Systems*, 45, 2, 242-250.

[23] Jahnke, I. 2009. Socio-technical Communities: From Informal to Formal? In: Withworth, B. (Hrsg.) *Handbook of Research on Socio-Technical Design and Social Networking Systems*. IGI Global Publisher, 763-778

[24] Karsten, H. und Jones, M. 1998. The Long and Winding Road: Collaborative IT and Organisational Change. In *Proceedings of Conference of the ACM 1998 Conference on Computer Supported Cooperative Work*, Seattle.

[25] Kotter, J.P. 1996. *Leading Change*. Harvard Business School Press, Boston.

[26] Koch, M. 2008. CSCW and Enterprise 2.0 - towards an integrated perspective. *In Proceedings 21st Bled eConference eTrust: Implications for the Individual, Enterprises and Society*.

[27] Koch, M., und Richter, A. 2009. *Enterprise 2.0 - Planung Einführung und erfolgreicher Einsatz von Social Software im Unternehmen.* Oldenbourg, München.

[28] Mans, R. 2010. *Bottom up: an overrated and underperforming social strategy (for lazy people).* Online: http://www.capgemini.com/technology-blog/2010/04/ bottom_ up_overrated _and_underperforming.php

[29] Markus, M. L. 1987. Toward a "critical mass" theory of interactive media: Universal access, interdependence and diffusion. *Communication Research*, 14, 5, 491.

[30] McAfee, A. 2009. *Enterprise 2.0: New Collaborative Tools for Your Organization's Toughest Challenges*. Mcgraw-Hill Professional, Boston.

[31] Mueller, J. und Stocker, A. 2010. Enterprise Microblogging at Siemens, Building Technologies Division: A Descriptive Case Study. In *Proceedings of 10th International Conference on Knowledge Management and Knowledge Technologies*, Graz.

[32] Orlikowski, W. J. 2000. Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, 11, 4, 404-428.

[33] Paroutis, S., und Saleh, A. 2009 Determinants of knowledge sharing using Web 2.0 technologies. *Journal of Knowledge Management*, 13, 4, 52-63.

[34] Pasmore, W., Franics, C., Haldeman, J., und Shani, A. 1982. Sociotechnical systems: A North American reflection on empirical studies in the seventies, *Human Relations*, 35, 12, 1179-1204.

[35] Pipek, V., und Wulf, V. 2009. Infrastructuring: Towards an Integrated Perspective on the Design and Use of Information Technology, *Journal of the Association of Information Systems*, 10, 5.

[36] Razmerita, L., Kirchner, K., und Suzdina, F. 2009. Personal knowledge management. The role of Web 2.0 tools for managing knowledge at individual and organizational levels.: *Journal of Knowledge Management*, Emerald Publishing.

[37] Richter, A., und Bullinger, A. C. 2010. Enterprise 2.0 – Gegenwart und Zukunft. Vorschlag einer Forschungsagenda. In *Proceedings Multikonferenz Wirtschaftsinformatik, Göttingen.*

[38] Richter, A., Mörl, S., Koch, M.; und Trier, M. 2011. Anwendungsszenarien als Werkzeug zur (V)ermittlung des Nutzens von Corporate Social Software. In Proceedings Wirtschaftsinformatik 2011, Zurich.

[39] Richter, A., und Riemer, K. 2009. Corporate Social Networking Sites – Modes of Use and Appropriation through Co-Evolution. In Proceedings *20th Australasian Conference on Information Systems*, Melbourne.

[40] Richter, A., und Vogl, G. 2009. SkiBaserl – Knowledge Management in High-Performance Sports. In *Proceedings 9th International Conference on Knowledge Management and New Media Technology*, Graz.

[41] Riemer, K., Frößler, F., und Klein, S. 2007. Real Time Communication – Modes of Use in Distributed Teams. In *Proceedings 15th European Conference on Information Systems*. St. Gallen.

[42] Riemer, K.und Richter, A. 2010. Tweet Inside: Microblogging in a Corporate Context. *In Proceedings 23rd Bled eConference eTrust: Implications for the Individual, Enterprises and Society*.

[43] Rosenstiel, L. v. 2000. *Grundlagen der Organisationspsychologie*. Schäffer-Poeschel Verlag, Stuttgart.

[44] Schubert, P., Williams, S. P. 2009. Constructing a Framework for Investigating and Visualizing ERP Benefits and Business Change. In *Proceedings of the 22nd International Bled eConference*, Bled.

[45] Schubert, P., Wölfle, R. 2006. eXperience-Methodik zur Dokumentation von Fallstudien. In: *Wölfle, R.; Schubert, P. (Hrsg.): Prozessexzellenz mit Business Software*. Hanser Verlag, München, 19-30.

[46] Schubert, P., Wölfle, R. 2007. The eXperience Methodology for Writing IS Case Studies. In *Proceedings of the Thirteenth Americas Conference on Information Systems*, Keystone.

[47] Senger, E., Österle, H. 2004. *PROMET – Business Engineering Cases Studies (BECS).* Arbeitsbericht BE HSG / BECS / 1. Institut für Wirtschaftsinformatik, Universität St. Gallen, Juni 2004.

[48] Stake, R.E. 1995. *The Art of Case Study Research.* SAGE Publications, London.

[49] Stocker, A. Tochtermann, K. 2009. Exploring the Value of Enterprise Wikis - A Multiple-Case Study. *Proceedings of the International Conference on Knowledge Management and Information Sharing*, Funchal.

[50] Stocker, A., und Tochtermann, K. 2010. *Wissenstransfer mit Wikis und Weblogs. Fallstudien zum erfolgreichen Einsatz von Web 2.0 im Unternehmen.* Gabler-Verlag, München.

[51] Stake, R.E. 1995 The Art of Case Study Research. SAGE Publications, London.

[52] Warta, A., und Richter, A. 2007. Die Medienvielfalt als Barriere für den erfolgreichen Einsatz von Wikis im Unternehmen am Fallbeispiel der Robert Bosch GmbH. In *Proceedings des Workshops Gemeinschaften in Neuen Medien*, Dresden.

[53] Wulf, V. 2009. Theorien sozialer Praktiken zur Fundierung der Wirtschaftsinformatik: Eine forschungsprogrammatische Perspektive, In: *Wissenschaftstheorie und Gestaltungsorientierte Wirtschaftsinformatik*, Springer/Physika, 211-224.

[54] Yin, R.K. 2003. *Applications of case study research*. Thousand Oaks, California: Sage Publications, Thousand Oaks.

8. ANHANG

#	Unternehmen, Land	Technologie (konkrete Umsetzung)	Ursprung / Konzept	Einführung (P = Promotion, E = Exploration)	
1	ABB, D	Wiki, Blog, Newsletter (Sharerpoint 3.0)	Pilotprojekt, Test versch. Tools, Nutzung nicht vorgegeben	**E, P**	"Wiki-Tour", Ansprache von Multiplikatoren
2	Accenture, weltweit	Social Networking Service (SNS) (Eigenentwicklung)	Pilotprojekt nach Vorbild Facebook und LinkedIN, „Business-Case“ unscharf	**E, P**	Evaluierung möglicher Nutzungsszenarien, div. Promotionen bei weltweiter Einführung
3	Bosch, D	Wiki (Confluence)	Mehrere unterschiedliche Plattformen wurden zusammengeführt	**E, P**	Zunächst Pilotprojekt bei Division (Diesel), dann weltweiter Roll-Out
4	CapGemini, weltweit	Microblogging (Yammer)	Einführung von kleiner Gruppe von Mitarbeitern vorangetrieben	**E**	Geschäftsführung (CTO) von Nutzung überzeugt, aber keine Vorgabe
5	Communardo, D	Microblogging (Eigenentwicklung)	Pilotgruppe mit Wordpress-Lösung, Rapid Prototyping	**E, P**	Potentiale während Entwicklung identifiziert, später Wunsch der Nutzung und Art der Nutzung durch GF klar
6	Deutscher Skiverband, D	Wiki (TWiki)	Pilotprojekt, ohne vorher klare Nutzungsszenarien	**E, P**	Nutzerworkshops zur Identifikation von Nutzungsszenarien, Schulungen
7	Fraport, D	Wiki (MediaWiki)	Interne Befragung zu WM-Instrumenten, Pilotprojekt	**E**	Ohne klare Nutzungsszenarien gestartet, Nutzung nicht vorgegeben
8	IBM, weltweit	SNS (Eigenentwicklung)	Evolution aus bestehender Lösung heraus, vielfältige Integration	**E**	Kontinuierliche Identifikation neuer Nutzungspotentiale
9	Infineon Technologies, AUT	Wiki (MediaWiki)	Bedarf durch Abteilungsleiter erkannt, teilweise Migration von anderem System	**P**	Vorstellung klar definierter angestrebter Nutzungsszenarien in den Abteilungen
10	IVM, AUT	Wiki (Perspective)	Kein Intranet vorhanden, Bedarf durch Abteilungsleiter erkannt und Projekt aufgesetzt	**E, P**	Nutzungsszenarien für Techniker und Verwaltung, Techniker haben effektive Nutzungspraktiken herausgebildet
11	Maschinenenfabrik Andritz, AUT	Wiki (TWiki)	Klar vordefinierter Einsatzfall, Bedarf durch Manager erkannt	**P**	Klar definierte Nutzungsszenarien, Ziel der Nutzung von Beginn an klar
12	Microsoft, AUT	Integrierte Blogs (Sharepoint)	Kein Intranet vorhanden, Kommunikationskonzept als Basis	**P**	Blog-Nutzung aus E-Mail-Nutzung ableiten, Schulungen, Flyer
13	Pentos, D	Blogging (Lotus Notes/Domino)	Bedarf durch Vorstand erkannt, simple Eigenkonzeption	**P**	Schulung durch Geschäftsleitung, finanzielle Honorierung
14	Raiffeisen, AUT	Wiki (JSPWiki)	Bedarf durch WM-Kerngruppe erkannt	**P**	Fest vordefinierte Inhalte, Artikel zur Nutzung und Zielen, Flyer
15	SAP, weltweit	SNS (Eigenentwicklung)	Pilotprojekt nach Vorbild Facebook und LinkedIN	**E, P**	Evaluierung möglicher Nutzungsszenarien, spätere Promotion
16	Siemens Building Technologies Division, CH	Microblogging-Service (Eigenentwicklung)	Pilotprojekt nach Vorbild Yammer, Integration in bestehende WM-Plattform	**E**	Individuelle Verwendung gewünscht, Akzeptanzmaßnahmen
17	Siemens, weltweit	SNS (Liferay)	Pilotprojekt, Migration bestehender Daten aus Vorgängersystem	**E**	Nutzer sollen selbst Zugang finden, freier Verwendungszweck
18	Spirit Link, D	Weblogs (Wordpress)	Gespräche der Geschäftsleitung mit befreundeter Organisation	**E, P**	Zuweisung von Verantwortlichen durch Geschäftsleitung nach Themen, Eigenständiges Bearbeiten
19	Synaxon, D	Wiki-Blog (Media-Wiki & eigenentwickelter Blog)	Überzeugung des Vorstands, Evolution aus bestehender Lösung	**E, P**	Mitarbeiter agieren in Eigeninitiative, entwickeln die Vision des Unternehmens weiter
20	TAO Beratungs- und Management, AUT	Wiki (MediaWiki)	Überzeugung eines Standortleiters, Pilotprojekt nach Vorbild Wikipedia	**E, P**	Schulungen, Ziel weitere Nutzungsszenarien zu finden
21	T-Systems MMS, D	Social Intranet (Confluence)	Heterogenes Projektteam, Test des Confluence Wiki	**E,P**	Pilotgruppe, iterative Freischaltung von neuen Nutzern

Investigating Early Adopters' Use of Location-based Social Networks: Implications for Local Businesses and Service Providers

Valentin Schöndienst
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin, Germany
+49 30 2093 5742
valentin@schoendienst.net

Linh Dang-Xuan
University of Münster
Leonardo-Campus 3
48149 Münster, Germany
+49 251 8338 118
linh.dang-xuan@uni-muenster.de

Oliver Günther
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin, Germany
+49 30 2093 5742
guenther@wiwi.hu-berlin.de

ABSTRACT
Social media has been defined by what happens on people's computer screens. But what happens when people turn off their computers and take social media to the real, physical world? Now, with recent advancements in mobile technology, early adopters build communities around the concept of 'check-ins'. They broadcast their location to friends, learn about other people's whereabouts, and share location-based information about bars, parks, cities, and virtually any kind of location. We present a study of 63 early adopters who use location-based social networks in their daily lives and analyze their behavior with respect to the impact on local businesses as well as service providers. Our results show that users derive real value from connecting information to location and indicate significant potential for customer-to-customer marketing. Further, our findings provide support for claims to include privacy and context-related constructs into technology acceptance theory.

Keywords
Location-Based Social Networks, Online Social Networks, Location-Based Services, Mobile Social Networks, Usage, Adoption

1. INTRODUCTION
The use of online social networks (OSNs) for information sharing is not a novelty. Twitter has built a billion-dollar company [63] around the question "What are you doing?" delivering real-time information about what is happening in people's lives all around the world. Now, innovative location-based social networks (LBSNs) develop rapidly around the concept of 'check-ins,' i.e. answering the question "Where are you?"

By connecting information to GPS-coordinates, LBSNs create a rich body of user-generated content around locations. People use their mobile phones to check-in to places varying from parks to bars to cities they visit. That way, they broadcast their whereabouts to others nearby as well as friends. Just as it became common to tag pictures or blog posts and comment on them on the "Web 2.0," people adopt LBSNs to mark locations, express opinions, and share know-how about places in the real world.

Location-sensitive functionality is integrated into OSNs in various ways. Generally, we define an LBSN as an OSN that provides location-sensitive features and is accessible through mobile devices.

Considering "all actual life is encounter" [10], knowing where friends are and finding interesting people nearby, literally adds a new dimension to computer-mediated communication and provides significant potential for social value, and consequently, business value.

Further, being able to attach user-generated information to virtually any location vests consumers with even more power - a trend that has already been witnessed on the internet [54] as brands and shops are exposed to publicly available reviews and comments. On the other side, businesses manage to turn LBSNs into a feedback channel are provided with a unique opportunity to gain customer insights.

Overall, pinning information and people to location provides tremendous potential for consumers, service providers, and local businesses. Already, regular social networking is one of the most important activities among mobile users [50]. Adding location-sensitivity is likely to increase the importance as studies project global revenues through LBSNs to amount to US$ 3.3 billion by 2013 [1]. However, due to technological limitations, the integration of location-sensitive functionality is still in its infancy. Therefore, little is known about LBSNs and how people (will) use geo-location features.

The literature on the Social Shaping of Technology (SST) describes the development of technology as an interactive process

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

in which (early) uses of an innovation determine future use [7, 42, 65]. Further, Diffusion of Innovations Theory [55] suggests that early adopters tend to be social leaders whose use of a technology will influence others.

Given the emerging importance of LBSNs, we investigate 63 early adopters to explore (1) why people use LBSNs, and (2) what effect the integration of location-sensitive features has on individual usage.

Extending our preliminary results [BLINDED Usage patterns on mobile networks, 2010], we derive practical implications for two major stakeholders: service providers and local businesses. Further, we examine our findings with respect to their implications for technology acceptance theory.

Considering the case of the LBSN provider "Brightkite," we perform an explorative study using qualitative data analysis. We chose Brightkite because (a) it provides users a multitude of location-features, and (b) because of its established user base that provides a unique opportunity to explore drivers and concerns of using LBSNs and to uncover the role location-features play.

To this end, we proceed as follows: In the following section, we review related work. Then, we give an overview on emerging LBSN concepts and describe Brightkite in more detail. In the subsequent sections, we present our approach and the results of our study. We conclude by deriving theoretical and practical implications from our findings and reflect on areas for future research.

2. RELATED WORK

Chen and Rahman [11] analyzed LBSNs from a technical perspective and identified a need for better privacy protection to handle "mash-ups" with other services. Li and Chen [39] investigated movement patterns of users of an LBSN to identify relationships between physical locations and gain insights on users' travel patterns. Further, Li and Chen [40] examined connections among LBSN users to derive friend recommendation methods.

Fusco et al. [20] conducted a comprehensive review on the research on OSNs, LBSs, and LBSNs. They found that in the absence of mature LBSN incarnations, the limitation of many studies "and prior research on LBSN technologies was the hypothetical nature of the research, or that the research took place within a controlled environment" (p. 9) [e.g. 4, 5, 13, 19, 24, 52].

However, a number of recent studies deal with actual implementations of LBSNs. Tsai and Kelley [61] implemented a Facebook application that shares a user's location with others. Then, they investigated the influence of system feedback on individual privacy managing. They find that users are more comfortable about sharing their whereabouts when they know who accesses this information. However, peer opinion and users' technical savviness contributed the most to participants' decision whether to continue to use the location sharing application or not.

In another study, field trials of a mobile micro-blogging application were conducted [62]. The application under study automatically shared users' location. The focus of the study was on the aspect of automated location disclosure and how automation affects individual usage behavior. The study showed that automated location sharing caused issues related to control, understanding, and privacy.

Fusco et al. [19] conducted four focus group sessions discussing the use of Google's automated location sharing service Latitude. Only two participants had actually tried an LBSN before, so the discussions were of hypothetical nature. Participants who would use such a service named monitoring and tracking of friends, family, and employees, keeping a travel journal, and fun as reasons to adopt. The majority of participants, however, indicated that they would not use an LBSN. These people perceived automated location sharing as intrusive and expressed concerns with respect to trust and privacy. In addition, they mentioned a lack of critical mass as well as technical issues.

Another study that deals with automated location disclosure presents preliminary results from 12 interviews with both users of Google Latitude as well non-users [51]. The authors argue for an integrated research approach that studies LBSNs in the context of other social networking and communication technologies.

Humphreys [27] explores the social and behavioral norms of Dodgeball, a text message-based LBSN. Among other things, the results indicate that users see an LBSN as a tool to enable and coordinate social interactions among loosely tied groups of friends. Humphreys' work makes an important contribution towards an understanding of emerging patterns of social interaction through LBSNs. Meanwhile, she acknowledges the rapid development in LBSNs and suggests investigating "ways in which people adopt and integrate these kinds of systems into their everyday lives" (p. 357) in future research endeavors. We intend to contribute to this research stream by exploring why people use LBSNs and what role location-features play.

In particular, we regard further research necessary as the current development shows that LBSNs emerge around the concept of check-ins rather than automated location sharing. Check-ins require users to pro-actively share their location which may lead to different attitudes towards uses of LBSNs as compared to automated LBSNs which were the focus of past studies.

3. LOCATION-BASED SOCIAL NETWORKS

3.1 Overview

Until recently, technological limitations did not allow for full-fledged incarnations of LBSNs. Now, a multitude of concepts and business models emerges around the idea of LBSNs. In the months between February and May 2010 alone, their number doubled to over one hundred [19, 57].

Among the more mature networks is Google Latitude which builds on the concept of (automated) tracking to let people share their current location with friends. Further, in early 2010, Google introduced Buzz which allows users to share geo-tagged status updates as well as other media content. In this sense, Buzz is similar to Twitter which launched a Geo-API in late 2009. Their API lets third parties build services on the basis of geo-tagged content posted on Twitter.

Foursquare and Gowalla apply the concept of check-ins which gives users control over when and where to reveal their location. The overall concept of the two competitors is the same. As people

check-in, they can leave notes and recommendations and see the ones other users left. For example, in a restaurant, users may leave a note on how they liked the food or suggest a nearby bar. Meanwhile, people learn about places their friends frequent. Furthermore, they include a game element as people collect points and earn badges or pins for checking in to locations or accomplishing tasks. Gowalla even lets people collect and trade virtual goods.

As of August 2010 Facebook integrates geo-location features into their platform [18], potentially bringing half a billion people [17] to location-based social networking. As on Brightkite, Facebook users can now share information about locations and broadcast their whereabouts to friends and to strangers who are at the same venue. Further, Facebook allows for the integration with other LBSNs and will also provide third-party applications to access users' location-data (upon the user's approval).

3.2 Brightkite

Launched in beta in April 2008, Brightkite is one of the largest LBSNs with approximately two million active users [35]. Through web interface, email, SMS and mobile applications, users check-in to locations to reveal their location at varying levels of granularity, from actual address to city to country to "somewhere in the world." Through the same channels, messages as well as pictures can be published. These are then linked to the location to which one checked-in most recently.

Users have profile pages showing a small picture and information such as name and gender. Further, profile pages feature the user's check-in history as well as the messages and pictures attached to the locations. Other users can comment on check-ins, messages and pictures.

In October 2009, Brightkite introduced an update along with several major changes. Most importantly, rather than following the example of Facebook, which builds a closed world that offers users no value outside their social graph, the concept of one-way friendships was introduced. Like on Twitter, one can follow other people's activities without their approval. This way, Brightkite creates a public space that provides a multitude of user-generated content connected to both people and specific locations. However, users may make individual check-ins, messages and pictures available only to their friends. In addition, Brightkite lets users cross-share check-ins, messages and geo-tagged pictures with Facebook, Twitter and Flickr.

4. DATA ANALYSIS

4.1 Methodology

The qualitative data obtained from our study of early adopters reflect people's "lived experience" and is therefore "fundamentally well suited for locating the meanings people place on the events, processes, and structures of their lives […] and for connecting these meanings to the social world around them" [46, p. 10]. Data analysis took place by using the open-coding [3] and content analysis methodology [45].

We choose these methods due to their ability to analyze data systematically and obtain a comprehensive view on underlying attitudes and behavioral factors of LBSN usage. We justify our inductive approach with the absence of research among actual users of modern LBSNs as described in the section on related works.

The data collection took place between December 2009 and January 2010. Brightkite users were asked to fill out a standardized online questionnaire. Participants were recruited through different channels within the Brightkite system. For the most part, users received a survey link via direct personal-message enquiries. In addition, some of the most active users were asked to post the survey link to recruit participants among their peers. Besides demographics and frequency of use, participants answered five open-ended questions using free text entry:

1. Why do you use Brightkite?
2. What reasons would you name to convince someone to use Brightkite?
3. In which situations do you use Brightkite?
4. What are your concerns about using Brightkite?
5. What reasons would you name to talk someone out of using Brightkite?

The first two questions aim at identifying drivers of LBSN usage. As we intend to explore a multitude of uses and gratifications, in addition to asking for personal reasons to use the system, we also checked for reasons that could drive other people to adopt LBSNs. This approach has also allowed us to diminish social desirability bias in the responses. The third question intends to find out when and where people access LBSNs, i.e. the context of using the system. The fourth and fifth questions were designed to identify possible inhibitors of using LBSNs.

63 users completed the questionnaire. The sample comprises 34 males and 29 females with a mean age of 33 years. Overall, participants were active users of the system. 59 participants stated they would use the system at least several times a week, most of them even several times a day (n=43).

Another method to explore uses and implications of emerging technologies is the analysis of focus groups where people discuss a topic based on stimuli and questions presented by a moderator. We chose our approach over focus groups as it allows us to grasp diverse insights from a larger number of LSBN users from various places around the globe.

4.2 Coding Procedure

Following [46], our analysis was performed in three concurrent flows of activity. First, we reduced the data by applying the procedure of open-coding [3, 14]. We derived conceptual categories by comparing and contrasting similar incidents and phenomena as articulated by participants regarding their usage of the LBSN in question. Second, the reduced data was displayed in the form of tables (see Table 1, 2 and 3) and diagrams to reveal overlaps and relative importance of categories. Third, based on the data displays, preliminary conclusions on the participants' use of the LBSN in question were drawn. If conclusions could not be verified based on the displays and original data set, categories were reevaluated.

Finally, we extracted a total number of 400 relevant quotations from the data and derived 22 conceptual categories with respect to reasons for using LBSNs, context of using LBSNs, and concerns

of using LBSNs. The quotations were assigned to the coding categories by two independent coders. Inter-coder reliability constituted 0.760 (p-value < 0.000) suggesting a high level of agreement between the coders [37].

Sometimes, participants mentioned the same category multiple times within as well as across questions. In fact, we expected redundancies for both drivers and inhibitors as we intentionally asked congeneric questions pairwise. Therefore, if a category was assigned multiple times throughout the answers of one participant, it was counted only once. Consequently, the count of a category (i.e. "n") equals the number of participants who mentioned it.

4.3 Results

(1) Reasons and concerns to use the system are closely related to (2) the system's location-sensitive features. Therefore, building on preliminary results [58], we examine both research questions concurrently.

4.3.1 Context

Participants use the system in their free time (17%) and alongside daily routines (38%) such as being "in the office" or "while traveling with public transportation." Meanwhile, the vast majority of participants stated to use the system on occasions like in "situations that have a significant impact on me" or when they feel the urge to communicate something "simply noteworthy" or "feel like venting."

Mostly, occasions are related to locations and refer to being in an "unusual location," "interesting places" or when "I see something eye catching or am in a place I find interesting." In particular, participants indicated to use the system "at new locations" like when being "in a new city" or "somewhere new for definite."

This shows that the LSBN's mobile aspect leads people to share more "me now" information compared to OSNs where people usually reflect on an occasion after it happened.

Categories describing the context of system usage are not mutually exclusive. In fact, 24% of the participants explicitly stated to use the system in "all situations," "everywhere possible" or "pretty much […] 24/7."

Table 1. Context

Category	n	
On Occasions	43	68%
Daily Routines	24	38%
All Situations	15	24%
Free Time	11	17%

4.3.2 Reason to Use

Meeting People

Meeting people emerged as central driver for participants to use the system. In fact, 44% of all participants use the system to connect to "new people who share the same interests and hobbies" or "meet new folks." For another participant the system is a way to "overcome my fear of meeting new people."

This contrasts interaction patterns among people on OSNs which were found to reflect or deepen relationships with users that people have met socially offline [36].

Participants did not always make it clear whether they were referring to real world encounters or meeting purely online. Half of all participants who indicated to use the system to meet people explicitly mentioned location-related factors. In line with our findings on the role of location in the creation of a community-feeling proximity did not emerge as a requirement to meet people, e.g. a participant "met some good friends from around the world."

However, a number of participants actually make use of the system's location-sensitivity to meet people nearby as "the concept of checking in to a place" enables users "to meet people nearby using the same service." For one participant it is "a great way to meet people in your area you may never otherwise meet" while for another one it "is super easy to find people close (nearby) to you and to engage them (drinks, bar, even a date!)" Others meet people through the system when attending events: "You go to a concert, you find out someone you know is there."

Sharing Information

Interesting patterns of usage emerge around the exchange of information. Sharing information in terms of updating status, writing messages, and commenting is regarded as a major reason to use the system. Participants share information as it is common practice on traditional OSNs, for instance to "capture thoughts and moments in text."

However, messages shared through the system are pinned to the sender's location and participants understand the value of the additional context to messages as they appreciate the "added dimension to sharing with my social network." The majority of participants who mentioned information sharing as a reason to use the system explicitly referred to location-aspects of the information.

In that sense, the location-context establishes an "information ground" where "information sharing emerges as a byproduct of social interaction" [15, p.2]. Participants share "information about locations with other people and friends" or leave "comments to locations checked-in." Participants seem to derive pleasure from writing messages knowing they are connected to their current location: "posting a lot of different things at locations possible" or "it's so interesting to post locations."

Partly, this motivation may be explained by the "performative function of saying that one is aligning oneself with a particular venue and its branding" [27, p. 349]. For example, one participant shares information when "in a place I find interesting or want to brag about."

Community

Even though one can selectively hide check-ins and messages from public access, the system is basically open. Yet, users appear to have a close community-feeling towards other members of the network. Participants praise "a great community, with cool people" who "make the site great" or "wonderful." Another participant names "the community itself, it's an amazing fun and caring crowd, from all over the world." The community-sense seems to emerge despite the network's public nature and geographical dispersion of its users. This is consistent with previous findings from research on LBSNs indicating that physical distance between regions does not necessarily correlate with how closely these regions' populations are connected [40].

However, location does play a role for the formation of the community-sense among users of the system. It is less geographical proximity that creates this sense, but rather the joy of finding "people all over the country and the world who are some of the nicest people I've NEVER met." One reason why participants even seem to appreciate geographical distance may lay in an interest to explore faraway places in order "to get a small taste of other people and cultures from around the globe."

Weaver and Morrison [64] describe the mass adoption of OSNs as "evolution in human social interaction" (p. 97). In this way, LBSN usage - as presented by the participants of our study - may be interpreted as "the logical extension of our human tendencies toward togetherness" (p. 100).

Keeping In Touch

In line with the research on OSNs without location-features [34, 36], participants emphasized the use of the system to keep in touch with people, e.g. to "catch up with friends" by "sharing just-in-time activities with friends." One participant even uses the system "to share my life with my friends."

Meanwhile, the system's location-features allow others to be in the loop "where friends go" or "friends currently are." Another participant finds it "nice (in a non-stalker way) to know where people are and see what they're up to."

Sharing Photographs

One out of four participants uses the system to share "photographs of things around me wherever I am." In particular, participants make use of the system's feature to geo-tag pictures and find it "great being able to add geographic metadata to my photographs so easily." Pinning photos to particular venues such as restaurants was mentioned by participants who find it "fun to post food pictures to specific locations."

Learning

As users generate a multitude of location-related content, a ubiquitous body of information is created around places. Even though some users' motivation to publish photos or textual information may be to show off, they can provide real value. For instance, participants "really enjoy seeing pictures of food from restaurants nearby" and utilize them for "learning about restaurants, etc., on the fly." In general, the system seems to provide social capital as it lets participants "learn from others" or "getting to know your city better."

Another participant uses the system to "learn more about what's happening out there. You may not have heard it on the news." The vast amount of information on the system is public and related to locations. This enables users "to learn about an area" as well as to "discover new places and see what people may think (i.e.: quick opinions of a restaurant or views from a hotel)." In particular, if "You come to a strange city, you find photos of places you'd like to see" and one can "search for different local places."

Fun

It is "very exciting" for participants to use the system. Fun was mentioned by 21% of the participants and is an important driver for participants to use the system who "use the service mostly to keep myself entertained during the day." One participant puts it in a nutshell: "It's social and fun."

Diary

In line with findings from studies on another mobile social network [27], participants appreciated the system's geo-tagging features "to track places you've been to" and keep "a sort of spatial diary." The system locates people automatically and lets them check-in and post notes or pictures with as little effort as the click of a button. This makes it very easy to keep "a record of where I've been and what I've been." Due to the integrated geo-tagging functionality, using the system to communicate and share information implicitly creates "a journal of my life, and it's great because not only am I able to log what I do on a daily basis but where I do them!"

Getting Response

Furthermore, participants benefit from using the system as they "see what people think of what you post" and are "hearing the responses." Not only, does the system provide information about locations, people also stated to utilize the community itself as a handy resource for knowledge on demand as participants use the system if they "need an answer to a question" or "to solicit feedback when you have a question."

Other Motivation

46% of the participants also articulated other motivation to use the system.

22% of the participants make use of the system's feature to cross-share information across multiple platforms. Participants "like to be able to post updates to Twitter and Facebook at the same time, as well as upload photos to both plus Flickr." That way, some participants use the system less for its network, but rather as a convenient tool to access other networks.

In fact, ease of use of the system's mobile application and website combined with other technology/design-related reasons were emphasized by 44% of the participants.

Other motivation to use the system includes interest in the technology itself. One participants use the system "because location-based social networking is the future and I am a trend setter."

Table 2. Reasons to Use

Category	n	
Meeting People	28	44%
Sharing Information	28	44%
Community	24	38%
Keeping In Touch	20	32%
Ease of Use	19	30%
Sharing Photographs	16	25%
Learning	14	22%
Cross-Sharing	14	22%
Fun	13	21%
Technology/Design-Related Reasons	9	14%
Diary	7	11%
Getting Response	5	8%
Other Motivation	29	46%

4.3.3 Concerns

Privacy

Privacy emerged as the single most important concern among participants. Related concerns were mentioned by 46% of all participants. For example, one participant said that the system "is nothing for people who are rather conservative regarding privacy" while another one would be concerned about "a female possibly posting sensitive information about herself."

Whereas privacy is an inherent issue on OSNs [e.g., 16, 25, 28, 38, 47, 56, 60], the LSBN's location-features raise particular privacy concerns. In fact, almost half of all participants who express privacy concerns are specifically concerned about risks related to the disclosure of location data as participants mentioned "geo-location privacy issues." Another participant recognizes that "Once in a while I get a twinge that someone could use this to figure out when you're on vacation, and loot your home."

Overall, most participants seem to be aware of privacy issues. Meanwhile, they stress personal responsibility of the user. One participant is aware that the system "maps out where you go... BUT" that it would be "really up to the user as to how specific the location is." Another participant is just as aware of potential privacy issues related to disclosing "location of house or school and whereabouts from your home" and recommends that "if you are concerned with privacy then be certain about those settings and controls."

Stalking

Closely related to privacy concerns, participants expressed fear of "stalkers, creeps, and such" or "crazy stalkers." Participants are worried that the system "could be good as a stalking tool" and "could help stalkers." Analogue to their attitude towards privacy on the system, participants are aware that "people with less than honorable intentions" could use "information they may find to stalk someone." But again, they refer to users' personal responsibility to share location data. One participant asks: "…but it's not really stalking if you choose to be found, right?"

Other Concerns

A few participants expressed fears of identity theft [8] and the "trustworthiness of other users." Other concerns dealt with critical mass of the system, e.g.: "small size of the community is the biggest problem."

LBSNs are still young and the technology has not reached maturity. Therefore, technology-related concerns regarding the mobile application, website, connectivity, and overall usability are frequently reported by participants. Meanwhile, a number of participants expressed management-related concerns. For example, participants complain about the company being "unresponsive to user concerns" as well as "horrible customer relations."

No Concerns

27% the participants explicitly stated to have "no concerns at all." The commitment to publish personal and location-related information seems to be a conscious decision as participants seem to be aware of potential privacy and stalking issues. However, they "don't foresee any real issues […] as long as you are sensible with what you divulge." Other participants have "no concerns. I can keep things as private as I want to" or are just "not the overly paranoid type."

Table 3. Concerns

Category	n	
Privacy Concerns	29	46%
Other Concerns, Technology	16	25%
Stalking	14	22%
Other Concerns	11	17%
Other Concerns, Management	7	11%
No Concerns	17	27%

5. CONCLUSION

5.1 Practical Implications

5.1.1 Service Providers

The LBSN market is getting highly competitive as a large number of LBSNs emerges quickly and major players like Google, Twitter, and Facebook add location-features to their products. Many LBSNs turn out to be "me too" applications and of Schapsis' [57] list of 116 LBSNs, nearly 10% already shut down. Service providers need to develop concepts with a distinctive value proposition to attract users and foster usage. Here, our investigation of early adopters provides some hints.

Generally, our study showed that participants derive real value from location-sensitive features and the contextual information they imply. Therefore, OSN providers can improve the user experience by providing ways to integrate location-context. Already, users of OSNs deal with the problem of information overload [30]. Building a bridge between the real, physical world and the vast amount of user-generated information on OSNs facilitates innovative algorithms to allocate and filter relevant information. For example, Facebook developed the "EdgeRank" algorithm to filter users' news feed based on how long ago some piece of content was created, an affinity score between the viewing user and the content creator, as well as the content type. Incorporating the viewer's as well as the content creation's location to the formula can increase the news feed relevance greatly, for example by showing news about places around a user, or pictures of places a user used to frequent. Exactly those use cases where important drivers for participants of our study to use LBSNs.

Integrating our research with Joinson's [34] findings on motives and uses of Facebook lets us derive further extensions that would improve user experience and engagement on OSNs: (1) "Keeping in touch" gets more interesting when people know from where friends write their updates, (2) "social surveillance" is brought to the next level when people can keep track of friends whereabouts, (3) people can "Re-acquire lost contacts" when they happen to be in the same neighborhood, (4) "Communication" can be initiated based on geographic proximity, and (5) geo-tagged "Photographs" let people share pictures based on where they were taken.

Users of the studied LBSN showed a strong community-sense. Interestingly, this feeling is not necessarily created through physical proximity. Rather, our results indicate that awareness of other people's whereabouts creates an emotional attachment - regardless of the geographic distance between them. That way,

OSNs can create customer loyalty by implementing location-features.

Meanwhile, meeting people is an important reason to use the system. For one thing, users connect with people online like on traditional OSNs without location-features. However, our study shows that people do also use location-based information to physically meet friends as well as new people.

In this context, LBSN providers need to consider privacy and security issues. For example, 'www.pleaserobme.com' launched as an online service which fetches people's check-ins from various LBSNs claiming to provide burglars with tips where to break in. While this is an attempt to bring attention to the threats connected to the disclosure of location data, we find that users are very well aware of the dangers and consciously weigh them against the benefits they obtain on LBSNs. Therefore, LBSN providers are advised to deal honestly with threats related to location-disclosure and offer granular privacy settings.

5.1.2 Local Businesses

Sharing information, getting response to questions, and gaining knowledge from user-generated content emerged as major usage categories and demonstrate that people derive real benefits on LBSNs. The current study shows that people particularly use location-based information to find out about nearby places and learn about new locations they travel to. Participants made it clear that they enjoy sharing information on occasions, especially when they visit a new place. When customers leave a comment about a location it is broadcasted to their friends and made available to everyone who wants to learn about places in the respective area. This is a unique opportunity for businesses to gain recognition throughout their customers' social graph and attract clientele in the vicinity. Given the power of social networks to spread news virally with enormous speed and reach [e.g., 35], local businesses are advised to provide incentives for customers to check-in and share their experiences on LBSNs.

Meanwhile, as consumers share their experiences with the products and services from different venues local businesses may investigate customer satisfaction by extending the classical approach [26] towards a location-aware element. Not only can businesses gain insights about how their customers experience their place, they can directly react to feedback and complaints. In addition, businesses may not only learn about their clientele, but also find out about potential customers who check-in to places nearby or visit competitors. Therefore, customer oriented businesses should monitor LBSNs and have mechanisms in place to react quickly to feedback regarding their own matters as well as incidents in the neighborhood.

5.2 Theory Implications

Our results on the use of LBSNs indicate the relevance of some constructs that are not captured by traditional technology acceptance theory.

There is a multitude of (potential) location-sharing applications [20], e.g. for the monitoring of employees [29] or patients [66], government surveillance [59], locating family members for safety [9, 12, 44], locating students at school [23] or socializing with friends [48, 51]. Certainly, an individual's attitude towards location-sharing differs subject to the context of the application. Therefore, Mallat et al. [43] argue that traditional technology acceptance theory may be extended with respect to the specific nature of the studied technology and show the significance of mobility and usage context on an individual's decision to use a mobile commerce application. Our findings underline the relevance of these constructs with respect to LBSN usage.

Further, we recognize the collaborative nature of LBSNs and find that privacy-related constructs play an important role for individual adoption. Therefore, the inclusion of privacy-related constructs may be necessary to explain LBSN adoption sufficiently. Krasnova et al. [32] already showed the significant influence of privacy on people's decision to use OSNs. While privacy issues on OSNs have been investigated from various angles [e.g., 16, 25, 28, 38, 47, 56, 60], LBSNs raise particular concerns with respect to the disclosure of location-information. Our results show that awareness of privacy risks and potential stalking issues can go hand in hand with LBSN usage and, consequently, location-sharing. This confirms the results from a hypothetical study on the usage of LBS [6].

It will be interesting to investigate interdependencies between context of use and privacy concerns. Already, studies showed that in situations of emergency, individuals have a higher willingness to give up some of their privacy [2, 44]. Our findings hint at Context as a potential moderator also in all-day situations. For example, participants who stated to use LBSNs to inform themselves about places when they are on a holiday may perceive the benefits of finding information about locations less useful when they are in their hometown.

Theories claiming the Internet would diminish social capital by drawing people away from family and friends and reducing interest in the local community [6, 49] are objected by our findings on LBSNs. Generally, the value of a social network rises as the number of its members increases [31]. As LBSNs continue to grow, fewer people will miss out on chances to physically meet friends or new people. At the same time, more auxiliary information is aggregated around locations. In this way, LBSNs are likely to play an increasingly important role in people's everyday life.

5.3 Limitations and Outlook

Since LBSNs are still in the early stages of development, our sample comprises only early adopters. The participants of our study illustrate how they integrate LBSNs into their lives and why they do so. However, usage patterns of early users may diverge from the way the majority deploys a technology later on (see [52]). On the other hand, various studies in the domain of the SST literature [e.g., 7, 42, 65], as well as the Diffusion of Innovations Theory [25] suggest that early adopters' usage shape future patterns of use.

LBSNs' tremendous growth numbers [53] suggest that the "early majority" [55] is about to follow and adopt. Future research should verify our findings on a broader scale and explore how people use LBSNs as the density of users and available location-based information increases.

Furthermore, the current study examined one particular concept of an LBSN. Research on different classes of LBSNs may reveal other uses and gratifications. For instance, examining gaming elements and the concept of virtual items on LBSNs also provides a rich field for future research.

6. REFERENCES

[1] ABIresearch. 2008. *Location-based mobile social networking will generate global revenues of $3.3 billion by 2013*. http://www.abiresearch.com/press/1204-Location-based+Mobile+Social+Networking+Will+Generate+Global+Revenues+of+%243.3+Billion+by+2013. Retrieved December 15, 2009.

[2] Aloudat, A., Michael, K. and Abbas, R. 2009. Location-Based Services for Emergency Management: A Multi-Stakeholder Perspective. In *Proceedings of the 8th International Conference on Mobile Business* (Dalian, China, 2009). ICMB 2009. DOI= http://dx.doi.org/10.1109/ICMB.2009.32.

[3] Bailey, C. A. 2007. *A guide to qualitative field research*. 2nd ed., Thousand Oaks, CA, Sage Publications.

[4] Barkhuus, L. 2004. Privacy in Location-Based Services, Concern vs. Coolness. *HCI 2004 workshop: Location System Privacy and Control* (Glasgow, UK, 2004).

[5] Barkhuus, L. and Brown, B. 2008. From awareness to repartee: sharing location within social groups. In *Proceedings of the 26th annual SIGCHI conference on Human factors in computing systems* (Florence, Italy, 2008). DOI= http://dx.doi.org/10.1145/1357054.1357134.

[6] Barkhuus, L. and Dey, A. 2003. Location-Based Services for Mobile Telephony: a study of user's privacy concerns. In *Proceedings of the INTERACT, 9th IFIP TC13 International Conference on Human-Computer Interaction* (Zurich, Switzerland, 2003).

[7] Bijker, W. E., Hughes, T. P., and Pinch, T. J. 1987. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA, MIT Press.

[8] Bilge, L., Strufe, T., Balzarotti, D. and Kirda, E. 2009. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web* (Madrid, Spain, 2009). DOI= http://dx.doi.org/10.1145/1526709.1526784.

[9] Brown, B. and Taylor, A. 2007. Locating Family Values: A Field Trial of the Whereabouts Clock. *Lecture Notes in Computer Science*, 2007, 4717/2007, 354-371. DOI= http://dx.doi.org/10.1007/978-3-540-74853-3_21.

[10] Buber, M. 1958. *I and thou*. 2nd ed., Edinburgh, T&T Clark.

[11] Chen, G. and Rahman, F. 2008. Analyzing privacy designs of mobile social networking applications. In *Proceedings of the IEEE/IFIP International Symposium on Trust, Security and Privacy for Pervasive Applications*. TSP 2008 (Shanghai, China, December 2008). DOI= http://dx.doi.org/10.1109/EUC.2008.156.

[12] Chou, L.-D., Lai, N.-H., Chen, Y.-W., Chang, Y.-L., Huang, L.-F., Chiang, W.-L., Chiu, H.-Y. and Yang, J.-Y. 2008. Management of mobile social network services for families with Developmental Delay Children. 10th *International Conference on e-health Networking, Applications and Services*. HealthCom 2008 (Singapore, 2008). DOI= http://dx.doi.org/10.1109/HEALTH.2008.4600115.

[13] Consolvo, S. and Smith, I. E. 2005. Location disclosure to social relations: why, when & what people want to share. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (Portland, Oregon, 2005). DOI= http://dx.doi.org/10.1145/1054972.1054985.

[14] Corbin, J. M. and Strauss, A. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology* 13, 1, 3-21. DOI= http://dx.doi.org/10.1007/BF00988593.

[15] Counts, S. and Fisher, K. E. 2008. Mobile Social Networking: An Information Grounds Perspective. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (Hawaii, 2008). DOI= http://dx.doi.org/10.1109/HICSS.2008.320.

[16] Dwyer, C., Hiltz, S. and Passerini, K. 2007. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In *Proceedings of the 13th Americas Conference on Information Systems*. AMCIS 2007.

[17] Facebook. 2010. *Facebook Official Statistics*. http://www.facebook.com/press/info.php?statistics. Retrieved February 16, 2010.

[18] Facebook Blog. 2010. *Who, what, when, and now… where*. http://blog.facebook.com/blog.php?post=418175202130. Retrieved August 20, 2010.

[19] Fusco, S. J., Michael, K., Michael, M. G. and Abbas, R. 2010. Exploring the Social Implications of Location Based Social Networking: An inquiry into the perceived positive and negative impacts of using LBSN between friends. *International Conference on Mobile Business* (Athens, Greece, 2010). DOI= http://dx.doi.org/10.1109/ICMB-GMR.2010.35.

[20] Fusco, S. J., Michael, K., and Michael, M. G. 2010. Using a Social Informatics Framework to Study the Effects of Location-Based Social Networking on Relationships between People: A Review of Literature. *IEEE International Symposium on Technology and Society*. ISTAS10. IEEE. DOI= http://dx.doi.org/10.1109/ISTAS.2010.5514641.

[21] Gadzheva, M. 2007. Privacy concerns pertaining to location-based services. *International Journal of Intercultural Information Management* 1 (2007). DOI= http://dx.doi.org/10.1504/IJIIM.2007.014370.

[22] Garfinkel, S. L. and Juels, A. 2005. RFID Privacy: An Overview of Problems and Proposed Solutions. *IEEE Security and Privacy* 3 (2005), 34-43. DOI= http://dx.doi.org/10.1109/MSP.2005.78.

[23] Glasser, D. J. and Goodman, K. W. 2007. Chips, tags and scanners: Ethical challenges for radio frequency identification. *Ethics and Information Technology* 9 (2007), 101-109. DOI= http://dx.doi.org/10.1007/s10676-006-9124-0.

[24] Grandhi, S. A., Jones, Q. and Karam, S. 2005. Sharing the big apple: a survey study of people, place and locatability. In *CHI 05 extended abstracts on Human factors in computing systems* (Portland, OR, USA, 2005). DOI= http://dx.doi.org/10.1145/1056808.1056928.

[25] Gross, R. and Acquisti, A. 2005. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the Workshop on Privacy in Electronic Society* (Virginia, USA, 2005). DOI= http://dx.doi.org/10.1145/1102199.1102214.

[26] Hill, N., Roche, G. and Allen, R. 2007. *Customer satisfaction: the customer experience through the customer's eyes*. Cogent Publishing, London, UK.

[27] Humphreys, L. 2007. Mobile social networks and social practice: A case study of Dodgeball. *Journal of Computer-Mediated Communication* 13, 1, 341-360. DOI= http://dx.doi.org/10.1111/j.1083-6101.2007.00399.x.

[28] Katherine, S. and Richter, L. H. 2008. Strategies and struggles with privacy in an online social networking community. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction* (Liverpool, UK, 2008).

[29] Kaupins, G. and Minch, R. 2006. Legal and ethical implications of employee location monitoring. *International Journal of Technology and Human Interaction* 2 (2006), 16-20.

[30] Koroleva, K., Krasnova, H. and Günther, O. 2010. 'STOP SPAMMING ME!' - Exploring Information Overload on Facebook. In *Proceedings of the 16th Americas Conference on Information Systems.* AMCIS 2010 (Lima, Peru, August 2010).

[31] Krasnova, H., Kolesnikova, E. and Günther, O. 2010. It won't happen to me!: Self-disclosure in Online Social Networks. In *Proceedings of the 15th Americas Conference on Information Systems*. AMCIS 2009 (San Francisco, CA, USA).

[32] Krasnova, H., Spiekermann, S., Koroleva, K. and Hildebrand, T. 2010. Online Social Networks: Why We Disclose. *Journal of Information Technology* 25, 109-125. DOI= http://dx.doi.org/10.1057/jit.2010.6.

[33] Kwak, H., Lee, C., Park, H. and Moon, S. 2010. What is Twitter, a Social Network or a News Media?. In *Proceedings of the 19th International World Wide Web (WWW) Conference* (Raleigh, NC, USA, April 26-30). DOI= http://dx.doi.org/10.1145/1772690.1772751.

[34] Joinson, A. N. 2008. 'Looking at', 'Looking up' or 'Keeping up with' people? Motives and uses of Facebook. In *Proceedings of the Conference on Human Factors in Computing Systems.* CHI 2008, ACM (Florence, Italy, April 5-10). DOI= http://dx.doi.org/10.1145/1357054.1357213.

[35] LA Times. 2009. *On Brightkite, ladies' night never ends*. http://latimesblogs.latimes.com/technology/2009/11/brightkite-girls.html. Retrieved February 27, 2010.

[36] Lampe, C. Ellison, N. B. and Steinfield, C. 2008. Changes in use and perception of Facebook. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work* (San Diego, CA, 2008). DOI= http://dx.doi.org/10.1145/1460563.1460675.

[37] Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1977), 159-174.

[38] Levin, A. and Foster, M. 2008. *The Next Digital Divide: Online Social Network Privacy*. Ryerson University, Ted Rogers School of Management, Privacy and Cyber Crime Institute.

[39] Li, N. and Chen, G. 2009. Analysis of a location-based social network. *International Symposium on Social Intelligence and Networking.* SIN (Vancouver, Canada. August 2009). DOI= http://doi.ieeecomputersociety.org/10.1109/CSE.2009.98.

[40] Li, N. and Chen, G. 2009. Geographic community analysis of mobile social network. *Workshop on Social Networks, Applications, and Systems*. SNAS (Boston, MA, USA).

[41] Lockton, V. and Rosenberg, R. 2005. RFID: The Next Serious Threat to Privacy. *Ethics and Information Technology* 7 (2005), 221-231. DOI= http://dx.doi.org/10.1007/s10676-006-0014-2.

[42] MacKenzie, D. and Wajcman, J. 2002. *The Social Shaping of Technology*. 2nd ed., Buckingham, UK, Open University Press.

[43] Mallat, N., Rossi, M., Tuunainen, V. K. and Öörni, A. 2006. The Impact of Use Situation and Mobility on the Acceptance of Mobile Ticketing Services. In *Proceedings of the 39th Hawaii International Conference on System Sciences HICSS* (Hawaii, USA). DOI= http://dx.doi.org/10.1109/HICSS.2006.472.

[44] Masters, A. and Michael, K. 2007. Lend me your arms: The use and implications of humancentric RFID. *Electronic Commerce Research Applications* 6 (2007), 29-39. DOI= http://dx.doi.org/10.1016/j.elerap.2006.04.008.

[45] Mayring, P. 2000. Qualitative content analysis. *Forum: Qualitative Social Research* 1, 2, 1-10.

[46] Miles, M. B. and Huberman, A. M. 1994. *Qualitative data analysis*. 2nd ed., Thousand Oaks, CA, Sage Publications.

[47] Mohammad, M. and Paul, C. O. 2008. Privacy-enhanced sharing of personal content on the web. In *Proceeding of the 17th international conference on World Wide Web* (Beijing, China, 2008). DOI= http://dx.doi.org/10.1145/1367497.1367564.

[48] Nan, L. and Guanling, C. 2009. Analysis of a Location-Based Social Network. *International Conference on Computational Science and Engineering* 2009. DOI= http://doi.ieeecomputersociety.org/10.1109/CSE.2009.98.

[49] Nie, N. H. 2001. Sociability, Interpersonal Relations, and the Internet: Reconciling Conflicting Findings. *American Behavioral Scientist* 45, 3, 420-435. DOI= http://dx.doi.org/10.1177/00027640121957277.

[50] NielsonWire. 2010. *The State of Mobile Apps*. http://blog.nielsen.com/nielsenwire/online_mobile/the-state-of-mobile-apps. Retrieved June 3, 2010.

[51] Page, X. and Kobsa, A. 2009. The Circles of Latitude: Adoption and Usage of Location Tracking in Online Social Networking. *International Conference on Computational Science and Engineering* 2009. DOI= http://dx.doi.org/10.1109/CSE.2009.195.

[52] Preece, J. 2004. Etiquette online: From nice to necessary. *Communications of the ACM*, 47, 4, 56-61. DOI= http://dx.doi.org/10.1145/975817.975845.

[53] ReadWriteWeb. 2010. *Location is hot: Foursquare traffic up 3x in 2 months*. http://www.readwriteweb.com/archives/location_is_hot_foursquare_traffic_up_3x_in_2_months.php. Retrieved February 27, 2010.

[54] Rezabakhsh B., Bornemann, D., Hansen, U. and Schrader, U. 2006. Consumer power: A comparison of the Old Economy and the Internet Economy. *Journal of Consumer Policy* 29 (2006), 3-36. DOI= http://dx.doi.org/10.1007/s10603-005-3307-7.

[55] Rogers, E. M. 1995. *Diffusion of innovations*. 4th ed., The Free Press, New York, NY, USA.

[56] Rosenblum, D. 2007. What Anyone Can Know: The Privacy Risks of Social Networking Sites. *IEEE Security & Privacy* 5 (2007), 40-49. DOI= http://doi.ieeecomputersociety.org/10.1109/MSP.2007.75.

[57] Schapsis, C. 2010. *Location Based Social Networks Links: A list of Location Based Social Networks*. http://bdnooz.com/lbsnlocation-based-social-networking-links. Retrieved August 20, 2010.

[58] Schöndienst, V., Dang-Xuan, L. and Günther, O. 2010. 'Checking-In' - Exploring the Usage of Location-based Social Networks. In *Proceedings of the 16th Americas Conference on Information Systems.* AMCIS 2010. Paper 445.

[59] Smith, G. D. 2006. Private eyes are watching you: with the implementation of the E-911 mandate, who will watch every move you make? (Telecommunications Act of 1996: Ten Years Later Symposium). *Federal Communications Law Journal* 58 (2006), 705-721.

[60] Snyder, J. and Carpenter, D. 2007. MySpace.com - A Social Networking Site and Social Contract Theory. *Information Systems Education Journal* 5 (2007), 3-11.

[61] Tsai, J. Y. and Kelley, P. 2009. Who's viewed you?: the impact of feedback in a mobile location-sharing application. In *Proceedings of the 27th international conference on Human factors in computing systems* (Boston, MA, USA, 2009). DOI= http://dx.doi.org/10.1145/1518701.1519005.

[62] Vihavainen, S. and Oulasvirta, A. 2009. "I can't lie anymore!": The implications of location automation for mobile social applications. *6th Annual International Mobile and Ubiquitous Systems: Networking & Services*. DOI= http://dx.doi.org/10.4108/ICST.MOBIQUITOUS2009.6847.

[63] Wall Street Journal. 2009. *Twitter's value is set at $1 billion*. http://online.wsj.com/article/SB125382643140938735.html. Retrieved February 27, 2010.

[64] Weaver, A. C. and Morrison, B. B. 2008. Social Networking. *Computer* 41 (2008), 97-100.

[65] Williams, R. and Edge. D. 1996. The social shaping of technology. *Research Policy* 25 (1996), 856-899. DOI= http://dx.doi.org/10.1016/0048-7333(96)00885-2.

[66] Xiao, Y. and Shen, B. 2006. Security and Privacy in RFID and application in telemedicine. *IEEE Communications Magazine* 44 (2006), 64-72. DOI= http://dx.doi.org/10.1109/MCOM.2006.1632651.

Innovationen im Mobile Government – Eine Analyse von Dienstattraktivitäten und Motivationen von deutschen Kommunen

Till J. Winkler
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Deutschland
till.winkler@wiwi.hu-berlin.de

Peter Ernst
Humboldt-Universität zu Berlin
Spandauer Straße 1
10178 Berlin, Deutschland
peter.ernst@student.hu-berlin.de

ABSTRACT (ZUSAMMENFASSUNG)

Städte und Gemeinden stehen unter einem immerwährenden Sparzwang. Gleichzeitig ergeben sich mit der wachsenden Beliebtheit des mobilen Internets innovative Anwendungsszenarien des Mobile Government. Diese Arbeit analysiert ein breites Spektrum an Mobile Government-Diensten aus der Sicht von kommunalen IT-Entscheidern und untersucht die organisationalen Faktoren, welche die wahrgenommene Attraktivität dieser Dienste beeinflussen. Hierzu wird auf Basis der Literatur sowie einer Serie von Interviews ein Modell bestehend aus drei Hauptdimensionen (*Effizienzstreben, Innovationsstreben* und *IT-Erfahrenheit)* abgeleitet. Wir testen das Modell empirisch mit Hilfe des Partial Least Square-Ansatzes in einer Erhebung mit 50 deutschen Städten und clustern die Fälle entsprechend der postulierten Dimensionen. Die Ergebnisse deuten darauf hin, dass sich die Motivationen in Städten mit hoher Ausprägung der wahrgenommenen Dienstattraktivität signifikant von denen mit geringer Ausprägung unterscheiden und dass die Innovatoren und IT-Erfahrenen unter den Städten in den nächsten drei Jahren deutlich stärker in mobile Dienste investieren werden als andere Kommunen.

Keywords (Schlüsselwörter)

Mobile Government, Mobile Dienste, Dienstattraktivitäten, Empirische Studie, Partial Least Squares, Clusteranalyse.

1. EINLEITUNG

Durch die fortschreitende technische Entwicklung ergibt sich ebenfalls ein positives Momentum für das Mobile Government. So wird erwartet, dass sich die Anzahl der Nutzer in Deutschland, die regelmäßig Internetfunktionen auf dem Mobiltelefon nutzen, von gegenwärtig 10 Millionen bis 2012 mehr als verdreifachen wird [4][6]. War Mobiles Government in der Vergangenheit im Wesentlichen beschränkt auf einfache Dienste wie SMS-Benachrichtigungen und isolierte Inselanwendungen [31] ergeben sich hieraus weiterführende Szenarien der Interaktion zwischen Verwaltung und Bürger, wie z.B. ortsbezogene Meldesysteme, mobile Bibliotheksausleihe oder intelligente Parkleitsysteme.

Im Rahmen dieser Arbeit soll die Begrifflichkeit von Mobile Government daher weitreichend gefasst und in Bezug auf verschiedene Formen von Mobilität verstanden werden, deren Umsetzung z.T. erhebliche Änderungen in Verwaltungsprozessen mit sich bringen kann. Zusätzliche Relevanz erhält dieses Thema durch solche IT-Vorhaben, die derzeit von nationaler Ebene auf Kommunen zukommen und eine enge organisatorische und technische Verzahnung mit dem Thema des mobilen Government aufweisen. Aktuelle Beispiele hierfür sind die Einführung der einheitlichen Behördenrufnummer 115, die Umsetzung der EU-Dienstleistungsrichtlinie, sowie die Einführung des elektronischen Personalausweises.

Zeitgleich zu derartigen strategischen Vorgaben sehen sich viele Kommunen einem massiven Kostendruck ausgesetzt – in Nordrhein-Westfalen stehen beispielsweise fast zwei Drittel der Kommunen im laufenden Jahr unter Haushaltssicherung [29]. Dies grenzt den Handlungsrahmen für innovative IT-Vorhaben deutlich ein und führt zu einer noch strengeren Überprüfung auf Wirtschaftlichkeit. Entsprechend leitet sich die Kernfrage dieser Arbeit wie folgt ab: Kann Mobile Government als weitere Ausbaustufe des E-Government generell einen Weg zu mehr Effizienz darstellen oder bleibt es ein Randthema für besonders kundenorientierte Verwaltungen, welche sich noch Raum für innovative Vorhaben freihalten können?

