User-Specific Semantic Integration of Heterogeneous Data: What Remains to be Done?

Technical Report ifi-2004.01

Patrick Ziegler
Database Technology Research Group, Department of Informatics, University of Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
pziegler@ifi.unizh.ch

Abstract

In data integration, autonomy of data sources is usually given higher priority than diversity of information needs of data end-users. However, data receivers strongly differ in their information needs and in their conceptual mental models of their particular application area.

In this paper, we review existing data integrations approaches for their compliance with the ASME criteria (Abstraction, Selection, Modeling, and Explicit semantics). Our goal is to assess whether existing data integration approaches provide suitable means for truly user-specific data integration from selected data sources. In particular, we investigate whether data from heterogeneous sources can be integrated in a way that it perfectly fits to a particular user's information needs, emphasizing his individual way to perceive a domain of interest. Additionally, we survey data integration approaches for their support for explicit representation of data semantics and for shielding users from technical-level heterogeneities of underlying data sources.

1 Introduction

In today's ever increasing abundance of online data sources, integration is becoming more and more indispensable in order not to drown in data while starving for information. In general, the goal of data integration is to combine data from different sources by applying a global data model and by detecting and resolving schema and data conflicts so that a homogeneous, unified view can be provided. The data sources to be integrated can be internal to an organization (e.g., data connected through a common intranet in an enterprise) or external (e.g., data available on the internet or extranet). The reason for data integration is twofold: First, given a set of existing data sources, an integrated view is to be created to facilitate data access and reuse through a single data access point. Second, given a certain information need, data from different complementing sources is to be combined to gain a more comprehensive basis to satisfy the information need.

There is a remarkable history of research projects in the area of data integration. The spectrum ranges from early multidatabase systems (like Multibase [16], Mermaid [26], or Pegasus [1]) over mediator systems (like TSIMMIS [8], Garlic [5], or DISCO [27]) to ontology-based integration approaches (like SIMS [2], OBSERVER [19], COIN [10], or KRAFT [21]). These approaches have in common that autonomy of the data sources to be integrated is considered to be of paramount importance; i.e., the degree to which a local data source can operate independently must not be reduced by the integration system.

Besides this autonomy of data sources, there is the often neglected autonomy and sovereignty of data receivers, i.e., human users and applications [10]. Data receivers are autonomous in the sense that they typically have different information needs and vary in the ways they perceive their particular domain of interest. Sovereignty of data receivers refers to the fact that using integrated data must be *non-intrusive* [24]; i.e., users should not be forced to adapt to any standard concerning structure and semantics of data they desire. Therefore, to take a "one integrated schema fits all" approach is definitely not a satisfactory solution.

In this paper, we review existing data integrations approaches for their compliance with the ASME criteria (Abstraction, Selection, Modeling, and Explicit semantics, see Sect. 2). Our goal is to assess whether existing data integration approaches provide suitable means for truly user-specific data integration from selected heterogeneous data sources while (1) explicitly representing data semantics and (2) shielding users from technical-level heterogeneities of underlying data sources. In particular, we investigate whether data from heterogeneous sources can be integrated in a way that it perfectly fits to a particular user's information needs, emphasizing his individual way to perceive a domain of interest.

2 The ASME Criteria for Evaluating Data Integration Approaches

In order to evaluate data integration approaches, some criteria are needed. We define a set of four criteria — the ASME criteria — for assessing whether a data integration system is capable of providing the means that allow data from heterogeneous sources to be integrated in a way that it perfectly fits to a particular user's information needs. That way, the user's individual way to perceive a domain of interest can be reflected in the integrated data. Additionally, we demand that data is provided in a semantically unambiguous and clear way. Moreover, we require that users are conceptually shielded from technical-level heterogeneities of data sources so that integration is possible without putting the whole burden of data integration solely on the user. In short, the acronym ASME stands for:

- ***A**bstraction* refers to shielding users from low-level heterogeneities and underlying data sources;
- ***S**election* means the possibility of user-specific selection of data and data sources for individual integration;
- ***M**odeling* corresponds to the availability of means to incorporate user-specific ways to perceive a domain of interest for which integrated data is desired in the process of data integration;
- ***E**xplicit semantics* refers to means for explicitly representing the real-world semantics of data.

These four assessment criteria were selected for the following reasons:

- First, without abstraction of users from technical issues of underlying data sources, there is no transparency. That is, users have to directly locate and access component data sources from which data is requested and have to cope with technical-level issues. Integration is considerably more demanding since no assistance in form of a conceptual, user-friendly interface is provided.
- Second, users may differ in their preference for data from different origins due to differences in their requirements for quality, reliability, etc. For truly user-specific data integration, this should be reflected by enabling users to select and combine data from individually selected data sources for later modeling of tailored views.
- Third, users are generally situated in different real-world contexts and widely differ in their way to conceptually structure their relevant application domains. Therefore, integrated data should be provided in a form that it fits a particular user's information needs, emphasizing his way to perceive a domain of interest.
- Fourth, without explicit statements on the intended semantics, users have to interpret data and schema items themselves, which is generally erroneous. Misinterpretations thus have to be expected. Even worse, important underlying assumptions concerning source data may be fully implicit due to the lack of explicit semantic metadata and documentation. Hence, such metadata should be provided in data integration to enable users to combine data in a semantically correct way.

3 Overview of Evaluated Data Integration Approaches

With respect to the ASME criteria, several areas of data integration approaches can be identified that can contribute to meeting these criteria. In many of the areas presented in this section, there is a plethora of approaches. Thus, we only present some representative approaches without claiming that our presentation is exhaustive. The first of these areas of relevant work are multidatabase languages and declarative integration languages.