Während in der unternehmensbezogenen Forschung Akzeptanzuntersuchungen häufig aus der Perspektive des IT-Entscheiders durchgeführt wurden, z.B. [3], sieht die E-Government Literatur bisher überwiegend den Bürger im Fokus der Adoptions-Entscheidung. Hierdurch wurden häufig solche Mobile Government-Anwendungen ausgeblendet, welche einen verwaltungsinternen Nutzen erzielen. Diese Arbeit ist insofern neuartig, als dass sie stellvertretend für die gesamte Kommune den kommunalen IT-Entscheider als Impuls- und Ideengeber für den Einsatz neuer Technologien in den Mittelpunkt der Betrachtung stellt. Daraus resultiert, dass Mobile Government als ein Bündel von Anwendungen und Anwendungsclustern verstanden sowie eine Vielzahl von möglichen Anwendungsszenarien berücksichtigt wird. Dieser Artikel leistet einen wichtigen Erklärungsbeitrag, indem aufgezeigt wird 1.) welches die organisationalen Faktoren und Rahmenbedingungen sind, die eine Kommune bei der Einführung von Mobilen Diensten beeinflussen, 2.) wie stark diese Faktoren das empfundene Potenzial von Mobile Government beeinflussen und 3.) welche Auswirkung dies auf das Investitionsverhalten der Kommune hat.

Der Artikel gliedert sich in sechs Abschnitte. Nach einer theoretischen Fundierung anhand verwandter Literatur und Hypothesenbildung im zweiten Abschnitt werden die Ergebnisse einer qualitativen Voruntersuchung in Abschnitt drei vorgestellt. Hierauf aufbauend wird in Abschnitt vier die Durchführung der empirischen Studie erläutert und deren Ergebnisse in Abschnitt fünf eingehend erörtert. Der letzte Abschnitt fasst den Gesamtkontext noch einmal zusammen, diskutiert praktische Implikationen und gibt einen Ausblick auf hierauf aufbauende Forschungsthemen.

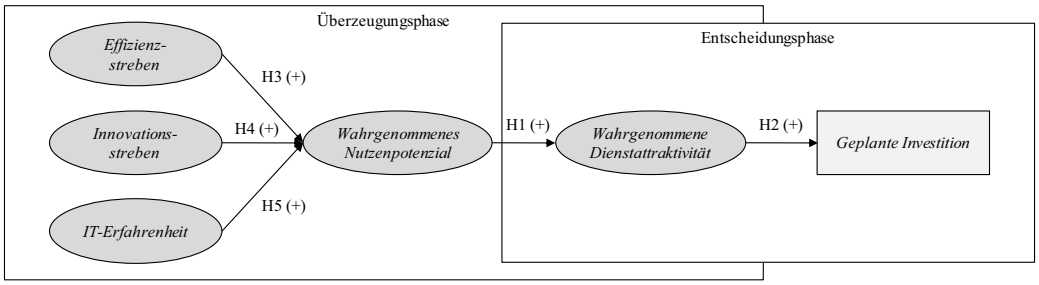


Abbildung 1 – Forschungsmodell

2. THEORETISCHE FUNDIERUNG UND HYPOTHESENBILDUNG

2.1 Formen des Mobile Government

Der Definition von Kuschu folgend [17] kann Mobile Government (M-Government) als eine Strategie und deren Umsetzung verstanden werden, die die Nutzung verschiedener Arten von drahtlosen und mobilen Technologien, Diensten, Anwendungen und Geräten umfasst, um für die beteiligten Akteure einen Mehrwert zu erzielen. Unter den Akteuren sind hierbei nicht nur die Bürger, sondern gleichsam Unternehmen sowie Angestellte der öffentlichen Behörden und Einrichtungen selbst zu verstehen. Gemäß der etablierten Terminologie kann daher auch im Bezug auf das Mobile Government die Unterscheidung in (mobiles) Government-to-Customer (G2C), Government-to-Business (G2B), und Government-to-Government (G2G) vorgenommen werden, wobei für die letztere Form auch der Begriff IEE (Internal Effectiveness and Efficiency) Anwendung findet [31].

Mobile Government kann somit als eine Erweiterung bzw. Untermenge des E-Government verstanden werden. Technologien, die im Mobile Government zum Tragen kommen, gehen allerdings weit über die Möglichkeiten der Telekommunikation hinaus: Drahtlose Netzwerke, Bluetooth, CCTV (Closed-Circuit Television), Ortungsverfahren, RFID (Radio Frequenz-Identifikation), biometrische Identifikation, Verkehrsüberwachung, Smart Cards und andere NFC-Anwendungen sind nur einige Beispiele für mobile (d.h. nicht-stationäre) Technologien, welche nicht notwendigerweise in Verbindung mit einem Mobiltelefon genutzt werden [16]. In einer Arbeit von Winkler et al. [32] werden acht Anwendungsbereiche für Mobile Government-Anwendungen im städtischen Kontext aufgezeigt, welche auf einem Kontinuum zwischen öffentlichem und privatwirtschaftlichem Nutzen angeordnet werden können. Hierunter fallen die Sektoren Verwaltung, Öffentliche Sicherheit, Bildung, Gesundheit, Verkehr, Umwelt und Infrastruktur, Tourismus und Kultur sowie Private und kommerzielle Anwendungen. Diese Einteilung soll im Rahmen dieser Arbeit als Framework zur Operationalisierung des Mobile Government-Begriffs verwendet werden.

2.2 Verhaltenstheoretische Fundierung

In der Literatur zu E-Government-Akzeptanz und -Nutzung kommen vorwiegend konzeptuelle Modelle zum Einsatz, die aus Rogers Innovations-Diffusionstheorie [27], Fishbein und Ajzens Theory of Reasoned Action [10] sowie Davis Technologie-Akzeptanz-Modell [7] abgeleitet werden. Letztere fokussiert auf konkrete Eigenschaften einer technologischen Innovation, wie Ease of Use und Perceived Usefulness, und ist daher im Rahmen einer ganzheitlichen Betrachtung von Mobile Government-Diensten nicht zur Anwendung geeignet. Wir wollen uns daher auf die beiden erstgenannten Ansätze beschränken.

Nach Rogers [27] ist der Innovationsbegriff definiert als die Akzeptanz einer Idee oder einer Praktik über die Zeit durch Individuen, die in Form von Kommunikationskanälen, sozialen Strukturen und einem System von Kultur und Werten verbunden sind. Der Prozess der Innovation durchläuft dabei fünf Phasen von Bewusstsein, Überzeugung, Entscheidung, Umsetzung und Bestätigung. Die vorliegende Arbeit unterstellt, dass sich die Entwicklungen im Bereich des mobilen Government zum gegenwärtigen Zeitpunkt, im Gegensatz zum E-Government, schwerpunktmäßig noch in den ersten drei Phasen befinden. Daher soll die Untersuchung auf die Phasen der Überzeugung und Entscheidung für (oder gegen) Mobile Government-Angebote abstellen.

Der Prozess der Entscheidung kann mit Hilfe der Theory of Reasoned Action (TRA) [10] genauer erklärt werden. Demnach ist die *Verhaltensabsicht* des Einsatzes einer bestimmten Innovation ein Mediator zwischen den (objektiven) *Einstellungen* des Entscheiders über diese Innovation und seinem tatsächlichen *Verhalten*. Trotz der Tatsache, dass die TRA ursprünglich zur Erklärung des Verhaltens einzelner Individuen entwickelt wurde und man annehmen darf, dass in der komplexen Struktur einer kommunalen Verwaltung mehrere Akteure an einer Entscheidung für oder gegen technologische Innovationen beteiligt sind, halten wir die TRA auch für eine Menge von Einstellungen und Verhaltensabsichten anwendbar, sofern diese hinreichend erfasst werden können. Damit folgen wir verschiedenen Beispielen in der Literatur wie z.B. [3]. Wir verzichten in diesem Fall jedoch auf den Einflussfaktor der *subjektiven Norm*, der für eine Gruppe von Individuen an Relevanz verliert [10].

Bezogen auf das Angebot mobiler Dienste in einer Kommune interpretieren wir die *Einstellung* der kommunalen Entscheider als ein aggregiertes Maß des *wahrgenommenen Nutzenpotenzials* mobiler Dienste, welches durch verschiedene organisationale Rahmenbedingungen in unterschiedlichem Maß beeinflusst wird (siehe Abschnitt 2.3). Die *Verhaltensabsicht* des Angebots mobiler Dienste soll operationalisiert werden als ein aggregiertes Maß der *wahrgenommenen Dienstattraktivität* konkreter Dienstangebote. Das *Verhalten* einer Verwaltung einen mobilen Dienst einzuführen, sollte letztendlich in der *geplanten Investition* reflektiert werden. Als Hypothesen können daher postuliert werden:

H1. Es besteht ein positiver Zusammenhang zwischen dem *wahrgenommenen Nutzenpotenzial* von Mobile Government insgesamt und der *wahrgenommenen Dienstattraktivität* konkreter Dienstangebote.

H2. Es besteht ein positiver Zusammenhang zwischen der *wahrgenommenen Dienstattraktivität* konkreter Dienstangebote und der *geplanten Investition* in mobile Dienste.

2.3 Einflussfaktoren des wahrgenommenen Nutzenpotenzials

Mobile Government ist erst seit kürzerer Zeit Gegenstand der Forschung, so dass bisher ein Mangel an empirischen Arbeiten in diesem Bereich festgestellt werden muss [17]. Zur Identifikation geeigneter Vorbedingungen und Einflussfaktoren des wahrgenommenen Nutzenpotenzials soll daher auf verwandte E-Government-Literatur und Arbeiten im Bereich der strategischen IT-Investitionsentscheidungen zurückgegriffen werden.

Im Bereich E-Government existiert eine Vielzahl empirischer Studien, die das Angebot und die Akzeptanz innovativer Verwaltungsdienste untersuchen. Für einen Überblick sei auf [25] verwiesen. Ein Großteil dieser Studien stellt den Bürger als Endnutzer in den Mittelpunkt. Untersucht werden individuelle Faktoren wie Vertrauen in das E-Government, IT-Erfahrung der Computernutzer (Skills) und IT-Affinität. Ebenso werden demografische Merkmale wie Geschlecht und Bildungsniveau als Schlüsselfaktoren betrachtet. Die Arbeit von Moon und Norris [21] stellt insofern eine Ausnahme dar, als dass sie organisationale Faktoren wie die Größe der Kommune (size) sowie die Form der Verwaltung (type of government) untersucht, um den unterschiedlichen Reifegrad der Einführung von E-Government-Diensten in US-amerikanischen Kommunen zu erklären [21].

Weitere organisationale Einflussfaktoren können der strategischen Information Systems- und IT-Alignment-Literatur entnommen werden. Die IT-Funktion in städtischen Verwaltungen kann insofern mit der in privatwirtschaftlichen Unternehmungen verglichen werden, als dass sie in Abhängigkeit der Geschäftsstrategie bestimmte Ziele und Anforderungen zu erfüllen hat. Somit gelten auch für die Einführung von Mobile Government-Diensten in Abhängigkeit von der Situation in der jeweiligen Kommune bestimmte strategische Vorgaben und Rahmenbedingungen. Zur Typologisierung derartiger IT Strategien haben sich in der Literatur mehrere, häufig dreigliedrige Modelle etabliert [8]. So leiten Sabherwal und Chan [28] aus der etablierten Miles and Snow Typologie für Geschäftsstrategien (Defenders, Analyzers, Prospectors) [20] im Bezug auf die IS Strategie die Attribute *IS for Efficiency*, *IS for Flexibility* sowie *IS for Comprehensiveness* ab. Zu einer ähnlichen Aufteilung gelangen später andere Autoren [28]. Wir kombinieren diese Trias von IT-strategischen Attributen mit den für das E- bzw. M-Government spezifischen Einflussfaktoren und postulieren die im Folgenden näher beschriebenen drei Dimensionen *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrenheit* als Einflussfaktoren der Mobile Government-Akzeptanz.

Effizienzstreben bezieht sich auf die Motivation der Kommune, durch die Einführung mobiler Dienste bestimmte Verwaltungsprozesse zu verbessern und letztendlich Kosten zu sparen. Je nach Haushaltslage der betrachteten Kommune, wird eine unterschiedlich starke Ausprägung dieser Dimension angenommen. Das *Effizienzstreben* spiegelt damit in erster Linie eine verwaltungsinterne, finanzielle Perspektive auf die Akzeptanz wider.

Innovationsstreben drückt den Grad aus, zu dem sich eine Kommune veranlasst sieht, ihr Dienstangebot im Sinne der Effektivität (aber nicht zwangsläufig der Effizienz) auszuweiten. Städte stehen nicht nur untereinander im Wettbewerb, sondern sehen sich auch verschiedensten Erwartungen ihrer Kunden ausgesetzt, vgl. [17]. Hier spielen ebenfalls die Akzeptanztreiber auf Ebene des Individuums, wie z.B. die zunehmende Technologie-Affinität, eine Rolle [25]. Neue, von mobiler Technologie unterstütze Dienstangebote können hier einen Beitrag leisten, ein Innovationsbedürfnis der Kunden zu befriedigen. Diese Dimension adressiert damit insbesondere die externe Perspektive.

IT-Erfahrung erfasst die IT-bezogenen Voraussetzungen des mobilen Government, die in einer Kommune gegeben sind. Hiermit ist sowohl die physische als auch die „softe“ Infrastruktur gemeint, wie Kushcu feststellt [17]. Genauso wichtig wie vorhandene Netzwerke und Systeme sind demnach institutionelle Arrangements und eine technologische Vision für das Mobile Government. Dies deckt sich mit der Erkenntnis von Tornatzky und Klein [30], dass Kompatibilität zu den vorhandenen Strukturen für eine Vielzahl von Innovationstypen eine entscheidende Bedeutung hat. Die hieraus entstehenden Synergien sind damit vor allem für solche Kommunen relevant, die bereits substanzielle Investitionen in E-Government Implementierungen getätigt haben [17]. Diese Dimension repräsentiert somit insbesondere die technologisch-organisationale Perspektive.

Hieraus ergeben sich die folgenden drei Hypothesen, die in dem in Abbildung 1 dargestellten Forschungsmodell zusammengefasst werden.

H3. Es besteht ein positiver Zusammenhang zwischen dem *Effizienzstreben* einer Kommune und dem *wahrgenommenen Nutzenpotenzial* von Mobile Government-Diensten.

H4. Es besteht ein positiver Zusammenhang zwischen dem *Innovationsstreben* einer Kommune und dem *wahrgenommenen Nutzenpotenzial* von Mobile Government-Diensten.

H5. Es besteht ein positiver Zusammenhang zwischen der *IT-Erfahrenheit* einer Kommune und dem *wahrgenommenen Nutzenpotenzial* von Mobile Government-Diensten.

3. QUALITATIVE VORUNTERSUCHUNG

3.1 Teilnehmer und Methode

Zur Überprüfung der Forschungshypothesen und Operationalisierung der postulierten Einflussdimensionen wurde eine qualitativ ausgerichtete Inhaltsanalyse in zwei Phasen durchgeführt [22]. Hierzu wurde ein Kategoriensystem entsprechend der in Abschnitt 2.1 beschriebenen Dimensionen und Variablen aufgestellt.

In der ersten Phase (Testphase) wurden drei Interviews mit Vertretern der Stadtverwaltung einer deutschen Großstadt bzw. einer stadtnahen Technologiestiftung geführt. Zum Zwecke der Triangularisierung der Hypothesen erschien es zu diesem Zeitpunkt sinnvoll, sowohl Vertreter von der fachlichen als auch der technischen Seite zu befragen. Tabelle 1gibt Aufschluss über die Positionen der einzelnen Interviewpartner. Auf Basis des Kategoriensystems wurde ein Interview-Leitfaden mit offenen Fragen und hypothetisierten Teilaspekten pro Kategorie entwickelt. Die Gespräche wurden als Präsenzinterviews von ca. 60 Minuten geführt und digital aufgezeichnet. Die Auswertung der Interview-Transkripte erfolgte mit Hilfe der Software Atlas.

Tabelle 1 – Interviewpartner und Stadtgrößen (in Einw.)

Phase	Position	Stadtgröße
1) Test-phase	Leiter IT Kompetenzzentrum	> 200.000
	Leiter Referat Medien, Inform. u. Kommunikationswirtschaft	> 200.000
	Leiter Bereich IKT	> 200.000
2) Codier-phase	Leiter der Stabsstelle E-Government	180.000
	Dezernent (u.a. für Personal, Organisation, EDV, Schulen)	70.000
	IT Organisator	60.000
	Leiter EDV	60.000

Im Sinne einer induktiven Vorgehensweise wurde das Kategoriensystem überarbeitet und Unterkategorien entwickelt.

In der zweiten Phase (Codierphase) wurden Interviews mit IT-Entscheidern von vier weiteren deutschen Kommunen geführt. Die Interviews konnten auf Basis des überarbeiteten Kategoriensystems mit einem leicht angepassten Leitfaden als Telefoninterviews von ca. 45 Minuten durchgeführt werden. Wie auch in der ersten Phase, wurde die Codierung der Interview-Transkripte von zwei Codierern unabhängig voneinander durchgeführt und im Falle von Nicht-Übereinstimmung nachträglich diskutiert. Nach Auszählen der insgesamt 488 Codes ergab sich eine Übereinstimmung von 71 Prozent, was einer guten Intercoder-Reliabilität entspricht [22].

3.2 Operationalisierung der Einflussfaktoren

Die aus der Inhaltsanalyse resultierenden Unterkategorien sowie die Anzahl der jeweiligen Nennungen sind in Tabelle 2 dargestellt. Da im Folgenden pro Dimension lediglich die drei Unterkategorien mit der häufigsten Anzahl an Nennungen weiter behandelt werden, beschränkt sich die Darstellung auf diese Einflussfaktoren. Weitere Kategorien die codiert wurden, bezogen sich beispielsweise auf die Verfügbarkeit von Partnern, themenspezifischen Netzwerken sowie Synergien mit anderen Kommunen.

Gemäß der Interviewergebnisse kann die Dimension *Effizienzstreben* in die Indikatoren *Verbesserung der Verwaltungstätigkeit*, *Win-Win Situation* und *Effizienzverbesserung* zerlegt werden. *Verbesserung der Verwaltungstätigkeit* beschreibt das Bestreben, mobile Dienste zur Verbesserung bestimmter Prozesse der Verwaltung einzusetzen. *Win-Win Situation* betont dabei den Aspekt, dass dabei nicht nur eine Erleichterung für die Kommune, sondern auch für den Kunden (Bürger, Unternehmen) geschaffen werden soll. *Effizienzverbesserung* kapselt solche Motivationen, die explizit aus dem Zwang zum Sparen der Kommune resultieren.

Innovationsstreben wird zum einen durch den Einflussfaktor *Externe Erwartung* wiedergegeben. Dies bezieht sich auf den Druck, den die Kunden der Verwaltung in Hinblick auf das Angebot innovativer Dienste ausüben. Zum anderen gehört hierzu der Indikator *Erweiterung des Dienstleistungsangebots*, welcher die Motivation zum Angebot neuer mobiler Dienste aus der Verwaltung heraus zusammenfasst. *Zunehmende Technikaffinität* beschreibt den Wandel in der Bevölkerung, zu dem mobile Dienste Teil des alltäglichen Lebens werden und somit auch von der Kommune entsprechend mobile Angebote erwartet werden.

Im Zusammenhang mit *IT-Erfahrenheit* erfasst *E-Government Plattform* den Grad, zu dem bestimmte Verwaltungsprozesse

Tabelle 2 – Unterkategorien der strategischen Dimensionen

Dimension	Unterkategorie	Nennungen
Effizienz-streben	Verbesserung der Verwaltungstätigkeit	17
	Win-Win Situation	14
	Effizienzverbesserung	12
Innovations-streben	Externe Erwartung	20
	Erweiterung des Dienstleistungsangebots	19
	Zunehmende Technikaffinität	18
IT-Erfahrenheit	E-Government Plattform	16
	Dienstleistungsstrategie	16
	IT-Strategie	9

bereits elektronisch abgebildet sowie online zugänglich sind und somit einfacher als mobiles Angebot zur Verfügung gestellt werden können. Im Zusammenhang hiermit wurde in den Interviews ebenfalls auf das Vorhandensein einer *Dienstleistungsstrategie* sowie einer *IT-Strategie* hingewiesen, welche eine behördenübergreifende Planung zukünftiger Dienstleistungsangebote unter Berücksichtigung von E-Government und M-Government Bestandteilen ermöglichen.

4. EMPIRISCHE STUDIE

4.1 Fragebogendesign

Zur Überprüfung des vorgestellten Modells wurde ein umfassender Fragebogen konzipiert und validiert. Zu Beginn des Fragebogens wurden Angaben zu der Kommune (Einwohner und kommunaler Haushalt) sowie demografische Angaben des Teilnehmers abgefragt. Die darauf folgenden Merkmale wurden jeweils über 7-Punkt-Likert-Skalen dimensioniert.

Die Einflussfaktoren *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrenheit* werden als formative Konstrukte aufgefasst, die sich aus den in 3.2 beschriebenen Einflussgrößen ergeben. Basierend auf der qualitativen Voruntersuchung wurde jede der Einflussgrößen mit der Skala „1=kein Einfluss“ bis „7=sehr hoher Einfluss“ operationalisiert (siehe Tabelle 3).

Schwerpunkt des Fragebogens bildete eine Liste von 60 möglichen mobilen Dienstangeboten, welche jeweils anhand einer kurzen Beschreibung entsprechend ihrer *wahrgenommenen Attraktivität* auf der Skala von „1=nicht attraktiv“ bis „7=sehr attraktiv“ für die Kommune bewertet werden sollten. Die Ideen für diese zum Teil sehr innovativen Anwendungsszenarien wurden aus wissenschaftlicher und angewandter Literatur extrahiert und gemäß der in Abschnitt 2.1 vorgestellten Anwendungsbereiche gruppiert. Tabelle 4 zeigt einen Auszug aus diesen Diensten.

Das insgesamt *wahrgenommene Nutzenpotenzial* von mobilen Diensten wurde ebenfalls auf Basis der in Abschnitt 2.1 vorgestellten Anwendungsbereiche operationalisiert und pro Anwendungsbereich mit den Skalenwerten „1=kein Nutzenpotenzial“ bis „7=sehr großes Nutzenpotenzial“ für die Kommune erfasst. Durch diese analoge Aufteilung sollte sichergestellt werden, dass das Verständnis des Mobile Government-Begriffs kongruent mit dem der bewerteten Dienstattraktivitäten ist und damit die Inhaltsvalidität des Konstruktes untermauert wird.

Tabelle 3 – Operationalisierung der Einflussfaktoren

Wie hoch schätzen Sie den motivierenden Einfluss folgender Faktoren ein, um drahtlose Kommunikationsdienste in Ihrer Kommune zu realisieren?	
EF1	Bedarf in der Verwaltung zur Verbesserung der Verwaltungsarbeit
EF2	Schaffen einer Win-Win Situation für Kommune und Anwender
EF3	Zwang zum Sparen durch Effizienzerhöhung
IN1	Erwartungsdruck von Bürgern und Wirtschaft zur Verbesserung der Verwaltung
IN2	Schaffen neuer Verwaltungsdienstangebote ermöglicht durch mobile Technologien
IN3	Zunehmende und Technikaffinität in der Bevölkerung
IT1	Verwaltungsprozesse, die neuen mobilen Diensten zugrunde liegen, sind bereits elektronisch abgebildet
IT2	In der Kommune existiert eine behördenübergreifende Planung zukünftiger Dienstleistungsangebote die die Nutzung moderner IT-Technologien und Kommunikationswege berücksichtigt
IT3	In der Kommune existiert eine behördenübergreifende IT-Strategie mit E- und M-Government-Bestandteilen

Die *geplante Investition* in mobile Dienste wurde direkt mit dem „geschätzten Investitionsvolumen der Kommune in mobile Dienste innerhalb der nächsten drei Jahre“ als 7-stufiges Item mit den Intervallgrenzen 0; 50; 100; 250; 500; 1.000; 5.000 und >5.000 T€ abgefragt.

Die inhaltliche Validität des Fragebogens wurde sorgfältig überprüft. Zunächst wurden die Attribute des Fragebogens durch mehrere Kollegen sowie Experten in Messtheorie und Statistik kontrolliert. Die anfängliche Version des Fragebogens wurde nach der von Hunt et al. [14] vorgeschlagen Methode in Präsenzterminen mit den an der ersten Inhaltsanalyse beteiligten Interviewpartnern pre-getestet, wobei sich nur noch minimale Änderungen in den Formulierungen der Dienstbeschreibungen und Einflussfaktoren ergeben haben.

4.2 Stichprobe

Die eigentliche Erhebung fand zwischen Mai und Juni 2010 statt und wurde als offene Online-Befragung durchgeführt. Aus der Teilnehmerliste einer der größten deutschen Konferenzen im Bereich des E-Government wurden elektronische Anschriften der (Ober-)Bürgermeister und IT-Verantwortlichen der 187 deutschen Kommunen mit mehr als 50.000 Einwohnern extrahiert sowie wo notwendig durch eine Internetrecherche vervollständigt. Da nicht davon ausgegangen werden konnte, dass die IT-Verantwortlichen in jedem Fall die richtigen Ansprechpartner für die Thematik des Mobile Government sind, wurde zunächst eine Einladung an die Bürgermeister mit Bitte zur entsprechenden Weiterleitung verschickt. Als Anreiz zur Teilnahme wurde den Teilnehmern lediglich das Zusenden der Studienergebnisse angeboten. Wenige Tage später wurden separate Erinnerungs-Emails an Bürgermeister und IT-Verantwortliche der Kommunen geschickt, die bisher nicht teilgenommen hatten. Bei Kommunen >100.000 Einwohner wurde zudem telefonisch nachgefragt. Rückfragen vor allem technischer Natur wurden ebenfalls telefonisch beantwortet.

Von den 187 angeschriebenen Kommunen haben 78 Vertreter begonnen den Fragebogen auszufüllen. Hiervon mussten 28 unvollständige Datensätze aussortiert werden, so dass 50 gültige

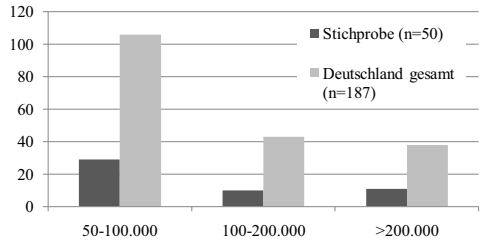


Abbildung 2 – Größe der Kommunen (Einwohner)

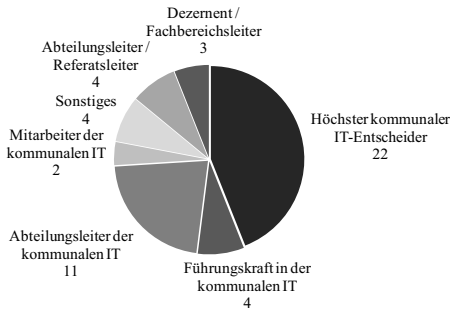


Abbildung 3 – Position der Studienteilnehmer

Antworten (Rücklaufquote 27 Prozent) in die Analyse einbezogen werden konnten. Hiervon entfielen 42 auf IT-seitige und 8 auf fachseitige Repräsentanten der jeweiligen Stadtverwaltungen. Die Verteilung der Einwohnerzahlen der teilnehmenden Städte sowie die Position der Studienteilnehmer sind in Abbildung 2 und 3 dargestellt. Nach den von Kromrey [15] vorgeschlagenen Kriterien kann die Erhebung als repräsentativ für den Raum Deutschland angesehen werden.

4.3 Deskriptive Auswertung der Dienstattraktivitäten

Die deskriptive Auswertung der Dienstattraktivitäten nach Mittelwert und Standardabweichung ergibt ein differenziertes Bild von den Präferenzen der IT-Entscheider. Im Bereich Bürger und Verwaltung wird beispielsweise mit der *mobilen Verwaltungsarbeit*, d.h. der mobilen Datenerhebung und medienbruchfreien Verarbeitung bspw. im Hygieneamt, ein verwaltungsinternes Mobile Government-Szenario mit M=4,86 als am attraktivsten bewertet.

Den mit Abstand höchsten Attraktivitätswert über alle Anwendungsbereiche bekommt das *mobile Feuerwehrunterstützungssystem*. Der mobile Zugriff auf Informationen wie Baupläne, Straßenkarten, Ereignis- und Objektinformationen, wie er in einigen Kommunen bereits im realisiert ist, wird offenbar von allen kommunalen Vertretern als äußerst sinnvolle Anwendung erachtet. Mit der zweithöchsten Bewertung insgesamt folgt hierdrauf die *digitale Authentifizierung*, welche in Kombination mit den Entwicklungen zum Elektronischen Personalausweis die gesicherte Identifikation bei Behördendienstleistungen über Telefon oder Internet ermöglichen soll.

Tabelle 4 zeigt die Dienstattraktivitäten pro Anwendungsbereich absteigend sortiert nach Mittelwerten der Bewertung (n=50). Aus Gründen der Übersicht wurden nur die besten drei sowie das am schwächsten bewertete Dienstangebot pro Bereich dargestellt. Für eine vollständige Beschreibung der Dienstangebote sei auf [9] verwiesen (auf Anfrage erhältlich).

Tabelle 4 – Dienstattraktivitäten pro Anwendungsbereich (Auszug)

Dienst	Mittel-wert	Standard-abweichung
Bürger und Verwaltung		
Mobile Verwaltungsarbeit	4,86	1,69
Mobile Bezahlung kommunaler Gebühren	4,50	1,68
Stadtinformationsdienst	4,48	1,59
Mobile Wahl	3,04	1,70
Öffentliche Sicherheit		
Mobiles Feuerwehrunterstützungssystem	5,57	1,53
Digitale Authentifizierung	5,14	1,71
Mobiles Polizeiunterstützungssystem	4,79	1,69
Personenortung	3,32	1,64
Infrastruktur und Umwelt		
Intelligente Gebäudesteuerung	4,80	1,58
Intelligente Straßenbeleuchtung	4,62	1,51
Luftverschmutzungs-Informationssystem	4,59	1,52
Intelligente Müllentsorgung	3,48	1,60
Bildung		
Elektronische Bibliothekskarte	5,04	1,69
Mobile Bibliothek	4,70	1,40
Elektronische Schülerkarte	3,91	1,74
Bildungsinformationssystem	3,06	1,63
Gesundheit		
Barrierefreie Navigation	4,16	1,76
Mediziner-Informationsdienst	3,69	1,84
Mobile Telemedizindienste	3,69	1,87
Angehörigen-Informationsdienst	3,21	1,72
Verkehr		
Mobile Bezahlung im ÖPNV	4,48	1,53
Parkleitdienst	4,42	1,75
Intelligentes Verkehrskoordinationssystem	4,25	1,59
Automatische Stadt-Maut	3,10	1,83
Tourismus und Kultur		
Mobiler Touristenführer	4,64	1,65
Mobile Ticketbuchung	4,35	1,66
Informationskioske	4,00	1,62
Mobiles Fernsehen	2,88	1,73
Privathaushalte und Soziales		
Drahtloses Breitband	4,74	2,02
Mobile Seniorendienste	3,81	1,68
Vernetzte Wohnwelt	3,58	1,67
Haustierortung	2,85	1,59

5. MODELLÜBERPRÜFUNG UND ERGEBNISDISKUSSION

5.1 Methode

Wir schätzen und testen das eingangs postulierte Modell mit Hilfe der varianzbasierten Partial Least Square-Pfadanalyse (PLS). Alle Berechnungen wurden mit den Softwareprodukten SPSS und SmartPLS durchgeführt. Der PLS Ansatz ist im vorliegenden Fall aus mehreren Gründen besonders geeignet. Zum einen ist die gegebene Fragestellung als eher explorativ zu bezeichnen, da kaum auf vorhandener Theorie aufgebaut werden kann. Gegenüber kovarianzbasierten Ansätzen stellt PLS keine Anforderungen an die Verteilung der zugrunde liegenden Variablen und ist besser geeignet, wenn der Fokus auf Theorieentwicklung liegt [5].

Zum Zweiten enthält das Forschungsmodell neben reflektiven Variablen ebenfalls formative Konstrukte, welche klassische kovarianzbasierte Verfahren nicht abbilden können [24]. So werden die Variablen *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrung* konstituierend durch die in Abschnitt 3.2 operationalisierten Indikatoren gebildet. Insbesondere in der Forschung zu Erfolgs- und Einflussfaktoren sind formative Konstrukte häufig besser geeignet, die kausale Wirkung zwischen Indikatoren und Konstrukt abzubilden [1]. Damit wird die latente Variable zu einem gewichteten Index ihrer Indikatoren. Die sich hieraus ergebenen Gewichte entsprechen den Beta-Koeffizienten eines Standard-Regressionsmodells und haben üblicherweise kleinere Absolutwerte als die Ladungen reflektiver Variablen. Es sei angemerkt, dass wir das Modell zusätzlich in zwei Varianten, sowohl mit allen Konstrukten reflektiv als auch formativ, getestet haben. In keinem der Fälle ergaben sich grundsätzliche Unterschiede an den Interpretationen der Pfadkoeffizienten und Signifikanzen, was sich mit den Beobachtungen von Albers [1] deckt.

Der notwendige Stichprobenumfang zum Testen von PLS-Modellen ist in der Literatur nicht unumstritten. Nitzl [23] argumentiert, dass PLS bereits bei Stichproben von 20 Fällen sinnvolle Ergebnisse liefern kann. Hingegen weisen Marcoulides und Saunders [18] darauf hin, dass Stichproben dieser Größe nicht dazu geeignet sind, zuverlässig schwächere Pfadkoeffizienten nachzuweisen. In Abhängigkeit der Freiheitsgrade des Models hat sich die Heuristik Chins [5] etabliert, nach der die Stichprobe mindestens 10 mal so groß sein sollte wie die größte Anzahl formativer Indikatoren einer latenten Variable bzw. die größte Anzahl der Prediktoren einer latenten endogenen Variablen. Beide Größen sind in dem vorliegenden Modell gleich drei, wonach mit n=50 die Heuristik erfüllt ist.

5.2 Evaluierung des Messmodells

Wir folgen dem Ansatz von Chin [5] und evaluieren Messmodell und Strukturmodell separat voneinander, um die allgemeine Güte des Modells zu erörtern.

5.2.1 Evaluierung der formativen Konstrukte

Die formativen Variablen *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrung* müssen auf Multikollinearität überprüft werden [24]. Hierzu werden pro Variable zum einen die Pearson'sche Korrelationsmatrix und zum anderen die Toleranz-Werte der Indikatoren in Regressionen untereinander berechnet. Im Ergebnis kann trotz zum Teil signifikanter Korrelationen von bis zu 0,6 aufgrund der hohen Toleranz-Werte, welche deutlich

über 0,1 liegen, für alle neun Indikatoren Multikollinearität ausgeschlossen werden [24], siehe Tabelle 5. Somit müssen diese Indikatoren nicht über Indexbildung weiter zusammengefasst werden. Da formative Indikatoren aufgrund ihrer inhaltlichen Relevanz in das Modell aufgenommen werden und somit keine Anforderungen der konvergente und diskriminante Validität gestellt werden [24], ist dieser Teil der Evaluierung hiermit abgeschlossen.

Tabelle 5 – Toleranzwerte der formativen Indikatoren

Effizienzstreben		Innovationsstreben		IT-Erfahrenheit	
EF1	0,619	IN1	0,624	IT1	0,503
EF2	0,634	IN2	0,662	IT2	0,390
EF3	0,845	IN3	0,484	IT3	0,391

5.2.2 Evaluierung der reflektiven Konstrukte

Die Evaluierung der reflektiven Variablen folgt der Logik von Homburg und Giering [13] in drei Untersuchungsschritten. Im ersten Schritt werden pro Variable explorative Faktoranalysen (EFA) mit den der Variable zugeordneten Indikatoren durchgeführt, um die Eindimensionalität zu überprüfen und gegebenenfalls Bereinigungen durchzuführen. Jeder extrahierte Faktor sollte mindestens 50 Prozent der Varianz der zugehörigen Indikatoren erfassen (AVE) und die Indikatoren jeweils eine Faktorladung von mindestens 0,4 aufweisen, um nicht aus der Analyse ausgeschlossen zu werden [13]. In den darauf folgenden Schritten werden konvergente und diskriminante Validität überprüft.

Die Variable *wahrgenommenes Nutzenpotenzial* wurde über acht Indikatoren, jeweils einem pro Anwendungsbereich, erfasst. Die EFA dieser acht Indikatoren ergibt, dass die wahrgenommenen Potenziale im Bereich Tourismus und Kultur (mit 0,78) sowie mit Einschränkung im Bereich Verkehr (0,51) auf einen zweiten Faktor laden. Offenbar haben städtische IT-Entscheider zu diesen nicht-originär in der Verantwortung der Verwaltung liegenden Bereichen eine divergierende Einstellung. Ein Grund hierfür könnte darin liegen, dass in beiden Bereichen stärker privatwirtschaftlich orientierte Angebote (z.B. mobiles ÖPNV-Ticketing, Carsharing sowie Touristenführer, mobiles Fernsehen etc.) als Dienste aufgeführt worden sind. Gemäß [13] folgern wir, dass diese Indikatoren nur beschränkt geeignet sind, die Einstellung bezüglich des Dienstepotenzials für Mobile Government als Ganzes zu erfassen und entfernen sie aus der Analyse. Der verbleibende eindimensionale Faktor erklärt durchschnittlich 64 Prozent der Varianz in den verbleibenden sechs Indikatoren.

Die zu erklärende Variable *wahrgenommene Dienstattraktivität* wird aus den bewerteten Dienstattraktivitäten abgeleitet. Hierzu werden zunächst durch Mittelwertbildung acht Indizes, einer pro Anwendungsbereich, gebildet, welche als aggregierte Indikatoren der *wahrgenommenen Dienstattraktivität* zugeordnet werden können. Um inhaltliche Kongruenz mit der Variable *wahrgenommenes Nutzenpotenzial* sicherzustellen, werden hier analog die Durchschnittswerte in den Bereichen Tourismus und Kultur sowie Verkehr entfernt. Die EFA ergibt einen Faktor der durchschnittlich 69 Prozent der Varianz in den sechs Indizes erklärt.

Die konvergente Validität der gewonnenen Konstrukte wird anhand der internen Konsistenz durch Cronbachs Alpha überprüft. Die Werte für beide Konstrukte liegen deutlich über den geforderten 0,7. Als weiteres Gütekriterium wird die Faktorreliabilität hinzugezogen, welche zeigt wie gut die Konstrukte durch die zugeordneten Indikatoren gemessen werden. Beide Werte liegen ebenfalls über dem Schwellwert von 0,6 [24], siehe Tabelle 6.

Tabelle 6 – Gütekriterien der konvergenten Validität

Konstrukt	AVE	Cronbachs Alpha	Faktor-reliabilität
Wahrgenommenes Nutzenpotenzial	0,626	0,899	0,921
Wahrgenommene Dienstattraktivität	0,709	0,930	0,944

Die diskriminante Validität gibt an, in welchem Ausmaß sich die Indikatoren der latenten Konstrukte voneinander unterscheiden. Hierzu sollte die Wurzel der durchschnittlich erfassten Varianz (AVE) eines Konstruktes größer sein als der jeweilige Korrelationskoeffizient eines Konstruktes mit allen übrigen Konstrukten des Modells [24]. Wie in Tabelle 7 deutlich wird, ist dieses Kriterium für alle latenten Variablen erfüllt. (Für formative Konstrukte ist dieses Kriterium nicht anwendbar und wurde nur zur Vollständigkeit im reflexiven Fall bestätigt.) Das zweite Kriterium der diskriminanten Validität besagt, dass die Faktorladungen der Indikatoren auf ihre jeweiligen latenten Variablen höher sein sollten als die Kreuzladungen auf andere Konstrukte. Dieses Kriterium ist ebenfalls erfüllt (Tabelle auf Anfrage erhältlich).

Tabelle 7 – Korrelation latenter Variablen (reflektiver Fall) mit erfasster Varianz auf der Diagonalen ($\sqrt{AVE}$)

	EF	IN	IT	NP	DA
Effizienzstr. (EF)	**0,790**				
Innovationsst. (IN)	0,390	**0,826**			
IT-Erfahrung (IT)	0,382	0,264	**0,888**		
Nutzenpot. (NP)	0,466	0,512	0,532	**0,791**	
Dienstattr. (DA)	0,428	0,355	0,412	0,812	**0,842**

5.2.3 Überprüfung auf methodenbedingte und demografische Verzerrungen

Um zu überprüfen, ob ein Großteil der festgestellten Varianz aus der Messmethode resultiert (Common-Method-Bias) wurde die Stichprobe dem One-Factor-Test von Harman [12] unterzogen. Der Test resultierte entsprechend der Anzahl der latenten Konstrukte in fünf Faktoren, wobei der erste Faktor lediglich 31 Prozent und nicht die Mehrheit der gesamten Varianz erklären kann. Dies spricht dafür, dass das Vorliegen einer methodischen Verzerrung nicht der ausschlaggebende Grund für die Zusammenhänge im Messmodell sein kann [26].

Um eine Verzerrung der Variablen durch demografische Merkmale auszuschließen wurde ein zweiseitiger Rangkorrelationstest nach Spearman zwischen den Variablen des Messmodells und der Anzahl der Einwohner sowie weiteren ordinalen Merkmalen des Befragten durchgeführt. Entgegen der Beobachtungen von Moon [21] finden wir keinen signifikanten Einfluss der Größe der Kommune auf die Messvariablen. Gleiches gilt für die demografischen Merkmale.

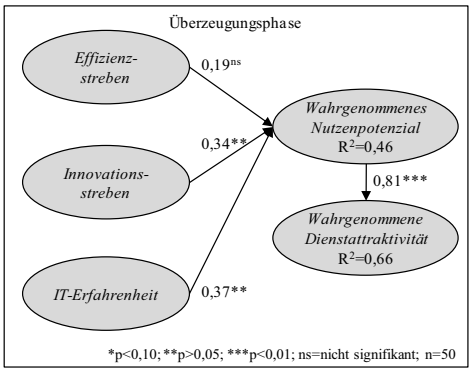


Abbildung 4 – Ergebnis der PLS-Analyse

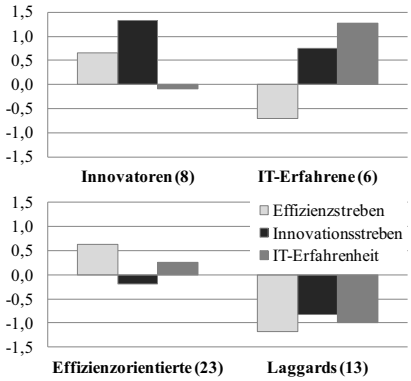


Abbildung 5 – Städtecluster und Motivationsprofile

5.3 Analyse des Strukturmodells

Die Ergebnisse der PLS Analyse sind in Abbildung 4 dargestellt. Die statistische Signifikanz wurde durch t-Tests auf Basis einer Bootstrap-Prozedur mit 5000 Resamples ermittelt. Zur Beurteilung des Modells können ähnlich wie in der einfachen Regression die erklärte Varianz R^2 sowie die Pfadkoeffizienten herangezogen werden [24].

Die Ergebnisse unterstützen die Hypothesen, dass das *Innovationstreben* (H4) sowie die *IT-Erfahrenheit* (H5) in positivem Zusammenhang mit dem *wahrgenommenen Nutzenpotenzial* von Mobile Government-Diensten stehen. Der Einfluss der *IT-Erfahrenheit* ist auf Basis der vorliegenden Daten nur als geringfügig stärker einzustufen. Dies kann so interpretiert werden kann, dass in erster Linie solche Kommunen für Mobile Government motiviert sind, welche ohnehin eine ausgeprägte IT-Ambition vorweisen. Mobile Government ist damit kein isolierter Teil der Technologiestrategie, sondern fügt sich in weitere Teile der E-Government Aktivitäten. Hierbei spielt der Anreiz neue Dienstleistungen anzubieten und auf sich ändernde Kundenerwartungen zu reagieren (*Innovationsstreben*) gleichsam eine große Rolle.

Für eine Einwirkung von *Effizienzstreben* (H3) kann dagegen, wenn auch knapp, keine signifikante Unterstützung gefunden werden. Dies bedeutet dass, obwohl zahlreiche Mobile Government-Anwendungen auf Prozessvereinfachungen und bilateralen Nutzen zwischen Bürgern und Verwaltung abzielen, dieses Potenzial in erster Linie offenbar noch nicht mit dem Gedanken der Effizienzverbesserung in Verbindung gebracht wird.

Der Einfluss dieser drei Faktoren auf die *wahrgenommene Dienstattraktivität* der dargestellten Dienstangebote wird durch das *wahrgenommene Nutzenpotenzial* mediiert. Zusammen erklären sie 46 Prozent der Varianz in dieser Variable, was vor dem Hintergrund der explorativen Grundausrichtung der Studie als gutes Nieveau der Erklärungsgüte gewertet werden kann [13]. Das *wahrgenommene Nutzenpotenzial* steht mit einem Pfadkoeffizienten von 0,81 in einem starken Zusammenhang mit der *wahrgenommenen Dienstattraktivität* (H1) und ist mit 66 Prozent der erklärten Varianz ein zuverlässiger Prädiktor hierfür. Dies spricht insbesondere für die hohe inhaltliche Kongruenz und statistische Validität dieser beiden Konstrukte.

5.4 Clusteranalytische Untersuchung

Die unterschiedliche Einwirkung der Einflussfaktoren, welche die Ergebnisse der qualitativen Voruntersuchung nur zum Teil widerspiegelt, deutet darauf hin, dass in der Stichprobe heterogene Motivationsstrukturen vorliegen könnten, wodurch in der Aggregation eines Regressionsmodells möglicherweise wesentliche Informationen verloren gehen [2]. Solche Informationen können durch eine a-posteriori Segmentierung der Stichprobe durch Aufteilung in kleinere Gruppen zurückgewonnen werden. Wir führen hierzu eine hierarchische Clusteranalyse nach Wards Linkage Methode basierend auf den (standardisierten) Faktorwerten der Variablen *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrenheit* durch. Anhand des Verlaufs des Distanzmaßes über die Clustering-Schritte beobachten wir einen starken Sprung des Distanzmaßes bei der Reduktion von vier auf drei Cluster. Damit legen wir die optimale Clusteranzahl auf vier fest. Die resultierenden Cluster unterschiedlicher Größe bezeichnen wir gemäß der dominierenden Mittelwertausprägungen der Einflussfaktoren als Innovatoren (n_1=8), IT-Erfahrene (n_2=6), Effizienzorientierte (n_3=23), und Laggards (n_4=13) wie in Abbildung 5 dargestellt.

Als **Innovatoren** können solche Kommunen bezeichnet werden, die sich als besonders empfänglich für die Erwartungen ihrer Kunden sehen und hieraus, neben dem Effizienzgedanken, eine Motivation zur Umsetzung neuartiger Mobile Government-Anwendungen ziehen. Offenbar muss es sich hierbei nicht notwendigerweise um besonders IT-erfahrene Kommunen handeln. Dagegen haben **IT-Erfahrene** Kommunen, der kleinste Cluster in der Stichprobe, eine gute Basis für derartige innovative IT-Vorhaben und dabei kein drängendes Effizienzstreben. Möglicherweise fehlt es diesen Kommunen aber zum Teil noch an attraktiven Angeboten (siehe auch Abschnitt 5.5). Den größten Teil der Kommunen kann man vermutlich vor dem Hintergrund der jeweiligen Haushaltsituationen als **Effizienzorientierte** einstufen, welche IT-Ausgaben auf das notwendigste begrenzen müssen und hierbei eher moderat im Bezug auf Mobile Government motiviert sind. Als **Laggards** können solche Kommunen mit einer unterdurchschnittlichen Ausprägung aller drei Mobile Government-Treiber bezeichnet werden. Diese Kommunen stehen im Innovationsprozess offensichtlich deutlich hinter den anderen drei vorgestellten Gruppen.

5.5 Zusammenhang zwischen Motivationsprofilen, wahrgenommener Dienstattraktivität und Investitionsverhalten

Zur Überprüfung des postulierten Zusammenhangs zwischen den Motivationsdimensionen der Kommunen und den wahrgenommenen Dienstattraktivitäten (H1) einerseits sowie den

Tabelle 8 – Städtecluster und Variablenmittelwerte

		Clustervariablen			Zielvariablen	
Cluster	**n**	**EF**	**IN**	**IT**	**DA**	**Invest.**
Innovatoren	8	0,65	1,33	-0,09	0,77	175.181€
IT-Erfahrene	6	-0,71	0,75	1,27	-0,08	158.333€
Effizienz-orientierte	23	0,62	-0,19	0,26	0,10	71.543€
Laggards	13	-1,17	-0,83	-0,99	-0,61	78.680€

geplanten Investitionen in mobile Dienste (H2) andererseits führen wir eine Varianzanalyse durch und testen post-hoc, ob sich die identifizierten Cluster (siehe Tabelle 8) im Bezug auf die Zielvariablen signifikant voneinander unterscheiden.

Es zeigt sich, dass im Bezug auf die Variable *Dienstattraktivität* die Nullhypothese gleicher Gruppenmittelwerte der Cluster mit $F=3,73$; $p=0,02$ widerlegt werden muss. Um zu ermitteln, zwischen welchen Clustern diese Unterschiede bestehen, werden Post-hoc-Tests nach Fisher LSD (least significant difference) durchgeführt. Bei einem Konfidenzniveau von $p<0,1$ zeigt sich, dass sich Innovatoren in der wahrgenommenen Dienstattraktivität ($M=0,77$) deutlich von Effizienzorientierten ($M=0,10$) und Laggards ($M=-0,61$) unterscheiden ($p=0,08$ bzw. $p=0,02$). Effizienzorientierte und Laggards unterscheiden sich ebenfalls signifikant untereinander ($p=0,04$). IT-Erfahrene Kommunen weisen bei einer leicht unterdurchschnittlichen Dienstattraktivität ($M=-0,08$) dagegen zu keiner Gruppe signifikante Mittelwertunterschiede auf, was für eine hohe Streuung innerhalb dieses Clusters spricht.

Die Varianzanalyse weist ebenfalls auf ungleiche Gruppenmittelwerte im Bezug auf die *geplante Investition* zwischen den Clustern ($F=2,54$; $p=0,07$). LSD Post-hoc-Tests zeigen, dass sich dabei Innovatoren ($M=175.181€$) signifikant von Laggards (78.680€) sowie Effizienzorientierten ($M=71.543€$) absetzen ($p=0,06$ bzw. $p=0,03$). Letztere liegen dabei ebenfalls deutlich unter dem Investitionsniveau der IT-Erfahrenen ($M=158.333€$; $p=0,09$).

6. SCHLUSSFOLGERUNG

6.1 Zusammenfassung

Die vorliegende Arbeit deutet darauf hin, dass bezüglich der Einstellung kommunaler IT-Entscheider ein Zusammenhang zwischen den organisationalen Rahmenbedingungen in einer Kommune und der wahrgenommenen Attraktivität von mobilen Diensten sowie der Investitionsabsicht besteht. Stellvertretend für die Summe von Meinungen und Einschätzungen in einer Kommune wurde verhaltenstheoretisch dahingehend argumentiert, dass derartige Rahmenbedingungen einen unterschiedlichen Grad an Motivation für oder gegen die Einführung von Mobile Government-Anwendungen erzeugen. Die organisationalen Rahmenbedingungen wurden auf Basis der Literatur sowie einer Serie von Interviews in einer Trichonomie von *Effizienzstreben*, *Innovationsstreben* und *IT-Erfahrung* typologisiert und operationalisiert. Die Ergebnisse einer empirischen Studie und PLS-Pfadanalyse konnten dabei nur den Einfluss der beiden letztgenannten Einflussfaktoren signifikant bestätigen, was vermuten lässt, dass sich Kommunen in den Ausprägungen dieser Einflussfaktoren unterscheiden. Diese Auffassung wird durch eine Clusteranalyse unterstützt, in der sich die untersuchten Kommunen in vier Typen von Motivationsprofilen einteilen lassen: Innovatoren, IT-Erfahrene, Effizienzorientierte sowie Laggards. Im Hinblick auf die zu erklärenden Variablen konnte gezeigt werden, dass Innovatoren und IT-Erfahrene planen, in den nächsten drei Jahren etwa das doppelte an Investitionen in Mobile Government-Dienste zu tätigen als andere Kommunen. Hierbei scheint es gerade den IT-erfahrenen Kommunen basierend auf der in der Studie durchgeführten Bewertung von 60 potenziellen Diensten noch an attraktiven Angeboten zu fehlen.

6.2 Implikationen für die Praxis

Aus den in der Studie gewonnenen Erkenntnissen ergeben sich nicht nur für die öffentliche Hand, sondern auch für private Dienste-Anbieter sowie die Politik wichtige Implikationen. Zunächst zeigen wir auf, dass Städte ebenso wie Unternehmen [27] heterogen sind in ihren Innovationsprofilen. Im Gegensatz zu Unternehmen stehen öffentliche Verwaltungen jedoch weitaus weniger im Wettbewerb. Daher erscheint es sinnvoll, dass die Erfahrungen, die innovative und stärker IT-erfahrene Kommunen machen, als Beispiele für Nicht-Innovatoren fungieren und ein Wissenstransfer stattfindet. Hierfür können Entscheider anhand der vorgeschlagenen Dimensionen und Motivationscluster eine Einordnung und Positionsbestimung der eigenen Kommune vornehmen. Trotz des thematischen Bezugs zum Mobile Government argumentieren wir, dass sich das vorgestellte Treibermodell auch auf weitere technologische und organisatorische Innovationen in der Verwaltung anwenden lässt.

Des Weiteren zeigt die Arbeit auf, welche Dienstangebote aus Sicht von kommunalen IT-Entscheidern für Investitionen am attraktivsten sind. Dies liefert nicht nur eine gute Orientierung für öffentliche Entscheider, sondern gleichsam eine wertvolle Marktinformation für Anbieter bspw. aus den Bereichen Software oder Telekommunikation, die in dieses junge Segment des E-Government vordringen. Bedenklich erscheinen dagegen die verhältnismäßig deutlichen negativen Ausprägungen im Motivationsprofil der Laggards gepaart mit der Erkenntnis, dass Effizienzstreben (bisher) kein signifikanter Treiber für Mobile Government zu sein scheint. Bei einem Großteil der Kommunen wird Mobile Government demnach bisher kaum als Weg zu mehr Effizienz in der Verwaltung betrachtet. Hier stellt sich die Frage, ob die Politik stärker bestimmte mobile Dienstleistungsangebote bzw. Standards forcieren sollte, wie es bspw. bei der Umsetzung der EU-Dienstleistungsrichtlinie oder dem elektronischen Personalausweis geschehen ist.

6.3 Limitationen und Ausblick

Einige Limitationen dieser Arbeit sollten bei der Interpretation berücksichtigt werden. Der Methoden-inhärenten Einschränkung der Generalisierbarkeit wurde versucht durch eine repräsentative Auswahl von Kommunen aus ganz Deutschland zu begegnen. Als zweites ist der, trotz der guten Rücklaufquote von 27 Prozent, verhältnismäßig kleine Stichprobenumfang zu nennen. Wie im Abschnitt 4.2 diskutiert, könnte durch eine größere Stichprobe möglicherweise eine stärkere Signifikanz der Modell-eigenen Faktoren (insbesondere des Effizienzstrebens) erreicht werden.

Des Weiteren wurde angenommen, dass die Einschätzung eines einzelnen Entscheiders pro Kommune die Summe der Einschätzungen hinreichend wiedergibt, was gerade in großen Kommunen zu Messfehlern führen könnte. Zudem stellt das im Rahmen der Hypothesenbildung generierte Modell zwangsläufig

eine starke Vereinfachung der Realität dar. Um den beiden letztgenannten Einschränkungen zu begegnen, wurden die Kommunen zusätzlich nach ihrer Bereitschaft zur Teilnahme an einer weiterführenden, qualitativ-ausgerichteten Untersuchung befragt. Hierdurch hoffen die Autoren, weitere Einflussfaktoren für das Mobile Government auch auf feingranularer Ebene identifizieren und der Forschung zugänglich machen zu können.

Trotz dieser Limitationen liefert diese Arbeit einen relevanten Forschungsbeitrag, indem sie als eine der ersten empirischen Arbeiten den Fokus der Mobile Government Adoptionsentscheidung auf die verantwortlichen Entscheider in den Kommunen legt. Dabei wurde der Mobile Government-Begriff ganzheitlich definiert sowie operationalisiert und die verwaltungsinterne, effizienzsteigernde Perspektive explizit nicht ausblendet. Die Arbeit erweitert damit die Endnutzer-bezogene E- bzw. M-Government-Forschung um eine organisational ausgerichtete Betrachtungsweise dieses aktuellen Forschungsgegenstands.

7. LITERATUR

[1] S. Albers & L. Hildebrandt (2006). 'Methodische Probleme bei der Erfolgsfaktorenforschung-Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells'. *Zeitschrift für betriebswirtschaftliche Forschung* **58**(1):2–33.

[2] K. Backhaus, et al. (2005). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Springer.

[3] A. Benlian, et al. (2009). 'Treiber der Adoption SaaS-basierter Anwendungen'. *Wirtschaftsinformatik* 51(5):414–428.

[4] Bitkom (2010). 'Internet per Handy erobert den Massenmarkt'. http://www.bitkom.org/de/presse/8477_64819.aspx, Zugriff am 8.8.2010

[5] W. W. Chin (1998). 'The partial least squares approach to structural equation modeling'. *Modern methods for business research* **295**:336.

[6] Computerwoche (2009). 'Zahl der mobilen Internetnutzer wird sich bis 2012 verdreifachen', http://www.computerwoche.de/netzwerke/mobile-wireless/1898628/, Zugriff am 8.8.2010

[7] F. D. Davis, et al. (1989). 'User acceptance of computer technology: a comparison of two theoretical models'. *Management science* pp. 982–1003.

[8] J. S. Denford & Y. E. Chan (2007). 'Reconciling IS strategic alignment approaches'. In *Proceedings of the 13th Americas Conference on Information Systems*, pp. 9–12.

[9] P. Ernst (2010). 'Ubiquitous City Dienste in deutschen Kommunen – Attraktivität, Realisierungsstand, Treiber, Hindernisse und Ziele aus kommunaler Sicht'. *Master's thesis.* Humboldt-Universität zu Berlin 2010.

[10] M. Fishbein & I. Ajzen (1975). 'Belief, attitude, intention and behavior: An introduction to theory and research'.

[11] S. Gopalakrishnan (1997). 'A review of innovation research in economics, sociology and technology management'. *Omega* **25**(1):15–28.

[12] H. H. Harman (1976). *Modern factor analysis*. University of Chicago Press.

[13] G. Homburg & A. Giering (1996). 'Konzeptualisierung und Operationalisierung Komplexer Konstrukte - Ein Leitfaden für die Marketing Forschung'. *Marketing ZfP* **18**(1):5–24.

[14] S. D. Hunt, et al. (1982). 'The Pre-Test in Survey Research: Issues and Preliminary Findings.'. *Journal of Marketing Research* **19**:269–73.

[15] H. Kromrey (2006). Empirische Sozialforschung. Utb.

[16] M. Kumar & O. P. Sinha (2007). 'M-government–mobile technology for e-government'. In *International conference on e-government, India.*

[17] I. Kushchu, et al. (2007). 'Introducing Mobile Government'. *Mobile government: an emerging direction in e-government* p. 1.

[18] G. A. Marcoulides (2009). 'A Critical Look at Partial Least Squares Modeling'. *Mis Quarterly* **33**(1):171–175.

[19] P. Mayring (2008). Qualitative Inhaltsanalyse: Grundlagen und Techniken. Beltz.

[20] R. E. Miles, et al. (1978). 'Organizational strategy, structure, and process'. *Academy of Management Review* **3**(3):546–562.

[21] M. J. Moon & D. F. Norris (2005). 'Does managerial orientation matter? The adoption of reinventing government and e-government at the municipal level'. *Information Systems Journal* **15**(1):43–60.

[22] K. A. Neuendorf (2002). *The content analysis guidebook*. Sage Publications, Inc.

[23] C. Nitzl (2010). 'Eine anwenderorientierte Einführung in die Partial Least Square (PLS)-Methode'. Tech. rep.

[24] G. Panten & S. Boßow-Thies (2007). 'Analyse kausaler Wirkungszusammenhänge mit Hilfe von Partial Least Squares (PLS)'. *Methodik der empirischen Forschung* pp. 311–326.

[25] H. Patel & D. Jacobson (2008). 'Factors Influencing Citizen Adoption of E-Government: A Review and Critical Assessment'. *16th European Conference on Information Systems* .

[26] P. M. Podsakoff & D. W. Organ (1986). 'Self-reports in organizational research: Problems and prospects'. *Journal of management* **12**(4):531–544.

[27] E. M. Rogers (1995). *Diffusion of innovation*. The Free Press New York.

[28] R. Sabherwal & Y. E. Chan (2001). 'Alignment between business and IS strategies: A study of prospectors, analyzers, and defenders'. *Information Systems Research* **12**(1):11–33.

[29] Städtetag NRW (2009). 'Gemeindefinanzbericht 2009 des Städtetages Nordrhein-Westfalen'.

[30] L. G. Tornatzky & K. J. Klein (1982). 'Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings'. *IEEE Transactions on engineering management* **29**(1):28–45.

[31] S. Trimi & H. Sheng (2008). 'Emerging trends in M-government'. *Commun. ACM* **51**(5):53–58.

[32] T. J. Winkler, et al. (2009). 'Wireless City Initiatives in Europe –Towards a Service-Oriented Approach'. In *Proceeding of the TAMoCo 2009*, pp. 36–47. IOS Press.

Automatische Reputationsmessung in der Wikipedia

Thomas Wöhner
Martin-Luther-Universität
Halle-Wittenberg
06099 Halle (Saale)
+49 345 55-23478
thomas.woehner@wiwi.uni-halle.de

Sebastian Köhler
Martin-Luther-Universität
Halle-Wittenberg
06099 Halle (Saale)
+49 345 55-23479
sebastian.koehler@wiwi.uni-halle.de

Ralf Peters
Martin-Luther-Universität
Halle-Wittenberg
06099 Halle (Saale)
+49 345 55-23471
ralf.peters@wiwi.uni-halle.de

ZUSAMMENFASSUNG

Die Wikipedia wird im Rahmen ihres offenen Zugangs von Autoren erstellt, die überwiegend anonym oder unter Pseudonym tätig und damit weitgehend unbekannt sind. Daraus resultiert eine Unsicherheit bezüglich der Güte der einzelnen Beiträge. Einen Lösungsansatz bieten automatische Reputationssysteme, die sich in den vergangenen Jahren als eigenständiges Forschungsgebiet etabliert haben. Durch diese Reputationssysteme wird die Reputation der Autoren anhand ihrer bisherigen Bearbeitungen automatisch berechnet. In der gegenwärtigen Forschung werden die zur Reputationsmessung vorgeschlagenen Metriken jedoch nur isoliert betrachtet und oft unzureichend bewertet, sodass sich deren Aussagekraft nur schwer abschätzen lässt. Im vorliegenden Beitrag werden insgesamt zehn Metriken vergleichend bewertet und durch Kombination anhand einer Diskriminanzanalyse zu einer effizienten Reputationsfunktion zusammengeführt. Die Metriken wurden der bestehenden Literatur entnommen und um eigene Vorschläge ergänzt. Die Analyse zeigt, dass die neu vorgeschlagene Metrik der *Effizienz der Bearbeitungen* besonders aussagekräftig ist.

Schlüsselwörter

Web 2.0, Wikipedia, Reputation, Reputationssystem, Persistenz, Qualität, Effizienz

1. EINLEITUNG

Unter dem Begriff Web 2.0 gewannen in den vergangenen Jahren Websites mit benutzergenerierten Inhalten (User Generated Content) zunehmend an Bedeutung [17]. Die besondere Relevanz des Web 2.0 zeigt sich beispielsweise an der Besucherstatistik der zehn weltweit meistbesuchten Websites, von denen fünf (Facebook, YouTube, Wikipedia, Blogger, Twitter) auf benutzergenerierten Inhalten basieren.[1] Ein weit verbreiteter Anwendungstyp innerhalb des Web 2.0 sind Wikis. Wikis sind Websites, deren Inhalt direkt im Browser durch die Internetbenutzer geändert werden kann. Die entsprechende Wiki-Syntax ist vergleichsweise einfach, sodass eine Bearbeitung ohne besondere technische Vorkenntnisse möglich ist. Wikis werden daher zur kollaborativen Erstellung von Websites eingesetzt [6].

Das weltweit größte und bekannteste Wiki ist die freie Online-Enzyklopädie Wikipedia, die im Januar 2010 mehr als *15* Millionen Artikel[2] enthielt und in mehr als *260* Sprachen[3] verfügbar ist. Entsprechend dem Wiki-Prinzip werden die Artikel ausschließlich durch die Internetbenutzer erstellt und unterliegen keiner Kontrolle durch Experten. Jede Änderung wird unmittelbar im World Wide Web veröffentlicht. Dieses offene Konzept führt auf der einen Seite dazu, dass sich viele Internetbenutzer an der Erstellung und Pflege der Wikipedia beteiligen. Durch die sogenannte *Weisheit der Vielen* [22] erfolgt eine zeitnahe Aktualisierung und Fehler werden zumeist schnell erkannt und korrigiert. Studien haben ergeben, dass die Wikipedia ein mit klassischen Enzyklopädien wie Brockhaus oder Britannica vergleichbares Qualitätsniveau erreicht [9, 13]. Auf der anderen Seite lassen sich durch die beschriebene Offenheit unerwünschte Änderungen wie Vandalismus, Spam und fehlerhafte Einträge aufgrund von Unwissenheit und Opportunismus nicht ausschließen [7, 25, 26]. Da die Bearbeitungen in der Wikipedia von weitgehend unbekannten Autoren vorgenommen werden, ist die Güte der einzelnen Beiträge nur schwer einzuschätzen.

Ein weit verbreiteter Ansatz, um Vertrauen zwischen unbekannten Teilnehmern zu schaffen, sind Reputationssysteme [20]. Reputationssysteme erfassen und bewerten das Verhalten der Teilnehmer in der Vergangenheit, um darauf aufbauend deren zukünftiges Verhalten abzuschätzen. Man unterscheidet dabei zwischen expliziten (benutzergetriebenen) und impliziten (automatischen) Reputationssystemen. Bei expliziten Reputationssystemen wird die Bewertung von den Nutzern des jeweiligen Systems abgegeben. Ein bekanntes Beispiel hierfür ist das Reputationssystem des Online-Auktionshauses eBay. Eine explizite Reputationsbewertung ist in der Wikipedia jedoch nicht praktikabel, da die MediaWiki-Software nicht darstellt, von welchem Autor ein Artikel bzw. einzelnen Textabschnitte erstellt wurden. Zur Ermittlung der Autorenschaft als Voraussetzung für eine benutzergetriebene Bewertung müsste daher zunächst die MediaWiki-Software erweitert oder die Versionsgeschichte manuell ausgewertet werden.

[1] http://www.alexa.com/topsites
[2] http://stats.wikimedia.org/DE/TablesWikipediaZZ.htm
[3] http://de.wikipedia.org/wiki/Wikipedia:Sprachen

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Im Kontext der Wikipedia haben sich in den vergangenen Jahren insbesondere implizite Reputationssysteme als eigenständiges Forschungsgebiet etabliert [1, 11]. Bei impliziten Reputationssystemen wird die Reputation automatisch berechnet. Im Falle der Wikipedia wird dabei durch Auswertung der Versionsgeschichte auf das Bearbeitungsverhalten der Autoren geschlossen und darauf aufbauend die Reputation errechnet. Die Schwierigkeit besteht dabei in der Auswahl geeigneter Metriken, die das typische Bearbeitungsverhalten von guten bzw. schlechten Autoren effektiv erfassen.

Reputationssysteme können in der Wikipedia unterschiedliche Funktionen erfüllen. Anhand des Reputationswertes lässt sich beispielsweise die Güte der Bearbeitungen eines Autors abschätzen [2] oder ein Qualitätsscore für gesamte Artikel berechnen. Auch kann das Reputationssystem genutzt werden, um Bearbeitungsrechte in der Wikipedia adäquat zu beschränken [1]. Des Weiteren kann das Reputationssystem die Autoren motivieren, sich mit qualitativ hochwertigen Bearbeitungen intensiv an der Wikipedia zu beteiligen [11].

Die Forschung zur automatischen Reputationsmessung in der Wikipedia steht bisher noch am Anfang. In den gegenwärtigen Publikationen werden bereits einzelne Metriken zur Reputationsmessung vorgeschlagen. Teilweise fehlt es aber an einer Bewertung der vorgeschlagenen Metriken. Darüber hinaus werden die einzelnen Metriken bisher nur isoliert untersucht.

Der vorliegende Beitrag adressiert diese Forschungslücke und untersucht zum einen zehn potentielle Metriken zur automatischen Reputationsmessung daraufhin, welche Einzelmetrik die größte Aussagekraft zur Reputationsmessung besitzt. Zum anderen werden die Metriken kombiniert und zu einer Reputationsfunktion zusammengeführt. Dahinter steht die Idee, dass durch eine solche Kombination verschiedene Facetten des Bearbeitungsverhaltens erfasst werden können. Dabei werden sowohl neu entwickelte als auch aus der Literatur bekannte Metriken berücksichtigt.

Zur Evaluation werden die Metriken anhand einer Diskriminanzanalyse zur Klassifikation zwischen schlechten (gesperrte Benutzer) und guten (nicht gesperrte Benutzer) Autoren in der Wikipedia herangezogen. Die Diskriminanzanalyse liefert die Reputationsfunktion mit maximaler Klassifikationsgüte, in die die Metriken linear mit einem Gewichtungsfaktor eingehen. Die ermittelten Gewichtungskoeffizienten spiegeln damit auch die Aussagekraft der einzelnen Metriken bei der Reputationsmessung wider.

Der Beitrag ist wie folgt aufgebaut. In Kapitel 2 wird der derzeitige Stand der Forschung zur automatischen Reputationsmessung in der Wikipedia vorgestellt. Die Beschreibung der untersuchten Metriken erfolgt in Kapitel 3. In Kapitel 4 wird die Evaluationsmethodik detailliert dargestellt. Nachfolgend werden im Kapitel 5 die Metriken anhand einer Diskriminanzanalyse evaluiert und kombiniert. Abschließend erfolgt in Kapitel 6 die Schlussbetrachtung.

2. STAND DER FORSCHUNG

Hinsichtlich der Qualitätsproblematik in der Wikipedia sind in den vergangenen Jahren zahlreiche Publikationen entstanden. In den Arbeiten von Potthast et al. [18], Smets et al. [21], Priedhorsky [19] und West et al. [27] wird Vandalismus in der Wikipedia untersucht. Bei der Vandalismusdetektion wird beispielsweise anhand der Bearbeitungskommentare oder durch den Vergleich von Hashwerten analysiert, inwieweit Bearbeitungen wieder rückgängig gemacht werden. Viégas et al. [25, 26] zeigen, dass Vandalismus in der Wikipedia in der Regel innerhalb von drei Minuten korrigiert wird.

Zahlreiche andere Publikationen untersuchen Metriken zur automatischen Qualitätsbewertung der Artikel in der Wikipedia [5, 8, 14, 15, 16, 24, 29, 30]. Als besonders effiziente Qualitätsindikatoren gelten die Länge eines Artikels [5] und Lebenszyklus-basierte Messgrößen wie der Umfang der durchschnittlichen persistenten (effektiven) Änderungen [16, 29].

Während die Forschung zur automatischen Qualitätsbewertung der Artikel bereits weit fortgeschritten ist und effiziente Metriken erarbeitet wurden, steht die automatische Reputationsmessung in der Wikipedia noch am Anfang. Anthony et al. untersuchen registrierte und anonyme Nutzer anhand der Bearbeitungshäufigkeit (Anzahl der Bearbeitungen) sowie anhand des Umfangs und der Effizienz (prozentualer Anteil der Bearbeitungen, die in der neusten Version des Artikels enthalten sind) der Bearbeitungen [3]. Stein und Hess analysieren als Reputationsmetrik die Mitarbeit des Autors bei der Erstellung von qualitativ hochwertigen Artikeln [23]. In den Publikationen von Adler et al. [1, 2] und Javanmardi et al. [11] werden zwei einander ähnliche Ansätze zur Reputationsmessung vorgeschlagen, die sich entsprechend der Ausführungen von Javanmardi et al. [11] hinsichtlich der Komplexität der Berechnung unterscheiden. Beide stellen jeweils eine Reputationsmetrik vor, die auf Grundlage der Persistenz von Bearbeitungen berechnet wird. Dabei bemisst sich die Persistenz anhand der Zeitspanne, in der eine Änderung in der Wikipedia Bestand hat.

3. METRIKEN ZUR REPUTATIONSMESSUNG

Im folgenden Abschnitt wird zunächst ein formales Modell der Wikipedia entworfen. Darauf aufbauend werden potentielle Metriken zur Reputationsmessung beschrieben und deren Berechnung anhand des formalen Wikipediamodells erläutert.

3.1 Modell der Wikipedia

Die Wikipedia besteht aus einer Menge von Artikeln $i = 0 \ldots n$. Bei der Bearbeitung eines Artikels i durch einen Benutzer entsteht eine jeweils neue Artikelversion $v_{i,j}$ mit $j = 0 \ldots m$ (Anzahl der Versionen eines Artikels), welche in der Versionsgeschichte abgespeichert wird. Die erste Version $v_{i,0}$ eines Artikels i wird in unserem Modell als leeres Dokument definiert. Diese Annahme weicht von der tatsächlichen Datenbank der Wikipedia ab, in der die erste Version bereits den Titel des Artikels und die erste Bearbeitung enthält. Diese erste Version entspricht der Version $v_{i,1}$ in unserem Modell. Die Definition einer vorgehenden Version $v_{i,0}$ als leere Version ist erforderlich, da so die Bearbeitung des ersten Autors im Modell formal korrekt erfasst wird. Der Autor $editor(v_{i,j})$ der Version $v_{i,j}$ bezeichnet den Benutzer, durch dessen Bearbeitung die neue Version entstanden ist.

Die im Folgenden erläuterten Metriken basieren überwiegend auf dem Unterschied zwischen verschiedenen Versionen eines Artikels. Die Ermittlung von Textdifferenzen kann prinzipiell mit unterschiedlicher Granularität erfolgen. Man kann zwischen Textdifferenzberechnungen auf Zeilen-, Wort- oder Zeichenebene unterscheiden. Je feingranularer die Textberechnung erfolgt, desto rechenaufwändiger ist die Kalkulation. In Analogie zu anderen Arbeiten der Wikipediaforschung [1, 11] werden die Textunterschiede in der vorliegenden Analyse auf Wortebene berechnet.

Dabei wird ein Wort als eine Zeichenfolge zwischen zwei Leerzeichen definiert. Die Änderung eines einzelnen Buchstabens eines Wortes wird somit als Löschen und Hinzufügen eines Wortes interpretiert. Die Berechnung der Textvergleiche erfolgt mit dem weit verbreiteten Algorithmus von Hunt und McIlroy [10], der beispielsweise im Linux-Programm diff implementiert ist. Der Algorithmus basiert auf der Zerlegung des Textes in einzelne Token. Hunt und McIlroy definieren in ihrer Publikation die einzelnen Zeilen eines Textes (Lines of Code) als Token und ermitteln so den Textunterschied auf Zeilenebene. In der vorliegenden Analyse werden zur Verbesserung der Genauigkeit die einzelnen Wörter des Textes als Token verwendet.

Der Differenztext *del(i,j,z)* aus der Version $v_{i,j}$ und einer früheren Version $v_{i,z}$ mit $z < j$ enthält alle zwischenzeitlich gelöschten Textabschnitte. Analog zu *del(i,j,z)* beinhaltet der Differenztext *add(i,j,z)* alle neu hinzugefügten Token. Die Differenz *diff(i,j,z)* gibt den Gesamtunterschied wieder und umfasst sowohl *del(i,j,z)* als auch *add(i,j,z)*. Die Differenz *diff(i,j,j-1)* bezeichnet den Unterschied einer Version $v_{i,j}$ zur direkten Vorgängerversion $v_{i,j-1}$ und quantifiziert damit die Bearbeitung des Autors *editor*($v_{i,j}$). Der gemeinsame Text zweier beliebiger Texte t_1 und t_2 wird mit *equal*(t_1,t_2) bestimmt. Die Wortanzahl eines beliebigen Textes t wird mit $|t|$ bezeichnet.

3.2 Metriken zur Reputationsmessung

Die in diesem Beitrag untersuchten Metriken zur Reputationsmessung lassen sich in die Kategorien Bearbeitungshäufigkeit, Gesamtbearbeitungsumfang, Persistenz, Umfang der Bearbeitungen, Beteiligung an Diskussionen und Beteiligung an qualitativ hochwertigen Artikeln untergliedern. Die Metriken sind entweder aus der bestehenden Literatur entnommen oder werden auf Grundlage der Ziele der Wikipedia neu vorgeschlagen. Bei der Bestimmung der Messgrößen werden mit Ausnahme der in Kapitel 3.2.5 vorgestellten Metrik nur Bearbeitungen im Artikelnamensraum berücksichtigt. Dieser Namensraum enthält die enzyklopädischen Artikel der Wikipedia. Bearbeitungen anderer Namensräume, wie Bearbeitungen an Benutzerseiten oder Textvorlagen, tragen nur indirekt zum Fortschritt der Wikipedia bei und werden deshalb von der Untersuchung ausgeschlossen. Tabelle 1 listet die im Beitrag untersuchten Metriken auf.

3.2.1 Bearbeitungshäufigkeit

Die Bearbeitungshäufigkeit ist ein vergleichsweise einfaches Kriterium, das beispielsweise von Anthony et al. [3] diskutiert wird. Es quantifiziert den Gesamtbeitrag eines Autors zur Wikipedia und bestimmt damit seine Erfahrung. Anthony et al. [3] definieren hierzu als Messgröße die Anzahl

(1) N^e_a

der Bearbeitungen eines Autors *a*. Im vorliegenden Beitrag wird als weitere potentielle Messgröße die Anzahl

(2) N^p_a

der Artikel, die durch den Autor *a* bearbeitet wurden, vorgeschlagen.

3.2.2 Gesamtbearbeitungsumfang

Ähnlich der Bearbeitungshäufigkeit ist der Gesamtumfang der Bearbeitungen eine neu vorgeschlagene, potentielle Messgröße zur Bestimmung der Autorenreputation, die die Erfahrung des Autors in der Wikipedia erfasst. Dieser Umfang lässt sich anhand der Differenzen aller Versionen eines Autors zur jeweiligen Vorgängerversion bestimmen. Die Metrik

(3) $N^w_a = \sum_i \sum_j diff(i,j,j\text{-}1)$ *mit* $editor(v_{i,j}) = a$

gibt die Anzahl der insgesamt von einem Autor *a* geänderten Wörter wieder.

Tabelle 1: Übersicht der untersuchten Metriken

Symbol	Bezeichnung	Quelle
$\mathbf{N^e_a}$	Anzahl der Bearbeitungen	[3]
$\mathbf{N^p_a}$	Anzahl der bearbeiteten Artikel	eigene Metrik
$\mathbf{N^w_a}$	Anzahl der insgesamt geänderten Wörter	eigene Metrik
$\mathbf{N^{pw}_a}$	Anzahl der persistent geänderten Wörter	eigene Metrik
$\mathbf{E_a}$	Effizienz	eigene Metrik
$\mathbf{Avg^w_a}$	durchschnittlicher Umfang der Bearbeitungen	[3]
$\mathbf{Avg^{pw}_a}$	durchschnittlicher Umfang der persistenten Änderungen	eigene Metrik
$\mathbf{max^{pw}_a}$	Umfang der größten persistenten Änderung	eigene Metrik
$\mathbf{N^{diss}_a}$	Anzahl der Bearbeitungen auf Diskussionsseiten	[12]
$\mathbf{N^{qh}_a}$	Beteiligung an qualitativ hochwertigen Artikeln	[23]

3.2.3 Persistenz

Die in der Literatur am häufigsten vorgeschlagenen Metriken zur Berechnung der Autorenreputation basieren auf der Persistenz der Bearbeitungen eines Autors [1, 3, 11]. Bei diesen Metriken wird bei einer hohen Verweildauer auf qualitativ hochwertige Änderungen geschlossen, da minderwertige Bearbeitungen frühzeitig von anderen Benutzern korrigiert werden [25, 26]. Bei der Berechnung der Persistenz ist sowohl das Einfügen als auch das Löschen von Text zu berücksichtigen, da auch das Löschen zur Verbesserung eines Artikels beitragen kann. Beispiele hierfür sind die Korrektur von Vandalismus und Spam-Einträgen.

In der Literatur werden unterschiedliche Berechnungsverfahren für die Persistenz der Bearbeitungen diskutiert, die jedoch verschiedene Nachteile aufweisen. Die Berechnungsverfahren von Adler et al. [1] und Javanmardi et al. [11] sind sehr rechenintensiv, da bei einer gegebenen Bearbeitung ein Textvergleich mit vielen Versionen durchzuführen ist. In der Publikation von Anthony et al. [3] wird die Persistenz daran bemessen, inwieweit die Bearbeitung in der jeweils neusten Version des Artikels enthalten ist. Diese Vorgehensweise weist den Nachteil auf, dass Löschungen nicht erfasst und somit nicht als persistente Änderungen erkannt werden.

In diesem Beitrag wird daher nicht auf eine Metrik aus der Literatur zurückgegriffen, sondern ein neuer Ansatz zur Bestimmung der Persistenz vorgeschlagen, der die Nachteile der bestehenden Verfahren vermeidet. Der Ansatz orientiert sich an den Überlegungen von Wöhner und Peters [29]. Wöhner und Peters analysieren in ihrer Publikation den Lebenszyklus von Artikeln und bezeichnen eine Änderung als persistent, falls die getätigte Bearbeitung bis zum Monatsende im Artikel verbleibt. Problematisch

hierbei ist jedoch, dass Änderungen am Ende eines Monats nur eine kurze Zeitspanne überdauern müssen und daher mit hoher Wahrscheinlichkeit als persistent klassifiziert werden. Aus diesem Grund wird im vorliegenden Beitrag als Persistenzkriterium eine einheitliche Mindestverweildauer verwendet, in der die Änderung nicht revidiert wird.

Bei der Festlegung der Mindestverweildauer sind zwei Aspekte zu berücksichtigen. Zum einen ist die Zeitspanne hinreichend groß zu wählen, damit unerwünschte Änderungen von der Wikipedia-Community innerhalb der Mindestverweildauer erkannt und revidiert werden. Das Erkennen und die Korrektur solcher unerwünschten Änderungen erfolgt in der Wikipedia in der Regel bereits innerhalb von drei Minuten [25, 26]. Anderseits darf die Zeitspanne nicht zu groß sein, da ansonsten wünschenswerte Änderungen im Rahmen der normalen Dynamik eines Artikels fälschlicherweise als nicht-persistent klassifiziert werden könnten.

Ausgehend von dieser Überlegung wurden Mindestverweildauern von einem Tag, zwei Tagen, zwei Wochen und zwei Monaten getestet, wobei die auf Basis der Persistenz berechneten Metriken nur geringfügig variierten. Die konkrete Ausgestaltung der Mindestverweildauer ist damit unkritisch für die ermittelten Ergebnisse. Die im Beitrag dargestellten Ergebnisse basieren auf einer Mindestverweildauer von zwei Wochen. Dieser Wert entspricht der durchschnittlichen Mindestverweildauer in der Untersuchung von Wöhner und Peters [29].

Im Unterschied zu den Verfahren von Adler et al. [1, 2] und Javanmardi et al. [11] wird in dem hier vorgestellten Ansatz die Persistenz nicht als metrische Größe interpretiert, die die exakte Zeitspanne misst, für die eine Änderung Bestand hat. Stattdessen wird eine Klassifikation in *persistente* und *nicht persistente* Änderungen vorgenommen. Im Gegensatz zu den Metriken von Javanmardi et al. [11] und Adler et al. [1] kann die Berechnung wesentlich schneller erfolgen, da deutlich weniger Textvergleiche benötigt werden. Bei Adler et al. [1] und Javanmardi et al. [11] muss bei einer gegebenen Änderung ein Abgleich mit allen folgenden Versionen durchgeführt werden, solange bis der entsprechende Textabschnitt nicht mehr vorhanden ist. Beim vorgestellten Ansatz bedarf es demgegenüber nur genau eines Vergleiches mit der entsprechenden Referenzversion. Der Ansatz von Anthony et al. [3] ist ähnlich schnell einzuschätzen, da auch hier nur ein Textvergleich erfolgt. Im Unterschied zu der hier vorgestellten Methodik berücksichtigt der Ansatz von Anthony et al. [3] jedoch keine Löschungen.

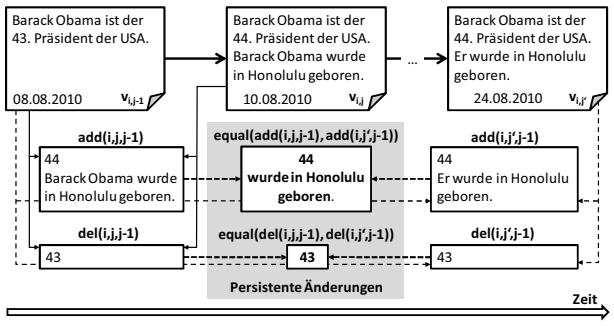


Abbildung 1: Persistente Änderungen am Beispiel

Auf Grundlage des hier vorgestellten Persistenzkriteriums wird die Anzahl der persistent geänderten Wörter eines Autors als neue potentielle Reputationsmetrik eingeführt. Die Berechnung vollzieht sich in mehreren Schritten, die in Abbildung 1 anhand eines fiktiven Beispiels veranschaulicht werden. Die Artikelversion, die nach der definierten Mindestverweildauer von zwei Wochen nach der Bearbeitung einer Artikelversion $v_{i,j}$ aktuell ist, wird im Folgenden mit $v_{i,j'}$ bezeichnet. Es werden zunächst für jede Bearbeitung des Autors sowohl die jeweils gelöschten *del(i,j,j-1)* als auch hinzugefügten Wörter *add(i,j,j-1)* berechnet. Anschließend werden die beiden Differenztexte *del(i,j',j-1)* und *add(i,j',j-1)* bestimmt, die die Änderung des Artikels innerhalb der Mindestverweildauer wiedergeben. Durch den Vergleich der berechneten Differenztexte *add(i,j,j-1)* und *add(i,j',j-1)* sowie *del(i,j,j-1)* und *del(i,j',j-1)* lässt sich ermitteln, inwieweit sich die Änderungen des Autors in den Änderungen des Artikels in der Mindestverweildauer wiederfinden und damit persistent sind. Somit lautet die Berechnungsvorschrift für die Anzahl der persistent geänderten Wörter eines Autors *a* wie folgt:

$$(4)\quad N^{pw}{}_a = \sum_i \sum_j (|equal(add(i,j',j\text{-}1), add(i,j,j\text{-}1))| + |equal(del(i,j',j\text{-}1), del(i,j,j\text{-}1))|) \text{ mit } editor(v_{i,j}) = a$$

Aufbauend auf der Persistenz $N^{pw}{}_a$ wird als weitere Metrik die Effizienz E_a eines Autors *a* vorgeschlagen. Die Effizienz E_a bezeichnet den Anteil der persistenten Änderungen $N^{pw}{}_a$ am Gesamtbearbeitungsumfang $N^{w}{}_a$ und wird mit

$$(5)\quad E_a = N^{pw}{}_a / N^{w}{}_a$$

berechnet. Geht man davon aus, dass das Verhalten eins Autors in der Vergangenheit auch Rückschlüsse auf das zukünftige Verhalten zulässt, kann die Effizienz als die Wahrscheinlichkeit interpretiert werden, mit der die Bearbeitung eines Autors von der Community akzeptiert wird.

3.2.4 Umfang der Bearbeitungen

Als weiteres Charakteristikum eines Autors wird in der Literatur der Umfang der einzelnen Bearbeitungen diskutiert [3]. Damit wird der Aufwand des Autors für seine Beiträge quantifiziert. Im Rahmen dieses Beitrages werden hierzu als Messgrößen der durchschnittliche Umfang der Bearbeitungen

$$(6)\quad Avg^{w}{}_a = N^{w}{}_a / N^{e}{}_a,$$

der durchschnittliche Umfang der persistenten Änderungen

$$(7)\quad Avg^{pw}{}_a = N^{pw}{}_a / N^{e}{}_a,$$

sowie der Umfang der größten persistenten Änderung

$$(8)\quad max^{pw}{}_a = \max_{i,j}(|equal(add(i,j',j\text{-}1), add(i,j,j\text{-}1))| + |equal(del(i,j',j\text{-}1), del(i,j,j\text{-}1))|) \text{ mit } editor(v_{i,j}) = a$$

eines Autors *a* betrachtet.

3.2.5 Beteiligung an Diskussionen

Kittur et al. [12] konnten in ihrer Studie feststellen, dass Beiträge auf den Diskussionsseiten eines Artikels zu Qualitätssteigerungen führen können. Aus diesem Grund wird mit der Anzahl der Bearbeitungen auf Diskussionsseiten

$$(9)\quad N^{diss}{}_a$$

die Beteiligung des Autors an Diskussionen als weitere Reputationsmetrik vorgeschlagen.

3.2.6 Beteiligung an qualitativ hochwertigen Artikeln

Stein und Hess [23] gehen davon aus, dass sich gute Autoren vorwiegend an qualitativ hochwertigen Artikeln beteiligen. Die Artikel, die den höchsten Qualitätsansprüchen genügen, werden in der Wikipedia als exzellente Artikel gekennzeichnet [28]. Stein und Hess verwenden daher als Reputationsmetrik das Verhältnis aus der Anzahl der Bearbeitungen an exzellenten Artikeln zur Anzahl an Bearbeitungen insgesamt. Neben den exzellenten Artikeln gelten in der Wikipedia die als lesenswert markierten Artikel ebenfalls als qualitativ hochwertig [28]. Im vorliegenden Beitrag wird deshalb in Anlehnung an Stein und Hess das Verhältnis

$$(10)\ N^{qh}{}_a = E^{qh}{}_a / N^e{}_a$$

aus der Anzahl der Änderungen an exzellenten und lesenswerten Artikeln $E^{qh}{}_a$ zur der Anzahl der Bearbeitungen insgesamt $N^e{}_a$ als weitere Reputationsmetrik vorgeschlagen.

4. EVALUATION

Die Ziele der Evaluation bestehen darin, zum einen die vorher beschriebenen Metriken vergleichend zu bewerten und deren Effizienz im Hinblick auf die Reputationsmessung zu beurteilen. Zum anderen soll eine Reputationsfunktion ermittelt werden, die die Metriken miteinander kombiniert und so eine effiziente Reputationsmessung gewährleistet. Im vorliegenden Abschnitt werden hierzu die Evaluationsmethodik und der verwendete Datenbestand vorgestellt.

Die Evaluation der vorgeschlagenen Metriken wird anhand des tatsächlichen Datenbestandes der Wikipedia durchgeführt, der als XML-Dump zur Verfügung steht. Der Datenbestand enthält sowohl die aktuellen Artikelversionen als auch die gesamte Versionsgeschichte. In der Versionsgeschichte sind die Quelltexte aller Artikelversionen inklusive HTML-Code und Wiki-Tags enthalten. Darüber hinaus beinhaltet die Versionsgeschichte zu jeder Bearbeitung weitere Metainformationen, wie den Bearbeitungszeitpunkt, den Benutzernamen des Autors und eventuelle Bearbeitungskommentare. Erfolgt die Bearbeitung durch einen nicht angemeldeten, anonymen Benutzer ist anstelle des Benutzernamens die jeweilige IP-Adresse gespeichert. Zur Evaluation wird der Datenbestand der deutschsprachigen Wikipedia vom 21. Januar 2008 ausgewertet, der zur Analyse in eine SQL-Datenbank importiert wurde. Der Datensatz enthält *646.099* enzyklopädische Artikel und *217.398* registrierte Benutzer.

Die Metriken zur automatischen Reputationsmessung werden im Folgenden anhand ihrer Trennschärfe bei der Klassifikation zwischen guten und schlechten Autoren bewertet. Dabei wird angenommen, dass eine Metrik die treffgenau gute und schlechte Autoren klassifiziert auch die Reputation der Autoren widerspiegelt. Die bei der Klassifikation erzielte Trefferquote wird daher als Maß für die Eignung einer Metrik zur Reputationsmessung interpretiert. Die Vorgehensweise, Metriken mittels einer Klassifikation zu evaluieren, ist in der Wikipediaforschung weit verbreitet und akzeptiert [5, 8, 11, 16, 24, 29, 30].

Als Grundlage für die Evaluation werden sowohl eine Menge guter als auch eine Menge schlechter Autoren benötigt, die dann anhand der Metriken zu klassifizieren sind. Hierzu wird auf Wikipedia-interne Benutzergruppierungen zurückgegriffen. Die Verwendung Wikipedia-interner Merkmale ist ebenfalls eine typische Vorgehensweise in der Wikipediaforschung. Auch bei der Evaluation von Metriken zur automatischen Qualitätsmessung werden oft Wikipedia-interne Bewertungen genutzt [5, 8, 16, 24, 29, 30].

Als Beispiele für schlechte Autoren werden gesperrte Benutzer verwendet. Dies sind Nutzer, die mutwillig Artikel zerstören oder gegen geltende Grundprinzipien der Wikipedia (Beachtung des Neutral Point of View, fairer Umgang untereinander) verstoßen und deshalb durch einen Administrator gesperrt wurden [28]. Alle anderen angemeldeten Benutzer können im Vergleich zu den gesperrten Benutzern als gute Autoren interpretiert werden. In der vorliegenden Analyse wird deshalb eine Klassifikation zwischen gesperrten Nutzern als schlechte Autoren und nicht gesperrten Benutzern als gute Autoren durchgeführt. Aus der Trefferquote dieser Klassifikation wird allgemein auf die Eignung der jeweiligen Metrik zur Reputationsmessung geschlussfolgert.

Im Unterschied zu dieser Vorgehensweise nutzen Javanmardi et al. [11] die Gruppe der Administratoren als Beispiel für gute Autoren. Administratoren sind spezielle Benutzer, die mit erweiterten Nutzerrechten ausgestattet sind und besondere Verwaltungsaufgaben wie das Sperren von Benutzern oder Wikipediaseiten und die Löschung von Seiten wahrnehmen [28]. Den Status eines Administrators erhält ein Benutzer, falls seine Kandidatur zum Administrator von der Community durch eine erfolgreiche Abstimmung bestätigt wird. Die Klassifikation zwischen Administratoren und gesperrten Benutzern ist im vorliegenden Beitrag jedoch nicht als Evaluationsmethodik geeignet. Bei der Kandidatur für den Status des Administrators wird erwartet, dass sich der entsprechende Benutzer über einen längeren Zeitraum an der Wikipedia beteiligt und mehr als *1.000* Bearbeitungen im Artikelnamensraum durchgeführt hat [28]. Somit sind bei Administratoren per Definition die Metriken der Bearbeitungshäufigkeit und des Bearbeitungsumfangs besonders intensiv ausgeprägt und das Evaluationsergebnis würde entsprechend verfälscht werden.

Die gesperrten Benutzer wurden anhand der Benutzerseiten identifiziert. Benutzerseiten sind Wikipediaseiten, auf denen sich registrierte Benutzer präsentieren können. Im Falle einer Benutzersperrung wird der ursprüngliche Inhalt von einem Administrator gelöscht und durch einen Hinweis auf die Benutzersperrung ersetzt [28]. In der Untersuchung wurden alle Benutzerseiten nach dem entsprechenden Sperrhinweis geparst.

Insgesamt enthält der Datensatz *1620* gesperrte Benutzer. Um die Gruppen der guten und schlechten Autoren in der Analyse mit gleichen Gewichten zu berücksichtigen, werden aus dem Datensatz zufällig *1620* nicht gesperrte Benutzer ausgewählt. Dabei werden alle registrierten nicht-gesperrten Benutzer mit der Ausnahme von Bots mit gleicher Wahrscheinlichkeit berücksichtigt. Insgesamt standen somit *215.642* Benutzer für die Stichprobenziehung zur Verfügung. Anonyme Benutzer werden bei der Klassifikation ausgeschlossen, da aufgrund der in der Regel dynamischen Vergabe von IP-Adressen keine dauerhaft eindeutige Zuordnung von Benutzern zu IP-Adressen gegeben ist. Die *3.240* ausgewählten Autoren sind an ca. *1,5* Millionen der insgesamt *26,3* Millionen Artikelversionen in der Wikipedia beteiligt. Damit werden durch die Analyse ca. *5%* des gesamten Artikelnamensraumes der Wikipedia ausgewertet.

Für die Autoren der Stichprobe wurden alle Messgrößen wie im vorherigen Kapitel beschrieben berechnet. Einfache Messgrößen, wie die Anzahl der Bearbeitungen $N^e{}_a$ oder die Anzahl der bearbeiteten Artikel $N^p{}_e$, lassen sich mit Hilfe einer einfachen SQL-Abfrage bestimmen. Zur Berechnung der komplexen Metriken wurde ein Satz einfacher Java-Programme implementiert.

Die Klassifikation anhand der vorgeschlagenen Metriken wird auf Basis einer linearen Diskriminanzanalyse [4] durchgeführt. Im Rahmen der Diskriminanzanalyse wird zunächst eine Diskriminanzfunktion der Form

$D(x_1...x_n) = a*x_1 + b*x_2 + \ldots + j*x_{10} + k$

berechnet.

Die Variablen x_1 bis x_{10} bezeichnen die Werte der untersuchten Metriken beim jeweiligen Autor, a bis j sind Gewichtungskoeffizienten und k ist die additive Konstante. Die Berechnung der Diskriminanzfunktion erfolgt so, dass eine bestmögliche Trennung beider Gruppen erreicht wird und sich die Funktionswerte der Diskriminanzfunktion in beiden Gruppen maximal unterscheiden.

Die Diskriminanzanalyse erfüllt damit beide einleitend genannten Evaluationsziele. Zum einen spiegeln die Gewichtungskoeffizienten die Bedeutung der einzelnen Metriken im Hinblick auf die Reputationsmessung wider. Dadurch wird der angestrebte Vergleich zwischen den verschiedenen Metriken realisiert. Zum anderen ermittelt die Diskriminanzanalyse eine effiziente Kombination der Metriken. Darüber hinaus wird im Rahmen der Diskriminanzanalyse eine Korrelationsanalyse durchgeführt und dadurch aufgedeckt, welche Metriken substituierbar sind.

In einem abschließenden Schritt wird anhand der Diskriminanzfunktion eine automatische Klassifikation zwischen nicht gesperrten und gesperrten Autoren vorgenommen und so die Güte der Diskriminanzfunktion quantifiziert. Hierzu werden zunächst die Verteilungen der Diskriminanzfunktionswerte in den beiden Gruppen berechnet. Daraus lässt sich bei gegebenem Diskriminanzfunktionswert eines Autors die Wahrscheinlichkeiten für dessen Zugehörigkeit zur Gruppe der gesperrten bzw. nicht gesperrten Autoren berechnen. Der Autor wird der Gruppe mit der höchsten Wahrscheinlichkeit zugeordnet. Die Diskriminanzanalyse wird im Rahmen der Evaluation mit Hilfe des Statistikprogramms *SPSS 17.0* durchgeführt. SPSS verwendet sowohl zur Bestimmung der Diskriminanzfunktion als auch zur Klassifikation die vollständige Stichprobe.

5. Ergebnisse

In diesem Kapitel werden die Ergebnisse der Analyse präsentiert. Zunächst werden die im Rahmen der Korrelationsanalyse aufgedeckten Abhängigkeiten der Metriken dargestellt. Danach erfolgt die vergleichende Evaluierung der Metriken. Abschließend wird die Diskriminanzfunktion als effektive Kombination der Metriken vorgestellt und bewertet.

5.1 Korrelationsanalyse

Die im Rahmen der Diskriminanzanalyse berechnete Korrelationsmatrix ist in Tabelle 2 dargestellt. Mit einem Korrelationskoeffizienten von *0,983* korrelieren die Anzahl der Bearbeitungen N^e_a und die Anzahl der bearbeiteten Artikel N^p_a sehr stark. Dies bedeutet, dass Autoren die viele Bearbeitungen durchführen auch viele verschiedene Artikel bearbeiten. Des Weiteren zeigt sich eine sehr starke Korrelation von *0,999* zwischen dem Gesamtbearbeitungsumfang N^w_a und der Anzahl der persistent geänderten Wörter N^{pw}_a.

Aufgrund der hohen Korrelationen zwischen N^e_a und N^p_a sowie zwischen N^w_a und N^{pw}_a sind die Variablen eines Paares miteinander substituierbar. Es ist daher sinnvoll, für die weitere Analyse jeweils nur eine Metrik stellvertretend für jedes Variablenpaar zu betrachten. Für die Bestimmung der Diskriminanzfunktion werden N^e_a und N^w_a gewählt, da deren Berechnung weniger aufwändig ist.

Häufig wird gegenüber der Wikipedia die Kritik geäußert, dass insbesondere Änderungen von erfahrenen Autoren von der Community akzeptiert werden. Diese Kritik kann durch die vorliegende Untersuchung widerlegt werden. Messgrößen, die die Erfahrung eines Autors quantifizieren (N^p_a, N^e_a, N^w_a, N^{pw}_a), korrelieren mit Korrelationskoeffizienten zwischen *0,044* bis *0,055* nur sehr gering mit der Effizienz E_a.

Tabelle 2: Korrelation der Metriken

Korrelation	N^e_a	N^p_a	N^w_a	N^{pw}_a	max^{pw}_a
N^e_a	1	0,983	0,561	0,559	0,321
N^p_a	0,983	1	0,517	0,517	0,292
N^w_a	0,561	0,517	1	0,999	0,719
N^{pw}_a	0,559	0,517	0,999	1	0,719
max^{pw}_a	0,321	0,292	0,719	0,719	1
E_a	0,052	0,054	0,044	0,047	0,039
Avg^w_a	0	0	0,012	0,008	0,008
Avg^{pw}_a	-0,01	-0,02	0,146	0,145	0,135
N^{qh}_a	-0,06	-0,07	-0,041	-0,04	-0,027
N^{diss}_a	0,434	0,38	0,383	0,371	0,198

Korrelation	E_a	Avg^w_a	Avg^{pw}_a	N^{qh}_a	N^{diss}_a
N^e_a	0,052	-0,004	-0,007	-0,062	0,434
N^p_a	0,054	-0,004	-0,017	-0,066	0,38
N^w_a	0,044	0,012	0,146	-0,041	0,383
N^{pw}_a	0,047	0,008	0,145	-0,041	0,371
max^{pw}_a	0,039	0,008	0,135	-0,027	0,198
E_a	1	-0,129	0,186	-0,129	0,053
Avg^w_a	-0,129	1	0,21	0,188	-0,005
Avg^{pw}_a	0,186	0,21	1	0,063	0,055
N^{qh}_a	-0,129	0,188	0,063	1	-0,043
N^{diss}_a	0,053	-0,005	0,055	-0,043	1

5.2 Vergleich der Metriken

In der Diskriminanzanalyse beschreibt die Strukturmatrix die Bedeutung der verwendeten Metriken bei der Klassifikation. Die Koeffizienten der Strukturmatrix geben die Korrelation der jeweiligen Metrik zur Diskriminanzfunktion an. Dementsprechend lässt sich auf Basis der Strukturmatrix die Aussagekraft der verwendeten Metriken für die Reputationsmessung vergleichend bewerten. Die Strukturmatrix ist in Tabelle 3 dargestellt.

Die Analyse zeigt, dass sich gute und schlechte Autoren besonders hinsichtlich der Effizienz E_a unterscheiden. Mit *0,965* erhält diese Metrik ein sehr hohes Gewicht in der Strukturmatrix. Im Vergleich zu allen anderen Metriken kann damit die Reputation eines Autors am besten beschrieben werden. In Tabelle 4 sind die Mittelwerte μ und die Standardabweichungen σ der untersuchten Metriken in den beiden Autorengruppen aufgelistet. Die Aussagekraft von E_a lässt sich auch an den entsprechenden Mittelwerten erkennen. So haben gesperrte Nutzer eine durchschnittliche Effizienz E_a von *23%* und nicht gesperrte Benutzer von *85%*. Die Analyse belegt daher, dass die Änderungen von schlechten Autoren zum größten Teil durch die Community verworfen werden,

wohingegen die Änderungen von guten Autoren mit einer hohen Wahrscheinlichkeit akzeptiert werden. Die Studien von Adler et al. [1], Javanmardi et al. [11] und Anthony et al. [3], die als Reputationsmetrik eine zu E_a ähnliche Metrik vorschlagen, werden somit durch die vorliegende Analyse bestätigt.

Tabelle 3: Vergleich der Metriken

	Korrelation zur Diskriminanzfunktion (Strukturmatrix)
$\mathbf{E_a}$	0,965
$\mathbf{Avg^{pw}_a}$	0,2
$\mathbf{N^{diss}_a}$	0,162
$\mathbf{N^{w}_a}$	0,147
$\mathbf{N^{e}_a}$	0,141
$\mathbf{N^{qh}_a}$	0,112
$\mathbf{Avg^{w}_a}$	-0,111
$\mathbf{max^{pw}_a}$	0,109

Mit Ausnahme der Effizienz E_a korrelieren alle anderen Metriken mit der Diskriminanzfunktion ähnlich stark. Die jeweiligen Korrelationskoeffizienten liegen im Bereich von *0,109* bis *0,2*. Damit kann neben der Effizienz allen weiteren untersuchten Metriken eine ähnliche Aussagekraft zugeschrieben werden. Diese ist jedoch im Vergleich zur Effizienz deutlich geringer. Die Mittelwerte der Metriken in den beiden Autorengruppen zeigen, dass alle Metriken bei den nicht gesperrten Nutzern höhere Werte als bei den gesperrten Nutzern aufweisen. Beispielsweise führen nicht gesperrte Nutzer im Durchschnitt persistente Änderungen $(N^{pw}{}_a)$ im Umfang von ca. *82.000* Wörtern und gesperrte Benutzer im Durchschnitt von nur ca. *1.600* Wörtern durch.

Tabelle 4: Mittelwerte und Standardabweichungen bei gesperrten und nicht gesperrten Benutzern

	gesperrte Benutzer		nicht gesperrte Benutzer	
	μ	σ	μ	σ
$\mathbf{N^{e}_a}$	65	598	2.774	12.629
$\mathbf{N^{p}_a}$	29	295	1.796	8.473
$\mathbf{N^{w}_a}$	3.116	22.476	88.346	419.374
$\mathbf{N^{pw}_a}$	1.631	18.262	81.988	403.512
$\mathbf{max^{pw}_a}$	167	942	3.705	22.506
$\mathbf{E_a}$	23,00%	33,00%	85,00%	19,00%
$\mathbf{Avg^{w}_a}$	219	904	44	159
$\mathbf{Avg^{pw}_a}$	11	47	32	43
$\mathbf{N^{qh}_a}$	7,00%	20,00%	13,00%	21,00%
$\mathbf{N^{diss}_a}$	16	109	128	454

Als einzige Ausnahme zeigt sich beim durchschnittlichen Umfang der Bearbeitungen $Avg^{w}{}_a$ ein umgekehrter Trend. Hier deutet ein kleiner Wert auf eine gute Reputation hin. Während nicht gesperrte Benutzer durchschnittlich *44* Wörter pro Bearbeitung ändern, ändern gesperrte Benutzer im Durchschnitt *219* Wörter. Bei der durchschnittlichen persistenten Änderung pro Bearbeitung $Avg^{pw}{}_a$ ist diese Tendenz nicht zu beobachten. Dies lässt sich damit erklären, dass gesperrte Benutzer umfangreiche Bearbeitungen wie das Löschen des gesamten Textes oder das Hinzufügen langer, unsinniger Textabschnitte vornehmen und die entsprechenden Bearbeitungen sehr schnell von anderen Wikipedia-Benutzern korrigiert werden.

Zusätzlich zur Strukturmatrix lässt sich ein Vergleich der Metriken realisieren, indem jede einzelne Metrik im Rahmen einer Diskriminanzanalyse isoliert betrachtet wird. Die Trefferquoten dieser Auswertung sind in Tabelle 5 abgebildet.

Tabelle 5: Trefferquoten der Einzelmetriken

	Trefferquote
$\mathbf{E_a}$	85,1%
$\mathbf{Avg^{pw}_a}$	67,3%
$\mathbf{N^{diss}_a}$	59,9%
$\mathbf{N^{w}_a}$	61,2%
$\mathbf{N^{e}_a}$	62,3%
$\mathbf{N^{qh}_a}$	58,4%
$\mathbf{Avg^{w}_a}$	56,8%
$\mathbf{max^{pw}_a}$	64,7%

Diese Analyse bestätigt die besondere Bedeutung der Effizienz E_a, die sich bereits aus der Strukturmatrix ergibt. Mit der Effizienz konnte die deutlich höchste Trefferquote von *85,1%* erzielt werden. Trotz der deutlichen Unterschiede der Mittelwerte μ zwischen gesperrten und nicht gesperrten Autoren (Tabelle 4) erreichen alle anderen Metriken eine vergleichsweise niedrige Trefferquote zwischen *56,8%* und *67,3%*. Dies deutet darauf hin, dass die deutlichen Unterschiede der Mittelwerte auf wenige Autoren mit besonders extremen Werteausprägungen zurückzuführen sind. Diese Auswertung belegt, dass sich die Metriken nicht für eine isolierte Verwendung in einem Reputationssystem eignen, da die Reputation ungenau erfasst wird. Diese sind Metriken hauptsächlich für eine kombinierte Reputationsfunktion anwendbar, um das Bearbeitungsverhalten der Autoren möglichst umfassend abzubilden.

5.3 Definition und Bewertung der Diskriminanzfunktion

Die Diskriminanzanalyse fügt die verschiedenen Metriken zu einer gemeinsamen Funktion zusammen. Die Gewichtung der einzelnen Metriken in der Diskriminanzfunktion entspricht dabei ihrer Signifikanz für die Klassifikation zwischen guten und schlechten Autoren.

Tabelle 6: Diskriminanzfunktion

	Gewichtungskoeffizient
$\mathbf{E_a}$	3,63979532
$\mathbf{Avg^{pw}_a}$	-0,00009770
$\mathbf{N^{diss}_a}$	0,00020281
$\mathbf{N^{w}_a}$	0,00000012
$\mathbf{N^{e}_a}$	0,00000437
$\mathbf{N^{qh}_a}$	1,15645448
$\mathbf{Avg^{w}_a}$	-0,00004466
$\mathbf{max^{pw}_a}$	0,00000104
Konstante	-2,12285923

Die von SPSS berechnete effiziente Diskriminanzfunktion ist in Tabelle 6 dargestellt. Aufgrund der bestehenden Korrelationen werden für einige Metriken auch negative Gewichtungskoeffizienten berechnet. Die ermittelten Koeffizienten sind nicht ausschließlich von der Aussagekraft der Metrik abhängig, sondern werden ebenfalls durch den jeweiligen Wertebereich beeinflusst. Da die Effizienz $(0 \leq E_a \leq 1)$ und die Beteiligung bei qualitativ hochwertigen Artikeln $(0 \leq N^{qh}_a \leq 1)$ im Vergleich zu den anderen Metriken sehr kleine Wertebereiche aufweisen (Tabelle 4), werden beide Metriken mit sehr hohen Gewichten berücksichtigt.

Anhand der Klassifikationsgüte lässt sich die Güte der Diskriminanzanalyse beurteilen. Tabelle 7 stellt das Ergebnis der Klassifikation auf Basis der Diskriminanzanalyse dar. Durch die berechnete Diskriminanzfunktion werden insgesamt *86,5%* der Autoren des Datensatzes richtig klassifiziert. Die True-Positive-Rate (TPR) bezeichnet den Anteil der richtig klassifizierten nicht gesperrten Benutzer und beträgt *93,5%.* Die False-Positive-Rate (FPR) als Anteil der fälschlicherweise als nicht gesperrt klassifizierten gesperrten Benutzer beträgt *20,6%.* Die hohen Trefferquoten zeigen, dass sich die Diskriminanzfunktion gut zur Reputationsmessung in der Wikipedia eignet.

Tabelle 7: Klassifikationsergebnis

	Anzahl	**Anteil**
TPR	1.514	93,50%
FPR	333	20,60%
Trefferquote	***2.801***	***86,50%***

Zur Analyse der Fehlklassifikationen wurden strichprobenartig einige falsch klassifizierte Testfälle ausgewählt und deren Änderungen in der Wikipedia manuell nachvollzogen. Bei gesperrten Nutzern wurde ergänzend das *Benutzersperr-Logbuch* ausgewertet. Darin sind alle Sperrvorgänge und die durch die Administratoren benannten Gründe für die Sperrung aufgelistet.

Die Fehlklassifikationen bei den gesperrten Nutzern lassen sich größtenteils darauf zurückführen, dass die Sperrung aufgrund von Meinungsverschiedenheiten bzw. Konflikten mit anderen Nutzern oder Administratoren ausgesprochen wurde. Die jeweiligen Autoren haben in der Regel vor der Sperrung umfangreiche Bearbeitungen mit einer hohen Effizienz vorgenommen, so dass ein entsprechend hoher Diskriminanzfunktionswert berechnet wird. Fehler bei den nicht gesperrten Benutzern sind darauf zurückzuführen, dass sich innerhalb dieser Gruppe auch schlechte Benutzer befinden, die beispielsweise Vandalismus begehen.

Durch die Kombination der verschiedenen Metriken in der Diskriminanzfunktion kann die Effektivität der Reputationsmessung gegenüber einer Einzelbetrachtung der Metriken verbessert werden. Im Vergleich zur Einzelbetrachtung der Effizienz (Tabelle 5) kann die Trefferquote um *1,4%* verbessert werden. Bei den anderen Metriken fällt dieser Unterschied wesentlich deutlicher aus.

Ein weiterer Vorteil der hier berechneten Diskriminanzfunktion zeigt sich am Beispiel der Administratoren. Diese als sehr gut geltenden Autoren erzielen die höchsten Diskriminanzfunktionswerte. Durch die Kombination der Metriken werden verschiedene Facetten ihres Bearbeitungsverhaltens abgebildet, sodass sich Administratoren hinsichtlich ihrer Reputation deutlicher von anderen Benutzern abgrenzen können.

6. SCHLUSSBETRACHTUNG

Im Rahmen der Schlussbetrachtung wird zunächst der Beitrag kurz zusammengefasst, anschließend die Untersuchung kritisch gewürdigt und ein Ausblick auf weitere Forschungsfragen gegeben.

6.1 Zusammenfassung

In diesem Beitrag wurden zunächst potentielle Messgrößen zur automatischen Messung der Autorenreputation vorgestellt. Die Metriken umfassen die Kategorien Bearbeitungshäufigkeit, Gesamtbearbeitungsumfang, Umfang der einzelnen Bearbeitungen, Beteiligung an qualitativ hochwertigen Artikeln, Beteiligung an Diskussionen und die Persistenz der Bearbeitungen. Die Metriken wurden aus der bestehenden Literatur entnommen und um eigene Vorschläge ergänzt. In der Kategorie der Persistenz wurde ein neuer Berechnungsansatz vorgestellt. Während bestehende Ansätze genau die Zeitspanne bestimmen, die eine Änderung Bestand hat, werden in diesem Beitrag Änderungen dann als persistent eingestuft, wenn sie für mindestens zwei Wochen Bestand haben. Dadurch wird die Anzahl der notwendigen Textvergleiche deutlich reduziert und die Berechnung kann wesentlich schneller erfolgen. Dies ist ein wichtiger Aspekt im Hinblick auf eine Implementierung in der Praxis, da seitens der Wikipedia insbesondere Performanceprobleme als Hürde bei der Einführung neuer Konzepte genannt werden.

Des Weiteren wurden erstmalig in der Wikipediaforschung zur Reputationsmessung verschiedene Metriken anhand einer Diskriminanzanalyse vergleichend evaluiert. Die neu vorgeschlagene Metrik der *Effizienz der Bearbeitungen* erhält dabei ein sehr hohes Gewicht und eignet sich demnach besonders gut zur Reputationsmessung. Die vorliegende Studie bestätigt damit die Forschungsarbeiten von Adler et al. [1, 2], Javanmardi et al. [11] sowie Anthony et al. [3], die eine vergleichbare Metrik vorgeschlagen haben.

Als weiteres Ergebnis wurden erstmals verschiedene Metriken durch die Diskriminanzanalyse zu einer Reputationsfunktion zusammengeführt. Bei der Gestaltung einer automatischen Reputationsmessung besteht ein Aspekt darin, das Bearbeitungsverhalten der Autoren möglichst vollständig zu erfassen. Durch die vorgeschlagene Kombination der Metriken können verschiedene Facetten des Bearbeitungsverhaltens der Autoren quantifiziert werden. Deshalb kann im Vergleich zur isolierten Betrachtung der Metriken, welche in den bestehenden Forschungsarbeiten angewendet wird, eine aussagekräftigere Reputationsmessung erreicht werden.

Anhand der berechneten Reputationsfunktion wird mit einer Trefferquote von *86,5%* zwischen guten (nicht gesperrten Benutzern) und schlechten Autoren (gesperrten Benutzern) unterschieden. Vor allem gute Autoren konnten mit einer True-Positive-Rate von *93,5%* sehr gut klassifiziert werden. Die Trefferquoten zeigen, dass die ermittelte Reputationsfunktion eine effektive automatische Reputationsmessung in der Wikipedia ermöglicht.

6.2 Kritische Würdigung und Ausblick

Die vorgestellte Studie zeigt, dass in der Wikipedia eine automatische Klassifikation von Benutzergruppen effizient umsetzbar ist. Diese Idee könnte zukünftig genutzt werden, um beispielsweise den Status des Administrators automatisch zu vergeben. Eine solche Vorgehensweise wird in der deutschen Wikipedia bereits beim Sichterstatus angewendet. Durch eine Diskriminanzanalyse zwischen Administratoren und angemeldeten nicht-gesperrten

Benutzern kann eine effektive Kombination von Metriken zur Vergabe des Administratorstatus berechnet werden.

Die hier eingeführte Metrik der Effizienz der Bearbeitung kann in weiteren Forschungsarbeiten mit den Ansätzen von Adler et. al. [1, 2] und Javanmardi et. al. [11] verglichen werden. Durch eine Implementierung dieser beiden Ansätze kann der Performancegewinn und die Aussagekraft im Vergleich zu den hier vorgestellten Verfahren quantifiziert werden.

Ferner kann untersucht werden, inwieweit sich ich das vorgeschlagene Reputationssystem im Hinblick auf die Effizienz der Reputationsmessung verbessern lässt. Hierzu können zunächst die Autoren anhand der Metriken geclustert werden, um so typische Bearbeitungsmuster aufzudecken. Beispielsweise ist es denkbar, dass einige Benutzer vorwiegend Korrekturen vornehmen oder Vandalismus beseitigen. Andere Benutzer wiederum schreiben die Artikel fort und fügen neuen Text hinzu. Für die identifizierten Cluster lassen sich möglicherweise spezielle Reputationsfunktionen definieren, die zu einer effizienteren Reputationsmessung führen. Die clusterbasierte Reputationsmessung kann beispielsweise mit Hilfe von Entscheidungsbäumen umgesetzt werden.

Des Weiteren soll das Reputationssystem prototypisch in die von der Wikipedia verwendete MediaWiki-Software integriert werden. In praktischen Tests sollen dann anhand des Prototyps die Wirkung auf das Benutzerverhalten und die Aussagekraft des Reputationssystems experimentell validiert werden. Darüber hinaus bietet ein Prototyp die Möglichkeit, eventuelle Performance-Einbußen aus der Reputationsberechnung abzuschätzen.

Das hier vorgestellte Reputationssystem kann in der Wikipedia in verschiedener Weise genutzt werden. So bietet das Reputationssystem eine Basis für ein effektives Rechtemanagement. Auch motiviert das Reputationssystem Autoren sich intensiv an der Wikipedia zu beteiligen. Eine weitere Anwendung besteht darin, anhand der Autorenreputation einen Qualitätsscore für die Artikel zu berechnen. Ein solches Verfahren wird in einer folgenden Publikation vorgestellt. Erste Untersuchungen zeigen, dass dadurch eine sehr effiziente Qualitätsmessung möglich ist.

7. LITERATUR

[1] Adler, B.T. und Alfaro, L. 2007. A Content-Driven Reputation System for the Wikipedia. In *Proceedings of the 16th International Conference on the World Wide Web* (Banff, Kanada, 08.-12. Mai 2007). WWW2007. ACM, New York, NY, 261-270. DOI=http://doi.acm.org/10.1145/1242572.1242608.

[2] Adler, B.T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I. und Raman, V. 2008. Assigning Trust To Wikipedia Content. In *Proceedings of the 2008 International Symposium on Wikis* (Porto, Portugal, 08.-10. September 2008). WikiSym 2008. ACM, New York, NY. DOI=http://doi.acm.org/10.1145/1822258.1822293.

[3] Anthony, D., Smith, S. W. und Williamson, T. 2007. *The Quality of Open Source Production: Zealots and Good Samaritans in the Case of Wikipedia*. Technical Report, TR2007-606. Department of Computer Science, Dartmouth College.

[4] Backhaus, K., Erichson, B., Wulff, P. und Weiber, R. 2008. *Multivariante Analysemethoden – Eine anwendungsorientierte Einführung*. Springer, Berlin und Heidelberg.

[5] Blumenstock, J.E. 2008. Size Matters: Word Count as a Measure of Quality on Wikipedia. In *Proceedings of the 17th international conference on World Wide Web* (Peking, China, 21.-25. April 2008). WWW08. ACM, New York, NY, 1095-1096. DOI=http://doi.acm.org/10.1145/1367497.1367673.

[6] Cunningham, W. und Leuf, B. 2001. *The Wiki Way. Quick Collaboration on the Web*. Addison-Wesley, Boston u.a.

[7] Denning, P., Horning, J., Parnas, D. und Weinstein, L.. 2005. Wikipedia risks. *Communications of the ACM*. 48, 12 (Dezember 2005), 152-152. DOI=http://doi.acm.org/10.1145/1101779.1101804.

[8] Dondio, P. und Barrett, S. 2007. Computational Trust in Web Content Quality: A Comparative Evalutation on the Wikipedia Project. *Informatica – An International Journal of Computing and Informatics*. 31, 2 (Juni 2007), 151-160. DOI=10.1.1.159.4166.

[9] Giles, G. 2005. Internet encyclopedias go head to head. *Nature*, 438, 7070 (Dezember 2005), 900-901. DOI=10.1038/438900a.

[10] Hunt, J. und McIlroy, M. 1975. *An algorithm for differential file comparison*. Computer Science Technical Report 41, Bell Laboratories.

[11] Javanmardi, S., Lopes, C. und Baldi, P. 2010. Modeling User Reputation in Wikipedia. *Journal of Statistical Analysis and Data Mining*. 3, 2 (April 2010), 126-139. DOI=10.1002/sam.10070.

[12] Kittur, A. und Kraut, R. E. 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In *Proceedings of the ACM 2008 Conference on Computer supported cooperative work* (San Diego, USA, 8.-12. November 2008). CSCW08. ACM, New York, NY, 37-46. DOI=http://doi.acm.org/10.1145/1460563.1460572.

[13] Kurzidim, M. 2004. Wissenswettstreit. Die kostenlose Wikipedia tritt gegen die Marktführer Encarta und Brockhaus an. *c't Magazin für Computertechnik*, 2004, 21 (Oktober 2004), 132-139.

[14] Lih, A. 2004. Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource. In *Proceedings of the 5th International Symposium on Online Journalism* (Austin, USA, 16.-17. April 2004). DOI=10.1.1.117.9104.

[15] Lim, E.P., Vuong, B.Q., Lauw, H.W. und Sun, A. 2006. Measuring Qualities of Articles Contributed by Online Communities. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (Hong Kong, 18.-22. Dezember 2006). WI '06, IEEE Computer Society, Washington, DC, 81-87. DOI=10.1109/WI.2006.115.

[16] Opuszko, M., Wöhner, T., Peters, R. und Ruhland, J., 2010. Qualitätsmessung in der Wikipedia: Ein Ansatz auf Basis von Markov-Modellen. In *Multikonferenz Wirtschaftsinformatik 2010* (Göttingen, Deutschland, 23.-25. Februar 2010). MKWI 2010, Universitätsverlag Göttingen, 705-716.

[17] O'Reilly, T. 2005. *What is Web2.0?* http://oreilly.com/web2/archive/what-is-web-20.html. Abruf am 02.12.2010.

[18] Potthast, M., Stein, B. und Gerling, R. 2008. Automatic Vandalism Detection in Wikipedia. In *Proceedings of the Advances in Information Retrieval - 30th European Conference on IR Research*. (Glasgow, UK, 30. März-3. April 2008). ECIR 2008, Springer, 663-668. DOI=10.1007/978-3-540-78646-7_75.

[19] Priedhorsky, R., Chen, J., Lam, S.K., Panciera, K., Terveen, L. und Riedl, J. 2007. Creating, Destroying, and Restoring Value in Wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work* (Sanibel Island, USA, 4.-7. November 2007). Group 2007, ACM, New York, NY, 259-268. DOI=http://doi.acm.org/10.1145/1316624.1316663.

[20] Resnick, P., Zeckhauser, R., Friedman, E. und Kuwabara, K. 2000. Reputation Systems. *Communications of the ACM*, 43, 12 (Dezember 2000). 45-48. DOI=http://doi.acm.org/10.1145/355112.355122.

[21] Smets, K., Goethals, B. und Verdonk, B. 2008. Automatic Vandalism Detection in Wikipedia: Towards a Machine Learning Approach. In *Proceedings of the AAAI Workshop, Wikipedia and Artificial Intelligence: An Evolving Synergy* (Chicago, USA, 13.-14. Juli 2008), WikiAI 2008, 43-48.

[22] Surowiecki, J.; 2004. *The Wisdom Of Crowds: Why The Many Are Smarter Than The Few And How Collective Wisdom Shapes Business, Economies, Societies And Nations*. Doubleday.

[23] Stein, K. und Hess, C. 2007. Does it matter who contributes: a study on featured articles in the german Wikipedia. In *Proceedings of the eighteenth conference on Hypertext and hypermedia* (Manchester, UK, 10.-12. September 2007). HT '07, ACM, New York, NY, 171-174. DOI=http://doi.acm.org/10.1145/1286240.1286290:

[24] Stvilia, B., Twidale, M.B., Smith, L.C. und Gasser, L. 2005. Assessing information quality of a community-based encyclopedia. In *Proceedings of the International Conference on Information Quality* (Cambridge, USA, 4.-6. November 2005). ICIQ 2005, 442–454. DOI=*10.1.1.78.6243*.

[25] Viégas, F., Wattenberg, M. und Dave, K. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Wien, Österreich, 24.-29. April 2004). CHI 2004, ACM, New York, NY, 575–582. DOI=http://doi.acm.org/10.1145/985692.985765.

[26] Viégas, F., Wattenberg, M., Kriss, J. und Ham, F. 2007. Talk before you type: Coordination in Wikipedia. In *Proceedings of the 40th Hawaii International Conference on System Sciences*. (Hawaii, USA, 3.-6. Januar 2007). HICSS 2007, IEEE Computer Society Washington, DC, 78-88. DOI=http://dx.doi.org/10.1109/HICSS.2007.511.

[27] West, A., Kannany, S. und Leez, I. 2010. *Detecting Wikipedia Vandalism via Spatio-Temporal Analysis of Revision Metadata*. Technical Reports (CIS). University of Pennsylvania.

[28] Wikipedia (Hrsg.). 2010. *Autorenportal*. http://de.wikipedia.org/wiki/Wikipedia:Autorenportal. Abruf am 02.12.2010.

[29] Wöhner, T. und Peters, R. 2009. Assessing the quality of Wikipedia articles with lifecycle based metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (Orlando, USA, 25.-27. Oktober 2009). WikiSym 2009, ACM, New York, NY. DOI=http://doi.acm.org/10.1145/1641309.1641333.

[30] Zeng, H., Alhoussaini, M., Ding, L., Fikes R.und McGuinness, D. 2006. Computing trust from revision history. In *Proceedings of the 2006 International Conference on Privacy, Security and Trust* (Markham, Kanada, 30. Oktober-1. November 2006). PST 2006, ACM, New York, NY. DOI=http://doi.acm.org/10.1145/1501434.1501445.

Track 7: Theorie und Methodik der Wirtschaftsinformatik

Stefan Klein und Volker Wulf

Editorial

Die Wirtschaftsinformatik ist forschungsmethodisch und paradigmatisch durch eine erhebliche Diversität der Ansätze gekennzeichnet. Status und Formen der Theoriebildung wie auch die Konzeptualisierung des Gegenstandsbereichs (das „IT Artefakt“) sind nach wie vor Gegenstand der Fachdiskussion. In der angelsächsischen Schwesterdisziplin ‚Information Systems‘ ist der bisher dominierende positivistisch ausgerichtete organisations- und verhaltenswissenschaftliche Grundkonsens einerseits durch eher interpretative Ansätze und andererseits durch den Ansatz der ‚Design Science’ und damit einer ingenieurwissenschaftlichen Perspektive in Frage gestellt worden. Die bisherigen Konzeptionen von ‚Design Science‘ thematisieren jedoch nur in unzulänglicher Weise die für die Wirtschaftsinformatik konstitutive Interaktion zwischen Organisation und Technik.

Vor diesem Hintergrund beabsichtigt dieser Track den forschungsmethodologischen und wissenschaftstheoretischen Diskurs innerhalb der Wirtschaftsinformatik zu beleben. Dabei galt unser spezielles Interesse Arbeiten, die für das Spezifikum der Wirtschaftsinformatik – die Gestaltung von IT-Systemen im Wechselspiel mit ihrer organisatorischen Aneignung – tragfähige wissenschaftstheoretische Grundpositionen formulieren und forschungsmethodische Orientierungen reflektierten.

Die skizzierten Grundfragen manifestieren sich in einer Reihe konkreter Herausforderungen, die auch jenseits des Diskurses auf der WI 2011-Tagung Bestand haben werden. Exemplarisch seien hier fünf Forschungsfoki genannt:

- Implikationen der Interaktion von Organisation und Technik für die Gestaltung von IS. Während der Design Science Ansatz diesbezüglich unterspezifiziert ist, ist dessen kritische Prüfung, Weiterentwicklung oder Ausdifferenzierung zu untersuchen.

- Forschungsmethoden für Längsschnittuntersuchungen. Aus forschungspragmatischen Gründen begnügen sich viele Untersuchungen mit Momentaufnahmen. Demgegenüber sind aber viele Fragen der Gestaltung und Aneignung von IT-Artefakten nur im Zeitverlauf sinnvoll zu klären.

- Möglichkeiten und Grenzen der Übertragbarkeit bzw. Generalisierbarkeit von Ergebnissen. Je nach wissenschaftstheoretischer Grundposition stellt sich das Problem in unterschiedlicher Schärfe: Je abstrakter das IT-Konstrukt konzeptualisiert wird, umso breiter ist sein potenzieller Einsatzbereich. Demgegenüber betonen konstruktionsorientierte und interpretative Ansätze den engen Bezug zum Anwendungskontext, der Übertragbarkeit nur sehr begrenzt ermöglicht.

- Wissenschaftstheoretische Herausforderungen prospektiver Aussagen. Technology Assessment, Foresight Studien oder Technologiefolgenforschung sind z.T. diskreditiert, es besteht aber erheblicher Bedarf Verlaufsformen von Technikentwicklung und -aneignung zu antizipieren. Die paradigmatischen, theoretischen und methodischen Grundlagen prospektiver Forschung verdienen deshalb eine breitere Diskussion.

- Praxis und Werte der Wissenschaft. Leitbilder wissenschaftlichen Arbeitens im Spannungsfeld inhaltsfreier Performancemessung und dem Primat wirtschaftlicher Relevanz.

Als Ergebnis der Veröffentlichung des Calls for Paper haben wir 24 Einreichungen erhalten. Diese sind von mindestens drei Gutachtern und Associated Editoren diskutiert und bewertet worden. Unter Berücksichtigung des Rebuttals der Autoren und weiterer Diskussionen zwischen Gutachtern, Associated Editoren, Track Chairs und den Conference Chairs haben wir uns entschieden, sechs Einreichungen anzunehmen, zwei davon mit Auflagen zur Überarbeitung. Das entspricht einer Annahmequote von 25%.

Die Einreichungen zu unserem Track decken ein breites Spektrum von Themen und Forschungsparadigmen ab, wie beispielsweise:

- Empirische Untersuchungen zum Publikations- bzw. Kooperationsverhalten,
- Neuartige konzeptionelle und methodische Ansätze für die Wirtschaftsinformatik sowie Transfer von Forschungsparadigmen aus verwandten Disziplinen,
- Literaturstudien zum Forschungsstand in Arbeitsfeldern wie Gender Studies oder Grounded Theory,
- Diskussion zu Fragen im Umfeld des Design Science Diskurses, etwa Generalisierbarkeit oder Verfahren der Verifikation.

Die erfreulich hohe Anzahl an Einreichungen zeigt, dass forschungsmethodischen Fragen auf erhebliches Interesse in der WI-Community stoßen. Dieses Faktum steht sicherlich auch in Zusammenhang mit der aktuellen Diskussion um die forschungsparadigmatische Orientierung der Wirtschaftsinformatik. Wir hoffen, dass unser Track und die in dessen Umfeld sich entfaltenden Diskurse zu einer Differenzierung der aktuellen Diskussion beitragen. Wir würden uns zukünftig vor allem eine Vertiefung der Diskussion wünschen, die sich den konzeptionellen und methodischen Hausforderungen der Gestaltung von IT-Systemen im

Wechselspiel zu bestehenden Arbeitspraktiken und organisatorischen Aneignungsformen stellt. Insbesondere an dieser Stelle besteht unserer Meinung nach noch erheblicher Forschungsbedarf.

Für ihre sehr kompetente und freundschaftliche Unterstützung bedanken wir uns ganz herzlich bei den Associated Editoren unseres Tracks: Christiane Floyd, Universität Hamburg; Ulrich Frank, Universität Duisburg-Essen; Franz Lehner, Universität Passau; Ulrich Müller-Funk, Universität Münster; Björn Niehaves, Universität Münster; Volkmar Pipek, Universität Siegen; Kai Riemer, University of Sydney; Markus Rohde; Universität Siegen; Kjeld Schmidt, Copenhagen Business School; Gunnar Stevens, Universität Siegen; Alexander Teubner, Universität Münster; Ina Wagner, TU Wien; Hannes Werthner, TU Wien und Oliver Wendt, TU Kaiserslautern.

Weitere Gutachten sind erstellt worden von Rainer Alt, Markus Bick, Bernhard Böhm, Alena Buchalcevova, Stephan Dlugosz, Waltraud Ernst, Annika Eschmeier, Werner Esswein, Rony G. Flatscher, Uri Gal, Dorina Gumm, Hardy Hanappi, Helmut Hauptmeier, Christian Huemer, Fischbach Kai, Ralf Klamma, Josef Küng, Sabine Matook, Dirk C. Mattfeld, Bernhard Nett, Barbara Paech, Susanne Patig, Jan Pawlowski, Margit Pohl, Jens Pöppelbuß, Corina Raduescu, Leonardo Ramirez, Michael Rebstock, Tim Reichling, Alexander Richter, Arno Rolf, Carola Schauer, Stefan Schellhammer, Michael Scholz, Petra Schubert, Stefan Seidel, Stefan Seifert, Vladimir Shekhovtsov, Alexander Simons, Stefan Smolnik, Stefan Stieglitz, Barney Tan, Ingolf Terveer, A Min Tjoa, Eitel von Maur, Ina Wagner, Maria Wimmer, Boris Wyssusek, Fahri Yetim und Thomas Zabel. Auch ihnen sei an dieser Stelle ganz herzlich gedankt.

Ein besonderer Dank geht an Benjamin Rensmann, Universität Münster, der den Review-Prozess umsichtig koordiniert und die Klippen des CMS erfolgreich umschifft hat.

Stilisierte Fakten in der gestaltungsorientierten Wirtschaftsinformatik – Allgemeine Potentiale und erste Erfahrungen

Constantin Houy
Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH und Universität des Saarlandes
Stuhlsatzenhausweg 3, Geb. $D3_2$
66123 Saarbrücken
+49-681-85775-5244
Constantin.Houy@iwi.dfki.de

Peter Fettke
Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH und Universität des Saarlandes
Stuhlsatzenhausweg 3, Geb. $D3_2$
66123 Saarbrücken
+49-681-85775-5142
Peter.Fettke@iwi.dfki.de

Peter Loos
Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH und Universität des Saarlandes
Stuhlsatzenhausweg 3, Geb. $D3_2$
66123 Saarbrücken
+49-681-85775-3106
Peter.Loos@iwi.dfki.de

ABSTRACT
Ein gestaltungsorientiert ausgerichteter Forschungsansatz gilt im Allgemeinen als zentrales Charakteristikum der deutschsprachigen Wirtschaftsinformatik (WI). Gemäß dieser Überzeugung sollen in der wirtschaftsinformatischen Forschung insbesondere innovative und nützliche Techniken wie Methoden, Software-Prototypen und ähnliche Artefakte zur Lösung betrieblicher Probleme gestaltet und erforscht werden, um Handlungs- und Erklärungswissen für die zukünftige Gestaltung von Informationssystemen zu generieren. Dieses Gestaltungswissen lässt sich beispielsweise in Form von Theorien oder technologischen Regeln darstellen. Obwohl die Fundierung der Wirtschaftsinformatik mit eigenen Theorien zunehmend an Bedeutung gewinnt, existieren bisher nur wenige Ansätze für die Theoriebildung und das Herausarbeiten technologischer Regeln. Vorliegender Beitrag möchte den wissenschaftstheoretischen Diskurs in diesem Bereich beleben, indem er Potentiale und Erfahrungen mit dem Konzept stilisierter Fakten aus der VWL im Kontext der Wirtschaftsinformatik sowie Vorschläge für Richtlinien und Handlungsempfehlungen für die Durchführung zukünftiger Studien präsentiert. Stilisierte Fakten ermöglichen die Entwicklung von Gestaltungswissen und stellen einen vielversprechenden wissenschaftstheoretischen Ansatz für die gestaltungsorientierte Wirtschaftsinformatik dar.

Keywords
Stilisierte Fakten, Gestaltungsorientierung, Design Science, Theoriebildung, technologische Regeln, Design Theorien.

1. EINLEITUNG

1.1 Ausgangssituation und Problemstellung
Die Gestaltung betrieblicher Informationssysteme steht allgemein im Zentrum des Forschungsinteresses der Wirtschaftsinformatik [35]. Folgt man dieser Auffassung, so soll wirtschaftsinformatische Forschung insbesondere innovative Techniken, wie Methoden, Modelle, Software-Prototypen sowie andere Artefakte hervorbringen, die zur Lösung von Problemstellungen in der Praxis beitragen können [3; 25]. Basierend auf den Gedanken, die Simon zu den Wissenschaften vom Künstlichen formuliert hat [42], erforscht die Wirtschaftsinformatik im Gegensatz zu den Naturwissenschaften nicht die „gegebene" Wirklichkeit, sondern sie entwickelt, gestaltet und erforscht neue „künstliche" Wirklichkeiten in Form innovativer Informationssysteme [14]. In diesem Kontext gewinnt die Entwicklung, Prüfung und Anwendung theoretischen sowie technologischen Wissens (Ursache-Wirkungs- bzw. Ziel-Mittel-Beziehungen) im Kontext von Artefakten und Techniken für die Gestaltung von Informationssystemen zunehmend an Bedeutung. Theoretisches und technologisches Wissen kann den ingenieurmäßigen Gestaltungsprozess von Informationssystemen maßgeblich unterstützen [19]. Neben dem Ziel der Gestaltung von Informationssystemen verfolgt die gestaltungsorientierte Wirtschaftsinformatik auch Erkenntnisziele [13].

In einem in WIRTSCHAFTSINFORMATIK / BISE publizierten Diskussionsbeitrag mit dem Titel „Was ist eigentlich Grundlagenforschung in der Wirtschaftsinformatik?" legen die Diskussionsteilnehmer Winter, Krcmar, Sinz, Zelewski und Hevner dar, dass „hinsichtlich der Grundlagenforschung in der Wirtschaftsinformatik nicht das ob fraglich ist, wohl aber das wie" [52]. Die Erforschung von Ursache-Wirkungs-Beziehungen sowie deren technologische Realisierung [7] ist für die Wirtschaftsinformatik von Bedeutung, da diese eine fundierte Gestaltung von Artefakten unterstützen können (Krcmar in [52]). In diesem Kontext weist Zelewski in der Diskussionsrunde auf die Potentiale des Konzeptes stilisierter Fakten („stylized facts") für die Grundlagenforschung und für Entwicklung und Prüfung grundlegender theoretischer Aussagen in der Wirtschaftsinformatik hin [52].

Stilisierte Fakten stammen aus dem Kontext der Volkswirtschaftslehre und bieten einen Ansatz zur Bündelung und Generalisierung von Erkenntnissen in einem Themengebiet, die anhand unterschiedlicher Methoden entwickelt wurden. Dies ermöglicht die Bildung, Prüfung sowie den Vergleich theoretischer Aussagen [24; 42]. Stilisierte Fakten unterstützen des Weiteren eine „Kartografierung" des vorhandenen Wissens sowie die Identifikation offener Fragen in einem Forschungsgebiet [50]. Im vorliegenden Beitrag wird argumentiert, dass sich das Konzept stilisierter Fakten als wissenschaftstheoretisches Konzept technologischer sowie theoretischer Forschung in der gestaltungsorientierten Wirtschaftsinformatik eignet. Vorhandenes konzeptionelles und empirisches Wissen zur Gestaltung und Wirkungen von Artefakten und

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Techniken der Wirtschaftsinformatik können anhand stilisierter Fakten erfasst und dokumentiert werden, z. B. als Design Theorien oder technologische Regeln (Ziel-Mittel-Aussagen) zur Unterstützung des Gestaltungsprozesses sowie in Form potentieller Theorien der Wirtschaftsinformatik im klassischen Sinne (Ursache-Wirkungsbeziehungen). Stilisierte Fakten wurden bisher vornehmlich in sozialwissenschaftlichen Disziplinen, wie der BWL und der VWL erfolgreich angewendet. Die Potentiale des Konzeptes für gestaltungsorientierte Technikwissenschaften wie die Wirtschaftsinformatik wurden bisher nicht untersucht.

1.2 Ziel des Artikels

Das allgemeine Ziel des vorliegenden Artikels ist es, an die laufende Diskussion um das Thema der theoretischen Fundierung der Wirtschaftsinformatik anzuknüpfen und insbesondere die Potentiale des Konzeptes stilisierter Fakten für die gestaltungsorientierte Wirtschaftsinformatik zu eruieren. Weiterhin werden Erfahrungen mit der Anwendung des Konzeptes sowie einige Vorschläge für Richtlinien und Handlungsempfehlungen für die zukünftige Anwendung erarbeitet und diskutiert. Zusammenfassend kann die zentrale Forschungsfrage des Beitrags folgendermaßen formuliert werden:

Welche Potentiale bieten stilisierte Fakten für eine ingenieurmäßige Gestaltung von Informationssystemen?

1.3 Forschungsansatz

Dieser Beitrag ist nicht das Ergebnis der Anwendung einer etablierten Forschungsmethode. Vielmehr wird ein innovativer Forschungsansatz gewählt, der theoretisch-konzeptionelle Überlegungen und empirische Erfahrungen mit der praktischen Anwendung des Konzeptes stilisierter Fakten in der Wirtschaftsinformatik sinnvoll miteinander kombiniert. Die Rolle und die Potentiale stilisierter Fakten für die Wissenschaftstheorie der Wirtschaftsinformatik werden auf Basis konzeptioneller Überlegungen argumentativ erarbeitet. Zusammenhänge zwischen stilisierten Fakten und zentralen Begriffen wie Theorie, Artefakt, Technik, technologische Regeln, Normen und Design Theorien werden erläutert. Darüber hinaus werden eigene Erfahrungen mit der Anwendung des Konzeptes der Ableitung stilisierter Fakten zu sechs bedeutenden Themenbereichen, Techniken bzw. Artefakten der Wirtschaftsinformatik gesammelt und beschrieben (Tabelle 1). Jede dieser Untersuchungen kann als eine (wissenschaftstheoretische) Fallstudie interpretiert werden. Aufbauend auf der Reflexion der Erfahrungen werden Handlungsempfehlungen für die zukünftige Anwendung aufgezeigt.

Tabelle 1: Übersicht über die durchgeführten Fallstudien

Nr.	Adressierte Techniken, bzw. Artefakte der Wirtschaftsinformatik	Anzahl der abgeleiteten stilisierten Fakten
1	Ereignisgesteuerte Prozessketten (EPK)	22
2	Unified Modeling Language (UML) im Allgemeinen	17
3	Software-Architekturen	23
4	Petri-Netze	29
5	Aktivitätsdiagramme der Unified Modeling Language (UML)	18
6	Enterprise Resource Planning (ERP) - Systeme	6

1.4 Aufbau des Artikels

Nach dieser Einleitung werden in Kapitel 2 das Basisverständnis von stilisierten Fakten und Begriffe in diesem Kontext eingeführt. In Kapitel 3 werden insbesondere die Bedeutung und die Potentiale stilisierter Fakten für die gestaltungsorientierte Wirtschaftsinformatik diskutiert, indem das Konzept auf die WI übertragen wird. Kapitel 4 beleuchtet anhand der Anwendungsbeispiele gesammelte Erfahrungen, Vorschläge für Richtlinien und Handlungsempfehlungen für die Anwendung des Konzeptes in der Wirtschaftsinformatik. Kapitel 5 resümiert den Beitrag.

2. DAS KONZEPT STILISIERTER FAKTEN

2.1 Basisverständnis

Stilisierte Fakten stellen Wissenselemente in Form generalisierter Aussagen dar, die bedeutende Eigenschaften und Zusammenhänge im Kontext eines untersuchten Phänomens beschreiben und empirisch breit gestützt sind [22]. Beispielsweise stellt die Phillips-Kurve, die den empirisch breit gestützten Zusammenhang zwischen der Entwicklung der Nominallöhne und der Arbeitslosigkeit in einer Volkswirtschaft beschreibt [24], ein typisches stilisiertes Faktum dar.

Im Gegensatz zu Hypothesen, die für die Theoriebildung anhand der hypothetisch-deduktiven Methode als „vorläufig unbegründete Behauptungen“ noch einen vagen Charakter haben [51], sind stilisierte Fakten Aussagen, die sich bereits in vielen Fällen bestätigt haben. Gleichzeitig handelt es sich bei stilisierten Fakten um pauschalisierte Aussagen, die auch in bestimmten Kontexten nicht gültig sind. An stilisierte Fakten wird nicht der Anspruch erhoben, gesetzesartige, streng deterministische oder kausale Aussagen zu repräsentieren. Vielmehr sollen interessante Phänomene im Gegenstandsbereich benannt werden, die in bestimmten Kontexten weitgehend zweifelsfrei akzeptiert sind. Dabei existieren entweder noch gar keine Theorien, die das stilisierte Faktum erklären können oder es wird akzeptiert, dass das stilisierte Faktum unabhängig von einer bestimmten Theorie auftritt. Mit anderen Worten sollte die Formulierung stilisierter Fakten möglichst keinen Bezug auf Konzepte bestimmter Theorien nehmen.

Das Konzept stilisierter Fakten geht nach herrschender Meinung auf Kaldor zurück, der es im Kontext der makroökonomischen Wachstumstheorie anwendete [27]. Kaldor argumentiert, dass im Rahmen der Modell- und auch der Theoriebildung im Allgemeinen die Erläuterung charakteristischer Merkmale („stylized facts“) eines zu untersuchenden Phänomens das bedeutendste Ziel darstellt. Durch die grundlegende Idee der Stilisierung ermöglicht das Konzept eine einheitliche Sicht („stylized view“) auf charakteristische Merkmale eines Gegenstandes.

Die Entwicklung stilisierter Fakten geschieht durch die Zusammenfassung grundsätzlich ähnlicher, sich aber möglicherweise im Detail unterscheidender Aussagen zu einer gebündelten und abstrahierten Eigenschaft eines Gegenstandes. Diese abstrahierte Eigenschaft sollte sich durch ein hinlängliches Maß an Repräsentativität und inhaltlicher Übereinstimmung der grundlegenden Arbeiten rechtfertigen lassen. Der durch subjektive Entscheidungen geprägte Ableitungsprozess kann durch den Nachweis eines entsprechenden Konsens von Experten, die im Forschungsfeld tätig sind, abgesichert werden, z. B. durch die Bewertung und Angabe der Anzahl der Quellen aus einer möglichst vollständigen Literatursammlung, die eine bestimmte Aussage stützen,

(*Konsensanalyse*) [50]. Es lässt sich eine Ähnlichkeit stilisierter Fakten zum Begriff der empirischen Generalisierungen feststellen, der die Zusammenfassung von empirischen Begebenheiten, die sich in mehreren Arbeiten gezeigt haben, in Form Mustern beschreibt. Allerdings lassen sich stilisierte Fakten nicht nur aus empirischen Arbeiten, sondern aus ganz unterschiedlichen Quellen gewinnen [42]. Grundlage können z. B. auch wissenschaftliche Artikel konzeptioneller Art, Expertendiskussionen, Praktikerinterviews, Lehrbuchwissen oder auch formale, qualitative oder quantitative Analysen anderer Quellenmaterialien z. B. wie Filmen, Bildern etc. sein. Stilisierte Fakten werden anhand eines systematischen Reviews der Quellen [12] erarbeitet. Bedeutende Unterschiede zur allgemein bekannten Methode des systematischen Reviews bestehen vor allem in der Abstraktion und Stilisierung der Inhalte durch das Ausblenden von Details und in der erhöhten Transparenz der Ableitung der stilisierten Aussagen durch einen mehrstufigen Generierungsprozess. Dabei sind einige Grundanforderungen zu beachten, die eine hohe intersubjektive Nachvollziehbarkeit der ermittelten Ergebnisse gewährleisten sollen [42]: (1) die Offenlegung des Generierungsprozesses, (2) eine methodenübergreifende Faktengenerierung, (3) eine möglichst vollständige Faktenmenge, (4) hinreichender Konsens zu jedem stilisierte Faktum und (5) die Offenheit der Faktenmenge. Zur Ableitung stilisierter Fakten wurden einige Vorgehensmodelle vorgeschlagen, die sich i. d. R. auf eine Ableitung aus Literaturquellen beziehen. Diese Modelle können allerdings in ähnlicher Form auch auf andere Quellentypen angewendet werden.

Schwerin kritisierte in seiner Dissertation das ursprüngliche Konzept Kaldors und erarbeitete darauf basierend einen grundlegend modifizierten Generierungsprozess für stilisierte Fakten [42]. Diese Operationalisierung wurde von Heine et al. unter Berücksichtigung typischer Vorgehensmodelle zur Erstellung von Metaanalysen für den betriebswirtschaftlichen Kontext weiter spezifiziert [22]. Das Vorgehensmodell von Heine et al. (Abbildung 1) wurde in weiteren Beiträgen im Kontext der BWL, z. B. in Weißenberger/Löhr [50] aufgegriffen, und wird auch den folgenden Ausführungen zugrunde gelegt.

Zunächst wird das zu untersuchende Phänomen begrifflich erfasst und das Problem, das durch die Anwendung des Konzeptes gelöst werden soll, formuliert, z. B. die Überprüfung eines theoretischen Aussagensystems. Im Anschluss daran werden anhand der für das Forschungsvorhaben definierten Begriffe die relevanten Quellen, recherchiert und eine zur Problemdefinition passende Quellensammlung angelegt. Zur Entwicklung der Quellensammlung sind die grundlegenden Erkenntnisse aus der Reviewforschung zu beachten. Im nächsten Schritt werden für die Problemlösung bedeutende Aussagen in den unterschiedlichen zugrunde gelegten Studien, die z. B. die zu untersuchende Theorie betreffen, extrahiert und eine Aussagensammlung erstellt. Im Anschluss werden korrespondierende Aussagen gruppiert und zu generalisierten Aussagen verdichtet. Im letzten Schritt werden diese Aussagen nochmals weiter verdichtet und zu stilisierten Fakten abstrahiert.

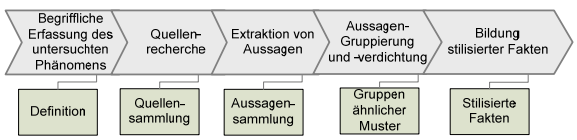


Abbildung 1: Generierungsprozess stilisierter Fakten [22]

2.2 Die Rolle stilisierter Fakten in der Theoriebildung

Theorien repräsentieren wissenschaftliches Wissen, das auf Basis wissenschaftlicher Standards gewonnen wurde. Der Wissensbegriff wird im Rahmen der allgemeinen Erkenntnistheorie nach wie vor kontrovers diskutiert, ohne dass ein allgemein akzeptiertes Verständnis vorliegt [21]. Im vorliegenden Beitrag wird Wissen im „traditionellen Sinne“ als wahre, begründete Meinung verstanden [11], das heißt, dass Meinungen unter Annahme überzeugender Begründungsstandards zu rechtfertigen sind sowie einem gewissen Wahrheitsanspruch genügen müssen. An dieser Stelle sei auf die Diskussion zu verschiedenen Begründungsstandards [44] sowie zu unterschiedlichen Wahrheitstheorien [16] verwiesen.

Vor diesem Hintergrund stellen Theorien Aussagensysteme dar, die eine Menge gesetzesartiger Aussagen (Wenn-Dann-Aussagen, nomologische Hypothesen) enthalten und vornehmlich der Erklärung und Prognose von beobachtbaren Phänomenen (Ursache-Wirkungs-Beziehungen) dienen [51]. Hypothesen als Grundelemente einer Theorie können untereinander logisch und systematisch vernetzt auftreten [30; 5]. Insbesondere im Kontext kulturwissenschaftlicher Forschungsdisziplinen wie der Wirtschaftsinformatik oder verwandte Disziplinen wie der BWL [54] können aufgrund der forschungsbereichimmanenten Komplexität Hypothesen meist nicht als kausale bzw. streng deterministische Zusammenhänge interpretiert werden [14]. Es handelt sich bei diesen Zusammenhängen häufiger um Korrelationen, die statistische Gesetzmäßigkeiten darstellen [51].

Zur wissenschaftlich fundierten Entwicklung, Überprüfung und zum Vergleich theoretischer Aussagensysteme werden verschiedene Ideen und Vorstellungen in der Erkenntnistheorie diskutiert. In seiner Arbeit „Logik der Forschung“ entwickelte Popper basierend auf der Kritik induktiver, verifikationistischer Forschungsansätze die kritisch-rationalistische Methode der Theorieprüfung, die insbesondere auf die Falsifikation von Hypothesen beruht und Fragen der Theoriebildung in den Bereich der Wissenschaftsgeschichte bzw. -psychologie einordnet [36]. Im weiteren Diskurs haben insbesondere die Arbeiten von Lakatos [29] und Kuhn [28] gezeigt, dass das wissenschaftstheoretische Erkenntnismodell von Popper zu erweitern und zu verfeinern ist.

Ein Kernpunkt der Erweiterung betrifft die Feststellung, dass auch die Erschütterung einer Hypothese aus logischen Gründen nicht zwangsläufig zum Verwerfen der gesamten Theorie führen muss. Beispielsweise können auch zusätzliche Hilfshypothesen verworfen werden oder der konkret betrachtete Anwendungsfall kann aus der Theorie ausgeschlossen werden. Die in wissenschaftlichen Theorien formulierten Aussagen werden mit Beobachtungen aus empirischen Untersuchungen, wie Experimenten bzw. mit anderen Theorien verglichen, um das komplette Aussagensystem oder nur einzelne Aussagen zu falsifizieren. Gegebenenfalls sind einzelne Aussagen zu revidieren und können anhand weiterer empirischer Daten erneut geprüft werden. Das Aussagensystem kann sich im Rahmen der Prüfung auch vorläufig bestätigen.

Im Folgenden wird aufgezeigt, dass stilisierte Fakten in diesem Kreislauf eine bedeutende Rolle spielen und die Theoriebildung an mehreren Punkten des Kreislaufs maßgeblich unterstützen können (siehe Abbildung 2). Stilisierte Fakten ermöglichen sowohl den Prozess der Bildung (I), der Prüfung (II) als auch den Vergleich von Theorien (III) [22; 42].

Werden stilisierte Fakten als bereits durch empirische Daten gestützte Hypothesen verstanden, so ermöglichen sie die grundlegende *Bildung von Theorien*. Der Zusammenhang, der in einem stilisierten Faktum formuliert wird, gilt dann als Hypothese, die im weiteren Prozess gemäß der Popperschen Forschungslogik nochmals zu überprüfen ist.

Werden stilisierte Fakten als breit gestützte empirische Daten interpretiert, so ermöglichen sie das *Testen von Theorien*. Die Aussage, die in einem stilisierten Faktum formuliert ist, kann dann als vorliegender, breit gestützter empirischer Befund mit den Hypothesen einer Theorie verglichen und dadurch die Theorie entweder vorläufig bestätigen, sie falsifizieren oder auf die Notwendigkeit der Anpassung von Hypothesen hindeuten.

Stilisierte Fakten können auch den *Vergleich theoretischer Aussagensysteme* unterstützen. Auch in diesem Kontext ermöglichen stilisierte Fakten als breit gestützte empirische Befunde, innerhalb einer Auswahl ähnlich aufgestellter Theorien diejenige Theorie zu identifizieren, die das in der Realität nachgewiesene Faktum am treffendsten darstellen, prognostizieren bzw. erklären kann.

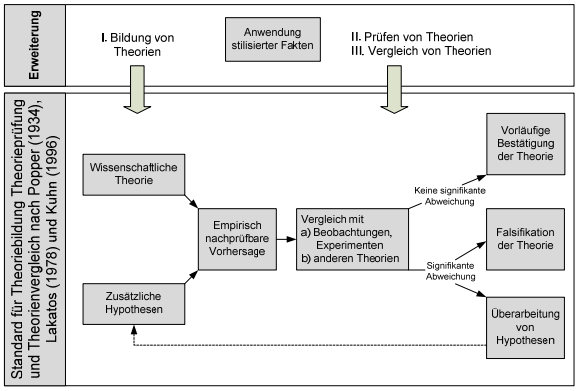


Abbildung 2: Bildung, Prüfung und Vergleich theoretischer Aussagesysteme mit stilisierten Fakten, in Anlehnung an [30]

Aufgrund dieser Potentiale stilisierter Fakten ist das Konzept nicht nur zur Theoriebildung und -prüfung in der VWL und dort insbesondere im Kontext der Politikberatung von Bedeutung. Stilisierte Fakten lassen sich auch in anderen empirisch ausgerichteten Realwissenschaften einsetzen. Vor diesem Hintergrund wurde das Konzept in den letzten Jahren auch vermehrt in der betriebswirtschaftlichen Forschung eingesetzt, z. B. in [50; 24].

3. Übertragung des Konzeptes auf die gestaltungsorientierte WI

3.1 Überblick über die Argumentation

Im Allgemeinen erscheinen die Potentiale stilisierter Fakten auch für die Wirtschaftsinformatik von Bedeutung. Allerdings ist nicht nur zu untersuchen, inwieweit die in anderen Disziplinen übliche Anwendung des Konzeptes stilisierter Fakten auch innerhalb der Wirtschaftsinformatik möglich ist. Vielmehr ist darüber hinaus zu klären, inwieweit Besonderheiten bei der Anwendung des Konzeptes in einer technikorientierten Wissenschaft wie der Wirtschaftsinformatik bestehen.

Im Folgenden wird die Bedeutung stilisierter Fakten für die gestaltungsorientierte Wirtschaftsinformatik argumentativ herausgearbeitet, indem sie im Kontext relevanter wissenschaftstheoretischer Begriffe und bedeutender praxisrelevanter Konzepte der gestaltungsorientierten Wirtschaftsinformatik erläutert werden. Dabei werden in der Diskussion folgende Aspekte berücksichtigt:

1. Theorien (Abschnitt 3.2),
2. Artefakte und Techniken (Abschnitt 3.3),
3. technologische Regeln (Abschnitt 3.4),
4. Standards und Normen (Abschnitt 3.5) und
5. Design Theorien (Abschnitt 3.6).

Da der Begriff der Design Theorie nicht einheitlich gebraucht wird, wird zunächst die Rolle stilisierter Fakten im Kontext von Artefakten, technologischen Regeln sowie Standards und Normen diskutiert, die jeweils wichtige Aspekte von Design Theorien darstellen. Auf dieser Grundlage kann dann der Zusammenhang zwischen stilisierten Fakten und Design Theorien erläutert werden.

3.2 Theorien und stilisierte Fakten

Stilisierte Fakten können in technikorientierten Wissenschaften wie der Wirtschaftsinformatik oder der Schwester-Disziplin Information Systems (IS) Research den Prozess der Bildung, Prüfung und des Vergleichs reiner erkenntnisorientierter Theorien ebenso unterstützen wie in anderen Realwissenschaften. Dies wird im Folgenden argumentativ und anhand von Beispielen aufgezeigt.

I. Theoriebildung

Stilisierte Fakten als empirisch gestützte Aussagen, die mehrfach in unabhängigen Quellen formuliert werden, können als Grundlagen von Hypothesen für wissenschaftliche Theorien der Wirtschaftsinformatik angesehen werden. Tatsächlich wurden bereits einige theoretische Modelle der IS Community und der Wirtschaftsinformatik basierend auf Aussagezusammenhängen entwickelt, die durch eine größere Zahl empirischer Primärerhebungen auf Basis unterschiedliche Forschungsmethoden (Fallstudien, Experimente etc.) gestützt werden. Diese werden häufig danach nochmals durch großzahlige Untersuchungen überprüft.

Als ein Beispiel für eine rein erkenntnisorientierte Theorie kann das *DeLone and McLean Information Systems Success Model* genannt werden [9], das in einem ersten Schritt auf der Basis der zentralen Aussagen von 100 Artikeln, die verschiedene empirische Forschungsmethoden, insbesondere Laborexperimente, Feld- und Fallstudien, einsetzen, und 80 konzeptionellen Beiträgen entwickelt wurde. Die in diesen empirischen Arbeiten erhobenen Daten und zentralen Beobachtungen wurden von DeLone und McLean verdichtet, um potentielle Ursache-Wirkungs-Beziehungen zwischen der Qualität von Informationssystemen und dem Erfolg ihres Einsatzes zu formulieren. So wird beispielsweise die generalisierte Aussage im *IS Success Model*, dass der Einsatz von Informationssystemen („use“) die Arbeitsleistung einer Organisation („organizational performance“) erhöht, aus 20 methodisch verschiedenen, empirischen Studien mit einem unterschiedlich operationalisierten Performance-Begriff abgeleitet (z. B. „profit“, „cost effectiveness“, „IS contribution to meeting goals“, „return on assets“ etc.).

Die verdichteten Aussagen, die die Grundlage für das *IS Success Model* bilden, stellen stilisierte Fakten im Sinne der vorliegenden Arbeit dar. DeLone und McLean explizieren zwar ihre Vorgehensweise, die sehr stark der Vorgehensweise zur Ableitung stilisierter Fakten ähnelt, ordnen diese aber nicht in ein wissenschaftstheoretisches Rahmenkonzept ein.

II. Theorieprüfung

Die Möglichkeit der Prüfung von Theorien anhand stilisierter Fakten ist im Kontext technikorientierter Wissenschaften ebenso gegeben. Beispielsweise überprüfen DeLone und McLean ihr *IS Success Model* zehn Jahre später nochmals anhand der zentralen Ergebnisse von 100 in der Zwischenzeit durchgeführten empirischen Studien, die sich der Überprüfung des Modells widmen [10]. Die zentralen Aussagen der einbezogenen Studien werden auch hier verdichtet bzw. generalisiert und werden als stilisierte Fakten im Sinne der vorliegenden Arbeit zur Theorieprüfung und zur Theorierevision verwendet. Basierend auf den abgeleiteten Aussagen revidieren DeLone und McLean das Modell gemäß dem oben aufgeführten Theoriezyklus, indem einige neue Hypothesen ergänzt und andere Aussagenzusammenhänge entfernt werden. Beispielsweise stützen fünf empirische Arbeiten den formulierten Zusammenhang zwischen der Qualität eines Informationssystems („system quality“) und Auswirkungen auf die Leistung einzelner Mitarbeiter („individual impact“).

III. Theorienvergleich

Stilisierte Fakten können als breit gestütztes empirisches Faktenwissen bei einem Vergleich unterschiedlicher Theorien die Identifikation derjenigen Theorie unterstützen, die diese in der Realität beobachteten Fakten am treffendsten darstellen, prognostizieren bzw. erklären kann. Beispielsweise erarbeitet Mellewigt im Kontext des Managements strategischer Kooperationen eine vergleichende Beurteilung des Resource-Dependence-Ansatzes, des Transaktionskostenansatzes und des Ressourcenorientierten Ansatzes anhand verdichteter konzeptioneller und empirischer Erkenntnissen, die den Charakter von stilisierten Fakten aufweisen [32]. Diese Arbeit stammt zwar aus dem Kontext der betriebswirtschaftlichen Forschung, insbesondere der Organisations- und Strategieforschung, hat allerdings aufgrund der besonderen Ausrichtung auf überbetriebliche Organisationsformen in der Telekommunikationsbranche bedeutende Implikationen für die Wirtschaftsinformatik.

3.3 Artefakte bzw. Techniken und stilisierte Fakten

Neben dem bedeutenden Ziel der Entwicklung von Theorien im Bereich von Informationssystemen (Erkenntnisorientierung) zielt die gestaltungsorientierte Wirtschaftsinformatik insbesondere auf die Gestaltung innovativer Artefakte und Techniken im Kontext von Informationssystemen (Technikorientierung) ab [15]. Die Artefaktgestaltung und Weiterentwicklung stellt nach Hevner et al. einen kontinuierlichen Zyklus dar [25], der dem Zyklus der Bildung und Prüfung von Theorien ähnelt. Diese beiden zentralen Perspektiven der gestaltungsorientierten Forschung in der Wirtschaftsinformatik werden in Abbildung 3 visualisiert.

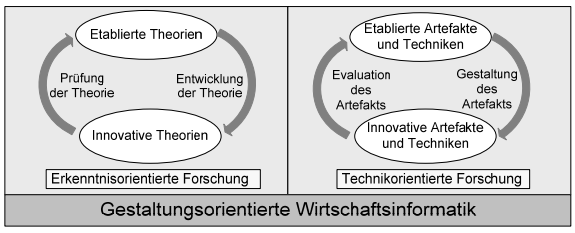


Abbildung 3: Erkenntnisorientierung und Technikorientierung in der gestaltungsorientierten Wirtschaftsinformatik

Unter der Gestaltung von Informationssystemen als Mensch-Aufgabe-Technik-Systeme [23] lässt sich gemäß dem in Abbildung 4 visualisierten Modell der Systemgestaltung die Weiterentwicklung eines Informationssystems als Gestaltungsobjekt verstehen [13]. Einem Gestaltungssubjekt stehen verschiedene Artefakte bzw. Techniken der Systemgestaltung zur Verfügung, wie z. B. Modellierungsmethoden, Prozessmodelle oder Datenbankmanagementsysteme, die es zur Umsetzung seiner Gestaltungsziele anwenden kann. Um eine ingenieurmäßige Gestaltung vornehmen zu können, benötigt der Systemgestalter spezifisches Gestaltungswissen zu relevanten Facetten der Systemgestaltung, z. B. zur tatsächlichen Verbreitung und Akzeptanz einer Technik in der Praxis, den Wirkungen, bzw. Nebenwirkungen ihrer Anwendung, den Kosten ihrer Anwendung und ähnliches. Dieses Gestaltungswissen ist nicht zwingend theoretischer Natur, ermöglicht dem Systemgestalter allerdings rationale Entscheidungen darüber, wie die Systemgestaltung durchzuführen ist.

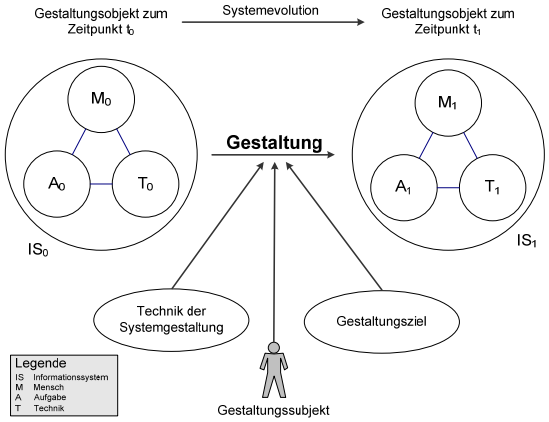


Abbildung 4: Modell der Systemgestaltung [13]

Auch dieses artefakt- bzw. technikrelevante Wissen kann durch stilisierte Fakten repräsentiert werden. Möchte der Systemgestalter beispielsweise prozessorientierte Informationssysteme in seine Organisation einführen, so benötigt er Wissen über geeignete Prozessmodellierungsmethoden, wie z. B. *Business Process Modeling Notation* (BPMN), ereignisgesteuerte Prozessketten (EPK), UML Aktivitätsdiagramme oder Petrinetze. Stilisierte Fakten über die charakteristischen Eigenschaften dieser Modellierungstechniken, wie z. B. Wirkungen, Nebenwirkungen, Reifegrad, Kosten und Effizienz der Anwendung etc. [13], ermöglichen rationale Entscheidungen darüber, welche Modellierungsmethode verwendet werden soll. Angenommen der Systemgestalter möchte für seinen Entscheidungsprozess insbesondere mehr über EPK erfahren, beispielsweise zur Verständlichkeit von EPK-Modellen. In diesem Kontext wären insbesondere folgende stilisierte Fakten interessant, die durch die angegebenen Quellen gestützt werden:

„*Die EPK ist leicht verständlich.*“ [46; 18; 41]

„*EPK ermöglichen intuitive grafische Prozessmodelle.*“ [46; 8; 47]

„*EPK-Modelle können mehrdeutig sein.*“ [41; 39; 45]

Der Systemgestalter kann sich anhand dieser stilisierten Fakten ein Bild zur Verständlichkeit der EPK machen und erkennen, welches Gestaltungswissen in den verwendeten Quellen vorliegt.

3.4 Technologische Regeln und stilisierte Fakten

Neben Theorien, die Ursache-Wirkungs-Beziehungen beschreiben, und artefakt- bzw. technikbezogenem Gestaltungswissen, das Aussagen über artefakt- bzw. technikspezifische Charakteristika darstellt, sind in der gestaltungsorientierten Wirtschaftsinformatik und in der IS Community auch Technologien im Sinne von handlungsorientiertem Regelwissen von Bedeutung. Solche Technologien stellen systematisch formulierte Aussagensysteme dar, die empirische Regularitäten des praktischen Handelns, also Handlungswissen in Form von Ziel-Mittel-Beziehungen (technologische Regeln), formulieren [7] und inhaltlich ausreichen, „um daraus glaubwürdige Empfehlungen für Entscheidungen über betriebliche Gestaltungsalternativen abzuleiten“ [54]. Technologische Regeln liegen i. d. R. als Aussagen der Form „Um X zu erreichen, führe Y aus“ vor [26; 38].

Die wissenschaftstheoretische Rolle von Technologien ist noch unklar und wird im Kontext der Wissenschaftstheorie weiterhin diskutiert [1; 6; 54]. Zu der Frage, wie technologische Aussagesysteme zu entwickeln sind, existieren unterschiedliche Ansichten, die auf alternative Technologieverständnisse zurückzuführen sind (einen Überblick über die Diskussion bietet Zelewski in [53]). Das *originäre* Technologieverständnis geht davon aus, dass technologische Regeln ohne zugrundeliegende Hypothesen einer Theorie existieren können, während das *derivative* Technologieverständnis davon ausgeht, dass technologische Regeln stets aus Theorien abgeleitet sein müssen. Eine solche Ableitung technologischer Aussagen aus Theorien sowie der umgekehrte Weg, also die Bildung von Theorien aus etablierten technologischen Aussagen gelten gemeinhin als problematisch und sind nur bedingt möglich. Einigkeit besteht allerdings bei beiden Auffassungen darüber, dass technologische Aussagensysteme wie Theorien zu entwickeln und ihre Zuverlässigkeit in Analogie zu dem oben dargestellten Zyklus zur Theorieentwicklung empirisch zu überprüfen ist [53].

Vor diesem Hintergrund wird im vorliegenden Beitrag argumentiert, dass stilisierte Fakten nicht nur bei der empirisch gestützten Bildung, Prüfung und dem Vergleich von Theorien, sondern auch bei der Bildung und Prüfung technologischer Regeln erfolgreich eingesetzt werden können. Da technologische Regeln ähnlich empirisch zu überprüfen sind wie die Hypothesen einer Theorie, können stilisierte Fakten analog auch die Entwicklung und Prüfung technologischer Regeln unterstützen.

Im Kontext der Informationsmodellierung liegen stilisierte Fakten z. B. in Form der Grundsätze ordnungsmäßiger Modellierung vor, die als technologische Regeln dienen können: „*Um die Strukturiertheit, Übersichtlichkeit und Lesbarkeit (Klarheit) von Prozessmodell zu erhöhen, sollten Prozessmodelle so wenige Elemente enthalten wie möglich.*“ [4; 33]

Im Kontext von EPK liegen in Anlehnung daran weitere, stärker konkretisierte technologischen Regeln in Form stilisierter Fakten vor, z. B.: „*Um die Strukturiertheit, Übersichtlichkeit und Lesbarkeit (Klarheit) von EPK-Modellen zu erhöhen, sollte der Kontrollfluss von oben nach unten ausgerichtet und die Prozessstränge nach dem Kriterium der Durchlaufhäufigkeit von links nach rechts angeordnet werden*“. [4]

Interessiert sich ein Systemgestalter, der in seiner Organisation EPK für die Geschäftsprozessmodellierung nutzt, für die Einführung eines Workflowmanagementsystems, so kann ihm folgendes stilisierte Faktum, das durch einige wissenschaftliche Arbeiten zu EPK gestützt wird, als technologische Regel dienen. „*Um EPK-Modelle automatisiert ausführen zu können, müssen die EPK-Modelle formalisiert werden (z. B. durch eine Transformation in Petrinetze).*“ [2; 34]

3.5 Standards bzw. Normen und stilisierte Fakten

Standards und technische Normen gelten im Allgemeinen als nützlich, da sie die Qualität und Austauschbarkeit von Gegenständen unterstützen können [40]. Dies gilt insbesondere auch im Kontext der Wirtschaftsinformatik, da durch die Bildung von Standards und Normen, wie z. B. das ISO OSI-Referenzmodell, die Nutzung rechnergestützter Informations- und Kommunikationssystemen überhaupt möglich wird. Des Weiteren werden zur Vereinfachung der Gestaltung von Informationssystemen und zur Verbesserung der Kommunikation zwischen Systemgestaltern auch andere technische Normen, wie Modellierungsmethoden (z. B. die UML), oder Referenzprozessmodelle (z. B. das SCOR-Modell für das Supply Chain Management) etc. vorangetrieben. Die Normierung dieser Techniken ermöglicht die Kommunikation der Beteiligten und ist die Grundlage für ein einheitliches Verständnis, z. B. von Datenstrukturen oder Unternehmensprozessen.

Der Begriff der Norm wird hier, wie bereits erwähnt, im Sinne einer technischen Norm verstanden, also im Gegensatz zu Rechtsnormen („Es ist verboten, über eine rote Ampel zu fahren.“) und sozialen Normen im Sinne moralischer Werturteile („Du sollst nicht töten.“). Technische Normen haben mit diesen anderen Normentypen allerdings gemeinsam, dass sie auch einen normativen Charakter haben. Das heißt, dass Systemgestalter anhand von technischen Normen erfahren können, welche Techniken, bzw. Technikvarianten man verwenden sollte, um das grundlegende Ziel der Einhaltung von Qualitätsstandards, bzw. den geglückten Austausch von Informationen für eine einfachere Kommunikation bzw. Kooperation zu unterstützen.

Stilisierte Fakten sind in diesem Kontext insbesondere hinsichtlich der Identifikation potentieller Standards und Normen von Bedeutung. Standards und technische Normen stellten vor ihrer Standardisierung bzw. Normierung i. d. R. breit akzeptierte Artefakte und Techniken dar, die basierend auf den mit ihrem Einsatz gemachten Erfahrungen weiterentwickelt wurden und aufgrund ihres erfolgreichen Einsatzes in der Praxis als standardisierungs- bzw. normierungswürdig erachtet worden sind. Stilisierte Fakten über charakteristische Eigenschaften von Artefakten und Techniken (Kapitel 3.3) können deshalb die Identifikation potentieller Standards und Normen unterstützen, indem sie Aussagen zur Leistungsfähigkeit und weitere bedeutende Eigenschaften eines Artefakts oder einer Technik bereitstellen.

3.6 Design Theorien und stilisierte Fakten

In der Wirtschaftsinformatik und insbesondere in der IS Community wird der Begriff der Theorie in den letzten Jahren im Vergleich zu dem oben explizierten Theoriebegriff (Aussagesystem über Ursache-Wirkungs-Beziehungen) ausgeweitet. Walls et al. prägten in den frühen 90er Jahren des letzten Jahrhunderts den Begriff der Design Theory [48], der mit der wachsenden Bedeutung des Design Science Paradigmas in der IS Forschung [25] in einer Reihe neuerer Artikel in den führenden IS Zeitschriften (MISQ, ISR, JAIS etc.) intensiv diskutiert und weiterentwickelt

wurde [49; 19; 20; 37]. Einige Publikationen entwickeln Design Theorien für unterschiedliche Themengebiete, verwenden den Begriff und seine Bestandteile aber nicht einheitlich [43; 31; 38].

Dienen erkenntnisorientierte Theorien im oben explizierten Verständnis vor allem der Beschreibung, Analyse, Erklärung und Prognose interessanter Phänomene, so werden diese von Design Theorien als Grundlage (sogenannte „kernel theories") verwendet, um gestaltende Handlungen insbesondere durch präskriptive, bzw. normative Aussagen zu unterstützen [19; 48]. Diese Aussagen werden bisher allerdings sehr unterschiedlich formuliert, z. B. als *technologische Regeln:* „*If you want to improve software processes in a situation where you believe that 'Best Practices' for an improvement area can be identified Then choose a Universally applicable model, such as CMM or Bootstrap. If not then choose a Situated model, such as Juran or Six Sigma*" [38].

Andere Arbeiten formulieren Gestaltungsempfehlungen als Aussagen, die eher den Charakter von *Hypothesen* haben, z. B.: „*The more consistent and concordant the processes and notation of a SIS [secure information system] design method is with common ISD [information system development] methods, the more successful the SIS design will be in practice*" [43], oder in Form einfacher *faktischer Gestaltungsaussagen*, wie z. B.: "*It is feasible and practicable to embed SIS methods into ISD methods*" [43].

Obwohl das Konzept von Design Theorien sehr interessante Ansatzpunkte für die wissenschaftlich fundierte Gestaltung von Informationssystemen anbietet, wird an dieser Stelle deutlich, dass der Begriff noch sehr uneinheitlich verwendet wird und noch kein eindeutiges Verständnis davon herrscht, wie die gestaltungsunterstützenden Aussagen einer Design Theorie aussehen sollten. Die drei gezeigten Beispielaussagen haben allesamt einen eignen, von den anderen Aussagen verschiedenen Charakter. Künftige Forschung in diesem Bereich sollte weiter aufklären, wie die Gestaltung von Informationssystemen durch Design Theorien unterstützt werden kann und die Vereinheitlichung des Verständnisses weiter vorantreiben.

Dass stilisierte Fakten allerdings erhebliche Potentiale für die Entwicklung von Aussagen, die in Design Theorien eine Rolle spielen, nämlich Theorien (kernel theories), artefakt- und technikbezogenes Gestaltungswissen, technologische Regeln und zu einem gewissen Grad auch bei der Findung normativer Aussagen im oben explizierten Sinne beitragen können, wurde im Lauf des Kapitels 3 herausgearbeitet. Welche Rolle Design Theorien in der Wirtschaftsinformatik und in der IS Community in Zukunft spielen werden, bleibt zu untersuchen. Allerdings ist die Entwicklung zentraler Bestandteile von Design Theorien auf Basis des Konzeptes stilisierter Fakten möglich.

4. Erfahrungen mit der Anwendung des Konzeptes stilisierter Fakten in der WI

4.1 Vorbemerkung

Im Folgenden werden erste Erfahrungen mit stilisierten Fakten in der Wirtschaftsinformatik präsentiert. Diese Erfahrungen wurden im Rahmen der Ableitung stilisierter Fakten in sechs Themengebieten der Wirtschaftsinformatik gemacht (siehe Abschnitt 1.3 und Tabelle 1). Basierend auf diesen Erfahrungen werden einige Empfehlungen und Vorschläge für Richtlinien für die Verwendung des Konzeptes in der Wirtschaftsinformatik erarbeitet.

Die Entwicklung stilisierter Fakten erweist sich als ein interessantes Verfahren, das wertvolle Erkenntnisse für unterschiedliche Zwecke hervorbringen kann. In den sechs durchgeführten Fallstudien wurden auf einer für die einzelnen Themengebiete repräsentativen Basis von Quellen stilisierte Fakten gebildet, die auch die Entwicklung größerer Aussagennetze (potentielle Ursache-Wirkungs-Gefüge, technologischen Regeln) ermöglichten. In einigen Fällen wurden auch potentiell widersprüchliche Aussagen identifiziert und abgeleitet. Die Offenlegung dieser Befunde kann sich sehr fruchtbar auf die unterschiedlichen wissenschaftlichen Diskurse in der Wirtschaftsinformatik auswirken.

Im Rahmen der konkreten Anwendung des Generierungsprozesses stilisierter Fakten (siehe Abbildung 1), der den einzelnen Fallstudien zugrunde gelegt wurde, stellen sich dem Anwender im Bereich der Wirtschaftsinformatik einige nicht unerhebliche Fragen zur Operationalisierung des Konzeptes. Einige dieser Fragen werden zwar in den grundlegenden Arbeiten zu stilisierten Fakten [42; 22] teilweise thematisiert, allerdings ist die Ableitung stilisierter Fakten in diesen Arbeiten vornehmlich auf die Theoriebildung und -prüfung ausgelegt und berücksichtigt nicht die Erfordernisse und Besonderheiten der Erarbeitung technikbezogenen Wissens, technologische Regeln etc., die im Rahmen wirtschaftsinformatischer Forschung einen anderen Umgang mit dem Quellenmaterial implizieren. Des Weiteren wurden interessante, noch nicht in der Literatur zu stilisierten Fakten behandelte Sachverhalte entdeckt, die im Folgenden beschrieben und zu denen einige Handlungsempfehlung präsentiert werden. Die drei adressierten Aspekte sind folgende:

1. Notwendigkeit der Anpassung der Extraktionsstrategie an die Zielsetzung,
2. Reliabilität stilisierter Fakten und
3. Fehlende Bewertungsstandards.

4.2 Notwendigkeit der Anpassung der Extraktionsstrategie an die Zielsetzung

Im Kontext der Wirtschaftsinformatik können stilisierte Fakten für unterschiedliche Zwecke eingesetzt werden. Im Rahmen der Fallstudien zeigte sich, dass je nach verfolgter Zielstellung die Extraktionsstrategie angepasst werden sollte. Grundsätzlich wäre es wünschenswert, bei der Bildung stilisierter Fakten sämtliches verfügbare Wissen zu einem bestimmten Gegenstand einfließen zu lassen. Da dies i. d. R. aber nicht eingehalten werden kann und eine möglichst enge Annäherung daran aus ökonomischer Sicht häufig nicht sinnvoll erscheint, ist bei der Auswahl der zugrunde gelegten Quellen nach befriedigenden Lösungen zu suchen.

So erscheint es im Kontext der Bildung neuer Theorien zu einem bestimmten Themengebiet sinnvoll, eine möglichst umfassende Menge von Quellen zu verwenden, um sämtliche relevanten Aspekte des Themengebietes im explorativen Prozess der Theoriebildung berücksichtigen zu können. Sollen stilisierte Fakten zum Zweck der Prüfung existierender Theorien dienen, so kann die Auswahl der verwendeten Quellen auf diejenigen beschränkt werden, die sich explizit mit der zu untersuchenden Problematik auseinandersetzen.

Bei der Ableitung stilisierter Fakten ist auch die qualitative Ausrichtung der abgeleiteten Aussagen an die Zielsetzung der Verwendung stilisierter Fakten anzupassen. Dies ist insbesondere bei der Ableitung stilisierter Fakten für die Entwicklung theoretischer bzw. technologischer Aussagensysteme zu berücksichti-

gen. Werden beispielsweise aus den Aussagen empirischer Literaturquellen im Rahmen des mehrstufigen Abstraktionsprozesses technologische Regeln abgeleitet, so gestaltet sich dieser Prozess deutlich unterschiedlich zu dem Prozess der Entwicklung von Hypothesen für eine potentielle Theorie. Die einzelnen Originalaussagen aus den zugrunde gelegten Quellen sind hinsichtlich der möglichen Zielsetzungen jeweils anders zu abstrahieren. Auch die abgeleiteten Aussagen unterscheiden sich deutlich voneinander Ein Wechseln der Strategie während des Ableitungsprozesses führt zu wenig brauchbaren Ergebnissen. Des Weiteren erweist sich eine Weiterentwicklung von Zwischenergebnissen für eine andere Zielsetzung als schwierig. Wird eine andere Zielsetzung verfolgt, so sollte wieder die Originalaussagen der ursprünglichen Quellen verwendet werden. Vor diesem Hintergrund ist es von Bedeutung, die Zielsetzung und die daran angepasste Extraktionsstrategie eindeutig zu klären und festzulegen.

4.3 Reliabilität stilisierter Fakten

Die gängige Literatur zu stilisierten Fakten weist auf die subjektive Prägung des Prozesses der Aussagenabstraktion und -verdichtung hin [42]. Dieser Prozess erfordert die Interpretation der verwendeten Quellen, die häufig in Form von natürlicher Sprache vorliegen. Dabei spielen unter anderem die Vorkenntnisse der jeweiligen Person eine Rolle, die die Quellen interpretiert. In diesem Kontext besteht dadurch die Gefahr einer von Experten und ihrer subjektiven Meinung geprägten Wissenschaft. Die Literatur verweist darauf, dass aufgrund dessen die intersubjektive Nachvollziehbarkeit der Ableitung stilisierter Fakten von besonderer Bedeutung ist. Dies soll u. a. durch die Offenlegung des mehrstufigen Prozesses der Extraktion und Gruppierung von Aussagen erreicht werden. Allerdings bleibt trotzdem die Reliabilität der erarbeiteten Endergebnisse unklar.

Um die Reliabilität sicherzustellen, sollten stilisierte Fakten von mehreren Personen unter Berücksichtigung der kompletten Menge der ausgewählten Quellen abgeleitet und die Ergebnisse miteinander verglichen werden. Im Idealfall kommen mehrere Subjekte zu ähnlichen Resultaten. Ist dies nicht der Fall, so sollten die beteiligten Personen strittige Punkte klären. Grundsätzlich scheint in diesem Kontext die Frage interessant, ob sich dieser Prozess auch automatisieren lässt. Aufgrund der Notwendigkeit der Interpretation von Aussagen und der Kontextabhängigkeit der Bedeutung von Zeichen und Symbolen erscheint dies bisher allerdings problematisch.

4.4 Fehlende Bewertungsstandards

Um die Bedeutung, die Aussagekraft und den Konsens zu einzelnen stilisierten Fakten abzubilden, werden diese in den gängigen Arbeiten anhand der Anzahl der Quellen, die die Aussage stützen, bewertet und darauf basierend in unterschiedliche Kategorien eingeteilt (*Konsensanalyse*). Weißenberger/Löhr legen für ihre Arbeit beispielsweise drei Kategorien fest, wobei stilisierte Fakten, die von vier oder mehr Quellen gestützt werden, in die Kategorie A („gut gestützt“), von zwei oder drei Quellen in die Kategorie B („durchschnittlich gestützt“) und mit einer Quelle in die Kategorie C („schwach gestützt“) eingeordnet werden [50]. Bei der Festlegung der nötigen Quellen pro Kategorie ist stets eine sinnvolle Relation zur Gesamtanzahl der verwendeten Quellen herzustellen.

Diese Methode lässt zwar grobe Rückschlüsse auf die Akzeptanz und Relevanz der Aussage in einer Community zu. Allerdings kann es als unstrittig gelten, dass häufiges Nennen einer offensichtlich falschen Behauptung ihren Inhalt nicht evidenter macht. Vor diesem Hintergrund wird vorgeschlagen, stilisierte Fakten anhand der Evidenz ihrer zugrunde gelegten Quellen zu bewerten. Dieser Ansatz, der häufig im Bereich der Evidenzbasierten Medizin verwendet wird, wurde auch schon in der Wirtschaftsinformatik zur Bewertung von Forschungsergebnissen diskutiert [13; 17].

In [17] wird beispielsweise ein sechsstufiges Klassifikationssystem verwendet, das verschiedene Evidenztypen von Experteninterviews (niedrigste Stufe) über Experimente, Fallstudien, qualitative und quantitative Querschnittsarbeiten sowie narrative Übersichtsarbeiten bis hin zu Metaanalysen (höchste Stufe) unterscheidet. Diese Übersicht ist allerdings auf empirische Arbeiten beschränkt. Das Evidenzstufenkonzept in [13] unterstützt eine Bewertung sämtlichen Wissens, also auch konzeptionell-argumentativer Quellen. Es werden fünf Evidenzstufen festgelegt, die von Stufe I („plausible Aussage ohne weitere Begründung“) über Stufe II („plausible Aussage, die mit rein konzeptionellen Überlegungen belegt wird“), III („Aussage, die durch beispielhafte Erfahrungen gestützt wird“) und IV („Aussage, die sich im Rahmen einer Vielzahl von Anwendungen bewährt hat.“) bis hin zu Stufe V („Aussage, die ohne Einschränkung gilt, bzw. die deduktiv aus anerkannten Aussagen abgeleitet werden kann.“) reichen. Im vorliegenden Beitrag wird vor diesem Hintergrund vorgeschlagen, stilisierte Fakten zusätzlich zur Konsensanalyse anhand der untersuchten Quelle mit der höchsten Evidenzstufe zu bewerten. Diese zusätzliche Information ermöglicht eine eindeutigere Bewertung stilisierter Fakten und lässt deutlicher Rückschlüsse auf ihre Qualität zu.

Im Falle der Ableitung stilisierter Fakten aus wissenschaftlicher Literatur kann es als sinnvoll angesehen werden, dass Aussagen, die in Topjournalartikeln bzw. Topkonferenzartikeln formuliert werden, ein höheres Gewicht haben, da diese Artikel in der Regel kritischer und nach strengeren Vorgaben begutachtet worden sind. Allerdings erscheint dies nur dann zweckmäßig zu sein, wenn stilisierte Fakten ausschließlich aus Literatur abgeleitet werden, für die eine Einordnung möglich ist, d. h. die in relevanten Ranking geführt wird. Da allerdings unterschiedliche Rankings verschiedene Journal und Konferenzen sehr unterschiedlich bewerten, bleibt fraglich, welches Ranking verwendet werden sollte. Eine Operationalisierung des Gewichtungsprozesses erscheint ebenso schwierig.

5. RESÜMEE UND AUSBLICK

Die vorliegende Arbeit dokumentiert grundlegende Potentiale und erste Erfahrungen mit dem wissenschaftstheoretischen Konzept stilisierter Fakten in der gestaltungsorientierten Wirtschaftsinformatik. Ein wesentlicher Beitrag, der im Rahmen dieses Artikels geleistet wurde, ist die Übertragung des Konzeptes stilisierter Fakten auf die gestaltungsorientierte Wirtschaftsinformatik und das Aufzeigen seiner Potentiale. Stilisierte Fakten können folgende zentralen Aufgaben und Ziele der wirtschaftsinformatischen Forschung unterstützen:

1. die Bildung, Prüfung und den Vergleich von Theorien,
2. die Entwicklung von wissenschaftlichem Gestaltungswissen über Artefakte und Techniken,
3. die Entwicklung und Prüfung technologischer Aussagensysteme,

4. die Identifikation potentieller Standards und Normen der Wirtschaftsinformatik sowie
5. die Entwicklung von Design Theorien.

Des Weiteren wurden Erfahrungen mit der Anwendung des Konzeptes präsentiert, noch offene Punkte zu seiner Operationalisierung in der Wirtschaftsinformatik aufgezeigt und diskutiert sowie einige Richtlinien und Handlungsempfehlungen zu seiner Anwendung angeboten.

Die Gewinnung stilisierter Fakten stellt ein anspruchvolles Verfahren dar, das erhebliche Potentiale für die gestaltungsorientierte Wirtschaftsinformatik und ihre wissenschaftstheoretische Fundierung anbieten kann. Es ermöglicht die grundlegende Entwicklung eigener Theorien der Wirtschaftsinformatik, die kumulative Weiterentwicklung von Gestaltungswissen und somit eine potentiell erfolgreiche Grundlagenforschung zur Gestaltung von Informationssystemen.

DANKSAGUNG

Die hier dargestellten Forschungsergebnisse wurden teilweise im Rahmen zweier Forschungsprojekte erarbeitet: „Prozessorientierter Web-2.0-basierter integrierter Telekommunikationsservice (PROWIT), gefördert durch das Bundesministerium für Bildung und Forschung (BMBF) (Förderkennzeichen FKZ 01BS0833), sowie „Pluralistische Beurteilung der Qualität von Unternehmensmodellen – Qualitätsdiskurse und Diskursqualität innerhalb der Wirtschaftsinformatik (PluralistiQue), gefördert durch die DFG (GZ LO 752/4-1). Die Autoren möchten sich weiterhin bei den anonymen Gutachtern für die konstruktiven Anmerkungen bedanken, die zur Verbesserung diese Artikel beigetragen haben.

LITERATUR

[1] Agassi, J. 1966. The Confusion between Science and Technology in the Standard Philosophies of Science. *Technology and Culture*. 7, 3 (1966), 348-366.

[2] Allweyer, T. 2007. Erzeugung detaillierter und ausführbarer Geschäftsprozessmodelle durch Modell-zu-Modell-Transformationen. In *EPK 2007 - Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten, Proceedings des GI-Workshops und Arbeitskreistreffens*. (St. Augustin 2007), 23-38.

[3] Becker, J. 1995. Strukturanalogien in Informationsmodellen - Ihre Definition, ihr Nutzen und ihr Einfluss auf die Bildung von Grundsätzen ordnungsmäßiger Modellierung (GOM). In *Wirtschaftsinformatik '95. Wettbewerbsfähigkeit, Innovation, Wirtschaftlichkeit*, W. König, Ed. Physica. Heidelberg, 133-150.

[4] Becker, J.; Rosemann, M. und Schütte, R. 1995. Grundsätze ordnungsmäßiger Modellierung. *Wirtschaftsinformatik*. 37, 5 (1995), 435-445.

[5] Bortz, J. und Döring, N. 2006. *Forschungsmethoden und Evaluation - für Human- und Sozialwissenschaftler*. Springer, Berlin et al.

[6] Bunge, M. 1967. *Scientific Research II: The Search for Truth*. Springer, Berlin.

[7] Chmielewicz, K. 1994. *Forschungskonzeptionen der Wirtschaftswissenschaft*. Schäffer-Poeschel, Stuttgart.

[8] Dehnert, J. 2001. Four Systematic Steps Towards Sound Business Process Models. In *Proceedings of the 2nd International Colloquium on Petri Net Technologies for Modelling Communication Based Systems*. (Berlin 2001), 55-64.

[9] DeLone, W. H. und McLean, E. R. 1992. Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*. 3, 1 (1992), 60-95.

[10] DeLone, W. H. und McLean, E. R. 2003. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*. 19, 4 (2003), 9-30.

[11] Detel, W. 2007. *Grundkurs Philosophie. Band 4: Erkenntnis- und Wissenschaftstheorie*. Reclam, Stuttgart.

[12] Fettke, P. 2006. State-of-the-Art des State-of-the-Art - Eine Untersuchung der Forschungsmethode „Review“ innerhalb der Wirtschaftsinformatik. *Wirtschaftsinformatik*. 48, 4 (2006), 257-266.

[13] Fettke, P.; Houy, C. und Loos, P. 2010. Zur Bedeutung von Gestaltungswissen für die gestaltungsorientierte Wirtschaftsinformatik. Konzeptionelle Grundlagen, Anwendungsbeispiel und Implikationen. *Wirtschaftsinformatik*. 52, 6 (2010), 339-352.

[14] Frank, U. 2006. *Towards a Pluralistic Conception of Research Methods in Information Systems Research*. Institut für Informatik und Wirtschaftsinformatik (ICB) der Universität Duisburg-Essen, Report No. 7. Essen.

[15] Frank, U. 2010. Zur methodischen Fundierung der Forschung in der Wirtschaftsinformatik. In *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, H. Österle; R. Winter; W. Brenner, Ed. St. Gallen, 35-44.

[16] Gloy, K. 2004. *Wahrheitstheorien - Ein Einführung*. A. Francke, Tübingen, Basel.

[17] Goeken, M. und Patas, J. 2010. Evidenzbasierte Strukturierung und Bewertung empirischer Forschung im Requirements Engineering. *Wirtschaftsinformatik*. 52, 3 (2010), 173-184.

[18] Green, P. und Rosemann, M. 2000. Integrated Process Modelling: An Ontological Evaluation. *Information Systems*. 25, 2 (2000), 73-87.

[19] Gregor, S. 2006. The Nature of Theory in Information Systems. *MIS Quarterly*. 30, 3 (2006), 611-642.

[20] Gregor, S. und Jones, D. 2007. The Anatomy of a Design Theory. *Journal of the Association for Information Systems*. 8, 5 (2007), 312-335.

[21] Grundmann, T. 2008. *Analytische Einführung in die Erkenntnistheorie*. Walter de Gruyter, Berlin.

[22] Heine, B.-O.; Meyer, M. und Strangfeld, O. 2007. Das Konzept der stilisierten Fakten zur Messung und Bewertung wissenschaftlichen Fortschritts. *Die Betriebswirtschaft (DBW)*. 67, 5 (2007), 583-601.

[23] Heinrich, L. J.; Heinzl, A. und Roithmayr, F. 2007. *Wirtschaftsinformatik – Einführung und Grundlegung*. Oldenbourg, München, Wien.

[24] Helfat, C. E. 2007. Stylized facts, empirical research and theory development in management. *Strategic Organization*. 5, 2 (2007), 185-192.

[25] Hevner, A. R.; March, S. T.; Park, J. und Ram, S. 2004. Design Science in Information Systems Research.

Management Information Systems Quarterly. 28, 1 (2004), 75-105.

[26] Iivari, J. 2007. A Paradigmatic Analysis of Information Systems As a Design Science. *Scandinavian Journal of Information Systems*. 19, 2 (2007), 39-64.

[27] Kaldor, N. 1961. Capital Accumulation and Economic Growth. In *The Theory of Capital, Proceedings of a Conference Held by the International Economic Association*. (London 1961),

[28] Kuhn, T. S. 1996. *Die Struktur wissenschaftlicher Revolutionen*. Suhrkamp, Frankfurt am Main.

[29] Lakatos, I. 1978. *The methodology of scientific research programmes*. Cambridge University Press, Cambridge, New York, Melbourne.

[30] Lauth, B. und Sareiter, J. 2005. *Wissenschaftliche Erkenntnis. Eine ideengeschichtliche Einführung in die Wissenschaftstheorie*. mentis, Paderborn.

[31] Markus, L. M.; Majchrzak, A. und Gasser, L. 2002. A Design Theory for Systems that Support Emergent Knowledge Processes. *MIS Quarterly*. 26, 3 (2002), 179-212.

[32] Mellewigt, T. 2003. *Management von Strategischen Kooperationen - Eine ressourcenorientierte Untersuchung in der Telekommunikationsbranche*. DUV, Wiesbaden.

[33] Mendling, J.; Reijers, H. A. und van der Aalst, W. M. P. 2010. Seven process modeling guidelines (7PMG). *Information and Software Technology*. 52, 2 (2010), 127-136.

[34] Moldt, D. und Rodenhagen, J. 2000. Ereignisgesteuerte Prozeßketten und Petrinetze zur Modellierung von Workflows. In *Visuelle Verhaltensmodellierung verteilter und nebenläufiger Software-Systeme, Proceedings des 8. Workshop des Arbeitskreises "Grundlagen objektorientierter Modellierung" (GROOM), Bericht Nr. 24/00-I*. (Münster 2000), 57-63.

[35] Österle, H.; Becker, J.; Frank, U.; Hess, T.; Karagiannis, D.; Krcmar, H.; Loos, P.; Mertens, P.; Oberweis, A. und Sinz, E. J. 2010. Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, H. Österle; R. Winter; W. Brenner, Ed. St. Gallen, 1-6.

[36] Popper, K. R. 1934. *Logik der Forschung*. Julius Springer Verlag, Wien.

[37] Pries-Heje, J. und Baskerville, R. 2008. The Design Theory Nexus. *MIS Quarterly*. 32, 4 (2008), 731-755.

[38] Pries-Heje, J. und Baskerville, R. 2009. A Design Theory for Managing Software Process Improvement. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST'09). Philadelphia, Pennsylvania.

[39] Recker, J.; Rosemann, M.; Indulska, M. und Green, P. 2009. Business Process Modeling - A Comparative Analysis. *Journal of the Association for Information Systems*. 10, 4 (2009), 333-363.

[40] Reimers, K. 1995. *Normungsprozesse - Eine transaktionskostentheoretische Analyse*. Gabler, Wiesbaden.

[41] Rittgen, P. 2000. Quo vadis EPK in ARIS? Ansätze zu syntaktischen Erweiterungen und einer formalen Semantik. *Wirtschaftsinformatik*. 42, 1 (2000), 27-35.

[42] Schwerin, J. 2001. *Wachstumsdynamik in Transformationsökonomien*. Böhlau Verlag, Köln, Weimar, Wien.

[43] Siponen, M.; Baskerville, R. und Heikka, J. 2006. A Design Theory for Secure Information Systems Design Methods. *Journal of the Association for Information Systems*. 7, 11 (2006), 725-770.

[44] Stegmüller, W. 1983. *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie, Band I, Erklärung, Begründung, Kausalität*. Springer, Berlin et al.

[45] van der Aalst, W.; Desel, J. und Kindler, E. 2002. On the semantics of EPCs: A vicious circle. In *EPK 2002 - Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten, Proceedings des GI-Workshops und Arbeitskreistreffens*. (Trier 2002), 71-79.

[46] van der Aalst, W. M. P. 1999. Formalization and verification of event-driven process chains. *Information and Software Technology*. 41, (1999), 639-650.

[47] van Dongen, B. F.; Jansen-Vullers, M. H.; Verbeek, H. M. W. und van der Aalst, W. M. P. 2007. Verification of the SAP reference models using EPC reduction, state-space analysis, and invariants. *Computers in Industry*. 58, 6 (2007), 578-601.

[48] Walls, J.; Widmeyer, G. und Sawy, O. E. 1992. Building an information systems design theory for vigilant EIS. *Information Systems Research*. 3, 1 (1992), 36-59.

[49] Walls, J. G.; Widmeyer, G. R. und A El Sawy, O. 2004. Assessing Information System Design Theory in Perspective: How useful was our 1992 Initial Rendition? *Journal of Information Technology Theory and Application*. 6, 2 (2004), 43-58.

[50] Weißenberger, B. E. und Löhr, B. W. 2007. Planung und Unternehmenserfolg: Stylized Facts aus der empirischen Controllingforschung im deutschsprachigen Raum von 1990 - 2007. *Zeitschrift für Planung & Unternehmenssteuerung*. 18, 4 (2007), 335-363.

[51] Wild, J. 1976. Theorienbildung, betriebswirtschaftliche. In *Handwörterbuch der Betriebswirtschaft*, E. Grochla; W. Wittmann, Ed. Stuttgart, Sp. 3889-3910.

[52] Winter, R.; Krcmar, H.; Sinz, E. J.; Zelewski, S. und Hevner, A. R. 2009. Was ist eigentlich Grundlagenforschung in der Wirtschaftsinformatik? *Wirtschaftsinformatik*. 51, 2 (2009), 224-231.

[53] Zelewski, S. 1994. *Das Konzept technologischer Theorietransformationen - eine Analyse aus produktionswirtschaftlicher Perspektive*. Universität Leipzig, Institut für Produktionswirtschaft und industrielle Informationswirtschaft, Report Nr. 1. Leipzig.

[54] Zelewski, S. 1999. Grundlagen. In *Betriebswirtschaftslehre*, H. Corsten; M. Reiß, Ed. Oldenbourg. München, Wien, 5-125.

Die Zielfunktion des Universitätslehrers der Wirtschaftsinformatik – Setzen wir falsche Anreize?

Peter Mertens
Prof. Dr.
Universität Erlangen-Nürnberg
Lange Gasse 20, 90403 Nürnberg
Tel. 0049(0)911-5302-493
mertens@wiso.uni-erlangen.de

ZUSAMMENFASSUNG

Gegenwärtig haben für die Beurteilung von Universitätslehrern allgemein und der Wirtschaftsinformatik speziell aus verschiedenen Anlässen (z. B. Lehrstuhlbesetzungen, Ranglisten, Aufnahme in die deutsche Exzellenzinitiative) die Veröffentlichungen in internationalen Zeitschriften besonderes Gewicht. Andere Fähigkeiten und Verdienste, die einen guten Hochschullehrer kennzeichnen, werden relativ gering geachtet. Entsprechend ändern sich die persönlichen Ziele und Prioritäten. Vor allem junge Wissenschaftler müssen dem Rechnung tragen, sodass es zu „Kollateralschäden", besonders zu Lasten der Lehre und des Praxisbezugs, kommt. Es werden zehn Ziele aufgeführt und nachteilige Wirkungen der einseitigen Gewichtung in der gemischten Zielfunktion für die Wirtschaftsinformatik erörtert. Der Referent plädiert für eine ausgewogenere Anreizstruktur, wobei auch die Ausgründung neuer Unternehmen aus der Universität honoriert wird. Er hinterfragt, ob „Internationalität" undifferenziert als Verdienst gewürdigt werden sollte.

Schlüsselwörter

Anreiz, Ausgründung, Internationalität, Praxistransfer, Rangliste, Universitätslehrer, Veröffentlichung

1. Einleitung

Starke Kräfte führen im Wissenschaftsbetrieb vieler, auch deutschsprachiger, Länder dazu, dass Zielvereinbarungen und damit gut definierte sowie operationalisierbare Ziele an Bedeutung gewinnen. Dies wiederum strahlt auf die Anreize aus, die der Wissenschaftlerin/dem Wissenschaftler geboten werden.

In diesem Beitrag wird zur Diskussion gestellt, ob die Anreizsysteme für Universitätslehrer der Wirtschaftsinformatik (WI) angemessen sind. Dabei soll die Sicht der Anspruchsberechtigten („Stakeholder des Wissenschaftsbetriebs") besonders beachtet werden ([2], [42], [50]).

2. Der Zehnkämpfer in der Universität

Der Katalog der Abbildung 1 listet stichwortartig Kriterien auf, nach welchen man einen Forscher und Lehrer an Universitäten beurteilen wird. In einer Analogie mag man sie als „Disziplinen eines Zehnkampfs" begreifen. Die Reihenfolge soll nichts über die Bedeutung sagen.

Die Gewichte bei der Gesamtbeurteilung fielen zu unterschiedlichen Zeiten jeweils anders aus.

In der Folge werden einige der Disziplinen des „Zehnkampfs" kommentiert:

Die Bedeutung der **Lehre** an Universitäten (D1) war in den Jahrzehnten, in denen ich als Hochschullehrer tätig war und in bescheidenem Maße noch bin, beträchtlichen Schwankungen unterworfen. Ein Extrem war die Auffassung „Wer einen Universitätsabschluss erwerben will, muss auch einem guten Wissenschaftler mit schlechter Didaktik folgen können". Als das andere Extrem mag man die Absicht der deutschen Bundesforschungsministerin werten, der Exzellenzinitiative Forschung eine solche für die Lehre folgen zu lassen.

D1.	Lehre, Betreuung von Studierenden, darunter Beschaffung von Praktikantenplätzen im In- und Ausland, Beschaffung von Stipendien im Rahmen des deutschen Nationalen Stipendienprogramms
D2.	Erarbeiten von Lehrbüchern und anderen Lehrmaterialien
D3.	Forschung und Ergebnistransfer in die Wissenschaft → Veröffentlichung in wissenschaftlichen Organen, auch international, Kongresse
D4.	Forschung und Ergebnistransfer in die Praxis → Veröffentlichung in Praktikerorganen, Fachbücher, Vorträge, (MBA-) Kurse, Kooperationsprojekte, Patente, Lizenzen
D5.	Forschung und Ergebnistransfer in Politik und Gesellschaft → Mitwirkung in Gremien, Politikberatung, Auftritt in Medien
D6.	Standortförderung, Hilfe bei Unternehmensgründung („Spin-offs") und deren Nutzung für die Ausbildung von Studierenden
D7.	Selbstverwaltung in der Universität
D8.	Selbstverwaltung in der Fachgemeinschaft → Gremien, Herausgeberkreise von Fachzeitschriften, vielfältige Gutachten, z. B. bei Besetzung von Professuren oder Preisverleihungen
D9.	Ausbildung des wissenschaftlichen Nachwuchses
D10.	Drittmittelakquisition

Abbildung 1: Die zehn „Disziplinen"

Eine Herausforderung, die es in dieser Form früher nicht gab, ist die in jedem Einzelfall sehr aufwändige Beschaffung von Praktikantenarbeitsplätzen im Ausland.

Die Exzellenzinitiative der deutschen Bundesregierung und verwandte Anreizsysteme haben wegen der starken Fokussierung auf internationale Publikationen (siehe Abschnitt 4) und wegen ihrer

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

Ausstrahlung auf die persönlichen Ziele von Hochschulpräsidenten und -rektoren zu Kollateralschäden geführt [24], zu denen auch die relative Geringschätzung der Lehre zählt. So wird Roman Inderst, Gewinner des Leibniz-Preises und damit des höchstdotierten deutschen Wissenschaftspreises, in der FAZ vom 21.12.2009 wie folgt beschrieben: „An der Lehre liegt ihm weniger. „Ich will vor allem das beste Forschungsumfeld", sagt Inderst – also mit so wenig … Lehrpflichten … wie möglich."" [41]

Auch das Erarbeiten von **Lehrmaterialien** (D2) war im Lauf der jüngeren Universitätsgeschichte unterschiedlich angesehen. Als ich Assistent war, galt es als unfein, aus neutraler Position heraus Material aus Publikationen anderer Wissenschaftler zusammenzutragen und in einem Lehrbuch didaktisch so aufzubereiten, dass man zur Lektüre keine Spezialkenntnisse brauchte. Später gab es eine Phase, in der sehr viele Hochschullehrer einfach aus ihrem Vorlesungsskript ein Lehrbuch ableiteten.

Publikationen in Praktikerorganen und damit verwandte Formen des Wissenstransfers wie etwa Referate auf großen Kongressen dieses Personenkreises (D4) und Fachbücher stehen nicht in hohem Ansehen. Mir sind mehrere Episoden bekannt, bei denen Mitglieder von Berufungskommissionen beim Vergleich von Veröffentlichungsverzeichnissen darauf hinwirkten, solche Aufsätze vorab zu eliminieren. Es fällt auf, dass auch dem Fach Information Systems in den USA in diesem Zusammenhang Vorwürfe gemacht werden. So schreiben Gill und Bhattacherjee: „Based on this assessment, we observe that the degree to which MIS research is informing its key external clients – practitioners, students an researchers in other disciplines – has declined over the years" [15].

Der **Auftritt in Medien** (D5) hat– zumindest wenn er sich häuft – zuweilen Naserümpfen zur Folge. Dies mag dann berechtigt sein, wenn ein Wissenschaftler sich nicht nur zu seinem Spezialgebiet äußert, sondern aufgrund seiner sogenannten Medientauglichkeit eine Art Allgemeinzuständigkeit suggeriert.

Der Wissenstransfer in die Privatwirtschaft und in die öffentliche Verwaltung über **Ausgründungen** („Spin-offs") (D6) erscheint dem Referenten für ein Fach wie die WI, zumindest für ihren konstruktionsorientierten Zweig, von außerordentlicher Bedeutung. Insbesondere trägt er der Notwendigkeit Rechnung, den Investoren in die Hochschulforschung zu einem Ertrag zu verhelfen und unfairen Umverteilungen (siehe dazu unten den Abschnitt über Internationalisierung) vorzubeugen.

Solche Ausgründungen haben auch große volkswirtschaftliche Bedeutung. So bauten zwischen 1977 und 2005 die bestehenden Unternehmen in den USA durchschnittlich eine Million Arbeitsplätze pro Jahr ab, während neu gegründete schon im ersten Jahr ihres Bestehens zusammen jährlich für drei Millionen Arbeitsstellen sorgten [53].

Als langjähriges Mitglied der Universität Erlangen-Nürnberg habe ich das Wettrennen zwischen Unternehmen aus vielen Ländern verfolgt, welches einsetzte, als es in unserer Technischen Fakultät mit Hilfe des OCF-Algorithmus' gelungen war, große Musikdateien so zu komprimieren, dass sie durch Netze übertragen werden konnten – heute als MP3-Verfahren bekannt. Eine Patentanmeldung war gescheitert, der in der Nähe ansässige Grundig-Konzern verkannte das Potenzial [10]. Der Erfolg versprechende Innovationspfad hätte nur über eine Instituts- oder Unternehmensgründung, verbunden mit der Sicherung der Betriebsgeheimnisse, führen können. (In diese Richtung ist man später auch gegangen.) Man mag im Geiste Revue passieren lassen, was geschehen wäre, hätte auf den Wissenschaftlern Dieter Seitzer und Karlheinz Brandenburg der heutige Druck zur raschen Publikation in einem internationalen Journal gelastet.

So nimmt es nicht Wunder, dass die Ausgliederung in verschiedenen Ländern an Bedeutung gewinnt, wie die folgenden Beispiele zeigen:

1. Große IT-Unternehmen wie IBM, SAP oder Siemens haben wiederholt demonstriert, welchen Rang sie dieser Spielart des Wissenserwerbs beimessen. So sagte etwa der bis 2010 amtierende Vorstandsvorsitzende der SAP AG, Apotheker: „Der Tanker braucht Schnellboote an seiner Seite, SAP will diese kaufen." (Abbildung 2) [16] Oracle und Cisco Systems haben in den vergangenen fünf Jahren zusammen mehr als 80 Unternehmen erworben [31]. Andy Rubin hat nach seinem Ausscheiden bei Microsoft das Unternehmen Android gegründet, um das Aufsehen erregende gleichnamige Betriebssystem zu entwickeln.

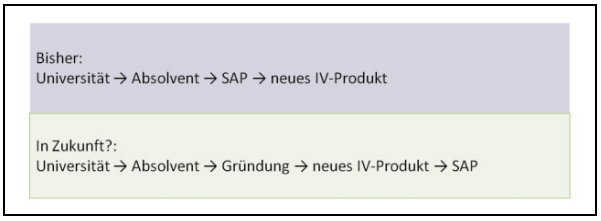


Abbildung 2: Traditionelle und alternative Abläufe beim Wissenstransfer in der WI

2. Für die englische Hochschulpolitik ist seit 2009 das Wirtschaftsministerium zuständig. Es drängt darauf, dass jeder Forscher die wirtschaftliche und gesellschaftliche Relevanz seiner Arbeiten erhöht, wobei er den Transfer in die Praxis nicht anderen überlässt, sondern mitgestaltet („Impact als Erfolgsindikator", „Hochschulutilitarismus") [30].

3. Sehr gezielt fördert Israel derartige Ausgründungen auf dem Gebiet der IT und hat damit einen Spitzenplatz in der Welt erreicht („Start-up-Nation") ([7], [48]).

4. Eine ebenso gezielte Strategie hat Österreich auf dem Sektor Biotechnologie entwickelt. Mit fast 3 Mrd. €/Jahr setzt die Branche mehr um als die deutsche Konkurrenz [47].

5. Das Ziel des Wissenstransfers durch Ausgründung lässt sich mit der Lehre gut vereinbaren. Als Modifikation des Humboldtschen Universitätsideals in Richtung „Einheit von Forschung, Lehre und Wissenstransfer" mag man das Konzept der National University of Singapore ansehen: Sie hat Niederlassungen an mehreren Orten in aller Welt gegründet (unter anderem Bangalore, Peking, Silikon Valley, Stockholm). Dort werden die Studenten in Zusammenarbeit mit neu gegründeten Unternehmen ausgebildet [26]. Ein ähnliches Konzept, allerdings lokal, hatte ich im Rahmen von studentischen Projekten umgesetzt – in der Regel zum Vorteil sowohl der Studierenden, der Gründer als auch der Universität. Das Projektstudium im zweiten Studienabschnitt fand zu einem beachtlichen Teil in Kooperation mit den aus meinem Bereich hervorgegangenen jungen Unternehmen statt. Vorteile waren u. a., dass die Gründer durch ihre frühere Assistententätigkeit Erfahrung beim Betreuen studentischer Arbeiten hatten und die Studierenden die reizvolle Atmosphäre in

einem solchen Betrieb kennen lernten. Eine verwandte Lösung in der WI der Universität Siegen beschreiben Fischer, Rohde und Wulf [11].

Selbst wenn nach einiger Zeit ein solches wissenschaftsnahes Unternehmen von einem ausländischen Konzern aufgekauft wird, ist (unter anderem durch die Versteuerung der Differenz zwischen Kaufpreis und Buchwert) eine Investitionsrendite für den deutschen Steuerzahler entstanden.

Die Gestaltung der jungen Unternehmen bedarf sorgfältiger Überlegungen: Wenn ein Lehrstuhlinhaber oder Institutsdirektor sich selbst nicht nur als Förderer, Mentor, Ratgeber, oder Aufsichtsrat, sondern als Unternehmer betätigt, kann rasch die Erfüllung anderer Aufgaben des Hochschullehrers leiden. Insofern ist zum einen die Gründung durch ehemalige Assistenten vorzuziehen. Dies war mir persönlich stets ein Anliegen, ich habe solche Initiativen mit Freude unterstützt und blicke mit Genugtuung auf die rund 1.000 so in Deutschland geschaffenen Arbeitsplätze. Ein sehr bedenkenswerter Weg ist Professoren in der Schweiz eröffnet. So hat sich z. B. Hubert Österle nach der Gründung eines Software-Unternehmens auf eine halbe Universitätsstelle zurückgezogen und aus den Erträgen seines Betriebs auch die Aufstockung der anderen halben Stelle zu einer Vollzeit-Professur mitfinanziert.

Die **Drittmittelakquisition** (D10) hat viel mit den anderen Anreizen zu tun. Man kann sogenannte „Drittmittelstärke" bzw. „Drittmittelerfolge" teilweise als Funktion der bisherigen Publikationen in wissenschaftlichen Organen ansehen (Disziplin 3), aber auch des Wissenstransfers in die Praxis (Disziplin 4), m. a. W.: Der Erfolg nährt den Erfolg. Insofern gelten bei der Bewertung die dort angestellten Überlegungen.

3. Die Eigenschaften des guten „Zehnkämpfers" in der Hochschule

Abbildung 3 enthält wesentliche Voraussetzungen, die unser „Zehnkämpfer" erfüllen muss, um im „Wettkampf" gut zu bestehen.

1. Didaktische Begabung, Fähigkeit der Wissensvermittlung bei zum Teil schwierigem Auditorium (z. B. Abiturientenanteil 40 %, Lockerung der Zugangsbedingungen zur Universität, Massenbetrieb)
2. Kenntnis des aktuellen Wissenstands bei raschem Wissensumschlag („Wissen, das der Bachelor bei seinem Abschluss hatte, ist schon überholt, wenn er seinen Master abschließt")
3. Kenntnis der neuesten Forschungsmethoden
4. Praxiserfahrung durch Tätigkeit in Betrieben
5. Auslandsaufenthalte
6. Fähigkeit zur Entwicklung von Prototypen und (zumindest in kleinem Maßstab) produktiven Systemen
7. Frustrationstoleranz bei der Mitwirkung an Reformen des tertiären Bildungssektors trotz kurzer Lebensdauer der Reformergebnisse

Abbildung 3: Voraussetzungen eines guten Universitätslehrers

Nicht selbstverständlich, aber sehr wichtig ist Eigenschaft 4: Wenn ein hoher Prozentsatz der Studierenden nach dem Bachelor erst einige Jahre in der Praxis arbeitet, bevor ein Master-Studium aufgenommen wird, kann der Dozent schlecht zu der Minderheit derer im Hörsaal gehören, die einen Betrieb nie „von innen" erlebt haben. In den Ausschreibungen zu WI-Professuren taucht diese Voraussetzung jedoch nur noch selten auf.

Diese ohnehin hohen Anforderungen werden derzeit noch dadurch überlagert, dass wir eine Belastungsspitze erleben, bedingt durch besonders starke Altersjahrgänge, durch Schulzeitverkürzung verursachte Doppeljahrgänge und den Wegfall der Wehrpflicht.

4. Problematische Anreize

Zurzeit legt man bei der Beurteilung von Wissenschaftlern, namentlich von Nachwuchsleuten, ein sehr großes Gewicht auf Publikationen in internationalen Zeitschriften und damit auf einen Teil der „Disziplin 3" gemäß Abbildung 1. So liest man z. B. in einer Stellenbeschreibung der FernUniversität Hagen vom 11.08. 2010 nach der Aufgabenbeschreibung nur zwei Kriterien: „Die Stelleninhaberin … soll in der Forschung mit Schwerpunkt in den Gebieten Produktion und Logistik durch Publikationen u. a. in angesehenen wissenschaftlichen Fachzeitschriften ausgewiesen sein." (Das zweite Kriterium betrifft die „Bereitschaft zu fachübergreifender Zusammenarbeit", die übrigen Merkmale sind nicht leistungsorientiert, wie z. B. Altershöchstgrenze, Bevorzugung von Frauen.)

Hiermit sind viele Probleme verbunden:

1. Es prägen sich Normstrukturen beziehungsweise -gliederungen aus, die nicht immer passen.

   Vor einigen Jahren hörte ich von einem amerikanischen Kollegen einen Vergleich: Angenommen, auf einem Kongress von Nahrungsmittelchemikern wäre sensationellerweise der Chefchemiker von Coca-Cola bereit, endlich die Jahrzehnte lang geheim gehaltene Zusammensetzung des Sirups zu beschreiben, und das Programmkomitee würde ihm zur Bedingung machen: „Legen Sie zunächst Ihre wissenschaftlichen Forschungsziele dar und beschreiben Sie, auf welchem Signifikanzniveau Sie Ihre Hypothesen testen!"

   Osterloh und Frey weisen darauf hin, dass derartige Standards den Ausbruch aus der Konformität erschweren [33, S. 23].

2. Aufsätze über die Beobachtung beziehungsweise Prüfung beziehungsweise Bewertung dessen, was *andere* erfunden und entwickelt haben („Empirischer Ansatz" der WI) besitzen bessere Chancen, von US-Zeitschriften angenommen zu werden, als solche, in denen eigene neue Ideen/Vorschläge/Konzepte/Prototypen behandelt werden (auch als Konstruktiver Ansatz bezeichnet). Letzteres gleicht – um im Bild zu bleiben – einem Hürdenlauf: In den verschiedenen Ausreifungsstufen werden die Manuskripte mit stets anderen Gründen abgelehnt (Abbildung 4).

   Bei dieser Einstellung hätte z. B. Wernher von Brauns bahnbrechendes Ideengut zur Weltraumfahrt, das diesen außerordentlich wichtigen Forschungs- und Entwicklungspfad eröffnet hatte, nicht publiziert werden können, ebenso nicht Zuses Beschreibung der Z1 oder Zemaneks des „Mailüfterls".

Phase	Entscheidung	Ablehnungsgrund
Konzeptionsphase	Ablehnung	„No proof of concept"
Prototypenphase (Machbarkeitsbeweis)	Ablehnung	„Only sample size one"
Realisierung in der Praxis	Ablehnung	„We do not publish How-I-dit-it-papers"
Breite Durchsetzung in der Wirklichkeit		
Beobachtung von außen (durch Dritte), Befragungen, Hypothesenprüfung	Annahme	

Abbildung 4: Gegenpositionen zu Konstruktionsarbeiten

3. Noch fehlt es an Belegen, ob die Praxis überhaupt in nennenswertem Maß Notiz von den sogenannten „Top Journals“ nimmt, zumindest in der WI. Die geringen Abonnentenzahlen, die sog. hochrangige wissenschaftliche Zeitschriften erzielen, sprechen dagegen. Parnas schrieb: „I am offended by discussions that imply that the journal is there to serve the authors rather than the readers.“ [40]

   Besondere Bedenken zu diesem Problemkomplex aus der Sicht der US-amerikanischen Schwesterdisziplin Information Systems tragen auf der Grundlage einer sehr interessanten Faktensammlung Gill und Bhattacherjee vor [15]. Ihre Besorgnis mag man wie in dem „Abstieg“ der Abbildung 5 veranschaulichen.

B und C jeweils in ihren ersten und zweiten Gutachten geäußert hatten (Abbildung 6).

Als der Aufsatz schließlich angenommen und in druckreife Form gebracht worden war, hat die Autorengruppe den Gesamtaufwand für das Begutachtungsverfahren (nicht die vielen Personenmonate für die eigentliche wissenschaftliche Arbeit, die Übersetzung ins Englische für eine Schwesterzeitschrift usw.) bilanziert.

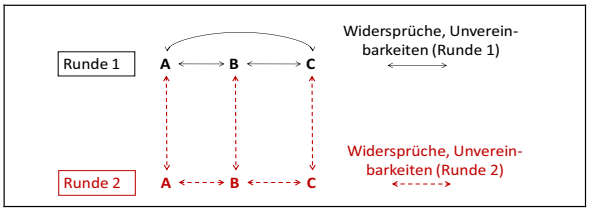


Abbildung 6: Widersprüchliche Gutachten

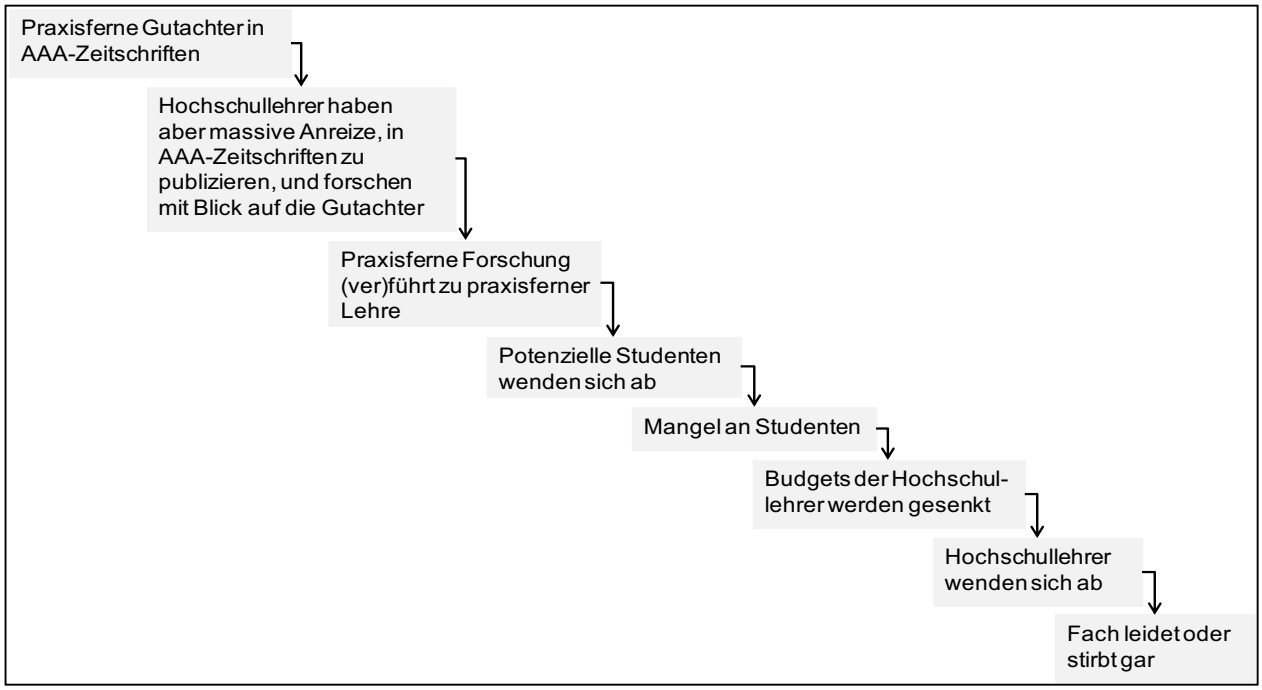


Abbildung 5: „Abstieg“

4. Den Wissenschaftlern entsteht enormer Aufwand bis hin zur Quälerei. Ich war selbst als Autor an einem Aufsatz beteiligt, in den vor allem zwei Nachwuchswissenschaftler außerordentlichen Fleiß und ebensolches Engagement „investiert“ hatten. Der Beitrag wurde vom Hauptherausgeber einer Zeitschrift drei Gutachtern A, B und C vorgelegt, wie es die Richtlinien dieser Zeitschrift verlangen. Wie so oft, erbrachte die Analyse der drei Gutachten Widersprüche und Unvereinbarkeiten zwischen A, B und C. Das Autorenteam hat sich sehr viel Mühe gegeben, in diesem Raum zu navigieren und eine Lösung in Gestalt eines beträchtlich überarbeiteten Textes zu finden. Diese zweite Einreichung wurde dann den gleichen Gutachtern präsentiert. Die Stellungnahmen von A, B und C zeigten erneut Widersprüche untereinander. In einigen Positionen divergierten aber auch die Ansichten, die A,

Es wurden die folgenden Personenstunden verbraucht:

Erste Überarbeitungsrunde:	365
Zweite Überarbeitungsrunde:	130
Dritte Überarbeitungsrunde:	160
Summe:	655 !!!!!!

Das entspricht 11 Wissenschaftlerwochen !!

Abbildung 7: Aufwand für „Navigation“ in Gutachtermeinungen

Was wäre erreicht worden, hätte man diese Ressource im Lehrbetrieb für ein Hauptseminar oder im Praxistransfer für eine Weiterbildungsmaßnahme genutzt?

5. Die vielfältigen Bestrebungen zur engeren Verknüpfung Wissenschaft – Praxis werden konterkariert. Die sog. High-

tech-Strategie der deutschen Bundesregierung sieht aber gerade eine solche engere Verknüpfung vor. Fächer, die wegen falscher Anreize hier nicht „mitspielen“, könnten im Rennen um Ressourcen bald zurückfallen.

6. Der Verzicht auf den Konstruktiven Ansatz in der Forschung beeinträchtigt die universitäre Lehre. Anders ausgedrückt: Das im deutschsprachigen Raum erfolgreich verwirklichte Humboldtsche Bildungsprinzip wird verdrängt. Damit liefert die Hochschule nicht genügend Nachwuchskräfte, die in der deutschsprachigen Wirtschaft und in der öffentlichen Verwaltung IT-Systeme entwickeln können.

7. Ein Wettbewerbsvorteil im deutschsprachigen Universitätswesen besteht darin, dass wir in der Regel über einige Jahre stabile Teams statt stark fluktuierender „Einzelkämpfer“ haben. Dieser könnte verlorengehen, würden wir zu stark US-amerikanisch geprägte Anreize und Organisationsformen übernehmen. Freilich sind auch Fälle bekannt, wo Professoren eines Departments in USA über viele Jahre hinweg freiwillig eine Forschungsgruppe bilden, die noch stabiler ist als eine deutsche „Lehrstuhlbesatzung“.

Eine Reihe weiterer Schwächen im Detail sind in Abbildung 8 aufgeführt (vgl. auch [20], [33, mit zahlreichen vertiefenden Literaturhinweisen], [40], [43]).

1. „Salamitaktik“ (Trend zu kleineren Artikeln) [35]
2. „Ghostwriting Boom“ [54]
3. Verstärkte Eigenzitate und „Zitierkartelle“
4. Aufforderungen an Autoren seitens einflussreicher Professoren, ihre Arbeiten zu zitieren
5. Ablehnung nach erster Durchsicht durch den Hauptherausgeber („Desk-Reject-Verfahren“)
6. Aufforderung durch Zeitschriftenherausgeber, zusätzlich Arbeiten zu zitieren, die in der eigenen Zeitschrift erschienen sind [32]
7. Die „Doppelt-blinde Begutachtung“ bleibt oft eine Fiktion, weil die Autoren die Gutachter mit beachtlicher Wahrscheinlichkeit ahnen können und umgekehrt. Irrtümer bei der Vermutung haben oft negative Folgen.
8. Oft extrem divergierende Gutachtermeinungen: „Peer review is a non-validated charade whose processes generate results little better than does chance.“ ([19], [33, S. 8], [44])
9. Eitelkeiten der Gutachter hinsichtlich Erwähnung eigener Arbeiten
10. Verzögerung und Blockade durch Herausgeber und Gutachter, um eigenen Vorsprung zu erzielen [25]
11. Schwache Korrelation von Gutachtermeinungen mit späteren Zitierungen
12. Abwertung von Monographien
13. Nicht-Berücksichtigung von Zitaten, die älter als zwei Jahre sind, im h-Index [12]
14. Bevorzugung von englischsprachigen Ländern [33, S. 11]
15. Statistische Verfälschungen durch den ersten Buchstaben des Familiennamens der Autoren (A > Z)
16. Zu große Anreize, stark ressourcenverzehrende internationale Konferenzen zu organisieren und zu besuchen, wodurch die Universitätslehrer sehr oft nicht für ihre Studierenden zu sprechen sind
17. Auswahl von Forschungsaktivitäten allein nach der Verfügbarkeit von Daten [32]
18. Vernachlässigung nationaler Forschungsanliegen, da in internationalen Zeitschriften schwer publizierbar. So findet man gegenwärtig z. B. kaum Arbeiten deutscher Wirtschaftsinformatiker zu Sachverhalten, die in Betrieben und in der Öffentlichkeit stark diskutiert werden, wie etwa ELENA oder die Neugestaltung des europäischen Zahlungsverkehrs

Abbildung 8: Schwächen der Beurteilung von Hochschullehrern über „Rankings“

Die vielen Nachteile einer Übergewichtung der Teildisziplin 3 führen in jüngerer Zeit zu skeptischen Äußerungen, für die die folgenden Beispiele aus sehr unterschiedlichen Quellen und vor sehr divergierendem Hintergrund stehen mögen:

1. Das Institut der Deutschen Wirtschaft (IW) hält es für bedenklich, dass in Deutschland die Patentanmeldungen im Verhältnis zu den Forschungs- und Entwicklungsausgaben sinken (in den 10 Jahren zwischen 1997 und 2007 sind die F&E-Ausgaben nach Schätzungen des IW um 60 Prozent gestiegen) [38]. Eine Teil-Interpretation könnte dahin gehen, dass die Anreize zum Patentieren im Vergleich zum frühen Publizieren zu gering geworden sind.

2. Es gibt Signale, dass in der Volkswirtschaftslehre gegenwärtig eine Trendwende von der Übergewichtung des Publizierens weg und zur Befassung mit wirtschaftspolitischen Tagesproblemen hin eingeleitet wird. Symptome sind die Kritik an der Ausrichtung des Deutschen Instituts für Wirtschaftsforschung weg von der Lösung von aktuellen wirtschaftspolitischen Problemen und der einschlägigen Politikberatung hin zum Publizieren [5] und Ergebnisse einer Umfrage, die die Financial Times Deutschland (FTD) mit dem Verein für Socialpolitik durchgeführt hat. Die FTD fasst ein Resultat so zusammen [14]: „Knapp 55 % der deutschen Ökonomen geben 2010 an, dass sie es für sehr wichtig finden, gute Kenntnisse der aktuellen Wirtschaftslage zu haben – das klingt für Außenstehende banal, für manchen Akademiker ist es jedoch eine kleine Revolution. Vor vier Jahren war es noch eine Minderheit, die so dachte. Das Kriterium ist jetzt sogar wichtiger als das Publizieren in (zumeist theoretischen) Fachzeitschriften, was jahrelang zum Fetisch erklärt wurde.“

3. Die große Anzahl der Schwachstellen, verbunden mit dem „Spiel gegen das System“, hat in jüngerer Zeit auch für Diskussionen in der DFG gesorgt, wie Äußerungen ihres Präsidenten Matthias Kleinert zeigen [35]. Es wird nach Möglichkeiten gesucht, den Akzent von der bibliometrisch festgestellten Quantität zu Qualitätskriterien zu verschieben. Offen bleibt, ob veränderte Regeln der DFG auf andere Institutionen, die Urteile über Universitätslehrer abgeben, wie z. B. Wissenschaftsministerien, ausstrahlen würden.
   Es dürfte nicht leicht sein, die Schwächen des „Peer-Review-Verfahrens“ über kleine Abhilfemaßnahmen hinaus (Beispiel: Gutachter dürfen nicht nur unterdurchschnittlichen Literaturbezug monieren, sondern müssen spezielle Hinweise auf von ihnen vermisste Arbeiten geben.) auszumerzen. Dies gilt mutatis mutandis für andere Kriterien. So können z. B. Dozenten bei der Lehrevaluation oft Rangplätze gewinnen, indem sie der Noteninflation Vorschub leisten.

5. Vorteile und Probleme der Internationalisierung

Ohne Frage hat eine gute deutsche Hochschule die Ziele, solide Kenntnisse der englischen Fachsprache zu vermitteln und ihren Studierenden die Chance zu bieten, Wissen im Ausland aufzusaugen, sei es durch Studium an Universitäten oder Praktika in Betrieben.

Internationalisierung der Wissenschaft gilt derzeit als ohne Wenn und Aber erstrebenswert. Wendungen wie „international tätig“,

„international publizierend“ oder „international renommiert“ haben als Leistungsindikatoren ein hohes Gewicht. (Besonders pointiert vertritt Wolfgang König diese Position, wenn er – vor allem von Nachwuchswissenschaftlern – fordert, als Erstes in einer oder mehreren internationalen Zeitschriften zu publizieren und erst dann, „wenn diese Pflicht erfüllt ist“, weitere Ziele ins Blickfeld zu nehmen, wie etwa den Wissenstransfer in die Praxis oder die Präsentation in deutschsprachigen Medien [21].) Es könnte sich in der Sprache der Entscheidungstheorie um „Routine- oder programmiertes Verhalten“ von Personen oder Organisationen ([18, S. 233-235], [49]) handeln, denn hier stoßen wir auf inkompatible Sachverhalte (Abbildung 9).

Um den Betrachtungsbereich auf die WI zu fokussieren, wollen wir die folgenden wissenschaftlichen Ergebnisse aussparen, wobei wir uns an das Patentwesen anlehnen:

1. Anweisungen an den menschlichen Geist ohne erkennbare Chance der baldigen kommerziellen Verwertung. Beispiel: OR-Algorithmen unterhalb der praxisrelevanten Größenordnungen (Reihenfolgeplanung bei vier Aufträgen auf drei Maschinen).
2. Wissenschaftliche Ergebnisse, die einen mit wenig Problemen und Risiken verbundenen sicheren Patentschutz erlauben. Beispiel: Eine neue RFID-Technik, die Störungen der Auslesung von Funketiketten durch Gegenstände, die zwischen dem Objekt und dem Leser liegen, vermeidet.
3. Abgegrenzter Wissensaustausch mit Ländern, durch den eine Äquivalenz oder gar ein eigener „Überschuss“ beim Geben und Nehmen erwartet werden darf. (Ob solche Relationen zwischen der US-Disziplin IS und der deutschsprachigen WI unterstellt werden kann, ist umstritten (vgl. dazu [15], [29] und [34]).) Jedoch ist darauf zu achten, dass vom bilateralen Austausch nicht Dritte profitieren, die in die Forschungsanstrengungen nicht investiert haben. Als Teil dieser Problematik ist die „künstliche“ Beteiligung am Wissensaustausch durch „Trivial-Arbeiten“ und Plagiate zu sehen.
4. Wissenschaftliche Ergebnisse, die im Prinzip ökonomisch verwertbar wären, bei denen dies aber aus humanitären Gründen nicht dem Markt überlassen werden soll. Diese Position ist in der WI wenig relevant, sie gilt vor allem für die Medizin. (Hier wären in der Zukunft Lösungen denkbar, bei denen die WHO für Wissenschaftler und Produktentwickler genügend finanzielle Anreize bietet, aber andererseits die Versorgung einschließlich der Logistik auch ärmerer Mitglieder der Weltbevölkerung mit planwirtschaftlichen Methoden sicherstellt.)

Abbildung 9: Inkompatible Sachverhalte

Nach Aussparen dieser vier Komplexe beschränken wir uns in der Folge auf wirtschaftlich verwertbare Ergebnisse der deutschsprachigen WI.

Legen wir auf die eine Seite der Argumentenbilanz:

1. Die Binsenweisheit, dass wir keine Rohstoffe, sondern nur Wissen und Können, meist geronnen in unseren Produkten, zu exportieren haben.
2. Die zunehmend enge Verzahnung von Forschung und Entwicklung, die z. B. 2010 durch eine Umorganisation auf der Führungsebene der SAP AG sichtbar wurde, als man die Forschungsabteilung dem Technologie-Vorstand zuordnete, um alle Technologie-Aktivitäten zu bündeln [39].
3. Die sogenannte Targeting-Strategie des stärksten Gegners im Ringen um die Weltmärkte, der Volksrepublik China. Sie will auf ca. 40 Feldern die führende Position erreichen [4] und benutzt dafür auch problematische Mittel (vgl. z. B. [9]).

   Eine ähnliche Strategie verwendet jetzt auch Russland, z. B. auf dem Gebiet der Verschlüsselung auf Hardware- und Software-Basis ([3], [23]). Ein neueres Beispiel für die Aggressivität Russlands ist das Bestreben, sich gegen Widerstände des Unternehmens und der Bundesregierung mit über 25% an der Infineon AG zu beteiligen, unter anderem um Wissen zu Pass- und Verschlüsselungstechniken zu übernehmen [27].

   Bei der Präsentation des Verfassungsschutzberichts für das Jahr 2009 hoben sowohl der deutsche Verfassungsschutzpräsident als auch der Bundesinnenminister hervor, wie sehr China und auch Russland versuchen, die deutsche Forschung auszuspähen. Hierfür werden auch Austauschprofessoren, Stipendiaten, Praktikanten und Kundschafter mit Touristenvisum eingesetzt, ebenso Cyberspionage ([52], [36]). Offenbar gründet man auch privatwirtschaftliche Unternehmen, die Identitätsdiebstahl anbieten ([13], [22]). Es ist mit einer beträchtlichen Dunkelziffer zu rechnen, weil betroffene westliche Unternehmen und Forschungsinstitute befürchten, sie könnten Nachteile erleiden, wenn sie derartige Erfahrungen preisgeben [45]. Deutsche Industriebetriebe, z. B. aus der Automobilbranche, bauen sogenannte Firewalls in die Kommunikationswege mit ihren eigenen fernöstlichen Niederlassungen, wozu sie teilweise auch durch die deutschen Exportkontroll-Vorschriften gezwungen werden. Die Fraunhofer-Gesellschaft muss darauf bedacht sein, dass sie nicht im Rahmen von Kooperationen mit der Privatwirtschaft deren Forschungsergebnisse mit eigenen verbindet und im Rahmen der Wissenschaftskommunikation in Konkurrenzländer transferiert, in welchen die wirtschaftliche Nutzung „außer Kontrolle gerät“. Dies wird bei Entscheidungen zu Auslandsniederlassungen stets berücksichtigt [8].
4. Die beeindruckende Analyse von Paul Samuelson, wonach die Ricardo-Mill-Theoreme zur „Win-win-Situation“ in der Globalisierung dann nicht herangezogen werden dürfen, wenn es einer Nation gelingt, den Wissensvorsprung einer anderen besonders *schnell* aufzuholen [46].

5. Die vor allem in Deutschland faktisch hohen Hürden, wenn man das Ergebnis wissenschaftlicher Forschung schützen möchte. Das gilt unter anderem für Softwaresysteme.

6. Die enormen Anstrengungen von Schwellenländern wie Indien, nicht nur simple Programmierarbeiten, sondern hochwertige Systementwicklung aus Industrieländern abzuwerben, das berühmte „Climbing up the value pyramid".

Abbildung 10 zeigt einen Pfad, wie er in etwa von der Volksrepublik China zur Entwicklung eines dritten großen Herstellers von Verkehrsflugzeugen neben Airbus und Boeing angestrebt wird (vgl. auch [1], [2], [42]).

Aufsehen erregte die US-Unternehmung Applied Materials, die Betriebsmittel zur Herstellung von Halbleitern, Solarpanels und Flachbildschirmen fertigt. Sie verlegt ein Forschungslabor von Santa Clara (Kalifornien) nach China [6]. Ähnlich ist die Umleitung von Forschungsgeldern der Continental AG zur Entwicklung von Software für Kfz-Elektronik, die früher zum Teil der Universität Hannover zugute kamen, hin zu Hochschulen in der Nähe des rumänischen Produktionsstandorts zu sehen.

Stufe	Deutschland	Schwellenländer
1	Produktion ->	
2	Entwicklung für Sch-Produkte ->	
3	Forschung für Sch-Produkte ->	
4	Entwicklung für H-Produkte ->	
5	Forschung für H-Produkte ->	
6	-> Forschung, Entwicklung, Produktion für H-Produkte*)	

Bemerkungen/Legende:
*) Dieses Stadium ist auf diversen Feldern schon erreicht. Bspw. erforscht die Bayer AG die Technologie hochwertiger Kunststofffolien für Mobiltelefone in Singapur [17].
Sch-Produkte: Produkte zum Einsatz in Schwellenländern
H-Produkte: Produkte zum Einsatz im Heimatland

Abbildung 10: Pfad zu höherer Wertschöpfung

Und legen wir auf die andere Seite unserer Argumentationsbilanz:

Den Anreiz, den wir setzen, damit ein interessantes Ergebnis der wissenschaftlichen Forschung, etwa zu Systemen, mit denen komplizierte internationale Liefernetze besser beherrscht werden können, möglichst rasch in einem internationalen Journal veröffentlicht wird, wo es die Konkurrenz ebenso rasch abholt. Haben wir dann im Interesse des deutschen Steuerbürgers gehandelt, der uns Wissenschaftler finanziert?

Zusammengefasst besteht aus der Sicht der deutschen Wissenschaft ebenso wie der Industriepolitik ein augenfälliger Widerspruch: Einerseits sollen im Rahmen der sogenannten High-tech-Strategie mit großem Abstimmungsaufwand Technologiefelder bzw. neue Bedarfsfelder definiert werden, auf denen Industrie bzw. Privatwirtschaft, Forschungsinstitute und Hochschulen kooperieren, „damit sie nicht ihre führenden Positionen in wenigen Jahren an Koreaner, Chinesen, Taiwanesen, Japaner oder Amerikaner abgeben" [37]. Andererseits werden die Wissenschaftler, z. B. im Kontext der Exzellenzinitiative, stark motiviert, ihre Resultate umgehend international öffentlich zu machen.

Aus den Forschungen zur Branchen-Software weiß man, dass Unternehmen ursprünglich für sich selbst entwickelte leistungsfähige, branchenspezifische IT-Anwendungen an unmittelbare Konkurrenten verkauft haben (z. B. Transworld Airlines sein Vielfliegersystem an Canadian), um einen Deckungsbeitrag zu den Entwicklungskosten zu erzielen [28]. Diese Betriebe sehen einen Vorsprung von einem halben Jahr als ausreichend an, bevor man sich eines Wettbewerbsvorteils begibt.

Die Konsequenz aus derartigen Überlegungen könnte sein, sich an traditionelle Publikationsgewohnheiten im Maschinenbau anzulehnen (Abbildung 11), d. h. z. B. über knappe Publikationen („Appetithappen") zunächst den Kontakt zu einheimischen Betrieben zu suchen, die das Forschungsresultat in die Praxis transferieren und gleichzeitig mit finanziellen und personellen Ressourcen sowie Wissen beziehungsweise Erfahrung in Gemeinschaftsprojekten mit Hochschulen die weitere Ausreifung vorantreiben. Zu präferieren sind die Ausgliederung von Unternehmen, die Weiterentwicklung der im wissenschaftlichen Raum erarbeiteten Prototypen in Transferprojekten gemeinsam mit Unternehmen oder der „Transfer über die Köpfe".

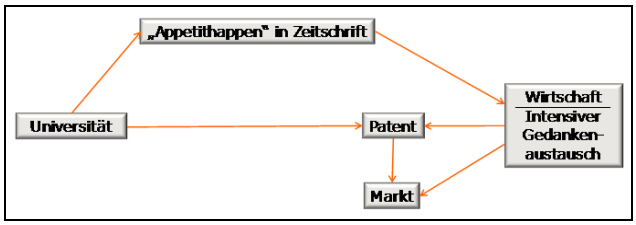


Abbildung 11: Verwertung von Forschungsergebnissen im Maschinenbau (traditionell)

6. Fazit

In der „multivariaten Zielfunktion" beziehungsweise beim „Zehnkampf" des Universitätslehrers der WI erscheint eine gleichmäßigere Gewichtung der Disziplinen geboten. Das Übergewicht von Publikationen bestimmten Stils (D3) in Verbindung mit der Internationalität ist zugunsten von Aktivitäten zu reduzieren, die dem Finanzier der Universitäten im deutschsprachigen Raum („Stakeholder") dienen, vor allem dem Transfer in die Praxis, auch durch Gründung von Unternehmen (D4, D6). In praxi könnte eine Instanz der deutschsprachigen WI, z. B. in Anlehnung an die Abbildungen 1 und 3, eine Standard-Prüfliste verabschieden, die Fachbereiche oder Berufungsausschüsse situations- und rollenabhängig modifizieren. Zusätzliche Fairness käme „ins Spiel", wenn der sparsame Umgang mit öffentlichen Gütern in Lehre und Forschung durch den Ausweis einer Art Input-Output-Relation honoriert würde. Die von der EU verlangte Einführung von Vollkostenrechnungen in der Universität wird die Quantifizierung erleichtern.

Die Politik der deutschen Bundesregierung, gleichzeitig, aber weitgehend unabhängig voneinander, das Publizieren in internationalen Zeitschriften im Rahmen der Exzellenzinitiative und den Wissenstransfer in die Praxis im Zuge von Technologieprogrammen sowie künftig die Exzellenz in der Lehre zu fördern, ist problematisch. Möglicherweise wäre die ganzheitliche Beurteilung des Universitätslehrers über alle Disziplinen des „Zehnkampfes" hinweg zweckmäßiger und wohl auch mit weniger öffentlichen Ausgaben verbunden.

Nicht zu unterschätzen bei all dem ist auch, dass es bei einer Überforderung des wissenschaftlichen Nachwuchses in marktnahen Fächern wie der WI bald dazu kommen könnte, dass sich zu

wenige unserer besten Absolventinnen und Absolventen für die Universitätslaufbahn entscheiden.

Es kommt hinzu, dass sich das Verhältnis zwischen Anforderungen, Risiken und Vergütung eher ungünstig entwickelt. Die anderen Aufgaben, Kriterien (Abbildung 1) und Voraussetzungen (Abbildung 3) sind geblieben, aber D3 fordert enorme Energien. Dabei hat der Universitätslehrerberuf in den vergangenen ca. 15 Jahren sehr an finanzieller Attraktivität verloren. Aus einer Studie des Statista-Instituts, der wiederum Zahlen des Statistischen Bundesamtes zugrunde liegen, kann man schließen, dass 2008 das durchschnittliche monatliche Nettogehalt eines Hochschullehrers nur etwa 10% über dem eines Gymnasiallehrers lag, dies bei einem wesentlich höheren Laufbahnrisiko der Nachwuchswissenschaftler, von weiteren Kriterien ganz abgesehen [51]. Trotz allem ist Resignation nicht am Platze. Auch andere Monopol- oder Oligopol-Anmaßungen, wie z. B. die, die DFG-Förderung nur auf angebliche „Hard-Science-Disziplinen" oder die Forschungssubventionen des Deutschen Bundesministeriums für Forschung und Technologie auf die Industrie zu konzentrieren und Dienstleistungsbetriebe nicht zu bedenken, blieben letztlich Episoden.

7. LITERATUR

[1] Arndt, M. und Einhorn, B. 2010. The 50 Most Innovative Companies. *BusinessWeek* vom 15.04.2010.

[2] Aspray, W., Mayadas, F. und Vardi, M.Y. (Hrsg.) 2006. *Globalization and Offshoring of Software.* A Report of the ACM Job Migration Task Force. O.O.

[3] Berke, J. 2010. Codes für Moskau. *Wirtschaftswoche* Nr. 27/2010, 56-57.

[4] Berke, J., Kamp, M., Kiani-Kress, R. und Seiwert, M. 2010. Im Rachen des Drachen. *Wirtschaftswoche* Nr. 12/2010, 40-47.

[5] Beyerle, H. und Fricke, T. 2010. Ökonomen sehen DIW auf Abwegen. *Financial Times Deutschland* vom 15.03.2010.

[6] Bradshier, K. 2010. China Drawing High-Tech Research From U.S. *The New York Times* vom 17.03.2010.

[7] Brooks, D. 2010. The Tel Aviv Cluster. *The New York Times* vom 11.01.2010.

[8] Persönliche Auskunft von H.-J. Bullinger, Präsident der Fraunhofer-Gesellschaft, 2010.

[9] Erd, R. und Rebstock, M. 2010. Produkt- und Markenpiraterie in China. Aachen.

[10] Finsterbusch, S. 2010. Tonmeister der digitalen Revolution, *FAZ* vom 06./07.03.2010, C3.

[11] Fischer, G., Rohde, M. und Wulf, V. 2007. Community-Based Learning. The Core Competency of Residential, research-based University. *International Journal on Computer Supported Collaborative Learning* 2 (2007) 1, 9-40.

[12] Flöhl, R. 2010. Die Bestseller in deutschen Kliniken. *FAZ* vom 21.04.2010, N1.

[13] Follath, E. 2010. Die Umarmung des Drachen. *SPIEGEL* Nr. 30/2010, 92-93.

[14] Fricke, T. 2010. Thomas Fricke – Bei Wirtschaftsexperten tut sich was, *Financial Times Deutschland* vom 25.06.2010.

[15] Gill, G. und Bhattacherjee, A. 2009. Whom Are We Informing? Issues and Recommendations for MIS Research from an Informing Sciences Perspective. *MIS Quarterly* 33 (2009) 2, 217-235.

[16] Hartmann, J. und Hildebrand, J. 2009. Radikaler Strategiebruch beim Softwareriesen SAP. *WELT ONLINE* vom 20. Juni 2009.

[17] Hein, C. 2010. Die dritte Welle. *FAZ* vom 17.07.2010.

[18] Heinen, E. 1971. Grundlagen betriebswirtschaftlicher Entscheidungen. Wiesbaden.

[19] Horrobin, D. 2001. Something Rotten at the Core of Science? *Trends in Pharmacological Sciences*, Vol. 22, No. 2, Februar 2001.

[20] Kieser, A. 2010. Die Tonnenideologie der Forschung. *FAZ* vom 09.06.2010, N5.

[21] König, W. 2010. Englische Spitzenveröffentlichungen als Pflicht und deutschsprachige Medien als Kür. *WIRTSCHAFTSINFORMATIK* 52 (2010) 4, 244-246, hier 246.

[22] Kreimeier, N. und Diethelm, V. 2010. Horch und Guck aus Moskau. *Financial Times Deutschland* vom 30.06.2010.

[23] Kuhn, T. 2010. Ewiger Kreislauf. *Wirtschaftswoche* Nr. 25/2010, 72.

[24] Leibfried, S. (Hrsg.) 2010. *Die Exzellenzinitiative. Zwischenbilanz und Perspektiven.* Frankfurt/New York.

[25] Lubbadeh, J. 2010. Stammzellforscher werfen Kollegen Fachartikel-Blockade vor. *SPIEGEL ONLINE* vom 03.02. 2010.

[26] Mahbubani, K. 2010. 5 Lessons America Can Learn From Asia About Higher Education. *The Chronicle* vom 07.03.2010.

[27] Maier, A., Clausen, S., und Klusmann, S. 2010. Russen buhlen bei Merkel um Infineon. *Financial Times Deutschland* vom 28.06.2010.

[28] Mertens, P. 1995. Application Templates: Faster, Better, and Cheaper Systems (Besprechungsaufsatz). *WIRTSCHAFTSINFORMATIK* 37 (1995) 2, 189-191.

[29] Mertens, P. 2005. Gefahren für die Wirtschaftsinformatik – Risikoanalyse eines Faches. In *Wirtschaftsinformatik 2005*. Ferstl, O.K., Sinz, E.J., Eckert, S. und Isselhorst, T. (Hrsg.), Heidelberg, 1733-1754.

[30] Metcalf, C. 2010. Hochschulutilitarismus nach englischem Muster. *FAZ* vom 04.02.2010.

[31] Narielvala, Y. und Beiten, M. 2010. Clever andocken. *Detecon Management Report* Nr. 1/2010, 8-13.

[32] Osterloh, M. 2010. Unternehmen Universität? *Neue Zürcher Zeitung* vom 19.05.2010.

[33] Osterloh, M. und Frey, B. S. 2010. *Academic Rankings and Research Governance*. unveröffentlichtes Manuskript, Zürich.

[34] Österle, H. et al. 2010. Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigour und Relevanz.* Österle, H., Winter, R. und Brenner, W. (Hrsg.), Nürnberg.

[35] O. V. 2010. Schluss mit der Salamitaktik. http://www.zeit.de/wissen/2010-02/dfg-publikationen-forschung, Abruf am 08.12.2010.

[36] O. V. 2010. Deutschland durch Spionage zunehmend bedroht. *FAZ* vom 22.06.2010.

[37] O. V. 2010. Ein Lichtstreif für die deutsche Industrie. *FAZ* vom 22.03.2010.

[38] O. V. 2010. China und USA hängen deutsche Tüftler ab. *SPIEGEL ONLINE* vom 27.03.2010.

[39] O. V. 2010. SAP tauscht Forschungschef aus. *FAZ* vom 07.08.2010.

[40] Parnas, D.L. 2007. Stop the Numbers Game. *Communications of the ACM* 50 (2007) 11, 19-21, hier 19.

[41] Plickert, P. 2009. An der Forschungsfront. *FAZ* vom 21.12.2009.

[42] Prabhu, J. und Williamson, P. 2010. Debate between Jaideep Prabhu and Peter Williamson. http://www.alumni.cam.ac.uk/news/cam/cam59/debate59/, Abruf am 08.12.2010.

[43] Ren, J., und Taylor, R. 2007. Automatic and versatile publications ranking for research institutions and scholars. *Communications of the ACM* 50 (2007) 6, 81-85.

[44] Rothwell, P. M. und Martyn, C. N. 2000. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain*, 123, 1964-1969.

[45] Ruch, M. 2010. Klar und deutlich. *Financial Times Deutschland* vom 05.07.2010.

[46] Samuelson, P. A. 2004. Where Ricardo and Mill Rebut and Confirm Arguments of Mainstream Economists Supporting Globalization. *Journal of Economic Perspectives* 18 (2004) 3, 135-146.

[47] Seiser, M. 2010. Begehrte Biotechnologie aus Österreich. *FAZ* vom 27.07.2010.

[48] Senor, D. und Singer, S. 2009. Start-up Nation: The Story of Israel's Economic Miracle. New York.

[49] Simon, H. A. 1959. *Administrative Behavior*. 2. Aufl., New York.

[50] Speckbacher, G., Wentges, P. und Bischof, J. 2008. Führung nicht-erwerbswirtschaftlicher Organisationen: Ökonomische Überlegungen und Folgerungen für das Hochschulmanagement. *Betriebswirtschaftliche Forschung und Praxis* 60 (2008) 1, 43-64.

[51] Statista-Institut 2008. *Gehaltsreport: Monatliches Bruttogehalt ausgewählter Berufsgruppen in Deutschland im Jahr 2008 (Platz 1-30)*. http://de.statista.com/statistik/daten/studie/74517/umfrage/monatliches-gehalt-nach-berufsgruppen-in-2008-%28platz-1-30%29/#stat, Abruf am 08.12.2010.

[52] Tillmann, S. 2010. Spionage wird zum Megaproblem. *Financial Times Deutschland* vom 22.06.2010.

[53] Wadhwa, V. 2010. Why Andy Grove Is Wrong About Job Growth - Reigniting job creation in the U.S: won´t happen by becoming more like China and India. *Businessweek* vom 09.06.2010. http://www.businessweek.com/technology/content/jul2010/tc2010079_953836.htm, Abruf am 08.12.2010.

[54] Wong, G. 2010. Rampant cheating hurts China's research ambitions. *The Boston Globe* vom 11.04.2010. http://www.boston.com/news/world/asia/articles/2010/04/11/rampant_cheating_hurts_chinas_research_ambitions/, Abruf am 08.12.2010.

The Artifact's Theory – A Grounded Theory Perspective on Design Science Research

Benjamin Müller
University of Mannheim
Chair of Information Systems IV (ERIS)
68131 Mannheim, Germany
+49 (621) 181-3607
mueller@eris.uni-mannheim.de

Sebastian Olbrich
University of Duisburg-Essen
Mercator School of Management
47057 Duisburg, Germany
+49 (171) 8025-001
sebastian.olbrich@arcor.de

ABSTRACT
The need for theoretical work in IS research has been pointed out frequently in past discussions on the cumulative tradition of our discipline. In order to distinct IS research from adjacent fields, the practice of designing artifacts plays a crucial role. On the one hand, the construction and evaluation of new and innovative artifacts solving real world problems is the core of our discipline. On the other hand, this designing aspect of IS research has been accused of not being linked sufficiently to theoretical considerations in the past. To help IS design researchers to address this challenge, our paper analyzes research generating theory based on the Grounded Theory Method. By identifying, analyzing, and reviewing such research we identify some process characteristics of grounded theory building that can help IS researchers to improve their design processes to also yield theoretical output. Thus, we aim to make an IS design researcher's work clearly distinguishable to the work of a software developer or a consultant. By also discussing the role of theory in this context, we intend to make a case for more theoretical work in IS design research.

Keywords
IS theory, grounded theory, design science, theory building.

1. INTRODUCTION AND MOTIVATION
The role of theory in scientific discovery is paramount. Providing a concise account of the "what, how, and why" of phenomena [119], theories offer the basis for the description, explanation, and the prediction of these. Moreover, theories can be used to facilitate the accumulation of knowledge over the course of scientific discovery. Such a cumulative tradition helps disciplines to advance their understanding of the subjects under investigation and to make that knowledge applicable in practice [66, 106]. Examples for the centrality of theories can be found in both the natural sciences (e.g., physics) and the social sciences (e.g., sociology and psychology) [1]. Beyond the documentation and accumulation of knowledge, theory is also a vehicle that can inform and guide the discovery and creation of new knowledge.

While natural sciences aim to generate universal laws and statements [85], generating theory in the information systems (IS) field is a particular challenge in light of permanently new and innovative technologies and a persistently changing attitude of users towards the application of that technology [49, 51]. In its young history, the IS discipline has earned harsh critique as some have described the investigation of information systems as a "mishmash of fuzzy thinking and incomprehensible jargon" [19, p. 90]. Put in less provocative terms, only few IS-specific theories have emerged [9]. Quite contrary, many of the phenomena observed in IS research have been explained using theories from neighboring disciplines such as sociology or psychology on the behavioral side and computer science or engineering on the technical side [3, 45, 91]. While this can be interpreted as a sign for the relevance of IS research by integrating knowledge at the intersection of its adjacent disciplines, many scholars – IS and non-IS – have complained about the lack of an IS-specific cumulative tradition of theory and emphasize the importance of generating IS-specific theories [e.g., 60, 116, 123]. Reasons for the importance of such powerful, general IS theories are the need for domain identity [6, 7], the legitimacy of the IS field in comparison to its neighboring disciplines [33, 69], and the "race for credibility" in the scientific discourse [115].

In this paper we take up the stance that IS has developed rapidly over the last decades and that IS research, as a discipline at the intersection of social, technological, and design sciences, has developed its own unique understanding of theory. This seems to be true more than ever when taking the discipline's recent turn towards the science of the artificial into account [95], that is, the paradigm of design science research (DSR). While the idea of DSR does already have a long-standing tradition in IS [e.g., 36, 70, 77, 114], this orientation of research on IS artifacts has experienced a huge surge in popularity following the seminal article by Hevner et al. [50]. While the approach has been refined in intense discussions [e.g., 2, 37, 54, 83, 120], researchers conducting DSR still face the challenge to show their theoretical contribution [89]; a particular challenge when trying to publish in some of the premier international journals of our discipline. Some describe DSR's missing link to theory as the largest issue of this research approach [83], probably because a designed artifact is mostly an individual or local solution of a specific problem [48, 79].

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

One recent approach to address this challenge has been introduced by Gregor [47]. She suggests that design could and should be understood as a science of the artificial and continues to explain what this implies for DSR and its theoretical contributions. The discussion tries to aggregate some of the disciplines considerations with respect to design theories at the theoretical heart of each artifact [46, 58, 86]. While these approaches aim at explaining the role of theory in designing, the question of how to explicitly integrate theoretical work into design endeavors remains largely under-investigated. One promising approach is using the grounded theory method (GTM), originally suggested by Glaser and Strauss [39]. For instance, Goldkuhl [43] suggests that design theories should be inherently grounded in the context of artifact development. Summarizing earlier discussions, Gregory [48] compared the paradigms of DSR and GTM. While he points to some important differences, he suggests that DSR and GTM do, indeed, have similarities that make the latter a promising candidate to overcome some of the theoretical shortcoming of the former and offer a basis for a comparison and integration of the two.

Building on the current discussion on theories in IS research [e.g., 45, 60, 116], our paper aims at investigating how theories can originate in IS research, specifically in the DSR context. To do so, we look at GTM as a potential source of IS-specific theories in a design project. As theories originating from a GTM-approach in an IS setting are likely not influenced by a priori use of theoretical considerations from outside the discipline, they provide a good opportunity to observe IS-specific generation of theory. In the long run, we believe that our research contributes to a common understanding of the role IS-specific theory plays in the investigation of the phenomena we study as a discipline; in short, what role *can, should* or *must* theory play in IS design research?

IS design researchers – especially young ones looking for tenure – seem to be in a publication challenge, that is, they need top-ranked publications to get tenure while top ranked journals still do not seem to treat design based research as equally valuable. As mentioned above, we will not address this topic in detail. However, in order to increase the chances of getting design based research published in leading international journals, we will look at GTM-based research that has been published in some of the top-ranked international journals of the IS discipline. We intend to use the analysis of this particular sample to understand some of the specificities of grounded research – both with respect to the research process as well as with respect to developing its theoretical contribution – that could help to make DSR a more successful contribution to the theoretical discourse of the IS community.

To approach this question, our paper is structured as follows. In the following two sections we have a brief look at theory in IS research in general and grounded theories in particular. Section four outlines our research design and discusses the assumptions of our literature review. Section five summarizes the results and provides a first interpretation. Based on these results, sections six and seven introduce our conclusions with respect to GTM's possible implication for DSR. The paper closes with an outlook on future research opportunities and a discussion of our current research's implications and contributions.

2. THE CONSTITUENTS OF THEORY

A first step towards understanding theory and its role in DSR is looking at the constituents of IS theory itself. Some authors have identified this as a rather challenging task since defining "theory" per se is not trivial [9, 34, 98, 116]. Looking at the literature from our adjacent fields, much emphasis is placed on concepts and their relations. In their seminal article on theory, Sutton and Staw [98] highlight the need to identify concepts and causal arguments in order to produce strong theory. In their review of various approaches to theoretical understanding in the IS field, Burton-Jones et al. [9] also highlight these constituents. Gregor [45] offers a broad discussion of theory in IS and identifies a set of constituents common to all theories. She concurs with the identification of constructs and relationships and adds the means of representation and the scope of a theory. Kaplan [56] points to theory's ability to explain why certain (empirical) patterns occur when a phenomenon is observed. Whetten [119] adds to this by suggesting that theoretical contributions go beyond the mere description of the what (i.e., concepts or constructs), but will specifically discuss the how (i.e., conceptual arguments) and why (i.e., causal arguments). This is also supported by Sutton and Staw [98] who emphasize that strong theories discuss the nature of causal relationships and identify their timing. Glaser and Straus [39] underline the importance of explaining why certain findings are observed. They describe the purpose of theory as to enable prediction and explanation of behavior and to support a discipline's theoretical advance.

Beyond constructs and relationships – closely related to the property of enabling predictions – the literature highlights the attribute of being falsifiable or testable as an important characteristic of theories [23]. Looking at the components of theory, Gregor [45] suggests that this property, however, is contingent on a theory's purpose. By looking at the interrelationships among theory types, she suggests that this characteristic seems to be more important for theories that go beyond the analysis and explanation of phenomena. These are two of the five types of theory that Gregor [45] introduces in her seminal work on the nature of theory in information systems research. Beyond these, called type 1 and 2 respectively, she describes theories of prediction (type 3), explanation and prediction (type 4), and analysis and design (type 5).

Across all theory types, Glaser and Straus [39] require a theory to be usable in solving practical problems. This indicates a link between the domain in which theory is deployed and the domain of the empirical and theoretical analysis underlying it. This is a property closely related to some of the basic characteristics of DSR [48, 120]. Artifacts are seeking solutions to real world problems which they are then directly related to [58]. Thus a closer look at the research approaches of GTM might help to identify some practices that could also be used to address the alleged theoretical shortcomings of DSR.

3. DEVELOPING GROUNDED THEORIES

In order to understand IS theories and how they are built, we want to look at theories that emerge from IS-specific research as a source of inspiration. Urquhart et al. [105] suggest that GTM is a good source for theories originating in the IS field. As one of the most renown techniques to generate theory, GTM was introduced through the work of Glaser and Strauss [39]. Later, GTM developed in two schools of thought [72]: an approach suggested by Strauss and Corbin [96] which focuses on the process of building grounded theories and a conceptualization advocated by Glaser [41] which is focused more on the underlying theory itself.

In the origins of GTM, Glaser and Strauss defined it as "the discovery of theory from data – systematically obtained and analyzed in social research" [39, p. 1]. The systematic approach to theory

development they suggested is characterized by a deep immersion in the empirical data on a given phenomenon of interest. The researcher then engages in an iterative process of discovery and formalization in which theoretical sampling leads to the addition of more and more "slices" of empirical data. This process continues until the theoretical categories under investigation and their properties are theoretically saturated, that is, additional empirical data does not lead to a change or extension in the new theory. The results are theories grounded in empirical observations that explain the relationships between the theoretical categories constituting a phenomenon. In a recent review of GTM literature, Urquhart et al. [105] identify four general characteristics of GTM: (1) focus on building theory, (2) no pre-formulated hypotheses, (3) joint data collection and constant comparison, and (4) theoretical sampling producing "slices of data."

Applications of GTM to build grounded theory (GT) are generally found in the context of qualitative research [76]. However, GTM is not just a description of how to code data but an approach to build theories. As such, GT is paradigmatically neutral [42] and, as a method, not bound to any prevailing ontological or epistemological position [105]. As an approach to investigate IS-related phenomena and to build respective theories, Scott [92] finds that GTM has reported strengths that qualify it to be employed in the process of scientific discovery in our discipline. For some time, IS scholars have engaged in a discourse on the right usage of GTM in IS research and how to maximize its potential to build GTs [e.g., 8, 103, 105]. They develop guidelines on working with GTM in the context of, for example, case studies [11] or action research [20]. While there are first steps towards an analysis of GTM in the context of design research [48, 89], a detailed analysis of what can be learned from GTM for DSR is missing as yet.

Taking this brief profile of GTM research into account, it seems like an interesting candidate to address our research question. In looking at GTM, we hope to identify practices and examples that illustrate GTMs potential to help with theory building in DSR, thus increasing DSR's theoretical contribution. We regard GTM to be a source of inspiration that provides an analogy for how the development of theories in a DSR-context can look like. Relevant work should therefore go beyond the mere use of GTM as an instrument for coding data [8] and should provide some form of a theoretical account [as, for example, described by the theory types identified in 45]. We thus design our research to identify papers that reflect these properties and do produce IS-specific GTs in order to look at what we can transfer to DSR research.

4. RESEARCH DESIGN

To identify such papers, we chose to review relevant publications from the IS discipline. In structuring our review, we rely on established guidelines for reviewing and synthesizing literature [12, 32, 117]. Literature reviews have been identified as a well suited approach to provide an overview of current work on a given concept in a series of disciplines [21, 75]. Especially their ability to aggregate and facilitate current knowledge as a basis for building new insights has been pointed out [102, 108].

For our work, we have reviewed the extended AIS senior scholar's basket of scholarly journals [90] known as the 'basket of eight.' The reason to focus our work on these eight journals is their acknowledged quality and their centrality in the international IS discipline. They are also likely to provide insight into what editors of these journals are looking for in a grounded theoretical contribution which we regard to be a good proxy for understanding the requirements towards DSR's expected theoretical contributions. All journals in the basket were covered from their first issue to the most recent issue available in the respective electronic databases (EBSCO, ScienceDirect, JSTOR, AISeL, and Ingenta-Connect). We also used the homepages of the respective journals or publishers (e.g., Wiley and Palgrave Macmillan) to ensure completeness and reliability of our search. Within the databases we conducted an extended search for articles that contain the phrase "grounded theory" in their title, abstract, or keywords. The rational for this approach is to exclude articles that only refer to GTM superficially or extend work of a previous article that was based on GTM. The resulting 27 articles were included in our detailed review and are listed in table 1.

Table 1: Overview of the articles reviewed in details

Journal	**Papers**
European Journal of Information Systems (EJIS)	[35, 109, 121]
Information Systems Journal (ISJ)	[10, 44, 57, 68, 93, 94, 105]
Information Systems Research (ISR)	[53]
Journal of the Association for Information Systems (JAIS)	[17, 20, 88, 110, 112]
Journal of Management Information Systems (JMIS)	[18, 82, 92, 122]
Management Information Systems Quarterly (MISQ)	[65, 78]
Journal of Strategic Information Systems (JSIS)	[55, 84, 99]
Journal of Information Technology (JIT)	[81, 113]

With these articles at hand, each author went through the papers separately. This review was conducted to assign the papers to one of three groups: (1) papers that use GTM to actually build GT, (2) papers that use GTM or elements of the method but do not build GT (e.g., work with a priori theoretical considerations or provide detailed empirical accounts of a phenomena using coding element of GTM), or (3) papers that deal with GT or GTM from a methodological standpoint [4, 49]. Aggregating our analyses, a total of 4 papers (out of 27) were rated differently by the authors. Primarily, we disagreed with respect to the categorization of papers in groups 1 and 2, that is, when to recognize a paper as such that actually contributes to the disciplines theoretical body. A joint discussion on the methodology – that is, the standards of our groups – and an additional review of the papers allowed us to resolve the discrepancies, ensured inter-rater reliability [100], and enabled aggregation of the results.

Looking at this methodological approach, we want to highlight an important property of our review: In the context of this report, the literature we sampled is not intended to be an exhaustive account of all the IS studies that follow GTM or that build theory. As mentioned in the introduction, instead of a complete overview and classification of such publications, we rather intend to select a set of exemplary articles that help us understand how theoretical contributions can be build from scratch in the context of IS research as an indication of how to extend current DSR approaches.

This leads us to acknowledge a set of limitations of our approach that might have an impact on our intended contribution. With

respect to the selection of journals we have chosen, we are aware that they are not all of the same age and have different target groups. This means that they are likely to show variances with respect to the number and nature of GTM-based papers they published. Focusing on the basket of eight is a limitation, too. With respect to journals outside the basket of eight, especially the Journal of Information Technology Theory and Application (JITTA) as well as the Scandinavian Journal of Information Systems (SJIS) published GT and GTM research. While such papers are not included in our review, we draw on their contributions to inform our approach and in our discussion and analysis of the review [e.g., 8, 43, 80, 103]. Beyond the journal-related issues, also our sampling of GT/GTM articles impacts our findings. First, some of the journals also published articles using GTM or producing GT in IS that are not labeled so in their titles, abstracts, or keywords [e.g., 30, 59, 87]. For reasons of consistency of our search criteria and overall feasibility of our study we decided to not include these in the review. However, we believe that this does not adversely impact our sample due to its exemplary nature. Finally, even though we conducted the analysis of the articles separately and read all papers carefully to ensure inter-coder reliability, a categorization of our colleagues' work always leaves room for potential misinterpretation or individual bias.

5. RESULTS AND DISCUSSION

In a first step, we reviewed the 27 articles that fall within the search criteria we defined to group them with respect to their theoretical contribution (table 2).

Table 2: Overview of the groups of GTM-articles

Journal	Gr. 1	Gr. 2	Gr. 3	#
EJIS	-	[35, 109, 121]	-	*3*
ISJ	[44, 93]	[10, 57, 68, 94]	[105]	*7*
ISR	-	[53]	-	*1*
JAIS	[17]	[20, 88, 110, 112]	-	*5*
JMIS	[18, 82, 92]	[122]	-	*4*
MISQ	[65, 78]	-	-	*2*
JSIS	[84]	[55, 99]	-	*3*
JIT	[81, 113]	-	-	*2*
Total	**11**	**15**	**1**	***27***

Group 1: papers that use GTM to build GT
Group 2: papers that use GTM or elements of it but do not build GT
Group 3: papers with a methodological viewpoint on GTs or GTM

To refine our observations, we extended our analysis of the theory building papers (group 1) by looking at the kind of theory they provide. We draw on Gregor's [45] discussion of the nature of theories in IS to distinguish between theories for (1) analysis, (2) explanation, (3) prediction, (4) explanation and prediction, and (5) design and action. This allowed for deeper insights into the theory building publications. As shown in table 3, all five types of theory were present with a noticeable peak of type 2 publications, that is, papers concerned with theories for explanation.

Table 3: Classifying group 1 based on Gregor [45]

Theory of …	Papers
... analysis	[93]
... explanation	[17, 65, 78, 82, 84, 92]
... prediction	[81]
... explanation and prediction	[18]
... design and action	[44, 113]
Total	**11**

Overall, our review shows that only one of the 27 papers from the basket of eight deals with GT/GTM from a methodological standpoint (group 3). Within the group 1 papers, over half the papers that actually build theory produce theories for analysis or explanation of a phenomenon. Only four papers out of the 27 we reviewed provide theories that allow for predictions of some kind, inform the design of artifacts, or prescriptively guide actions. Even more surprising, 15 out of the 27 papers we reviewed only revert to GTM as a methodology but do not explicitly use it to build theories. Most of these papers go to great length to provide detailed empirical accounts or use GTM to code data to work with a set of pre-established hypotheses. While some would argue that the latter is a misconception of GTM altogether [97], we feel that all of these papers provide rich insight into the phenomenon they study and that the application of GTM allows for a deep emersion in the materials the authors analyze. Nevertheless, none of the papers in this second group of our review made an additional effort in framing its results in a way that would make it an explicit theoretical contribution to IS research in its own right.

In the context of our investigation our primary interest are the papers published in group 1. As stated before, group 1 contains papers that actually build theory that is grounded directly in empirical observations from an IS-specific context. Through our analysis of these papers we made two interesting observations. First, only few of these papers explicitly express that they are producing theory. Based on a thorough analysis of their empirical work, most papers aggregate their findings into an abstract representation that summarize the factors and relationships the authors identify in their work. Only few of them state the theoretical nature of their findings. Second, and possibly related, all of the group 1 papers produce theory in a very specific context. While this has impacts on their generalizability [62], they still qualify as "early" or "small" substantive theories which describe a specific area of inquiry [45]; a property quite related to theories that are likely to emerge from a specific design's context.

6. THE THEORY BUILDING PROCESS

Glaser and Strauss [39] highlight that any interesting observation or lived experience can be a source for empirical material to build a theory with – a GTM characteristic that seems to be especially suitable for design science projects since the DSR paradigm makes similar assumptions.

Looking at the group 1 papers, the authors of the studies we analyzed all use their detailed empirical material to gradually define and refine their theoretical understanding of the phenomena they study; as suggested by GTM [39]. Some authors use an initial review of relevant literature to inform and motivate the research [17, 18, 65, 82, 84, 92, 93].These reviews are, however, not designed to develop an a priori understanding of the phenomenon, but to frame the context of the research and enable the researchers to focus their attention in the field. Thus, such a review is not a violation of the GTM principles [22, 104] and could also help researchers to better frame the problem they are trying to solve and to identify the relevant problem aspects and how an IS artifact would support their interaction.

Looking at the sources of empirical data used to generate theory, all of the studies rely on either open or semi-structured interviews. Going beyond this, some studies also use documented informal

discussions [44, 82] during their time in the field as a source of data extending the official statements gathered during the interviews. Follow ups (e.g., via telephone) for clarification [44, 82, 113] are also an established approach to increase the validity and reliability of observations. Participant observations [18, 44, 78, 82, 92] or analytical field notes taken by the researcher [65, 84] can be another important instrument to ensure that the researchers have gathered information comprehensively and exhaustively. Some studies also used strategies beyond the immediate interactions of the subjects in the field. An example is the analysis of both internal and external documents [44, 78, 84, 92]. Such an analysis can also include various press releases [17], blogs [17], or records of meetings [18], even though these sources were less prominent in our sample. The same is true for the use of questionnaires to gather additional information from study participants [18] or the implementation of focus groups [81]. Within the studies we analyzed, action research or action learning [82] were not used frequently. With the data at hand, one study summarizes its observations as a concise case study [65] in which the authors are developing a comprehensive narrative that incorporates all their data and observations from the field. With respect to DSR, we think that all of these strategies are currently used in the context of (software) design. While the materials are often used to identify requirements or to conduct evaluations, they could easily be leveraged to also identify underlying theoretical considerations.

Looking at the way data was analyzed to support theory building some studies follow the Straussian strand of GTM [13, 96, 113] and use respective coding schemes [17, 82]. Other studies use more general inductive-hypothetic research strategies [18, 44, 78, 84, 92]. All the studies used intertwined data analysis and theory building. Some had rather short cycles in which both occurred almost simultaneously (e.g., after each interview individually), others relied on a more granular approach by abstracting their findings towards theory only after one instance had been completed (e.g., theoretical analysis of the materials only occurred after one entire company had been captured).

With respect to actually building the theory, the studies mostly followed a process theoretical understanding. The authors used their empirical materials and their understanding thereof to identify causes, contexts, intervening conditions, actions, interactions, and the consequences of phenomena [17, 18, 44, 65, 78, 81, 82, 92]. With respect to the framing of these results, the studies differed according to the type of theory they develop. While some only focus on a deep analysis and description of the problem [93], others use their theoretical understanding to give advice to future system designers [17]. Either the insights were used as an extension of existing theoretical models [18] or the improved understanding of processes and their context was used to suggest an own conceptual framework [44, 65, 78, 81, 82, 84, 92, 113]. Some authors even go beyond the mere description and make recommendations on how to improve these processes [44, 113].

Beyond the 11 group 1 papers, especially the group 3 paper by Urquhart et al. [105] is of particular interest when analyzing how to enrich DSR using GTM. The authors are providing very detailed and elaborate guidelines on how to conduct GTM-based research in IS and offer an up-to-date view on important methodological publications and examples of IS studies using GTM.

Table 4: The emergence of a grounded theory

Stage	**Purpose**
Codes	Identifying anchors that allow the key points of the data to be gathered (open coding)
Concepts	Collections of **codes** of similar content that allows the data to be grouped (axial coding)
Categories	Broad groups of similar **concepts** that are used to generate a **theory** (selective coding)
Theory	A collection of explanations that explain the subject of the research

Taken altogether, table 4 shows how GTM can be understood as a process of carving out the theory form observations while collecting and coding of data occur simultaneously. By adding additional observations, mere descriptions of single observations (codes) gradually become more general (concepts and categories) until, ultimately, a theory emerges [e.g., 105]. Beyond the methodological literature on GTM in a general context [e.g., 13, 41, 96], there also is some literature that suggests guidelines for techniques and application and provides practical examples of IS-specific GTM research [c.p., 74, 105]

7. THE ROLE OF THEORY IN DSR

Looking at the group 1 articles to better understand how they create theory and what the role of theory in their projects is reveals some interesting patterns. Reverting back to the principal understanding of a theory's components and comparing it to the results of our review, we believe that theories are going through a process of maturing as they help IS researchers to document and accumulate knowledge. Such a process in which theories are maturing is not uncommon in other disciplines. Holström et al. [52] suggest a similar process in operations management research and in the administrative sciences, van Maanen [107] highlights the need for a strong basis of descriptive narratives before being able to build strong theories. But, as Mintzberg points out [73], data alone does not generate theory, only researchers do. Quoting Weick [118], Sutton and Staw [98] highlight that knowledge grows by extension and that providing accounts of small but comprehensible events is a chance to build cumulative theory. We believe this to be a good link to DSR. At first, designs of an artifact are likely to be isolated, single possible solutions to an individual real-world problem [48]. Repeated instantiations of an artifact and its systematic introduction and constant evaluation in several different contexts are an excellent way to add additional grounding to the artifact and its theoretical core [24].

In IS research, Lee and Hubona [63] support this view by highlighting that there are two general forms of validity of theories. The formative validity of a theory is achieved through theory building and describes a theory's property to adequately capture a phenomenon's concepts and their relations. A theory's summative validity means that it survives repeated empirical testing and that its external validity grows as the theory is able to model or predict more and more instances of the phenomenon. The interrelationships among theory types identified by Gregor [45], as well as the fact that the ability to predict and guide action are only covered by theory types three through five, also suggest that such a process of maturing is present in IS theories. Looking at our review, GT-based work seems to produce the detailed empirical accounts needed to build substantive theories as a starting point for the development of mid-range or even grand theories.

Building such theories starts with the invaluable work of providing interesting empirical accounts [25]. This helps initial IS theories to have fit and relevance [39] and to be current and interesting in their domains [5, 16, 67]. In our review we were able to identify some examples of such accounts. The works by King [57] and Work [121] (both group 2) document the extensive analyses the authors have conducted in their respective areas and offer a great opportunity to influence the design of respective artifacts and develop a respective theoretical core at the same time.

Once the understanding of the phenomena grows beyond good stories alone, the identification of constructs and relations that are present beyond individual cases will help to build theories [27, 28]. For example, Day et al. [17] (group 1, type 2) use their insights from investigating the disaster response to hurricane Katrina to identify a set of factors that explain how information flows are impeded in extreme cases. The authors also highlight some implications this understanding could have on the design of future disaster recovery systems in general. Pauleen [82] (also group 1, type 2) offers a detailed discussion how leaders facilitate relationship building in virtual teams.

Such explanation in IS-specific substantive theories will enable a certain degree of prediction. A good example is Palka et al. (group 1, type 3) who describe their contribution as follows: "The outcome is a grounded theory of mobile viral marketing with respect to the consumer and his social network, decomposing the mobile viral effect and identifying the determinants of reception, usage, and forwarding of mobile viral content" [81, p. 172]. Such an understanding will, in turn, give practitioners an understanding and some control of the situations they find themselves confronted with [8]. We believe this to be important to highlight that while IS research accumulates knowledge on a certain subject, these insights should also be integrated back to the larger body of (IS-specific) theory. This corresponds to both the suggestions of enfolding extant literature often recommended in the context of case-based theory building [26, 38] as well as to the requirement of DSR to contribute back to the knowledge base [50]. A GTM example is the paper by de Vreede et al. [18] (group 1, type 4) in which the authors use a grounded analysis of the acceptance of a group support system in an African context to identify constructs refining the technology acceptance model [14, 15]. In our study, however, most studies use multiple sites or instances and only one uses a single site [17]. This is a property not fully compatible with DSR as most designs focus on individual contexts and specific problems they intend to solve [48].

Once theories incorporate explanation and prediction, they might be ready to inform the design of IS artifacts by serving as kernel theories [50] to design theories [46, 111]. From our set of reviewed articles, Webb and Gallagher [113] (group 1, type 5) are suggesting a methodology for multimedia systems development which they ground in a study of development processes across 16 companies. Abstracting from an individual or local focus, their work certainly points direction to draw broader, more general conclusions – i.e. generalize DSR.

Observing the behavior of actors and systems as they are confronted with the effects of the artifact in the field will then serve as an opportunity to extract knowledge relevant to extend, refine, or even reject theories [58]. This way, in terms of GTM and DSR, instantiations of IS theories in practice by means of designing artifacts can serve as an additional "slice of data" in the process of theoretically sampling more data to work towards theoretical saturation. Such an approach will enable the empirically grounded extension and maturation of IS-specific theories. Glaser also provides guidelines how to develop formal theory based upon prior substantive theories [40]. We suggest that it could be this process of transforming behavioral theories into kernel theories for design theories and using the design to feed back into theory development and extension that might be a remedy for DSR's alleged lack of theoretical contributions. Many of the group 1 papers we looked at build their contributions from scratch – that is, they are not based on prior empirical or theoretical work but grounded in the observed data. This fact also concurs with the basic assumptions of the DSR paradigm. On top of that, this paper shows some interesting opportunities for future investigations with respect to the theoretical work in DSR. It is intended to help design researchers to frame their research results in a way that makes them a theoretical contribution and to leverage the rich empirical materials many of us gather in a way that supports the IS-specific accumulation of knowledge.

Of course such an approach raises further questions on the nature of theory in IS research like, for example, "can a generalized, isolated, and IS-specific theoretical body exist after all?" or "what is the specific role of IS theories at the intersection of computer and social sciences?" Gregor's [45] seminal article has provided an influential discussion on the nature of theory in the IS discipline. Since the discussion is still relatively young, we feel that these questions have not been sufficiently answered to date. Given our findings, we would like to emphasize the need for a discussion on how DSR related work can contribute the emerging body of theoretical knowledge in our discipline.

8. CONCLUSION

Given the IS-specific background of our analyses, our research presents the example of the grounded theory method applied to developing theories in IS research. We conducted a literature review which shows that some authors rely on GTM to produce IS-specific theory. We look forward to some of the work we present in this paper helping IS design researchers to understand the principal components of theory that constitute a respective contribution. Moreover, our analysis also highlighted the important role of theory in the process of discovering, extending, and refining, or in short maturing IS theory in general and in the context of designing in particular.

Looking at the findings presented in the last section, we concur with Gregory in highlighting that "developing grounded substantive theory about IT artifact use and its relationships with human behavior and the organizational environment may provide the means for IT artifact evaluation […] and leverage a design science research project to make a theoretical contribution to the knowledge base to go beyond the local solution of a problem and the implementation of an IT artifact" [48, p. 11]. We believe that the considerations on the processes of conducting the research project and of developing grounded theory we highlighted above are an important step towards understanding how DSR and GTM can actually be integrated and to promote the role of theory in DSR-related work. While the methodological discussion on GTM in the IS discipline is still relatively young [8, 103, 105], we expect that our review and discussion offers scholars a first insight into the subject and shows the potential of GTM to theoretically extend DSR. Beyond what we discussed here, general descriptions of

GTM-based research and respective guidelines can be found in a variety of sources – both IS-specific [e.g., 31, 64, 74, 105] and outside the IS discipline [e.g., 22, 71, 72, 97, 101].

Our contribution to this discussion can, however, only be a first step into a more thorough analysis of the potentials and implications of theory building in the context of DSR. We believe the analogies between GTM and DSR at are the foundation of our results have some interesting implications with respect to the next steps in this discourse. However, the results need to be carefully evaluated in light of our study's limitations we discussed earlier. As mentioned in the paper's section on research design, the selection of papers we used for our review is supposed to be an exemplary illustration of grounded theory building that could also occur in DSR. To strengthen this argument, future research will have to more thoroughly explore the epistemological and methodological underpinnings of GTM and DSR in order to determine whether an integration of the two can go beyond the analysis of analogies as introduced here.

Beyond the study's considerations on theory, also an explicit step towards an analysis of seminal DSR papers would be necessary to develop a concise set of recommendations that could guide DSR scholars in their theory building efforts. This could help to produce a better and more detailed understanding of how to integrate theory building into designing explicitly. Possibly, a more detailed distinction of a-priori grounding design work theoretically, doing theoretical work while designing, and re-integrating theoretical contributions to the larger body of knowledge ex-post (e.g., through evaluation) can also result in a more refined understanding of how to increase DSR's theoretical contributions. This might help to improve the distinction between mere design as consultants' work and DSR as the task of the scientists [120].Moreover, the fact that IS-specific theory does not necessarily have to be built using GTM needs to be considered. Gregor [45] suggests a wide selection of methods that can be used in the context of building any of her five theory types. As other methods can also produce IS theories, a more inclusive selection of theories produced in the IS context is a logical next step in continuing and extending our research. Analyzing some of the theories identified by the wiki on theories used in IS research [91] could be an interesting opportunity to better understand theory building in IS.

In further developing our understanding of design as a distinct source of knowledge in the IS discipline, we want to conclude by highlighting some of the potentials we see in enabling DSR to also produce relevant theories. In IS research, there are technological artifacts on the computer science side and behavioral research paradigms on the social science side [29]. We assume that bridging this gap is one of the core challenges of the IS field and that IS-specific theory can help to do so. Lee [61] supports this view by pointing to the value of examining the interaction between technological and social systems. Emphasizing the role of the IT artifact in theorizing in IS research, Orlikowski and Iacono [79] highlight how IS theories can inform research and practice on how understanding behavioral processes influences the design of IT artifacts and how these, in turn, impact behavioral processes. Gregor [47] points to the need for theories that help us understand the links between the natural sciences, the social sciences, and the sciences of the artificial to bridge this gap.

However, current IS research seems to be reluctant to rely on the early and often small theories that originate from IS-specific contexts. Quite to the contrary, many IS researchers seem to be more enthused by relying on (grand) theories from our adjacent disciplines [116]. DSR can help to offer viable alternatives by providing detailed theoretical accounts of the phenomena encountered during artifact design and evaluation. Explicitly discussing the role of theory and integrating respective recommendations into IS research approaches might help DSR to mature and to "catch up" in the race for credibility. Moreover, it could help IS researchers to refine their expectations towards DSR – that is, improve our understanding of what theoretical contributions based on design are and what role they can play in the context of IS research. Beyond this, a more refined conceptualization of theories and their role can help IS (design) researchers to frame their results in a way that makes them a valuable theoretical contribution. This will make even the smallest pieces of empirical observations accessible to building a cumulative tradition in IS and make them valuable contributions to build, extend, or refine IS-specific theories. On top of that, this should help to get more of the valuable design work conducted published in our discipline's premier journals, thus making it accessible for the whole IS community.

REFERENCES

[1] Atmanspacher, H. On the Role of Theoretical Work in the Sciences. Freiburg, Germany, 2007.

[2] Baskerville, R. What Design Science is not. *European Journal of Information Systems*, 17, 5 (2008), 441-443.

[3] Baskerville, R.L., and Myers, M.D. Information Systems as a Reference Discipline. *MIS Quarterly*, 26, 1 (2002), 1-14.

[4] Becker, J., Niehaves, B., Olbrich, S., and Pfeiffer, D. Forschungsmethodik einer Integrationsdisziplin - Eine Fortführung und Ergänzung zu Lutz Heinrichs "Beitrag zur Geschichte der Wirtschaftsinformatik" aus Gestaltungsorientierter Perspektive. In, Becker, J., Krcmar, H., and Niehaves, B., (eds.), *Wissenschaftstheorie und Gestaltungsorientierte Wirtschaftsinformatik*, Münster, Germany: Institut für Wirtschaftsinformatik - Westfälische Wilhelms-Universität Münster, 2008, pp. 5-26.

[5] Benbasat, I., and Zmud, R.W. Empirical Research in Information Systems: The Practice of Relevance. *MIS Quarterly*, 23, 1 (1999), 3-16.

[6] Benbasat, I. Editorial Notes. *Information Systems Research*, 12, 2 (2001), iii-iv.

[7] Benbasat, I., and Zmud, R.W. The Identity Crisis within the IS Discipline: Defining and Communicating the Discipline's Core Properties. *MIS Quarterly*, 27, 2 (2003), 183-194.

[8] Bryant, A. Re-Grounding Grounded Theory. *Journal of Information Technology Theory and Application*, 4, 1 (2002), 25-42.

[9] Burton-Jones, A., McLean, E.R., and Monod, E. In Pursuit of MIS Theories: Process, Variance, and Systems. *Working Paper*, Atlanta, GA, USA: Department of Computer Information Systems, Robinson College of Business, Georgia State University, 2004.

[10] Calloway, L.J., and Ariav, G. Designing with Dialogue Charts: A Qualitative Content Analysis of End-User Designers' Experiences with a Software Engineering Design Tool. *Information Systems Journal*, 5, 2 (1995), 75-103.

[11] Carroll, J.M., and Swatman, P.A. Structured-Case: A Methodological Framework for Building Theory in Information Systems Research. *European Journal of Information Systems*, 9, 4 (2000), 235.

[12] Cooper, H.M. Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews. *Knowledge in Society*, 1 (1988), 104-126.
[13] Corbin, J., and Strauss, A. Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology*, 13, 1 (1990), 3.
[14] Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13, 3 (1989), 319-340.
[15] Davis, F.D., Bagozzi, R.P., and Warshaw, P.R. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35, 8 (1989), 982-1003.
[16] Davis, M.S. That's Interesting! - Towards a Phenomenology of Sociology and a Sociology of Phenomenology. *Philosophy of the Social Sciences*, 1, 2 (1971), 309-344.
[17] Day, J.M., Junglas, I., and Silva, L. Information Flow Impediments in Disaster Relief Supply Chains. *Journal of the Association for Information Systems*, 10, 8 (2009), 637-660.
[18] de Vreede, G.-J., Jones, N., and Mgaya, R.J. Exploring the Application and Acceptance of Group Support Systems in Africa. *Journal of Management Information Systems*, 15, 3 (1998), 197-234.
[19] Dearden, J. MIS Is a Mirage. *Harvard Business Review*, 50, 1 (1972), 90-99.
[20] DeLuca, D., Gallivan, M.J., and Kock, N. Furthering Information Systems Action Research: A Post-Positivist Synthesis of Four Dialectics. *Journal of the Association for Information Systems*, 9, 2 (2008), 48-71.
[21] Denyer, D., and Tranfield, D. Using Qualitative Research Synthesis to Build an Actionable Knowledge Base. *Management Decision*, 44, 2 (2006), 213-227.
[22] Dey, I. *Grounding Grounded Theory: Guidelines for Qualitative Inquiry*. San Diego, CA, USA: Academic Press, 1999.
[23] Doty, D.H., and Glick, W.H. Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling. *The Academy of Management Review*, 19, 2 (1994), 230-251.
[24] Dubois, A., and Gadde, L.-E. Systematic Combining: An Abductive Approach to Case Research. *Journal of Business Research*, 55, 7 (2002), 553-560.
[25] Dyer, W.G., and Wilkins, A.L. Better Stories, Not Better Constructs, to Generate Better Theory: A Rejoinder to Eisenhardt. *Academy of Management Review*, 16, 3 (1991), 613-619.
[26] Eisenhardt, K.M. Building Theories from Case Study Research. *Academy of Management Review*, 14, 4 (1989), 532-550.
[27] Eisenhardt, K.M. Better Stories and Better Constructs: The Case for Rigor and Comparative Logic. *Academy of Management Review*, 16, 3 (1991), 620-627.
[28] Eisenhardt, K.M., and Graebner, M.E. Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal*, 50, 1 (2007), 25-32.
[29] Evermann, J., and Tate, M. Constructs in the Mist: The Lost World of the IT Artifact. *30. International Conference on Information Systems (ICIS 2009)*, Phoenix, AZ, USA: Association for Information Systems, 2009.
[30] Feller, J., Finnegan, P., Fitzgerald, B., and Hayes, J. From Peer Production to Productization: A Study of Socially Enabled Business Exchanges in Open Source Service Networks. *Information Systems Research*, 19, 4 (2008), 475-493.
[31] Fernández, W.D. The Grounded Theory Method and Case Study Data in IS Research: Issues and Design In, Hart, D., and Gregor, S., (eds.), *Information Systems Foundations: Constructing and Criticising* Canberra, Australia: School of Business and Information Management, The Australian National University, 2005, pp. 43-60.
[32] Fettke, P. State-of-the-Art Des State-of-the-Art: Eine Untersuchung der Forschungsmethode „Review" Innerhalb der Wirtschaftsinformatik. *Wirtschaftsinformatik*, 48, 4 (2006), 257-266.
[33] Frank, U. Towards a Pluralistic Conception of Research Methods in Information Systems Research. In, Adelsberger, H., Chamoni, P., Dorloff, F., Echtle, K., Eicker, S., Frank, U., Goedicke, M., Kollmann, T., Müller-Clostermann, B., Pohl, K., Rathgeb, E.P., Unland, R., and Zelewski, S., (eds.), *ICB-Research Reports*, Essen, Germany: Institut für Informatik und Wirtschaftsinformatik (ICB), 2006, pp. 84.
[34] Freese, L. Formal Theorizing. *Annual Review of Sociology*, 6 (1980), 187-212.
[35] Galal, G.H. From Contexts to Constructs: The Use of Grounded Theory in Operationalising Contingent Process Models. *European Journal of Information Systems*, 10, 1 (2001), 2-14.
[36] Gasparski, W.W., Gregory, S.A., Foqué, R., and Strzalecki, A. Contemporary History of Design Science. *Design & Systems General Applications of Methodology - Praxiology* (1995), 15-24.
[37] Gehlert, A., Schermann, M., Pohl, K., and Krcmar, H. Towards a Research Method for Theory-Driven Design Research. In, Hansen, H.R., Karagiannis, D., and Fill, H.-G., (eds.), *9. Internationale Tagung Wirtschaftsinformatik (WI 2009)*, Vienna, Austria: Österreichische Computer Gesellschaft, 2009, pp. 441-450.
[38] Gibbert, M., Ruigrok, W., and Wicki, B. What Passes as a Rigorous Case Study? *Strategic Management Journal*, 29, 13 (2008), 1465-1474.
[39] Glaser, B.G., and Strauss, A.L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL, USA: Aldine, 1967.
[40] Glaser, B.G. *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley, CA, USA: Sociology Press, 1978.
[41] Glaser, B.G. *Basics of Grounded Theory Analysis: Emergence vs. Forcing*. Mill Valley, CA, USA: Sociology Press, 1992.
[42] Glaser, B.G. *The Grounded Theory Perspective: Conceptualization Contrasted with Description*. Mill Valley, CA, USA: Sociology Press, 2001.
[43] Goldkuhl, G. Design Theories in Information Systems - a Need for Multi-Grounding. *Journal of Information Technology Theory and Application*, 6, 2 (2004), 59-72.
[44] Goulielmos, M. Systems Development Approach: Transcending Methodology. *Information Systems Journal*, 14, 4 (2004), 363-386.
[45] Gregor, S. The Nature of Theory in Information Systems. *MIS Quarterly*, 30, 3 (2006), 611-642.

[46] Gregor, S., and Jones, D. The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, 8, 5 (2007), 313-335.
[47] Gregor, S. Building Theory in the Sciences of the Artifical. *4. International Conference on Design Science Research in Information Systems and Technology*, Philadelphia, PA. USA, 2009.
[48] Gregory, R.W. Design Science Research and the Grounded Theory Method: Characteristics, Differences, and Complementary Uses. *18. European Conference on Information Systems (ECIS 2010)*, Pretoria, South Africa, 2010.
[49] Heinrich, L.J. Forschungsmethodik einer Integrationsdisziplin: Ein Beitrag zur Geschichte der Wirtschaftsinformatik. *NTM International Journal of History and Ethics of Natural Sciences, Technology and Medicine*, 13, 2 (2005), 104-117.
[50] Hevner, A.R., March, S.T., Park, J., and Ram, S. Design Science in Information Systems Research. *MIS Quarterly*, 28, 1 (2004), 75-105.
[51] Hirschheim, R., and Klein, H.K. Crisis in the IS Field? A Critical Reflection on the State of the Discipline. *Journal of the Association for Information Systems*, 4, 5 (2003), 237-293.
[52] Holmström, J., Ketokivi, M., and Hameri, A.-P. Bridging Practice and Theory: A Design Science Approach. *Decision Sciences*, 40, 1 (2009), 65-87.
[53] Hunter, M.G., and Beck, J.E. Using Repertory Grids to Conduct Cross-Cultural Information Systems Research. *Information Systems Research*, 11, 1 (2000), 93-101.
[54] Iivari, J. A Paradigmatic Analysis of Information Systems as a Design Science. *Scandinavian Journal of Information Systems*, 19, 2 (2007), 39-64.
[55] Irani, Z., Love, P.E.D., and Jones, S. Learning Lessons from Evaluating Egovernment: Reflective Case Experiences That Support Transformational Government. *The Journal of Strategic Information Systems*, 17, 2 (2008), 155-164.
[56] Kaplan, A. *The Conduct of Inquiry*. New York, NY, USA: Harper & Row, 1964.
[57] King, S.F. Case Tools and Organizational Action. *Information Systems Journal*, 6, 3 (1996), 173-194.
[58] Kuechler, B., and Vaishnavi, V. On Theory Development in Design Science Research: Anatomy of a Research Project. *European Journal of Information Systems*, 17, 5 (2008), 489-504.
[59] Lederer, A.L., and Mendelow, A.L. The Impact of the Environment on the Management of Information Systems. *Information Systems Research*, 1, 2 (1990), 205-222.
[60] Lee, A.S. Editor's Comments - Research in Information Systems: What We Haven't Learned. *MIS Quarterly*, 25, 4 (2001), v-xv.
[61] Lee, A.S. Editor's Comments - MIS Quarterly's Editorial Policies and Practices. *MIS Quarterly*, 25, 1 (2001), iii-vii.
[62] Lee, A.S., and Baskerville, R.L. Generalizing Generalizability in Information Systems Research. *Information Systems Research*, 14, 3 (2003), 221-243.
[63] Lee, A.S., and Hubona, G.S. A Scientific Basis for Rigor in Information Systems Research. *MIS Quarterly*, 33, 2 (2009), 237-262.
[64] Lehmann, H.P. A Grounded Theory of International Information Systems *School of Business and Information Management*, Auckland, New Zealand: University of Auckland, 2001.
[65] Levina, N., and Vaast, E. Innovating or Doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *MIS Quarterly*, 32, 2 (2008), 307-332.
[66] Lewin, K. The Research Center for Group Dynamics at Massachusetts Institute of Technology. *Sociometry*, 8, 2 (1945), 126-136.
[67] Lewis, M.W., and Grimes, A.I. Metatriangulation: Building Theory from Multiple Paradigms. *Academy of Management Review*, 24, 4 (1999), 672-690.
[68] Lundell, B., and Lings, B. The 2g Method for Doubly Grounding Evaluation Frameworks. *Information Systems Journal*, 13, 4 (2003), 375-398.
[69] Lyytinen, K., and King, J.L. Nothing at the Center? - Academic Legitimacy in the Information Systems Field. *Journal of the Association for Information Systems*, 5, 6 (2004), 220-246.
[70] March, S.T., and Smith, G.F. Design and Natural Science Research on Information Technology. *Decision Support Systems*, 15, 4 (1995), 251-266.
[71] Martin, P.Y., and Turner, B.A. Grounded Theory and Organizational Research. *Journal of Applied Behavioral Science*, 22, 2 (1986), 141-157.
[72] Melia, K.M. Rediscovering Glaser. *Qualitative Health Research*, 6, 3 (1996), 368-378.
[73] Mintzberg, H. An Emerging Strategy of "Direct" Research. *Administrative Science Quarterly*, 24, 4 (1979), 582-589.
[74] Müller, B., and Olbrich, S. Developing Theories in Information Systems Research - the Grounded Theory Method Applied. In, Dwivedi, Y.K., Wade, M., and Schneberger, S.L., (eds.), *Information Systems Theory: Explaining and Predicting Our Digital Society*, New York, NY, USA: Springer, forthcoming.
[75] Mulrow, C.D. Rationale for Systematic Reviews. *British Medical Journal*, 309, 6954 (1994), 597-599.
[76] Myers, M.D. Qualitative Research in Information Systems. *MIS Quarterly*, 21, 2 (1997), 241-242.
[77] Nunamaker, J.F., Chen, M., and Purdin, T.D.M. Systems Development in Information Systems Research. *Journal of Management Information Systems*, 7, 3 (1990), 89-106.
[78] Orlikowski, W.J. Case Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development. *MIS Quarterly*, 17, 3 (1993), 309-340.
[79] Orlikowski, W.J., and Iacono, C.S. Research Commentary: Desperately Seeking the "IT" In IT Research - a Call to Theorizing the IT Artifact. *Information Systems Research*, 12, 2 (2001), 121-134.
[80] Ovaska, P., Rossi, M., and Smolander, K. Filtering, Negotiating and Shifting in the Understanding of Information System Requirements. *Scandinavian Journal of Information Systems*, 17, 1 (2005).
[81] Palka, W., Pousttchi, K., and Wiedemann, D.G. Mobile Word-of-Mouth a Grounded Theory of Mobile Viral Marketing. *Journal of Information Technology*, 24, 2 (2009), 172-185.
[82] Pauleen, D.J. An Inductively Derived Model of Leader-Initiated Relationship Building with Virtual Team Members. *Journal of Management Information Systems*, 20, 3 (2003), 227-256.
[83] Peffers, K., Tuunanen, T., Rothenberger, M.A., and Chatterjee, S. A Design Science Research Methodology for Infor-

mation Systems Research. *Journal of Management Information Systems*, 24, 3 (2008), 45-77.
[84] Petrini, M., and Pozzebon, M. Managing Sustainability with the Support of Business Intelligence: Integrating Socio-Environmental Indicators and Organisational Context. *The Journal of Strategic Information Systems*, 18, 4 (2009), 178-191.
[85] Popper, K. *The Logic of Scientific Discovery*. London: Unwin Hyman, 1980.
[86] Pries-Heje, J., and Baskerville, R. The Design Theory Nexus. *MIS Quarterly*, 32, 4 (2008), 731-755.
[87] Ransbotham, S., and Mitra, S. Choice and Chance: A Conceptual Model of Paths to Information Security Compromise. *Information Systems Research*, 20, 1 (2009), 121-139.
[88] Ribes, D., and Finholt, T.A. The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, 10, 5 (2009), 375-398.
[89] Rohde, M., Stevens, G., Brödner, P., and Wulf, V. Towards a Paradigmatic Shift in IS: Designing for Social Practice. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST '09)*, Philadelphia, Pennsylvania, USA: New York: ACM, 2009, Article No.15, 2009.
[90] Saunders, C., Avison, D., Davis, G., Ein-Dor, P., Galletta, D., Hirschheim, R., and Straub, D. Senior Scholars' Basket of Journals. AIS Senior Scholars Forum Subcommittee on Journals, 2006.
[91] Schneberger, S., and Wade, M. Theories Used in IS Research. 2007.
[92] Scott, J.E. Facilitating Interorganizational Learning with Information Technology. *Journal of Management Information Systems*, 17, 2 (2000), 81-113.
[93] Seeley, M.E., and Targett, D. A Senior Executive End-User Framework. *Information Systems Journal*, 7, 4 (1997), 289-308.
[94] Siau, K., Tan, X., and Sheng, H. Important Characteristics of Software Development Team Members: An Empirical Investigation Using Repertory Grid. *Information Systems Journal*, online article in advance of print, published electronically December 2007 (2007).
[95] Simon, H.A. *The Science of the Artificial*. Cambridge, MA, USA: The MIT Press, 1996.
[96] Strauss, A.L., and Corbin, J.M. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA, USA: Sage, 1990.
[97] Suddaby, R. From the Editors: What Grounded Theory Is Not. *Academy of Management Journal*, 49, 4 (2006), 633-642.
[98] Sutton, R.I., and Staw, B.M. What Theory Is Not. *Administrative Science Quarterly*, 40, 3 (1995), 371-384.
[99] Tingling, P., and Parent, M. An Exploration of Enterprise Technology Selection and Evaluation. *The Journal of Strategic Information Systems*, 13, 4 (2004), 329-354.
[100] Tinsley, H.E., and Weiss, D.J. Interrater Reliability and Agreement of Subjective Judgments. *Journal of Counseling Psychology*, 22, 4 (1975), 358-376.
[101] Turner, B.A. The Use of Grounded Theory for the Qualitative Analysis of Organizational Behaviour. *Journal of Management Studies*, 20, 3 (1983), 333-348.
[102] Urbach, N., Smolnik, S., and Riempp, G. The State of Research on Information Systems Success. *Business & Information Systems Engineering*, 1, 4 (2009), 315-325.
[103] Urquhart, C. Regrounding Grounded Theory - or Reinforcing Old Prejudices? A Brief Reply to Bryant. *Journal of Information Technology Theory and Application*, 4, 3 (2002), 43-54.
[104] Urquhart, C., and Fernandez, W. Grounded Theory Method: The Researcher as Blank Slate and Other Myths. *27. International Conference on Information Systems (ICIS 2006)*, Milwaukee, WI, USA, 2006, pp. 457-464.
[105] Urquhart, C., Lehmann, H., and Myers, M.D. Putting the 'Theory' Back into Grounded Theory: Guidelines for Grounded Theory Studies in Information Systems. *Information Systems Journal*, online article in advance of print, published electronically May 2009 (2009).
[106] van de Ven, A.H. Nothing Is Quite So Practical as a Good Theory. *The Academy of Management Review*, 14, 4 (1989), 486-489.
[107] Van Maanen, J. Some Notes on the Importance of Writing in Organizational Studies. In, Cash, J.I., and Lawrence, P.R., (eds.), *The Information Systems Research Challenge*, Boston, MA, USA: Harvard Business School, 1989, pp. 27-33.
[108] Viering, G., Legner, C., and Ahlemann, F. The (Lacking) Business Perspective on SOA – Critical Themes in SOA Research. In, Hansen, H.R., Karagiannis, D., and Fill, H.-G., (eds.), *9. Internationale Tagung Wirtschaftsinformatik (WI 2009)*, Vienna, Austria: Österreichische Computer Gesellschaft, 2009, pp. 45-54.
[109] Volkoff, O., Strong, D.M., and Elmes, M.B. Understanding Enterprise Systems-Enabled Integration. *European Journal of Information Systems*, 14, 2 (2005), 110-120.
[110] Wales, R.C., Shalin, V.L., and Bass, D.S. Requesting Distant Robotic Action: An Ontology for Naming and Action Identification for Planning on the Mars Exploration Rover Mission. *Journal of the Association for Information Systems*, 8, 2 (2007), 75-104.
[111] Walls, J.G., Widmeyer, G.R., and El Sawy, O.A. Bulding an Information System Design Theory for Vigilant Eis. *Information Systems Research*, 3, 1 (1992), 36-59.
[112] Webb, B., and Mallon, B. A Method to Bridge the Gap between Breadth and Depth in IS Narrative Analysis. *Journal of the Association for Information Systems*, 8, 7 (2007), 368-381.
[113] Webb, B., and Gallagher, S. Action in Context and Context in Action: Modeling Complexity in Multimedia Systems Development. *Journal of Information Technology*, 24, 1 (2009), 126-138.
[114] Weber, R. Toward a Theory of Artifacts: A Paradigmatic Base for Information Systems Research. *Journal of Information Systems*, 1, 2 (1987), 3-19.
[115] Weber, R. *Ontological Foundations of Information Systems*. Melbourne, Australia: Coopers & Lybrand, 1997.
[116] Weber, R. Editor's Comments - Theoretically Speaking. *MIS Quarterly*, 27, 3 (2003), iii-xii.
[117] Webster, J., and Watson, R.T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26, 2 (2002), xiii-xxiii.

[118] Weick, K.E. Agenda Setting in Organizational Behavior: A Theory-Focused Approach. *Journal of Management Inquiry*, 1, 3 (1992), 171-182.
[119] Whetten, D.A. What Constitutes a Theoretical Contribution? *Academy of Management Review*, 14, 4 (1989), 490-495.
[120] Winter, R. Design Science Research in Europe. *European Journal of Information Systems*, 17, 5 (2008), 470-475.
[121] Work, B. Patterns of Software Quality Management in Tick-it Certified Firms. *European Journal of Information Systems*, 11, 1 (2002), 61-73.
[122] Zahedi, F., Van Pelt, W.V., and Srite, M. Web Documents' Cultural Masculinity and Femininity. *Journal of Management Information Systems*, 23, 1 (2006), 87-128.
[123] Zmud, B. Editor's Comments - "Pure" Theory Manuscripts. *MIS Quarterly*, 22, 2 (1998), xxix-xxxii.

Strategies for Creating, Generalising and Transferring Design Science Knowledge – A Methodological Discussion and Case Analysis

Philipp Offermann
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin, Germany
+49 30 8353-58439
philipp.offermann@telekom.de

Sören Blom
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin, Germany
+49 30 8353-58123
soeren.blom@telekom.de

Udo Bub
Deutsche Telekom Laboratories
Ernst-Reuter-Platz 7
10587 Berlin, Germany
+49 30 8353-58410
udo.bub@telekom.de

ABSTRACT
Design Science Research has been well accepted as part of Information Systems Research. The discussion about the research process and the structure of design theories has been going on for some time. While research has been done on the relation between design theories and other types of theories, not much has been said about how design knowledge can be re-used. Other disciplines refer to such re-use as "generalisation" and "transfer". We define a three-level separation of design abstraction (short-, mid-, and long-range) and show how knowledge re-use strategies operate between and within them, as well as how they relate to generalisation and transfer. Each strategy is supported by a case from an existing publication, showing that the types of design theories and the research strategies can be found in practice. We argue that these research strategies can provide guidelines to researchers and reviewers for planning, performing and evaluation Design Science Research.

Keywords
Design science research, knowledge creation, knowledge transfer, generalisation, research strategies.

1. INTRODUCTION
Design Science Research (DSR) is a recent mode of research in Information System Research (ISR). It is distinct from natural science, as it is concerned with the creation of artefacts relevant to practitioners, whereas natural science is "aimed at understanding reality" [18]. A sizeable amount of research has been done in recent years to outline and detail the methodological principles of design science in ISR. Papers on DSR, such as the seminal paper by Hevner et al. [13], have been published in prestigious journals, illustrating the increasing acceptance of DSR as a valid research paradigm within the community [10; 15; 23; 29; 30].

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

As in any other scientific discipline, Information System Research aims at developing knowledge based on evidence. A desired attribute of such knowledge (usually represented by theories in the natural science paradigm) is generalisability. This means that a theory is valid "[...] in a setting different from the one where it was empirically tested and confirmed." [16]. But even within the natural science paradigm the achievability of generalisation is debated, especially with regard to qualitative research methods [11; 16]. Here, the concept of transferability is introduced [11] to contrast generalisability. Settings might be similar, especially when research involves social dimensions, and insights might be transferred from one to the other but still be far from (statistical) generalisability.

The discussion about creation of design knowledge has so far either focussed on the construction (and sometimes evaluation) of a single design theory or the discussion on how design science as a field is connected with other parts of ISR [13]. The publications of design theory structures acknowledge the use of theories and other types of knowledge [10; 32], but do not explicitly examine the relation to other design theories. From our own design experience and knowledge of design examples, it appeared unintuitive that designs should be that "monadic", also considering that other fields practice knowledge re-use and other strains of design discuss the matter [4].

The aim of this paper is to identify re-use and creation strategies of design knowledge. Studying similar discussion outside of DSR, we saw that "generalisability" suggests different levels of knowledge and "transferability" abstraction (at least in some sources) and describe a lateral movement between settings. We adopted this for designs by introducing three levels of abstraction, in which short-range designs refer to individual solutions, mid-range designs refer to the notion of design theories and long-range designs refer to "schools of thought" and paradigms. Furthermore we employed the "purpose and scope" concept of [10] to distinguish if designs addressed a similar or different problem. We identified knowledge creation and re-use strategies on this basis as transformations of design knowledge that occur between or within abstraction levels and purpose and scope. For each strategy we present a published example to show how the strategy looks like in practice.

The remainder of this paper is organised as follows: first, we introduce the concepts of generalisability and transferability form

behavioural Information Systems Research. Then, we introduce Design Science Research with its different types of output. Based on the types of design identified, we present strategies to create, generalise and transfer knowledge. For each strategy, we present an ISR publication where the strategy has been realised. Finally, we discuss the approach and draw a conclusion.

2. KNOWLEDGE CREATION IN INFORMATION SYSTEMS RESEARCH

To understand the knowledge creation process in Information System Research, it is helpful to understand what knowledge is typically created trough research in this discipline. The Merriam Webster dictionary defines knowledge as "the fact or condition of knowing something with familiarity gained through experience or association" [19]. Scientific knowledge, then, is a subset of this knowledge; namely the knowledge that has been gathered through following scientific methods. While different types of knowledge might exist (Hevner et al. [13] list theories, frameworks, instruments, constructs amongst others), theories are at the core of this knowledge, as the behavioural side of IS research is focussed on "the development and justification of theories" [13].

Theories, as all scientific knowledge, are supposed to be supported by evidence. Usually, though, the theories make claims about a range of settings or phenomena that are larger than the instances in which the theory has been tested. It is not possible to test instances that do not exist at a given point in time and for practical reasons it might not be possible to ever test every possible instance. This very circumstance is what makes theories and scientific knowledge in general valuable: Being able to reason about a phenomenon that is new, or has not been tested yet, based on prior experiences. "Academic knowledge involves the quest for general or 'covering' laws and principles concerning the fundamental nature of things. The more context free, the more general and the stronger the theory." [1].

2.1 Generalisability

The activity that gives knowledge a wider use is called "generalise" as in "give wider use to something: to use something in a wider or different range of circumstances, or be used in this way" (Encarta World English Dictionary). "Generalisability is a major concern to those who do, and use, research." [16]. As a body of knowledge should contain knowledge useful to the whole discipline, generalising is an important research activity.

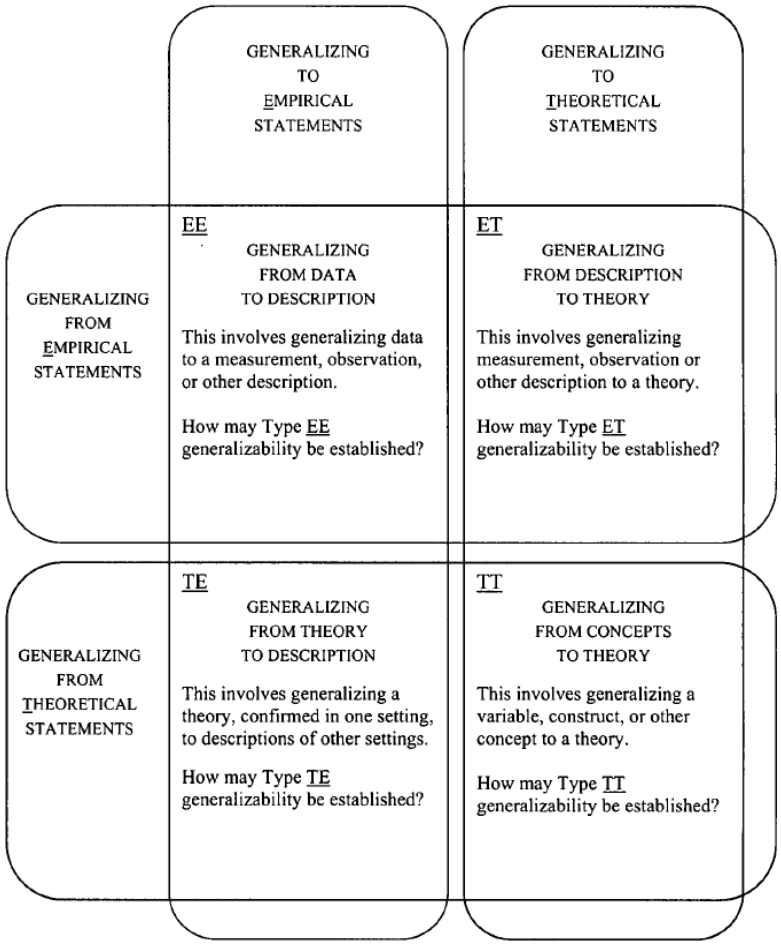


Figure 1: Four Types of Generalizing and Generalizability [16].

Lee and Baskerville [16] analysed generalisability in Information Systems Research. While generalisability is well established in quantitative research, they argue that qualitative researchers also call for generalisability. However, they see their interpretation of generalisability to be too heavily influenced by quantitative research methods and even there the statistical meaning of "generalisability" to be often misinterpreted. Generalisability of sample points in quantitative research is done to a sample estimate, not to the corresponding population characteristic [16]. A generalisation from a sample to population characteristics is not possible. Accordingly, a theory that was developed from case studies cannot be generalised to other cases where the theory has not been tested.

Lee and Baskerville [16] propose a generalisability framework that contains four types of generalising and generalisability (see figure 1). Type "EE" generalises from empirical statements to empirical statements. Examples are the generalisation of data to measurement and the generalisation of a measurement beyond the sample from which the data was collected. Type "ET" generalises from empirical statements to theoretical statements. Examples are the generalisation from measurement to theory and generalising a theory beyond the sample from which the theory was derived. Type "TE" generalises

from theoretical statements to empirical statements. This happens when a theory is tested in a setting the theory has not been tested before, e.g. a practitioner using a theory in an enterprise. Finally, type "TT" generalises from theoretical statements to theoretical statements. This occurs when theoretical propositions are developed based on concepts.

2.2 Transferability

Generalisability, however, is only one of the terms used for scientific inquiries for the aspect of applicability [11]. For naturalistic inquiries, "generalizations of the rationalistic variety are not possible because phenomena are intimately tied to the times and the contexts in which they are found" [11]. However, to ensure applicability, transferability of the research results should be achieved: "Yet these facts do not obviate the possibility that some transferability between two contexts may occur because of certain essential similarities between them." [11]

While Lee and Baskerville discuss generalisability both for positivism and for interpretivism, Travis [27] identifies transferability rather than generalisability as a goal in interpretivist research. She clarifies that transferability is not generalisability. Transferability does not give a precise prediction about the applicability of the findings to a different sample. Rather, transferability enables the utilisation of the findings in a different setting. "Transferability represents the degree of transfer 'between sending and receiving contexts' where the sending context is that of the researcher or inquirer." [27] It is left to the scientist doing the transfer to determine the validity in the new context. To achieve transferability, a thick description of the context is needed. "If the thick descriptions demonstrate an essential similarity between two contexts, then it is reasonable to suppose that tentative findings of Context A are also likely to hold in Context B (although, to be safe, an empirical test of that presumption should be made)." [11] Lee and Baskerville do not mention "transferability" at all. Perhaps they subsume "transferabilty" under "generalisability".

3. DESIGN SCIENCE RESEARCH

From our point of view, it is intuitively clear that generalisation and transfer are relevant for design science. In this community, the objects under study, as well as the outputs of research, are the designs themselves. If designs would not be general but merely particular solutions to particular, everyday problems (e.g. of one company), it would be of little relevance to the research community.

3.1 Relevance and generality of Design Science knowledge

Hevner et al. [13] see IS research to act between the environment and the knowledge base (see figure 2). The research has to be applicable in the appropriate environment, and at the same time provide additions to the knowledge base. This knowledge base is used to generate new designs by abduction [28]. As "knowledge becomes 'relevant' when it is context specific" [1] to fulfil business needs, an artefact designed needs to be as specific in respect to people, organisations and technology as possible. The more adapted a design is to a specific setting in practice, the more relevant it is, as instantiations are easier to generate. On the other hand, the more specific a design is, the narrower the scope and the less likely to find a case for another instantiation.

Design science is about creating designs ("design as an artifact" [13] (Guideline 1)) that solve "important and relevant business problems" [13] (Guideline 2). The types of artefacts observed in publications are system design, method, language/notation, algorithm, guideline, requirements, pattern and metric [21].

The research output should be described as a design theory [10; 32]. The publication of a design theory is not the solution itself, e.g. a running software systems or an optimised business process. Rather, a design theory contains the design of a solution, e.g. software architecture or a business process model. If a solution is based on a design, it is an "instance" of the design. A design theory also contains the purpose and scope of the design, specifying for which context instances are supposed to be useful and the utility to be expected respectively.

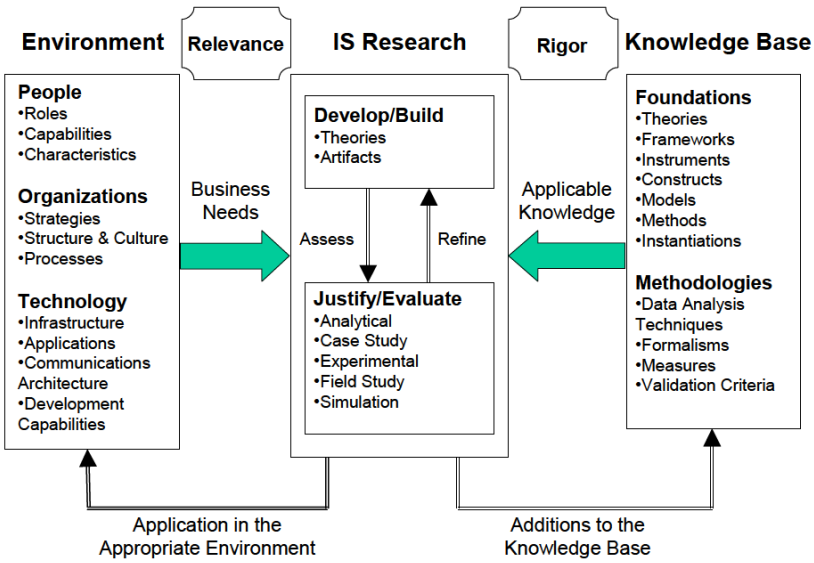


Figure 2: Information Systems Research Framework [13].

There seems to be some awareness of the relevance of the level of abstraction in the community. However, generalisation and transfer have not received much explicit attention. The most explicit statements about generalising that we could find were: "The design scientist must be able both to generalize the findings and demonstrate a theoretical contribution." [14] and "Design-science research holds the potential for three types of research contributions based on the novelty, generality, and significance of the designed artifact." [13]. Other

Table 1: Types of design according to range of scope.

Design type	Definition	Role in design	Role in research	Examples
Short-range design	Design for a specific setting	An instance (system implementation, method enactment) can directly be derived from the design	First-of-a-kind solution to a relevant problem.	The specification for a CRM system; the software development process for a company
Mid-range design	Design for a specific type of setting	The design can be used to create a short-range design for a particular solution of the same problem domain	Identification of relevant design elements for a particular problem domain	eXtreme Programming, TOGAF, Rational Unified Process, relational database design
Long-range design	General insights about a type of design approach	Educational, as a starting point for dealing with a problem, illustrating a particular design "world-view"	Inform more specific designs	SOA, Object-Orientation, relational data-management, agile software development

Design Research disciplines outside of ISR discuss generalisability and transferability [4; 5].

The limited awareness might also be due to the limited discussion of epistemology in design science [24]. We think that the focus of Rohde et al. [24] onto social practice with the corresponding position of ontological idealism and a consensus theory of truth will lead to knowledge that is only valid in a specific organisation. In that case, no generalisation or aggregation of knowledge would be possible, rendering any prescriptive theory pointless. While we debate the exclusiveness of their opinion, we agree that it is a vital part of design science.

3.2 Types of Designs according to their range

Analogous to the work in [16], we want to identify the different inputs and outputs of generalisation and transfer for design. We believe that this is a first step to better understand what these processes mean for Design Science Research in ISR. We do so by focussing on designs as inputs and outputs.

In the field of sociology, Merton [20] introduces the concept of "theory of the middle range", implying that there are different levels of theory, relative to their distance to empirical observations and stating that "middle range theories" are preferable to more general "total systems of sociological theory" [20]. Merton specifies: "Middle-range theory involves abstractions, of course, but they are close enough to observed data to be incorporated in propositions that permit empirical testing" [20] This concept has been mentioned by other DSR scholars (e.g. [8; 14]), but it remains to be discussed how design research can generate knowledge by moving between different levels of design. To do so, we will first introduce three different "ranges" of design and will discuss how generalisation and transfer happens between these types of design in the next section.

Winter discusses a "tradeoff between the level of solution generity and the problem scope" [33]. Design research needs to produce "situational artefacts" [33] that cover a range of problems that then can be adapted to a problem at hand. We will call this "adapted" type of design *short-range design*: the design is only applicable to the particular situation (e.g. it contains company-specific features). Such a design might be the architecture of a company's CRM system or a company's software development process.

A more general design, that is valid not only for a specific setting, but for the whole type of settings will be called *mid-range design*. Borrowing from Merton's concept of mid-range theory, we postulate that mid-range designs form the basis for a whole range of (situation-specific) short-range designs permitting empirical testing through instantiations. As these designs are of general interest, they form the most common design type observed in research, the "situational artefacts" of Winter [33]. When looking at the design theory literature, especially Walls et al. [32] and Gregor and Jones [10], this level of abstraction seems to be what design theories are supposed to address.

From these mid-range designs, general design principles can be extracted. They are not bound to particular situations and are not bound by individual designs but capture the "fundamentals" from which certain types of designs can be constructed. We will call these principles *long-range designs*. Long-range designs can become paradigms that shape the research agenda of part of a research community and lend their concepts to many different mid-range designs. They often can be found in introductory-level courses and textbooks. An overview of the three design types is presented in table 1.

3.3 Epistemological positions for different types of Designs

Becker and Niehaves [3] propose an epistemological framework, with which they want to capture different philosophical assumptions of IS research efforts. The framework consists of five epistemological questions and offers different answers for each.

The first question asks what the "object of cognition" is. Designers attempt to solve problems in "the real world" and solve problems by creating / manipulating "real-world" artefacts. This corresponds to a position of "Ontological realism". On the other hand, the process of identifying what a problem is usually incorporates the views of the researcher and different stakeholders, which might introduce aspects of the problem based on their perception. This would point more to a position of transcendental idealism or "Kantianism" in the words of [3]. The second question asks what the "relationship between cognition and object of cognition" is. In the context of design this refers to how we can learn about properties of the problem (analysis) and of solution artefacts (validation). "Objective" measures can be

Table 2: Research strategies in Design Science.

Strategy	Affected types	Approach	Research contribution	Validation
Explore new problem	Short-range	Invent design for new problem	First-of-a-kind design offers first design insights on new problem	At least one real-life instance validates utility
Validate mid-range design	Mid-range to short-range	Create new short-range design and validate its utility	Increased generalisability of utility statement	Use of accepted evaluation strategies
Generalise to mid-range design	Short-range to mid-range	Analyse commonalities and differences of short-range designs with comparable purpose and scope and find generalised representation	Captures generalised knowledge in terms of common design elements about a problem domain	The process of identifying similarities and finding generalised representation of concepts. Demonstration of applicability of new mid-range design by creating a new short-range design from it.
Apply out of scope	Mid-range to short-range	Derive short range design from mid-range design and change it to work for new problem	Indication that mid-range design might cover wider scope and possibly first-of-a-kind design	At least on real-life instance validates utility in a setting outside of the original scope
Synthesise mid-range design	Mid-range to mid-range	Analyse commonalities and differences of mid-range designs with comparable purpose and scope and find generalised representation	Make mid-range design better transferable and possibly increase utility	The process of identifying similarities and finding generalised representation of concepts. Demonstration of applicability of new mid-range design by creating a new short-range design from it
Combine designs	Mid-range to mid-range	Merge designs with adjacent purpose and overlapping scope	Create a design with a more comprehensive purpose	The process of combining the designs. Demonstration of applicability of new mid-range design by creating a new short-range design from it
Extract long-range design	Mid-range to long-range	Analyse commonalities and differences of mid-range designs from the same domain and identify common principles	Captures design principles that apply to a whole class of problems	The process of identifying the design principles

made, but phenomena can also be interpreted based on individual predispositions. Hence, both realism and constructivism play a role. The third question asks what true cognition is. "Truth" concerns both the results of analysis and validation. Both correspondence theory (for "objective" measurements) as well as consensus theory (for interpretative results) are relevant. The fourth question asks about the sources of cognition. Designers, as discussed previously both observe and interpret and either activity informs the other. Observations might change interpretation, but interpretation "guides" the observation. This is best captured by "Kantianism". The final question asks how the cognition can be achieved methodologically. In design, a variety of approaches might have to be employed. This includes induction, wherever knowledge is gained by generalizing individual observations; deduction can be necessary if the design is derived from a reference framework and hermeneutics might be employed when a designer transfers knowledge between contexts, and needs to re-interpret statements in the new situation.

The different levels of abstraction in the three ranges of design reflect a difference in the nature of knowledge that is captured on each range and therefore differ in their epistemological position. The above description captures short-range designs as we understand it. Mid-range designs, which according to [20] should also lead to testable hypotheses, are also largely capture by the description. Nevertheless, mid-range designs always refer to short-range designs, either as the source from which a mid-range design is constructed, or as the result, which is derived from a mid-range design. This increases the role of ontological idealism, as the dependence on the concepts of other individuals grows. It also introduces the necessity to harmonise different designs, which leads to a semantic theory of truth [3]. Long-range designs are even further away from the real word. The objects have no direct representation outside of the mind. Their "truth" can only be established consensually, and they represent "a priori" knowledge. This kind of knowledge has large influence when a researcher approaches a new short- or mid-range design, as it

provides the categories which shape the perception of the problem at hand.

4. PROPOSAL OF STRATEGIES FOR KNOWLEDGE CREATION, GENERALISATION AND TRANSFER

Our approach of identifying strategies consisted of three steps:

1. Generating candidates, based on the framework of design levels and "scope and purpose",
2. Trying to find at least one published example for each candidate,
3. Pruning the candidate list of those strategies for which we could not find an example.

For our design-level model to be meaningful, we expected to find at least one strategy on each level and/or between each level. The distinction between "similar scope" vs. "different scope" would have to be found in at least one strategy. We have identified seven strategies: one within the short range, three between short- and mid-range, of which ones stretches different scopes, two within mid-range and one between mid- and long-range, as described in the remainder of this section.

At the beginning of any design research effort, a problem is identified for which no solution is available. The problem has not been solved before and cannot be solved by some trivial variation of some other solution. Such a problem would either occur unexpectedly while researching some other problem or as the starting point of a deliberate venture, but in both cases the main focus would be to solve this particular problem, resulting in a short-range design. The design process is influenced by experiences with earlier designs and general theories. The first design would show that the problem does have a solution. Insight gathered during the design process can deliver first insights on the nature of the problem and the range of design alternatives. Hence, **exploring a new problem** is a valid research strategy in early stages of research.

The goal of mid-range designs, either created through generalisation or as an immediate, genuine creation, is to be able to derive specific, short-range designs from them to address concrete problems. Whenever a new short-range design is derived from a mid-range design and it successfully solves the problem at hand, it **validates the mid-range design**. The mid-range design proposes that its application within a certain scope of situations will yield a certain utility and the more situations a design has been shown to work, the more likely it is considered to work for similar new problems. This mode comes closest to the notion of "generalisation" in quantitative science.

Once several short-range designs exist for an existing problem, it might be possible to **generalize** them **to a mid-range design**. For this strategy, a number of short-range designs are analysed for commonalities. The mid-range design is formulated in a way that is more general than the grounding short-range designs. An example for this activity is the creation of design patterns (e.g. [7]). Patterns are parts of re-usable software designs, derived from many individual solutions. The act of generalisation is in itself a creative, design-based act: Elements of the short-range designs must be identified for inclusion, and other elements must be added or changed to work in a more generalised context. Also, the designs need to be considered holistically to grasp their intentions and idiosyncrasies.

A mid-range design might also be transferred to a short-range design outside of the original scope (**apply out of scope**). The designer might realise similarities between the problem at hand and the solution, even if the latter does not claim to solve this problem. As the resulting short-range design is out of scope, the claimed utility of the mid-range design might not materialise. If the solution does show to be successful, it is an indicator that the scope of the mid-range design can be extended.

To create better designs, existing mid-range designs with an overlapping purpose and scope might be analysed to identify the strong points of each and to create a new mid-range design that combines these, in other words, **synthesizing a mid-range design**. This would typically happen whenever several approaches with the same intention have emerged and it becomes obvious that they are not fundamentally different. Since a synthesised design is nevertheless a new design, validation becomes necessary.

To enlarge the scope and/or purpose of a design, designs might be combined into a single design that is more comprehensive (**combine designs**). To arrive at a new design, similar elements in the original designs need to be harmonised and the interface between the individual source designs needs to be defined.

Finally, short-range and/or mid-range designs might be analysed to extract common principles that solve a certain class of problems - to **extracting a long-range design** from these inputs. This would typically happen when the research in a field has progressed and several, possible synthesised mid-range designs have emerged. Those designs would have very basic common assumptions, even if the designs themselves differ from each other. A coherent set of such assumptions would form the paradigm resulting in a long range design. An overview over the strategies can be found in table 2.

The strategies imply a certain order of applicability for each research topic, depending of the topic's maturity, as shown in figure 3. For a new topic, no designs are available. Therefore, exploring a new problem and creating a short range design is a viable strategy. Alternatively, a mid-range design might be created from scratch and validated. As time goes by and the topic remains relevant, several short-range designs become available. Then, mid-range designs can be generalised. In a next step, these mid-range designs can be validated to increate the generalisability of their utility. As time moves on and more and more mid-range designs become available, the design can be synthesised to enhance the utility of the design. Also, as the topic develops and the problem changes, mid-range designs can be applied out of scope or be combined to solve the changed problem. Finally, from a set of mid-range designs of various types (system architecture, pattern, method, etc.) but similar topics, a long-range design can be extracted.

5. CASE STUDIES

For each strategy, we present a paper that uses the strategy to demonstrate that the types of design and the strategies are not merely theoretical constructs but can also be found in research.

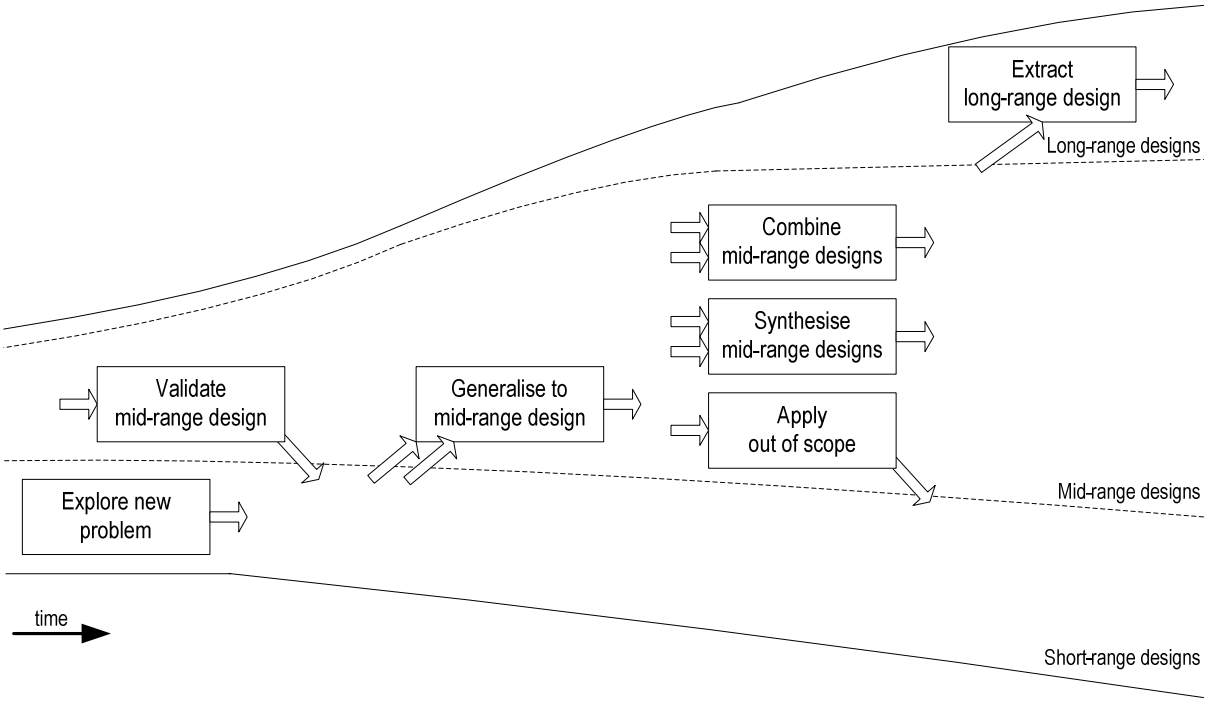


Figure 3: Sequence of design strategy application as design theories accumulate.

5.1 Explore new problem

A good short-range design for a new problem has been published by Spiekermann et al. [26] in their article "SkillMap – A Social Software for Knowledge Management – From Concept to Proof". They start by identifying shortcomings in current designs: "... knowledge management systems (KMS) have failed to fulfil the organizational promises with which they were first introduced." [26] They then identify a theory that promises to better explain human behaviour relevant to knowledge management: "In their seminal work on KM, they accumulate a number of enabling conditions that foster 3rd generation knowledge creation and sharing in companies. These include: intention, autonomy, variety, creative chaos and redundancy." [26] They continue to present a design along the enabling conditions and then present their solution: "This article presents a software called skillMap that was built as a proposition for how 3rd generation knowledge management tools could function." [26] They present the graphical user interface and the architecture of the system. To evaluate user perception, the "user experience through the GUI and the activation of intrinsic motives such as fun and curiosity" were measured.

The paper presents a typical short-range design. Based on theory, a first-of-kind solution is created. The design presented is an abstraction of the solution and therefore tightly bound to the solution. It explains how the different parts of the solution support the enabling conditions. Once more third-generation knowledge management tools have been implemented and the decisive elements are better understood, the designs might be generalised into a mid-range design that is independent of a specific solution.

5.2 Validate mid-range design

Balijepally et al. [2] took the pair programming technique as a mid-range design and evaluated its utility. While pair programming is only one of the techniques that form a software development method, it is independent of any specific method and can therefore be seen as a mid-range theory. Before presenting their own evaluation, they present the results of prior studies on the utility of pair programming. As there is no official utility statement for pair programming, each evaluation has a different operationalisation of utility. The main topics evaluated were software quality, development effort and task complexity. Balijepally et al. [2] then present their research model that evaluates software quality, programmer satisfaction and confidence in performance. They conduct a laboratory experiment with students. They found "that pair performance typically cannot exceed the performance of its best member working individually". They also found that "pairs were more satisfied than both the best and the second-best members of nominal pairs" and that "pairs were more confident in their performance, compared to the second-best members of nominal pairs, but not the best members".

While Balijepally et al. [2] did not create a new design, they evaluated a design to increase the generalisability of the design's utility. Design evaluation is an essential part of Design Science Research, and while it might be argued if evaluation is actually design, without an evaluation the utility of a design cannot be demonstrated.

5.3 Generalise to mid-range design

A good mid-range method design to manage the company-wide application architecture that is generalised from the short-range designs for "Credit Suisse Financial Services", "Die Mobiliar" and "HypoVereinsbank" has been published by Hafner and Winter [12] in their article about a management method for the company-wide applicaton architecture. The aim of the paper is to design a consolidated method. After discussing general requirements, they discuss four existing methods for the management of application architectures. They then present three company case studies where the implemented management process has been analysed. Based on the existing methods and the case studies, a consolidated method is derived.

The method published by Hafner and Winter [12] looks at three short-range designs extracted from case studies to identify common elements and idiosyncrasies. They then consolidate the common elements into a method that focuses on the elements that are likely to be important for all instances of the method, leaving out elements that are specific to a certain case. Thereby transferability is increased, as it is more likely to focus on the relevant elements.

5.4 Apply out of scope

The UN/CEFACT Unified Modeling Methodology (UMM) is a mid-range design in form of a method to model B2B processes. It specifies how to model interactions and the information exchanged between different entities. The scope of the method uses B2B interactions. Dietrich [6] wanted to see if the UMM also has utility when used internally by companies, outside the scope of the original design. To verify his thesis, he applied the UMM to processes internal to a company. He created a short-range design for a German capital investment company. By instantiating the design, he was able to demonstrate that the UMM has utility outside of its declared scope.

The transfer of design knowledge out of scope will regularly be performed in practice, where practitioners use any design knowledge available to solve a given problem, despite the declared scope of the design. Dietrich [6] is one of the few researchers who scientifically performed this transfer and demonstrate the usefulness of the transferred design.

5.5 Synthesise mid-range design

Offermann and Bub [22] looked at existing mid-range methods to design systems according to the service-oriented architecture (SOA). For each method, they discussed which parts of the SOA and of the software development lifecycle are covered. They continued to identify weaknesses of the existing methods. Based on the existing methods, they then proposed a new mid-range method to design SOA-systems that overcomes the weaknesses. For the relevant activities in the new method, they discuss which existing SOA-method it is based on.

To incrementally advance knowledge in an established domain, synthesising an improved mid-range design from existing mid-range designs is common practice in research. The publication of Offermann and Bub [22], is a good example for such a synthesis. Usually, after the synthesis, the new mid-range design is validated to ensure its utility and to demonstrate that the new design is indeed an improvement.

5.6 Combine designs

A combined mid-range design has been published by Sowa et al. [25] in their paper "Integrated Information Security Risk Management – Merging Business and Process Focused Approaches". They introduce the existing approaches Business Oriented management of Information Security (BORIS) and Operational Risks in Business and IT (ORBIT). The focus of BORIS is "to handle ... business oriented ISM issues" [25], while ORBIT aims "to control operational risks in business processes in regard to information technology" [25]. The authors then propose a merged design as a generic data model for the integrated information security risk management. The integration point of the two approaches is the control management where in both cases a scorecard evaluation of security and risk drivers is done. The merged design fulfils all requirements on information security management.

The design of Sowa et al. [25] combines business driven and process oriented information security risk management into a single design. The new design has a more general purpose and a larger scope than the original designs. Because it fulfills all the requirements, it can be expected to yield a higher utility than each of the original designs.

5.7 Extract long-range design

Legner and Heutschi [17] published a survey of service-oriented architecture (SOA) design principles. They analysed nine publications from the domain of SOA and extracted ten design principles grouped in four classes: "interface orientation", "interoperability", "autonomy / modularity" and "business suitability". For each of the nine publications they identify which design principles are mentioned. The publications do not focus on a discussion of the design principles. Rather, they propose method and/or technologies for SOA that are directly transferable to solve a company problem. The design principles extracted are more general and form the basis for more specific SOA designs.

By extracting general design principles from concrete design, the scope of the design is not changed. However, such design principles do not directly solve any business problem and therefore also have no direct utility. They form the basis for a whole class of solutions. Other such examples of design principles are object-orientation with a whole set of supporting technology (e.g. c++), modeling notations (e.g. UML) and methods (e.g. xxx) and relational data models (MySQL, entity-relationship-model, database normalisation).

6. DISCUSSION

The concept of generalisation in positivist research is used to infer from a sample to another sample e.g. taken by future researchers. For design theories, this kind of generalisation can be applied to a utility statement, affirming the utility of a design for a set of problems [31]. For example, by creating more instances of a design and measuring the utility of the instance, the probability that the next researcher or practitioner creating an instance will also find the utility increases. This notion of generalisation assumes that the observed variables are to some extent similar and remain unchanged within the whole population of observed entities, so that observing a subset of entities can plausibly approximate the conditions in the population. This mode of

knowledge creation was only one ("validate mid-range design") of the seven we propose in this article. In design science "'artificial phenomena' have to be created by the researcher" [14]. Designs capture not only the structure of theses artificial phenomena, but also the goals and intentions of those creating them. When creating a solution for a problem at hand, a design will guide the creator but will leave room for and will actually require creativity and considerable thought. As such any design is open to change and adaptation and the original design does not necessarily exclude other uses. Three of our strategies acknowledge this (combine, synthesize mid-range design and apply out of scope). Creation of something new, without reference to an existing design, is also possible at any time (create new short-range design). The long-range design, finally, is not so much a theory as a world view for a particular set of design problems. As seen in the example above, the principles are valuable, but neither lead to any specific design nor make any specific testable predictions. Nevertheless, they have value. For example, they can offer a basis for education and to evaluate critical parts in an unfamiliar design. While the three levels of abstraction were deduced theoretically, we were able to find published examples on all levels, as can be seen in table 1 and section 4.

The fact that the majority of our strategies are concerning mid-range design is coherent to other views on design theory (e.g. [9]). This is explainable through the re-use potential of mid-level designs: Short-range designs solve the immediate problem at hand but contain many details that are idiosyncratic for the solution. It might offer inspiration and insights for designers that have similar problems at hand, but at the outset it is unclear whether the amount of help gained through the design outweighs the effort to understand and then discard design parts irrelevant to the new solution. Long-range designs, on the other hand, might not be available for a particular kind of design; and even if available, it might be the problem itself might not be addressed by a general design principle or the translation of such a principle into specific design is not straight-forward. Also, as already mentioned, long-range designs are hardly verifiable.

7. CONCLUSION

While the research process in design science is well established, it is still unclear how design knowledge can be created in a cumulative way that goes beyond individual solutions to individual problems. The problem arises from the requirement to create designs that are relevant to practice but at the same time contribute to the knowledge base. In this paper we presented three types of designs that offered different levels of abstractions in terms of distance between solution support as codified in a design and problems.

We then presented strategies how to create knowledge based on the different abstraction levels and how to generalise and transfer designs. By means of presenting published examples of all seven strategies, we demonstrated that the theoretically derived design types and strategies do not only exist on paper, but can be found in practice. The strategies help researchers in identifying and performing Design Science Research projects as they offer criteria to categorise their design, depending on a research topic's maturity. By focussing not only on the creation of new designs for specific business problems but by also developing existing designs to become more general, we hope to contribute towards design science becoming a more cumulative science which can move forward.

The strategies proposed by us are relevant to all types of Design Science Research outputs. Therefore, the operationalisation of the strategies remain rather generic. We have only included research strategies for which we could find examples. It is possible, and we expect further strategies to be identified. For certain types of outputs more specific guidelines might be given. For example, when combining mid-range method designs, there might be integration points such as the method's role models or the artefacts produced by the methods a researcher could look into to create the combined design. Both extending and detailing strategies offer opportunities for future research.

8. REFERENCES

[1] Aram, J.D. and Salipante Jr., P.F. 2003. Bridging Scholarship in Management: Epistemological Reflections. British Journal of Management 14, 189-205.

[2] Balijepally, V., Mahapatra, R., Nerur, S., and Price, K.H. 2009. Are Two Heads Better than One for Software Development? The Productivity Paradox fo Pair Programming. MIS Quarterly 33, 1, 91-118.

[3] Becker, J. and Niehaves, B. 2007. Epistemological perspectives on IS research: a framework for analysing and systematizing epistemological assumptions. Information Systems Journal, 17, 197-214.

[4] Chow, R. 2008. Case Transfer Vs Case Study: An Evaluation of Case Study as a Method for Design Research. In Swiss Design Network Symposium.

[5] Chow, R. and Jonas, W. 2010. Case Transfer: A Design Approach by Artefacts and Projection. Design Issues.

[6] Dietrich, J. 2008. Nutzung von Modellierungssprachen und -methodologien standardisierter B2B-Architekturen für die Integration unternehmensinterner Geschäftsprozesse. Gito.

[7] Gamma, E., Helm, R., and Johnson, R.E. 1994. Design Patterns - Elements of Reusable Object-Oriented Software. Addison-Wesley Longman, Amsterdam.

[8] Gregor, S. 2006. The Nature of Theory in Information Systems. MIS Quarterly 30, 3, 611-642.

[9] Gregor, S. 2009. Building Theory in the Sciences of the Artificial. In DESRIST'09 ACM, Malvern, PA, USA.

[10] Gregor, S. and Jones, D. 2007. The Anatomy of a Design Theory. Journal of the Association for Information Systems 8, 5, 312-335.

[11] Guba, E.G. 1981. Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. Educational Communications and Technology Journal (ECTJ) 29, 2, 75-91.

[12] Hafner, M. and Winter, R. 2005. Vorgehensmodell für das Management der unternehmensweiten Applikationsarchitektur. In Wirtschaftsinformatik 2005: eEconomy - eGovernment - eSociety, O. Ferstl Ed. Physica, Bamberg.

[13] Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. Design Science in Information Systems Research. MIS Quarterly 28, 1, 75-105.

[14] Holmström, J., Ketokivi, M., and Hameri, A.-P. 2009. Bridging Practice and Theory: A Design Science Approach. Decision Sciences 10, 1, 65-87.

[15] Kuechler, B. and Vaishnavi, V. 2008. On theory development in design science research: anatomy of a research project. European Journal of Information Systems 17, 489-504.

[16] Lee, A.S. and Baskerville, R.L. 2003. Generalizing Generalizability in Information Systems Research. Information Systems Research 14, 3, 221-243.

[17] Legner, C. and Heutschi, R. 2007. SOA Adoption in Practice - Findungs from Early SOA Implementations. In 15th European Conference on Information Systems (ECIS 2007), St. Gallen.

[18] March, S.T. and Smith, G.F. 1995. Design and natural science research on information technology. Decision Support Systems 15, 251-266.

[19] Merriam-Webster Online Dictionary 2010. knowledge, August 17, 2010, http://www.merriam-webster.com/dictionary/knowledge.

[20] Merton, R.K. 1968. Social Theory and Social Structure. The Free Press, New York.

[21] Offermann, P., Blom, S., Schönherr, M., and Bub, U. 2010. Artifact Types in Information Systems Design Science – A Literature Review. In Global Perspectives on Design Science Research Springer, Berlin, 77-92.

[22] Offermann, P. and Bub, U. 2009. A Method for Information Systems Development according to SOA. In AMCIS 2009 Proceedings, Paper 108.

[23] Purao, S., Baldwin, C., Hevner, A., Storey, V.C., Pries-Heje, J., Smith, B., and Zhu, Y. 2008. The Sciences of Design: Observations on an Emerging Field. Communications of the AIS 23, 523-546.

[24] Rohde, M., Stevens, G., Brödner, P., and Wulf, V. 2009. Towards a Paradigmatic Shift in IS: Designing for Social Practice. In DESRIST'09, Malvern, PA.

[25] Sowa, S., Tsinas, L., Lenz, H., and Gabriel, R. 2009. Integrated Information Security Risk Management - Merging Business and Process Focused Approaches. In Business Services: Konzepte, Technologien, Anwendungen - Band 1, H.R. Hansen, D. Karagiannis, and H.-G. Fill Eds., 327-336.

[26] Spiekermann, S., Meyer, B., Hertlein, M., and Lattke, T. 2009. skillMap - A Social Software For Knowledge Management - From Concept To Proof. In Business Services: Konzepte, Technologien, Anwendungen - Band 2, H.R. Hansen, D. Karagiannis, and H.-G. Fill Eds., 297-306.

[27] Travis, J. 1999. Exploring the Constructs of Evaluative Criteria for Interpretivist Research. In 10th Australasian Conference on Information Systems, 1037-1049.

[28] Vaishnavi, V. and Kuechler, W. 2004/5. Design Research in Information Systems, January 20, 2004, last updated August 16, 2009, http://desrist.org/design-research-in-information-systems.

[29] Vaishnavi, V.K. and Kuechler Jr., W. 2007. Design Science Research Methods and Patterns: Innovating Information and Communication Technology. Auerbach Publications.

[30] Van Aken, J.E. 2004. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. Journal of Management Studies 41, 2, 219-246.

[31] Venable, J.R. 2006. The Role of Theory and Theorising in Design Science Research. In DESRIST 2006, Claremont, CA.

[32] Walls, J.G., Widmeyer, G.R., and El Sawy, O.A. 1992. Building an Information System Design Theory for Vigilant EIS. Information Systems Research 3, 1, 36-59.

[33] Winter, R. 2008. Design science research in Europe. European Journal of Information Systems 17, 470-475.

Focusing on Values in Information Systems Development: A Critical Review of Three Methodological Frameworks

Fahri Yetim
University of Siegen,
Information Systems and New Media,
Hölderlinstr. 3, 57076 Siegen,Germany
Tel.: +49-(0)271-740-3019
Fahri.Yetim@uni-siegen.de

ABSTRACT
Values play a significant role in Information Systems Development (ISD). This paper presents a critical analysis of three methodological frameworks which aim at systematically considering values in the development of systems. The analysis focuses on their main goals, value concepts, and activities. In addition, this paper discusses some challenges and controversial issues with respect to the design for values and suggests an agenda for future research.

Categories and Subject Descriptors
D.2.10 [**Software Engineering**]: Design – *methodologies.* H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors.*

General Terms
Design, Human Factors, Theory.

Keywords
Value Sensitive Design, Information Systems Development, Design for Motivation, Methodology Review, Research Agenda.

1. INTRODUCTION
The term value has different meanings [24]. It stands not only for the material or monetary worth, but also for the importance or usefulness of something. Moreover, values in plural can also mean "principles or standards of behavior" or "one's judgment of what is important in life". People may find many things important, and what they value may influence their decisions when acquiring, designing, or using information systems: For buying a software system, decision makers have to believe that the system is worth being acquired, that it has value. For designing a system, designers need to make decisions which human values (e.g., privacy versus security) should be supported by /or embodied in the system. For using a system, users need to trust the system that it supports or does not violate their core values.

According to Klein and Hirschheim [16], values define the ultimate 'good' at which some design projects aim. A system's 'good' may refer to technical, economic, aesthetic standards or a variety of other social 'goods' such as equity, peace, conservation of natural resources, etc. Often design values are implicit, only partially defined and possibly not fully agreed upon by the stakeholders of a systems development project. Even so they will guide the design and implementation of an IS application. Values have individual, cultural and ethical dimensions, and differences in value preferences often involve conflicts and require a resolution or tradeoffs. Miller et al. [21] point out some consequences of unaddressed value tensions which range from lack of appropriation by disadvantaged groups to more severe consequences such as system sabotage. For example, values held by the group may conflict with those held by the individual, as in the case of open calendaring systems which have supported the group awareness of others' activities and presence at the expense of the individuals' privacy [25]. In addition, values supported by the system may be at odds with those promoted by the organization's culture and reward structure, as was the case with the failed adoption of Lotus Notes in a consulting firm whose organizational structure rewarded competition rather than collaboration [23]. Moreover, value tensions may lead to system sabotage, as in the case of the Virtual Kitchen system [15], which was designed to increase sociality among employees by linking several kitchens at the workplace with continuous video and audio. Stakeholders (presumably with privacy concerns) placed notes in front of cameras and, at times, completely disconnected the system [cf, 21]. All these examples illustrate that values matter to people and thus deserve a careful treatment in systems development projects.

In the field of Information Systems, values in relation to information technology have already gained explicit attention in different approaches [9,14,16,17,18,19,20,26,28,30,31]. They range from descriptive approaches, which aim at understanding the interaction between values and technologies, to design oriented approaches, which aim at purposively supporting human values through system design. Supporting values through design has emerged within different areas, including Computer Ethics, Social Informatics, Computer Supported Cooperative Work, and Participatory Design. These research strands envisioned an ideal world in which technologies not only promote instrumental values such as functional efficiency, reliability, and ease of use, but also the substantive social, moral, and political values, such as privacy,

10th International Conference on Wirtschaftsinformatik,
16th - 18th February 2011, Zurich, Switzerland

justice, and autonomy [8,9]. Thus, the goal has been to design systems that embody values to which designers, users, other stakeholders, and the surrounding society are committed. Yet, putting these ideals into practice and designing values in mind is not straight forward. Taking values into consideration during design requires incorporating diverse and frequently far-flung areas of knowledge and know-how into the design processes that are not normally conceived as elements of the design toolkit [8]. There is a need for explicit guidelines, or methodologies, for reliably embodying values in information systems.

Yet, only a few approaches aim to provide methodological guidelines for systematically identifying and accounting for values in the development of systems. The goal of this paper is to critically review three approaches and analyze their basic concepts and activities. The three approaches are selected as they have the following common characteristics: First, they seek to be proactive to influence the design of technology early in and throughout the design process. Second, they enlarge the arena in which values arise to include not only the work place, but also the education, the home, commerce, online communities, and the public life. Third, they contribute to the development of a methodology.

The analysis of the approaches seeks to understand their basic building blocks, that is, (a) what types of values are considered, and (b) what methodological steps or value-specific activities are suggested. Based on the results, I will also reflect on the commonalities and differences of these approaches. Finally, I will discuss some significant controversial as well as open issues when designing with values in mind and will suggest a research agenda. In this way, this paper contributes to the current discussion on value sensitive agenda within the information systems field. The critical review may be of value for researchers, who may take up some research gaps and advance the research on values one step further, as well as for practitioners, who may use the analysis for making an informed choice among available approaches.

2. THREE APPROACHES

2.1 VSD Framework

Friedman et al [12] defined *Value-Sensitive Design* (VSD) as an approach to the design of technology that accounts for human values in a "principled and comprehensive" manner "throughout the design process." VSD follows a common use of the term value wherein a value refers to what a person or group of people consider important in life. It assumes that certain values are universally held, although the way in how such values play out in a particular culture at a particular point in time can vary considerably. VSD distinguishes between values of ethical import and stakeholders' values. Some values of ethical import are explicitly supported in system design (e.g. fairness, accountability, democracy) and embedded in the product, independent form whether all stakeholders uphold them or not. In addition, VSD considers also stakeholders' values which are important to some but not necessarily to all of the stakeholders (e.g., environmental sustainability and walkable neighborhoods).

VSD offers a three-part framework in which conceptual, empirical, and technical investigations are applied iteratively. *Conceptual investigations* comprise philosophically informed analyses of the central constructs and issues under investigation. For example, how does philosophical literature conceptualize certain values (e.g. trust, privacy, informed consent)? Who are the direct and indirect stakeholders affected by the design at hand? How should we engage in trade-offs among competing values in the design, implementation, and use of information systems (e.g., autonomy vs. security, or anonymity vs. trust)? Value Sensitive Design takes up these questions under the rubric of conceptual investigations.

Empirical investigations may focus on the analysis of the social context in which the technical artifact is situated as well as on the evaluation of a particular design. Thus, the entire range of quantitative and qualitative methods used in social science research is potentially applicable here, including observations, interviews, surveys, experimental manipulations, collection of relevant documents, and measurements of user behavior and human physiology. Empirical investigations can focus, for example, on questions such as: How do stakeholders apprehend individual values in the interactive context? How do they prioritize competing values in design trade-offs? Are there differences between espoused practices (what people say) compared with actual practice (what people do)? How do organizations appropriate value considerations in the design process (for example, what are the organizations' motivations, reward structures, and economic incentives?)?

Technical investigations focus on the analysis of how technology performs and in what way its design supports values. VSD adopts the position that technologies provide a value suitability that follows from the properties of the technology. That is, a given technology is more suitable for certain activities and thus more readily supports certain values while rendering other activities and values more difficult to realize. In one form, technical investigations focus on how existing technological properties and underlying mechanisms support or hinder human values. In the second form, technical investigations involve the proactive design of systems to support values identified in the conceptual investigation.

To date, VSD is being applied in a wide range of research and design contexts. For example, the web browser case study [10] began with a conceptual investigation of the value of *informed consent* by drawing on diverse literature. With a conceptualization for informed consent in hand, they conducted a retrospective analysis of existing technical mechanisms such as the cookies and web-browser technology and redesigned the browser. In another case dealing with the design of simulation software UrbanSim for supporting urban planning, Borning et al. [1] started with conceptual investigations and distinguished between moral values such as fairness, accountability, democracy and stakeholder values such as environmental sustainability and walkable neighborhoods. As part of supporting the democratic process, they decided that the model should allow different stakeholders to articulate the values that are most important to them, and evaluate the alternatives in light of these values. Other case studies explore different sets of values and illustrate other ways to employ the VSD methodology [11, 21].

There are different ways to enter into a VSD process. Friedman et al [12] provide some guidelines for practicing VSD, suggesting: (1) to start with a value, technology, or context of use; (2) to identify direct and indirect stakeholders; (3) to identify harms and benefits for each stakeholder group; (4) to map harms and benefits onto corresponding values; (5) to conduct a conceptual investigation of key values; (6) to identify potential value

conflicts; and (7) to integrate value considerations into one's organizational structure. In addition, they suggest heuristics for both interviewing stakeholders and technical investigations.

2.2 VAP Framework

Another methodological framework that aims at fostering value integration into the design process is the so called Values At Play (VAP) methodology [6,7,8]. The methodology has so far been applied in the context of game design, particularly in the case of RAPUNSEL, which was designed to promote interest and competence in computer programming among girls. Here, like in VSD, diverse sets of values, including ethical, social and political values are in focus. The main hypothesis is that value can be integrated into the design and - like in VSD, VAP also argues for considering philosophical, empirical and technical mode of investigations for considering values in design. In contrast, the VAP methodology consists of three main activities for systematically incorporating values in the design process: *discovery, translation, and verification*. They are meant to be followed iteratively.

(1) Discovery is the activity in which designers identify the values that are relevant to or inform a design project. Values can be identified in the initial stages of a given project as well as each iterative stage of development. There are several sources where designers and researchers seek for values, which are for example, the explicitly stated project goals, the hypotheses generated by the team to achieve those goals, the values expressed in prior empirical work, including related technical systems, values present in the design environment (academia, commercial, activist, etc.) and values held by individual members of the design team [7]. In the context of the RAPUNSEL project, for example, Flanagan and Nissenbaum [7] started with a preliminary list of relevant values (Cooperation, Creativity, Gender Equity, and Authorship). They also identified values (such as equity, empowerment) expressed in the purpose of the RAPUNSEL project, which were formulated "to address gender inequities". Values in project goals tend to be 'higher-order' values and are perceived as ends in themselves. Other values emerged when specifying design features, for example, in the RAPUNSEL game, designers opted for a reward system for reinforcing cooperation by providing rewards for sharing. Players gained status by sharing and earning points. Further emerging values are the values of designers. In RAPUNSEL, "diversity" was important to design a team which was then included in the list of explicit values. To team members, diversity meant expanding the general activity of programming across boundaries of age, gender, and ethnicity, and also fostering a diverse range of approaches to learning. Other obvious sources of values are users or other stakeholders. Finally, social and political values generate background constraints, for example, concerning issues such as how much privacy a log-in system offers, what is shared publicly, etc.

(2) Translation is the second main activity in which designers "translate" the relevant values in system design. It comprises three sub-activities: Operationalization, implementation, and resolving value-conflicts. *Operationalization* of values involves articulating value concepts – which are often understood only in abstract terms – in operationally accessible, practical terms, in order to relate them to design features. For example, in RAPUNSEL, the designers opted for defining the value of "Gender Equity" in practical terms as 'girl friendly' features and designed along the lines of partner conversations and prior research. They included a chat system since studies had shown that teenage girls are deeply engaged in instant messaging and chat as a means for higher levels of computer use. Cooperation, as another value, was translated through development of robust mechanisms for sharing program code. Operationalizing values require a jump from 'concept' to 'feature'. Flanagan and Nissenbaum [7] state that: "The leap between the ideal value and the feature could sometimes seem like a leap of faith." (p. 185).

Implementation transforms the operationalized values into concrete design specifications and then to lines of codes.

Resolution of value-conflicts is another complex sub-process. Flanagan and Nissenbaum [7] considered two key strategies: "dissolving conflict" and "values trade-off". That is, designers either seek ways, through creative re-design, to satisfy both values simultaneously, or decide to trade one value off in favor of the other. For example, conflicts with respect to using or not using sexualized female characters may be dissolved by avoiding human characters and using animals and abstract characters. In contrast, in values trade-offs offering sexualized characters to attract the players to an educational game might be considered preferable to their not playing the game at all. Nevertheless, resolving values remains, in general, a difficult problem.

(3) Verification is the third main activity in which designers assess whether and to what extent they have successfully implemented target values in a given system. Here, several qualitative and quantitative methods can be employed to explore diverse modes of verification, including critical reflection and analysis, testing and user studies in controlled settings, formal and informal interviews, pre- and post attitudinal surveys, etc.

2.3 WCD Framework

Cockton [2] proposed an initial framework for *Value-Centered Design* (VCD), which was later renamed to *Worth-Centered Design* (WCD) [3]. He argues that *value* should not be understood only in commercial or moral terms and preferred to use the term *worth* to articulate the focus on development of the *worthwhile*, i.e. things that will be valued. This approach is not limited to human values as countable nouns – as in VSD. Rather, it is based on the sense of value which is an uncountable noun. According to Cockton, creating a (nameless) *value* (i.e., something worthwhile) is different from creating *values* (something to believe in). Not only can values as beliefs be worthwhile, but also other things. WCD focuses the development on things that are valued or *worthwhile*.

Cockton further states that worth is a motivator and that designing worth means to design things that will *motivate* people to buy, learn, use or recommend an interactive product. The motivations of individuals and social groupings reveal what is worthwhile (and thus valuable). WCD involves the following phases [3]:

(1) Worth as a Requirement: The emphasis on worth within WCD has moved from the simple expression of the intended value to add the elicitation of what individuals and groups consider to be worthwhile. This involves a more focused study of needs, wants, and unfelt needs. They can be identified through existing approaches such as ethnography, interviews and prototyping. In addition, cultural probes can be used to expose values. Prototyping is viewed as vital to validate 'worth as requirements' in a timely and reliable manner. According to Cockton [3],

wherever possible, worth should be expressed using words and images of users and other stakeholders.

(2) Worth Design: This second process involves – as a key activity – the building of prototypes, in order to assess earlier whether a new design can deliver anything worthwhile. Cockton [2] suggests as much evaluation planning as possible prior to any design to provide evaluation criteria that can be used to compare design alternatives in the design process. In addition, he suggests writing *worth delivery scenarios,* which relate envisioned design features to the delivery of something worthwhile for all stakeholders. For example, for a university web-site, value or worth delivery scenarios would explain how a proposed design would deliver *appropriate, adequate and effective help with choice of course and university*, and how this in turn would achieve high levels of student recruitment"[2, p.1294]. According to Cockton [3], creating a Worth Map is a useful method to express complete means-end chains from design elements to human elements, by creatively re-expressing ideas about technical possibilities and about what motivates people (because it is of value).

(3) Evaluation: WCD focuses its evaluation on assessing the impact of user experience and the performance on achievement of intended worth. Worth is achieved in the world and endures after interaction. One exception is transient individual worth in the form of pleasure in entertainment systems, which must be measured during interaction. In other cases, impact must be assessed in the world, which requires a broader range of measures and instruments that may well have to be embedded in the system itself.

(4) Iteration: This is the fourth separate process, which aims to improve negative impacts on worth by revisiting and repeating any other process. Iteration is considered as a distinct process within, rather than an overall attribute, of systems development. Iteration requires the involvement of the whole project team, and not just evaluators, since everything can be iterated and everyone must iterate.

Finally, regarding the application of the WCD framework, it should be noted that a tried and tested WCD framework has not been developed yet. The works published so far focus mainly on theoretical discussions and the illustration of the framework with the help of imagined use scenarios.

2.4 Summary and Comparison of the Approaches

As summarized in Table 1, the three approaches described so far consider values or worth as *ends* of design and add new activities to existing development methodologies. There are some commonalities as well as differences between these approaches with respect to the value concepts and methodological steps.

Table 1. Summary of the Approaches

Approach	Goals & Values	Activities/Methods
VSD (Friedman et al [11,12])	Integrating ethical values and design/ considering a set of ethical values	*Conceptual investigations* Conceptual investigation of values; Identifying direct and indirect stakeholders; Identifying benefits and harms for each stakeholder group; Mapping benefits and harms onto corresponding values. *Empirical Investigations* Analysis of the social context in which the technical artifact is situated; How stakeholders apprehend individual values; How they prioritize competing values in design trade-offs; Understanding differences between espoused practice (what people say) compared to the actual practice (what people do); How organizations appropriate value considerations in the design process; Evaluation of a particular design. *Technical Investigations* Proactive design of systems to support values identified in the conceptual investigation; Analysis of how technological properties and underlying mechanisms support or hinder human values.
VAP (Flanagan et al. [6,7,8])	Integrating values and design/ considering a set of social and ethical values	*Discovery of values* Creating a list of values from sources including: Explicitly stated project goals, prior empirical work, related technical systems, application environment, design team, prototyping and user testing. *Translation of values* Operationalization, Implementation and Resolving of Value- Conflicts. *Verifying values* Checking if the desired values are embedded in the system.
VCD (Cockton [2,3,4])	Designing worthwhile systems/ Users and designers values	*Worth as Requirement* Identification of needs and wants *Worth Design* Creating worth delivery scenarios, Prototypes *Evaluation* Value impact analysis. *Iteration* Repetition of any other process

Concerning the value orientation the VSD and VAP in the core are focused on human values of ethical import and also consider other social and individual values. For example, VSD starts with values of ethical import, independent from whether all stakeholders uphold them, and then extends considerations to other stakeholders' values, that is, to things that some stakeholders value irrespective of moral obligations. This means that VSD and VAP are more concrete with respect to relevant human values and expand out from an initial focus on ethical values, In contrast, WCD is not rooted in moral considerations and has a more open genesis. It starts with the *worthwhile*, that is, whatever some people value somewhere, individually or collectively, irrespective of ethics, or the approval of others. This means that WCD is more abstract with respect to the value's manifestations in the world by the adoption of a neutral word "worth", which makes WCD also broader than VSD and VAP with respect to their consideration of relevant outcomes.

Concerning the methodological steps or activities, both the VSD and VAP frameworks emphasize the relevance of empirical, philosophical, and technical modes of inquiry to the sound inclusion of values in design. Whereas the VAP framework organizes the main questions and activities under the rubric of identification, translation and verification, VSD presents them under the rubric of conceptual, empirical, and technical investigations. Conceptual investigations include analyses of the values and potential value tensions. Empirical investigations involve assessing the stakeholders' experience of the value-oriented features of a system. Technical investigations are concerned with the design of a system. It should also be mentioned that the VSD framework has been criticized by Le Dantec et al [19], arguing that it does not prescribe a unique perspective on the design process (which is largely left open ended). In addition, as several applications of VSD start with a list of "values of ethical import" identified through conceptual investigations, this gives rise to a further critique that VSD privileges known values over the discovery of values present in the situated context. For the discovery of values, so the argument, empirical investigations need to come at the beginning of the investigation. On the other hand, one may be content with the fact that VSD is open for different possibilities to enter into value sensitive activities [31]: one may start with a list of values, or with contexts, or technology. Moreover, VSD and VAP do not aim to substitute other system development methodologies but rather focus on value related activities which in practice have to be integrated in the chosen development methodology.

In contrast to VSD and VAP, WCD is a rather neutral approach, which may start by brainstorming and which might be receptive to all ideas about technical possibilities, and about what motivates people, because it is of value [4]. After identifying what is valued by the users (e.g., by ethnography or interviewing users), the developers can take over an active role in developing a system that creates worth. In addition, the focus on worth in WCD shifts the attention to investigating the *impact,* i.e., to the assessment of the achieved value/worth in the world.

Finally, concerning the applications of the three approaches, it should be mentioned that the application of VSD has so far been illustrated in different contexts, whereas the application of VAP remained limited to game design cases. .In contrast, WCD is underdeveloped and has not been tested in any application yet.

3. RESEARCH AGENDA: CHALLENGING AND CONTROVERSIAL ISSUES

So far I have described three approaches and discussed their commonalities and differences. The approaches focus on many aspects and issues of a value or worth centered design. Yet, there are still open issues with respect to the elicitation, expression and validation of worth/value. In this final section, I will reflect on some of the challenging and controversial issues, in order to point out some relevant future research areas that need explicit attention to move the research on value/worth centered design one step further.

The first challenging and controversial issue is **how to start**, i.e. which investigation should be conducted first. Should we start by considering a set of relevant values in the design and then conduct conceptual investigations (literature analysis) to understand the chosen value concepts? Or should we start with empirical investigations in the use context to identify/discover local values and express those using local terms? Should we start with a predefined classification of values or should the classification of values be divined from the empirical work? Some authors see virtue in both, drawing conceptual clarity and normative justification from theoretical works in moral and political philosophy, while supplementing these with knowledge about actual interpretations and value commitments of populations relevant to the technologies under study [8]. In contrast, Le Dantec at al. [19] argue for starting with an empirical understanding of local values and expressing them by using local terms. Yet, they also acknowledge that values of ethical import can be used as an analytic tool with respect to the locally expressed values: This means that empirical investigation can shape the understanding of values and the conceptual investigation may become a tool through which the designer can reflectively evaluate the values presented through the empirical investigation.

A second challenging issue is **where to stop**, i.e., the justification of the boundaries. A key aspect of value or worth centered design is its focus on direct and indirect stakeholders. Although there are methods for involving stakeholders, ordinary citizens, or their representatives (e.g., focus groups, public forums, online discussion groups, and open calls for participation), there still remains the difficulty of determining where to cut off ever-broadening circles of involvement in the public discourse: the local community, the country or the whole world? For example, a web-based system designed for one culture or society can also be used by other cultures. Should other cultures articulate their interests and value orientations? Even though we carefully considered as many aspects of the situation as possible, the problem is always where do you stop? At some point, an exclusionary judgment must be made about who should participate in any particular discussion for the desired discussion to occur or what values should be focused on [31]. Critical researchers advocate for reflecting on boundary issues. For Ulrich [27], boundary judgments determine which facts and value considerations count as relevant, and, conversely, value judgments drive the definition of boundaries. Hence, value and boundary judgments are mutually dependent, and critique should focus on both.

The third challenging issue is **how to justify** the selection of values, i.e., who should decide and in what kind of decision

process. When tensions between values of designers and users emerge, whose values take precedence? As also noted by Flanagan et al [8], the commitment to specific values, if it is to rise above dogma, requires justification, i.e. explanation of why these values are important in the broader social and political context and, also, why and how they are relevant to the design project at hand. The philosophical mode inquiry can contribute to this effort by articulating the rationale behind, or the justification for, commitments to particular values in a given system. Traditional moral and political theories are a source of explanation, and when conflicts among values result from a specific design choice, they may guide sound resolutions, or reasonable tradeoffs. Yet, from a discourse-ethical perspective [13], those affected need to participate and reflect on which values should or should not be promoted. Hence justification of values from this perspective requires discourse mechanisms [29, 30, 31]. In such discourses, designers of a patient record system, for example, may draw on moral and political theories to persuade hospital administrators that privacy is important, or necessary, and should be protected even if the cost rises as a result.

The fourth challenging issue is **how to represent** values, in order to communicate them in design discourses. A sound grasp of value terms is one of the necessary links between values and specific design features [8]. Concrete definitions would relieve the burden on designers, allowing them to draw on existing ones. The representation of values is important as such representations cannot only include abstract and concrete definitions, but also pro and contra arguments for or against a value as well as links to related design features. The explicit representation of values would enable the transparency and deliberation on them. An approach to represent values can be guided by the research on design rationale to provide a kind of value rationale for design, i.e. a template, which can evolve in design discourses through critiques, comments, additions, and revisions of the design participants.

The fifth challenging issue is **how to resolve** value conflicts. The challenge is that value tensions cannot only be between values of stakeholders, but also between values of users and values embedded in IT, as well as between values and other design goals. In addition, when group member values diverge from the general IT, values held by the group members, or the values embedded in a technology diverge from the general IT values held by a group, then there will be conflicts associated with the introduction and use of the technology [20]. There are only a few practical methods for addressing tensions among a diverse types of values as they unfold during the design and deployment process [7, 12]. According to Flanagan et al. [8], where practice requires decision, even in the absence of philosophical resolution, a sound alternative is to turn to empirical investigation of relevant populations, ascertaining their commitments and preferences through such mechanisms as surveys, interviews, testing under controlled conditions, and observations in the field. Yet, to resolve value conflicts remains a difficult task. Moreover, the value system of users may change over time and may need to adapt the system to meet their values and may involve new value conflicts. Hence, the relevant research issues are: what new methods can be envisioned to help designers to deal with different types of value tensions in a principled way? And how should we deal with changes and the values of unforeseen users?

The sixth challenge concerns the issue of **how to verify** or evaluate the inclusion of values in design. Flanagan et al. [7, 8] note that verifying the inclusion of values introduces additional complexity: first, not only the successful implementation of a value in a specific component is of relevance, but also whether its implementation does not detract from other design goals. Second, it is not easy to grasp what it means for a value to be implemented in a system (e.g., claiming that a system is 'privacy-preserving' or autonomy enhancing'). This difficulty arises partly from the less concrete nature of value concepts and partly because the means by which values are embodied are often more diverse. Third, although values may be related to specific system features, they may also emerge, indirectly, as a property of the system's interaction with the contextual setting in which it operates. A final complexity involves the fact that the impact of some values may be experienced immediately, while others may emerge only in the long term. Therefore, we can conclude – in line with Flanagan et al. [8], that the verification phase of a project is likely to produce only partial results.

The seventh challenge deals with **how to educate** designers to enable them to deal with complex value issues in design. This issue is relevant, because developing systems with attention to value designers need to engage simultaneously with distinct areas of knowledge and their respective methodologies. Flanagan et al. [8] describe this challenge as follows: "Design and engineering projects must incorporate contextual knowledge about values and, where such knowledge is not readily available; designers will need to grapple directly with questions about the relevant values. Not only does this lie outside the usual boundaries of engineering expertise but is attainable through modes of inquiry, such as systematic analysis of values, unfamiliar in the technical and scientific environments. Achieving technical design that soundly incorporates values requires not only competence in the technical arts and sciences, but also a reflective understanding of the relevant values and how these values function in the lives of people and possible groups affected by the systems in question. Within the academy, systematic reflection on values generally takes place in humanistic areas, such as moral and political philosophy, as well as in empirical and theoretical social sciences" (p.324). Hence, there is a challenge concerning the qualification of designers, i.e., enabling them to engage actively with scientific and technical results, absorb relevant philosophical reflections on values and also to consider the results of the empirical investigation of values in relation to the individual and his/her societies.

Finally, it should also be mentioned that to focus on the worthwhile and to take a motivational perspective is very promising. In fact, there are some research efforts that provide theoretical guidance to the design of interactive systems [32] or aims to develop methodological frameworks for the design of motivating systems [5]. A motivational perspective may function as a framework to unite various design approaches (such as cognitive or usability centric, affective and emotional, or value centered design) to represent a holistic picture of issues in information systems development and use.

4. CONCLUSION

In this paper we started with the assumption that values matter to people and that unaddressed value tensions may have negative consequences on the implementation and use of information

systems. Then, we critically reviewed three approaches that aim to systematically consider values in the development of systems. The purpose of the review is to encourage a debate about the methodological frameworks, so that they may mature through the crucible of discourse within the IS community. Focusing on values in ISD faces several further challenging and controversial issues, as discussed briefly. These issues provide further contexts and starting points in a future research agenda for moving value-based ISD forward.

The critical review of three approaches indicates that, despite the differences concerning the views on values, the "value of value" has been already acknowledged in ISD. The approaches reviewed do not aim to substitute existing ISD methodologies. Rather they make incremental contributions toward a value sensitive ISD, by developing value-related activities and methods, including methods for identifying, designing and evaluating values. They need to be integrated in the existing ISD methodologies. The review also makes clear that focusing on values throughout the ISD involves the framing of requirements in terms of intended value or worth, creative designing in terms of envisaged value/worth and evaluating in terms of achieved value/worth. Moreover, in all these activities, there must be a sufficient openness/flexibility to allow the extensive co-creation of value/worth and the appropriation by a wide range of stakeholders.

The challenging and controversial issues mentioned call for a particular attention, i.e. to provide more guidance on effective empirical instruments to identify values as well as to developing tools and methods applicable in design practice. Tools are needed to enable all stakeholders to articulate and reflect on their values, to relate them to design goals and features, and to communicate them in design discourses. Using such tools throughout the ISD would allow for the integration of value considerations into the full range of existing and emerging ISD practices and would support the creation of desirable systems for future users.

5. ACKNOWLEDGMENTS

This work has been partly supported by the EU-funded project INSEMTIVES - Incentives for Semantics (www.insemtives.eu, FP7-ICT-2007-3, Contract Number 231181).

6. REFERENCES

[1] Borning, A.; Friedman, B.; Davis, J. and Lin, P. (2005). Informing public deliberation: Value sensitive design of indicators for a large-scale urban simulation. In *Proc. ECSCW 2005*, 449-468.

[2] Cockton, G., "A development framework for Value-Centred Design", Proceedings of the *CHI 2005 Extended Abstracts*, ACM Press, 2005, pp. 1292-1295.

[3] Cockton, G., "Designing worth is worth designing", Proceedings of *NordiCHI Conference*, ACM Press, 2006, pp. 165-174.

[4] Cockton, G., "Designing worth – connecting preferred means to desired ends," *Interactions*, 2008, ACM, July+August, pp. 54-57.

[5] Cuel, R.; Morozova, O.; Rohde, M.; Simperl, E.; Siorpaes, K.; Tokarchuk, O.; Wiedenhöfer, T.; Yetim, F.; Zamarian, M. (2010). Motivation Mechanims for Participation in Human-driven Semantic Content Creation. *International Journal of Knowledge Engineering and Data Mining*,*(Forthcoming).*

[6] Flanagan, M, Howe, D. C., & Nissenbaum, H. (2005). Values at Play: Design Tradeoffs in Socially-Oriented Game Design. *Proceedings of CHI 2005.* New York: ACM Press, 751-760.

[7] Flanagan, M. & Nissenbaum, H. (2007). A game design methodology to incorporate social activist themes. Proceedings of CHI 2007. New York: ACM Press, 181–190.

[8] Flanagan, M.; Howe, D.; and Nissenbaum, N. (2008). "Embodying Values in Technology: Theory and Practice," In: Information Technology and Moral Philosophy, Jeroen van den Hoven and John Weckert (eds.) Cambridge: Cambridge University Press, 2008, 322-353.

[9] Friedman, B. (1997)(ed.). *Human Values and the Design of Computer Technology.* Cambridge University Press and CSLI, New York, NY and Stanford, CA, 1997.

[10] Friedman, B., Howe, D. C., & Felten, E. (2002).Informed consent in the Mozilla browser: Implementing Value-Sensitive Design.*Proceedings of the Thirty-Fifth Annual Hawai'i International Conference on System Sciences*.

[11] Friedman, B.; Kahn, P.; Hagman, J.; Severson, R. And Gill, B. (2006a). The watcher and the watched: Social judgments about privacy in a public place. *Human-Computer Interaction 21(*2) 2006, 233-269

[12] Friedman, B.; Kahn, P.; and Borning, A (2006b). Value Sensitive Design and Information Systems. In P. Zhang & D. Galletta (eds.), *Human-Computer Interaction and Management Information Systems: Foundations*. M.E. Sharpe, New York, 348-372.

[13] Habermas, J. (1990). *Moral Consciousness and Communicative Action* Polity Press, Cambridge.

[14] Hirschheim R., & Klein, H. K. (1994). Realizing Emancipatory Principles in Information Systems Development: The Case for ETHICS. *MIS Quarterly,* 18**,** 83-109.

[15] Jancke, G., Venolia, G. D., Grudin, J., Cadiz, J., and Gupta, A. Linking public spaces: technical and social issues. In *Proc. of CHI* (Seattle, WA, USA, Mar. 31 - Apr. 5, 2001). ACM, N.Y., N.Y, 2001, 530-537.

[16] Klein, H. and Hirschheim, R. (2001) Choosing Between Competing Design Ideals in Information Systems Development. *Information Systems Frontiers* 2, 75-90.

[17] Kling, R., (1978), "Value-Conflicts and Social Choice in Electronic Funds Transfer Developments", Communications of the ACM, 21(8), 1978, pp 642-657.

[18] Kujala, S. and Väänänen-Vainio-Mattila, K. (2009). Value of Information Systems and Products: Understanding the Users' Perspective and Values. JITTA: Journal of Information Technology Theory & Application, 9, 4, 23-39.

[19] Le Dantec, C.A.; Poole, E.S. & Wyche, S.P. (2009). Values as Lived Experience: Evolving Value Sensitive Design in Support of Value Discovery. Proceedings of the CHI, April 7th, 2009, Boston, MA, USA, pp. 1141-1150

[20] Leidner, D.E. and T. Kayworth (2006). "Review: A Review of Culture in Information Systems Research: Toward a Theory of Information Technology Culture Conflict," *MIS Quarterly* (30)2, pp. 357-399.

[21] Miller, J.; Friedman, B.; Jancke, G. and Gill, B. (2007). Value tensions in design: The value sensitive design, development, and appropriation of a corporation's groupware system. In *Proc. of GROUP 2007,* 281-290.

[22] Mumford, E. (1983). Designing Human Systems: The ETHICS Method. Manchester Business School, Manchester, United Kingdom.

[23] Orlikowski, W. J. Learning from Notes: Organizational Issues in Groupware Implementation. In *Proc. of the 1992 ACM Conf. on CSCW* (Toronto, Canada, Oct. 3 - Nov. 4, 1992). N.Y., N.Y.: ACM, 1992.

[24] Oxford Dictionaries Online. Available at: http://oxforddictionaries.com

[25] Palen, L. Social, individual, and technological issues for groupware calendar systems. In *Proc. of CHI* (Pittsburgh, PA, USA, May 15-20, 1999). ACM, N.Y., N.Y., 1999, 17-24.

[26] Sellen, A.; Rogers, Y.; Harper, R. and Rodden T. (2009). Reflecting Human Values in the Digital Age. *Communications of the ACM* 52(3) 2009.

[27] Ulrich, W. (2000). Reflective practice in the Civil Society: the contribution of critically systemic thinking. *Reflective Practice* (1:2) 2000, 247-268.

[28] Yetim, F. (1998). Interkulturalität und Informatische Gestaltung – Eine Interdisziplinäre Annäherung. *Informatik-Spektrum* 21 (1998), 203-212.

[29] Yetim, F. (2006). Acting with Genres: Discursive-Ethical Concepts for Reflecting on and Legitimating Genres. *European Journal of Information Systems* 15(1), 54-69.

[30] Yetim, F. (2010). Taking Universal Perspective in Design: A Plea for two reflective Principles and Mechanisms. *Hawaii International Conference on System Sciences (HICSS-43)*, Kauai, Hawaii

[31] Yetim, F. (2011). Bringing Discourse Ethics to Value Sensitive Design: Pathways Toward a Deliberative Future. Submitted to: *AIS Transactions on Human-Computer Interaction.*

[32] Zhang, Ping (2008). Motivational affordances: Fundamental reasons for ICT design and use. *Communications of the ACM*, 51(11), 145-147.

Theoretical and Experimental Insights into Decentralized Combinatorial Auctions

Georg Ziegler
Technische Universität München
Boltzmannstr. 3
85748 Garching, Germany
zieglerg@in.tum.de

Tobias Scheffel
Technische Universität München
Boltzmannstr. 3
85748 Garching, Germany
scheffel@in.tum.de

ABSTRACT

Combinatorial Auctions (CAs) are promising to increase social welfare by enabling bidders to express their valuation on any combination of items. A major issue of many CAs is the requirement to optimally solve the $\mathcal{NP}$-hard Combinatorial Allocation Problem. To release a centralized auctioneer from that computational burden he can shift it to the bidders. One of the few discussed decentralized auctions is PAUSE, in which bidders suggest new allocations to the auctioneer. In our theoretical analysis we examine the bidders' bid complexity and determine a worst case bound concerning efficiency, if bidders follow a profit maximizing strategy. Based on these results we conduct computational experiments with different bidding and computation strategies, and analyze their impact on efficiency, auctioneer's revenue and auction runtime. Surprisingly, even if agents deviate from the optimal bid price calculation, PAUSE still achieves high levels of efficiency and auctioneer's revenue compared to the Combinatorial Clock auction.

Keywords

combinatorial auctions, bidding agents, computational experiments

1. INTRODUCTION

The Internet allows for the exchange of complex preference profiles and laid the foundation for the design of new market mechanisms. The promise of these mechanisms is to increase economic welfare by allowing market participants to reveal more comprehensive information about cost structures or utility functions. In the last decade, a growing literature in management science and information systems is devoted to the design of such smart markets [**?**, **?**], with combinatorial auctions (CAs) emerging as a pivotal example [**?**]. Allocation of spectrum licenses for wireless communication services [**?**], transportation [**?**] and industrial procurement [**?**, **?**] are not nearly all domains in which there is an increasing requirement and usage for multi-item auction mechanisms allowing complex bids. However, the auctioneer of a CA faces the $\mathcal{NP}$-hard Combinatorial Allocation Problem (CAP) [**?**], for which an optimal solution is untraceable in larger instances, although the availability of computing power grows.

Decentralizing the CAP is the approach of the ***P**rogressive **A**daptive **U**ser **S**election **E**nvironment* (PAUSE) auction proposed by [**?**]. In PAUSE bidders submit not only their own bids, the desired packages of items and the price, but have to propose a new allocation including their new bids and existing bids, being better than the current provisional allocation. Checking bid validity and publishing accepted bids remains the auctioneer's only tasks. Another simplification for the auctioneer is that there is no need for a price calculation mechanism in the iterative process like in most other iterative CAs.

There is only little work in the literature on decentralized auctions, therefore, we study PAUSE theoretically and experimentally. Our theoretical analysis shows the growing complexity for the bidders in PAUSE and gives a worst case bound concerning efficiency, if bidders follow a certain strategy. The determination of a lower bound in CAs has to our knowledge not been done and published yet, but it reveals important insights what can go wrong concerning bidder behavior, value models and auction rules. In this context we analyze PAUSE with computational experiments.

[**?**, **?**] developed some sophisticated bidding strategies for distributed auctions, however, in our experiments we focus on more simple strategies in which bidders reveal as little as possible about their valuations. Further, we use another value model with more items, in which the advantage of not having to calculate the CAP optimally is more decisive. To compare and benchmark we run computational experiments with the Combinatorial Clock auction (CC), which is a centralized CA and known for its sparse need of solving the CAP [**?**].

2. THE PAUSE AUCTION

PAUSE especially concentrates on achieving the following properties: It should permit bidders to submit any combinatorial bid they choose (*fully combinatorial*) and allow losing bidders to clearly see why they lost (transparent). Furthermore it should allow the auctioneer to determine the winner easily for auctions of any size and achieve high auctioneer payoffs. The basic idea of PAUSE is to place the computational burden of evaluating synergies on the bidder claiming

10^{th} International Conference on Wirtschaftsinformatik,
$16^{th} - 18^{th}$ February 2011, Zurich, Switzerland

those synergies. This leaves the auctioneer to simply check that a bid is valid. He no longer faces the CAP. PAUSE is a multi-round, multi-stage CA decentralizing the CAP [?]. A PAUSE auction with m items has m stages.

Stage 1 consists of a *Simultaneous Ascending Auction* (SAA) [?] on all items. During this stage bidders can only place individual bids on items - no package bidding is allowed. The stage ends when bidding ends and the auctioneer determines the provisional allocation by simply choosing the best bid on every item.

In each round of a successive stage $h = 2, 3, \ldots, m$ a bidder is required to submit a composite bid (denoted by X^{CB}), which covers all items and includes only disjoint package bids each of maximum cardinality of h. Bidders are allowed to use bids that other agents have placed in previous rounds. The bid price $p(X^{CB})$ of a composite bid is the sum of its package bid prices. For each new package bid in a composite bid, the bidder has to outbid the currently winning composite bid by the minimum increment ϵ. After each round the auctioneer declares the highest composite bid as the provisional allocation and registers the highest submitted package bids in the database. A stage ends when bidding finishes. At the end of each stage h, all agents know the best bid for every subset of size h or less so far.

3. THEORETICAL RESULTS

For our theoretical analysis we assume bidders follow a *straightforward* strategy, by bidding on the package which yields the highest possible payoff at current prices. Since there is no known equilibrium bidding strategy in PAUSE this assumption is justified by the typical use in game theoretical analysis and as it seems natural since bidders reveal as little information as possible keeping the chance for high profits. We assume further that the straightforward bidders do not consider a combination of their package bids, since they are able to bid on those combinations in a single package bid in later stages, thus avoiding a possible exposure problem, which would leave a bidder winning a package of items at prices he is not willing pay.

Let $\mathcal{K} = \{1, ..., m\}$ denote the set of items and $\mathcal{I} = \{1, ..., n\}$ the set of bidders indexed by i. In general, bidders have different valuations for packages $S \subseteq \mathcal{K}$. Let $v_i(S) > 0$ indicate the valuation of bidder i for package $S \subseteq \mathcal{K}$. Each bidder i has a demand set $\mathcal{D}_{i,h} := \{S : v_i(S) \geq p_j(S), i \neq j \wedge |S| \leq h\}$, i.e. it contains all packages S for which bidder i has a higher valuation than the price of the current highest bid from another bidder j ($p_j(S)$) and the cardinality of S must not be greater than h. If bidders want to determine the ask-price for a package S, they have to calculate the price $(p\left(X^{CS}(S)\right))$ of a set of complement disjoint bids, not overlapping with S and covering all items in $\mathcal{K} \backslash S$.

[?] designed PAUSE under the premises of an OR-bidding language, meaning a bidder can win more than just one of his bids, and super-additive valuation functions. We adopted these assumptions in our analysis of the *Bid Determination Problem* (BDP) and the worst case efficiency bound.

Definition 1. The Bid Determination Problem: To maximize bidder i's current payoff $\pi_i \in \mathbb{R}_0^+$, he has to bid on the package(s) S determined by:

$$\max_{S \in \mathcal{D}_{i,h}} \left(v_i(S) - p\left(X^{CB}\right) + p\left(X^{CS}(S)\right) - \epsilon\right) \geq \pi_i$$

The inequation ensures that bidder i bids on package(s) S only, if the prospective payoff will not be less than his current payoff. The optimal determination of $p\left(X^{CS}(S)\right)$ is $\mathcal{NP}$-hard, as it is a CAP on the complementary set, which has to be calculated for every package $S \in \mathcal{D}_{i,h}$ to determine the straightforward bid.

The following example in Table 1 shows valuations of bidders in $\mathcal{I} = \{1, 2\}$ for the items in $\mathcal{K} = \{1, 2\}$ and sketches the PAUSE auction process with straightforward bidders. PAUSE does not achieve the efficient allocation indicated by the asterisks, but terminates with 51.5% efficiency.

	1	2	1, 2	$p(X^{CB})$	π_1	π_2
v_1	100^*	0	103			
v_2	0	100^*	103			
Stage1	1_1	1_2	0	2	99	99
Stage2	0	0	3_1	3	100	0
	0	0	4_2	4	0	99
	...					
	Termination		103_1	103	0	0

Table 1: Bidders' valuations and auction process - an example of low efficiency in PAUSE

THEOREM 1. *PAUSE terminates with an allocation that is at least $1/m$ efficient, if all bidders follow the straightforward strategy and have superadditive valuations.*

Proof: The proof leans towards the example in Table 1. Given the premises stated in the theorem, inefficiencies can only occur in PAUSE, if the auction terminates allocating big packages, although disjoint subsets of them would support the efficient allocation.

Lets assume stage 1 terminates with bids

$$p_i(\mathrm{i}) \quad = \quad \max_{i \neq j} v_j(\mathrm{i}) + \epsilon \quad \forall i \in \mathcal{I} \tag{1}$$

W.l.o.g. these bids can be considered to support the efficient allocation. The current auctioneers revenue $\Pi_{h=1}$ would be $\sum_i p_i(\mathrm{i})$.

In order to terminate with another allocation we demand no improvement on any of these individual bids. That means once any of these bids $p_i(\mathrm{i}) \notin X^{CB} \Rightarrow \exists S \in \mathcal{K}$ which applies to

$$\begin{aligned} & v_i(S) - (p(X^{CB})) > v_i(\mathrm{i}) - (p(X^{CB}) - p(X^{CS}(\mathrm{i}))) \\ \wedge \quad & |S| \leq h \end{aligned} \tag{2}$$

i.e. bidder i has a better alternative than bidding on the individual item i once his provisional payoff drops to zero.

If $v_i(S)$ is part of the final allocation, we want $p(X^{CS}(S))$ to be as small as possible considering the worst case. Thus we determine $S = \mathcal{K}$. That means as long as $h < m$ every

bid $p_i(\mathrm{i})$ for all i is part of the composite bid, which further means that no new bids are submitted before stage m. In stage m the following must apply:

$$\exists i \in \mathcal{I} \text{ with } v_i(S) - (\Pi_m) > \pi_i \tag{3}$$

Since in this case bidder i bids on the package S, all other bidders $j \in \mathcal{I} \backslash \{i\}$ have a current payoff $\pi_j = 0$ and thus also the following inequation must hold:

$$v_j(S) - (\Pi_m + \epsilon) > v_j(\mathrm{j}) - (p_j(\mathrm{j}) + p(X^{CS}(\mathrm{j})))\forall j \neq i \tag{4}$$

Efficiency is then calculated by

$$E(X^{CB}) = \frac{\max_i v_i(S)}{\sum_i v_i(\mathrm{i})} \tag{5}$$

To determine the worst case efficiency we need to minimize the numerator and maximize the denominator. Thus we can determine w.l.o.g. $v(S) = v_i(S)$ and $v(\mathrm{i}) = v_i(\mathrm{j})\forall i$.

Since the most strict condition on $v_i(S)$ is

$$v_i(S) \quad > \sum_j p_j(\mathrm{j}) + v_i(\mathrm{i}) - b_i(\mathrm{i}) + \epsilon \tag{6}$$

the worst case efficiency results in:

$$\begin{array}{rcl} \min_v E(X^{CB}) & = & \min_v \frac{v_i(S)}{\sum_i v_i(\mathrm{i})} \\ & \overset{\epsilon=1}{=} & \frac{m+v(\mathrm{i})+1}{m \cdot v(\mathrm{i})} \\ & \overset{v(\mathrm{i})\to\infty}{=} & \frac{1}{m} \end{array} \tag{7}$$

Note assuming a bid increment $\epsilon = 1$ the equations 7 only apply if the valuation $v(\mathrm{i})$ is sufficiently large, i.e. depending on m this valuation must be greater than 2 or 1 respectively. □

While such situations which lead to $1/m$ efficiency can be considered degenerated cases that will not happen too often in practice, it is very likely to achieve high efficiency on average with more realistic value models.

4. EXPERIMENTAL DESIGN

To analyze the impact of our theoretical results on the outcome of the PAUSE auction in realistic settings, we conduct computational simulations, which consists of three main components. A value model, which defines valuations of all packages for each bidder, auction formats, which define the rules, and bidding agents, who follow certain strategies.

4.1 Value Model

We use a 3 x 6 ***Real Estate*** value model that is based on the *Proximity in Space* model from the Combinatorial Auction Test Suite (CATS) in [**?**]. Our model contains two different bidder types, one big bidder, interested in all items, and five smaller bidders. Each small bidder is interested in a randomly determined preferred item, all horizontal and vertical neighbors and their respective neighbors. This means small bidders are interested in 6 to 11 items with local proximity to their preferred item. An example is shown in Table 2, in which the preferred item of a small bidder is Q, and all gray shaded items in the proximity of the preferred item have a positive valuation. For

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q*	R

Table 2: The value model with the preferred item Q of a small bidder. All his positive valued items are shaded.

each bidder we draw the baseline item valuation $v_i(k)$ from a uniform distribution separately. Complementarities occur upon vertical and horizontal adjacent items based on a logistic function to determine package valuations: $v_i(S) = \sum_{C \in P} \left(\left(1 + \frac{a}{100\left(1+e^{b-|C|}\right)} \right) * \sum_{k \in C} v_i(k) \right)$, with P being the partition of S containing maximal connected packages C. This complementarity structure takes the lack of economies of scale with small packages and a saturation effect with larger packages into account. For our experiments we choose $a = 320$ and $b = 10$ for the big bidder and $a = 160$ and $b = 4$ for all small bidders, and draw the baseline valuations for the big bidder on the range $[3, 9]$ and for the small bidders on the range $[3, 20]$.

4.2 Auction Formats

We analyze two different auction formats in our economic environment. The PAUSE auction, as described in Section 2 with a minimum increment of 3 and the CC auction [**?**].

The CC auction is also a multi-round auction, in which bidders are able to place new bids in every round according to new calculated linear ask prices. The price for a package is simply the sum of item-prices. Bidders use an OR-bidding language. Prices for all items are initially zero. In every round bidders identify a package of items, or several packages, which they offer to buy at current prices. If two or more bidders demand an item then its price is increased by the minimum bid increment of 1 in the next round. This process iterates. In a simple scenario in which supply equals demand, the auction terminates and the items are allocated according to the current round bids. If at some point there is excess supply for at least one item and no item is overdemanded, the auctioneer determines the winners to find an allocation of items that maximizes his revenue by considering all submitted bids. If the solution displaces a bidder, who was active in the last round, the prices of items in the corresponding bids rise by the bid increment and the auction continues. The auction ends when no prices are increased and bidders finally pay their bid prices for winning packages.

4.3 Bidding Agents

In PAUSE we use two different bidding strategies and two different approaches to determine the bid price. As introduced in Section 3 we implement the straightforward (BR) bidding strategy, and a *Greedy* bidding strategy that allows the agents to reduce their demand set to one package calculated by $\max(v_i(S)/|S|), \forall S \in \mathcal{D}_{h,i}$ in every stage. As shown by our theoretical analysis the optimal calculation of the corresponding complement set $X^{CS}(S)$ is $\mathcal{NP}$-hard, therefore we explore two different types of calculating it, an optimal (oCS) and a heuristic (hCS) approach. We propose the following heuristic, with $k\left(X^{CS}\right)$ denoting the set of items covered by the bids in X^{CS}:

1) $X^{CS} := \emptyset$
2) `while` $k\left(X^{CS}\right) \neq \mathcal{K} \backslash S$
$$X^{CS} = X^{CS} \cup \arg \max_{p_i(T)|_{T \subseteq \mathcal{K} \backslash \left(S \cup k(X^{CS})\right)}} p_i(T)$$

We start with an empty complement set X^{CS}, determine all active bids not overlapping the current considered package S, choose the bid with the highest price and add it to our complement set X^{CS}. Then we determine the next bid, not overlapping S and $k(X^{CS})$ with the highest bid price. We repeat until our complement set covers all items of $\mathcal{K} \backslash S$.

For our experiments with the CC auction we use the straightforward bidder and a *heuristic* bidder (*5of20*) bidding on 5 of his 20 best packages in every round, more details to this in [**?**]. Additionally we implemented a *preselect* bidder (*pres10*) who determines his 10 most valuable packages before the auction starts, and bids in each round on all of them applying to $v_i(S) \geq p(S)$.

5. RESULTS

We run 50 simulations for every of the 4 bidding agents in PAUSE and for the 3 different bidding strategies in CC. All experiments run on an Intel Core2Duo processor with 2.67 GHz, 4 GB of RAM, Windows Vista and the open source IP solver "lp_solve".

A primary measure for the quality of an auction mechanism is the allocative efficiency (X^* denotes the best allocation and $v_i(X)$ bidder i's valuation for the allocation X):

$$E(X) = \frac{\sum_{i \in \mathcal{I}} v_i(X)}{\sum_{i \in \mathcal{I}} v_i(X^*)}$$

A further measure is the auctioneer's revenue share:

$$R(X) = \frac{\sum_{i \in \mathcal{I}} p_i(X)}{\sum_{i \in \mathcal{I}} v_i(X^*)}$$

Bidders revenue share is:

$$B(X) = E(X) - R(X)$$

As expected by our theoretical analysis, straightforward bidding in PAUSE with more items and higher competition leads to a better efficiency than the lower bound. We find that BR_{oCS} agents achieve in many cases a solution near the efficient one (Figure 1) and a high auctioneer's revenue (Table 3). In PAUSE all considered agents are able to find a highly efficient solution, even *Greedy* agents, who generate only ∼60% final bids compared to BR agents. Surprisingly, calculating the complement set X^{CS} with our heuristic (hCS) leads only to a small deviation in all measures (except the runtime) from the results with agents calculating X^{CS} optimally.

RESULT 1. *Determining the complement set X^{CS} suboptimally has only a small impact on the auction outcome.*

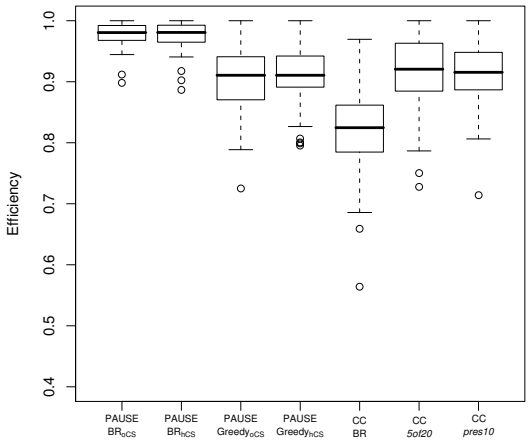


Figure 1: Auction efficiency with different bidding strategies and auction formats

In contrast to PAUSE, the CC auction mostly ends in allocations with lower efficiency and auctioneers' revenue. We suspect mainly the high number of unsold items (Table 3) to lead to such inefficiencies, together with the bigger size of winning packages (∅6.5 with BR agents vs. ∅5.03 with BR_{oCS} agents vs. ∅5.45 in efficient solutions) and the lower number of final bids. To analyze the pure impact of unsold items we ran additional simulations with CC auctions, in which we enforce the agents to bid in the first round on all items they are interested in and found, that the efficiency increases to 89.93% on average with BR agents.

RESULT 2. *An auction mechanism forcing agents to bid also on smaller packages, guides them in solving their coordination problem.*

RESULT 3. *CC needs fewer rounds to clear than PAUSE auctions.*

This results from the only moderate increasing of the allowed package size and from the package increment vs. the linear item increment in CC auctions.

Concerning bidders' calculation complexity shows Figure 2 that with an increasing number of items a small BR_{oCS} agent is interested in, the required calculation time in the auction increases exponentially. We omit the result of the big bidder, who needs around six hours (particularly ∅ 21.451

	PAUSE				CC		
	BR_{oCS}	BR_{hCS}	$Greedy_{oCS}$	$Greedy_{hCS}$	BR	$5of20$	$pres10$
∅ Efficiency in %	97.71	97.52	90.54	91.01	81.81	91.70	90.95
∅ Auctioneers' revenue in %	88.02	88.44	73.62	73.54	76.22	87.96	88.68
∅ Bidders' revenue in %	9.69	9.08	16.92	17.48	5.59	3.74	2.27
∅ Rounds	126.98	127.74	101.48	101.32	43.14	47.02	44.88
∅ Unsold items	0.00	0.00	0.00	0.00	3.96	1.66	1.50
∅ Auction runtime in sec.	22714.29	2166.12	26.85	25.37	44.81	45.34	11.10
∅ Number of final bids	54.95	55.07	33.33	33.14	35.49	145.35	52.63
∅ Size of winning packages	5.03	5.60	2.87	2.75	6.50	5.34	10.58

Table 3: Summary of simulation results

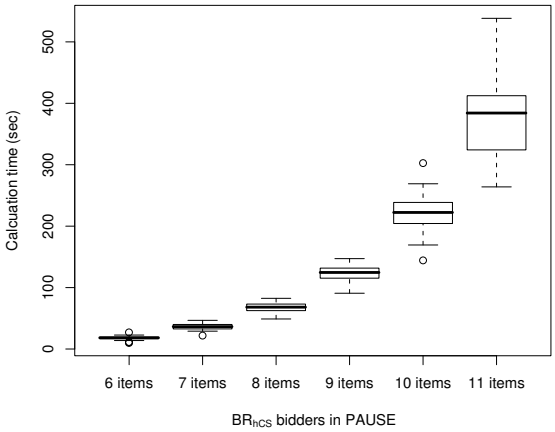


Figure 2: Bidders' required time over the auction dependent on the number of items they are interested in

sec.) per auction. The boxplot further exhibits, with 7 items or more of interest, a single BR_{oCS} agent in PAUSE requires more calculation time than the complete CC auction process.

Result 4. *Determining the straightforward bid in PAUSE drastically increase the bidders' complexity.*

Comparing the BR bidders in PAUSE with the $pres10$ bidders in CC or the $Greedy$ bidders in PAUSE with the BR bidders in CC we find the following result.

Result 5. *With a similar number of active bids, PAUSE leads to higher efficiency.*

PAUSE collects package bids of every size due to the restrictions of the package size in every stage. This helps to find allocations with high revenue, while in CC more bigger sized package bids are collected which often overlap with each other and so prohibit a "good" allocation.

6. CONCLUSION

We provide a deeper theoretical insight in the decentralized PAUSE auction and present experimental results of two different auction mechanisms. We analyzed effects of the straightforward bidding strategy in PAUSE. First we discover following this strategy leads to a growing bid determination complexity, as bidders are not allowed to submit new package bids without embedding them in a new allocation. Secondly if all bidders follow the straightforward strategy, we determine a worst case bound of $1/m$ efficiency.

Since our theoretical analysis promises better efficiency and auctioneer's revenue by the use of more realistic value models, we conducted computational experiments to verify this prediction. We used an agent-based system to compare different bidding strategies and auction mechanisms and find straightforward bidding with optimal bid price determination in PAUSE leads to very high efficiency and auctioneer revenue. Surprisingly, deviating from the optimal bid price determination does not have a significant impact on the auction outcomes, while the auction runtime is reduced drastically. The comparison to the CC auction exhibits that PAUSE is a better guide solving the bidders' coordination problem since it collects different sizes of package bids.

PAUSE shows some desirable properties, however, before taking it to the field it needs further research concerning bidder behavior and auction rules.

7. ACKNOWLEDGMENTS

Financial support from the Deutsche Forschungsgemeinschaft (DFG) (BI 1057/3-1) is gratefully acknowledged.

8. REFERENCES

[1] M. Bichler, A. Davenport, G. Hohner, and J. Kalagnanam. Industrial procurement auctions. In P. Cramton, Y. Shoham, and R. Steinberg, editors, *Combinatorial Auctions*. MIT Press, 2006.

[2] M. Bichler, P. Shabalin, and A. Pikovsky. A computational analysis of linear-price iterative combinatorial auctions. *Information Systems Research*, 20(1):33–59, 2009.

[3] Y. Caplice, C.and Sheffi. Combinatorial auctions for truckload transportation. In P. Cramton, Y. Shoham, and R. Steinberg, editors, *Combinatorial Auctions*. MIT Press, Cambridge, MA, 2006.

[4] P. Cramton. *Auctioning the Digital Dividend*. Karlsruhe Institute of Technology, 2009.

[5] P. Cramton, Y. Shoham, and R. Steinberg, editors.

Combinatorial Auctions. MIT Press, Cambridge, MA, 2006.

[6] P. Cramton, Y. Shoham, and R. Steinberg. Introduction to combinatorial auctions. In P. Cramton, Y. Shoham, and R. Steinberg, editors, *Combinatorial Auctions*. MIT Press, Cambridge, MA, 2006.

[7] J. Gallien and L. Wein. A smart market for industrial procurement with capacity constraints. *Management Science*, 51:76–91, 2005.

[8] F. Kelly and R. Steinberg. A combinatorial auction with multiple winners for universal service. *Management Science*, 46(4):586–596, 2000.

[9] K. Leyton-Brown, M. Pearson, and Y. Shoham. Towards a universal test suite for combinatorial auction algorithms. In *ACM Conference on Electronic Commerce*, pages 66–76, 2000.

[10] K. McCabe, S. Rassenti, and V. Smith. Smart computer-assisted markets. *Science*, 254:534–538, 1991.

[11] B. Mendoza and J. M. Vidal. Bidding algorithms for a distributed combinatorial auction. In *Proceedings of the on Autonomous Agents and Multiagent Systems*, 2007.

[12] B. Mendoza and J. M. Vidal. Approximate bidding algorithms for a distributed combinatorial auction (short paper). In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems*, 2008.

[13] D. Porter, S. Rassenti, A. Roopnarine, and V. Smith. Combinatorial auction design. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 100:11153–11157, 2003.

[14] M. H. Rothkopf, A. Pekec, and R. M. Harstad. Computationally manageable combinatorial auctions. *Management Science*, 44:1131–1147, 1998.

[15] L. D. C. M. M. P. H. R. J. J. Sandholm, T. and D. Begg. Changing the game in strategic sourcing at procter & gamble: Expressive competition enabled by optimization. *Interfaces*, 36(1):55–68, 2006.