One way to provide user-specific data integration is to offer a declarative integration language, such as the multidatabase manipulation language MSQL [18]. MSQL is an extension of SQL and allows access and manipulation of data from autonomous relational databases that are mutually not integrated. MSQL supports collective identifiers (names shared by several relations or attributes) and semantic variables (variables whose domains cover several names of attributes, relations, or databases). MSQL was later extended to MSQL+ [20]. In MSQL+, additional features to define global virtual database objects and their mapping to local database objects are introduced to provide improved mechanisms to deal with semantic heterogeneity.

The multidatabase language SQL/M [14] supports integration of related entities (classes, relations) from one or more component database systems into virtual classes that form a global schema. Each virtual class has a list of given queries to populate the virtual class with virtual instances from the component database systems. SQL/M mainly considers attribute conflicts and provides mechanisms for unit transformation and scaling.

SchemaSQL [15] is an extension of SQL and offers the possibility to query and manipulate both data and metadata in relational multidatabase systems. To restructure data, views can be defined whose schema is dynamically dependent on the data provided by component database systems. Additionally, SchemaSQL supports horizontal (column-wise) and vertical (row-wise) aggregation of data elements.

FRAQL [23] is a query language for object-relational database federations. It is not targeted to be an end-user query language but as an intermediate language to define integrated views. FRAQL is an extension of SQL and uses the object-relational data model. It supports definition of federations, access of metadata in queries, restructuring of query results, and resolving of integration conflicts. FRAQL can be extended with user-defined data types and functions and supports dynamic addition of new databases.

A second area of integration approaches abstract from the local data sources by providing a conceptual layer on top of which data integration can be performed.

InfoQuilt [25] is an agent-based integration system that allows to semantically request and correlate information from different heterogeneous sources. With IScapes (i.e., semantic information requests), the user is abstracted from the characteristics and structure of the data sources. IScapes are specified using concepts from several ontologies. Given an IScape, execution plans can be generated to retrieve and integrate relevant data. The primary focus of InfoQuilt is on human-assisted knowledge discovery and support for decision making.

Unity [17] is a data integration system for relational databases. Semantics of local databases are expressed with terms from a predefined dictionary. For each database, an XML-based specification document called X-Spec is created that encodes the local schema with terms from the dictionary and additional metadata. Unity then constructs an integrated view by combining local views.

KIND [11] is a model-based mediator that incorporates semantic models of information sources into the integration process. Views of data sources are defined at the conceptual rather at the structural level. Additionally, domain maps (semantic nets of concepts and relationships) with formal semantics are used to mediate between data sources of different origins ("worlds").

INDUS [6] is a federated, query-centric approach to data integration. An ontology that links the different data sources according to the user's point of view has to be provided by the user as an input to INDUS. Then, data from the sources can be extracted and stored in a relational database that is structured according to the specified ontology.

Third, object-oriented virtual integration approaches can be useful for user-specific data integration. These approaches enable the user to express specific views and ways to compose integrated data objects.

In TSIMMIS [8], a set of tools is provided to assist users in information processing and integration. In TSIMMIS, there is no global schema but purpose-specific mediators are semi-automatically created to provide access to selected data portions. Using OEM as the global, object-oriented data model, the meaning of objects is described with tags. Wrappers translate data from structured and semi-structured data sources into OEM objects for querying with the SQL-like query language OEM-QL.

Garlic [5] is a heterogeneous multimedia information system that provides integrated access to text, image, audio, and video data. In Garlic, an extension of ODMG-93 [7] is used as the global data model. Using wrappers, local data sources provide wrapper schemas ("interfaces") for the data they provide for integration. Using complex objects, data from different sources can be combined and related in new ways. Data access for end-users is enabled by a graphical query/browser interface.

ViewSystem [13] is an object-oriented programming environment with dedicated information integration operators. ViewSystem uses VODAK as its common global data model and supports different types of relationships between classes, such as specialization, generalization, grouping, and aggregation. On top of autonomous data sources, semantically enriched VODAK schemas are supplied. By using them and other shared views, user-specific views can be defined by declaring derived classes from existing ones. Then, queries against these personal views can be asked with an object-oriented query language.

Fourth, ontology-based integration approaches that support explicit, formal semantics. These approaches provide means to explicitly represent the intended semantics of data items.

SIMS [2] is an ontology-based mediator system. SIMS uses a domain model as a single ontology that provides class descriptions and subclass/superclass relationships. All the data sources are described with information source models that map the contents of these sources to the single global ontology. Then, queries expressed in terms of the global ontology can be asked.

OBSERVER [19] is a multi-ontology approach to data integration. In OBSERVER, the content of each data source is described by one or more ontologies. Queries are formulated in terms of user-selected ontologies. OBSERVER then uses ontological inferences to classify a user query and to determine relevant data sources. The query can then be translated to the query languages of these data sources.

KRAFT [21] is an agent-based system for knowledge fusion. As OBSERVER, it is a multi-ontology approach: For each data source, a local ontology is specified and there is a shared, global ontology. For mismatches between local ontologies and the global ontology, ontology translations are defined that specify how an expression using terms from a source ontology has to be translated into an expression using terms from a target ontology. That way, agents in a

KRAFT network can communicate to fuse information.

COIN [10] is a hybrid approach. In COIN, the semantics of both, data provided by data sources and data expected by users formulating queries are explicitly represented in so-called contexts. These contexts are built using terms from a common, shared ontology. User queries are rewritten by a context mediator into semantically equivalent mediated queries and subqueries can then be sent to the local data sources.

Fifth, there are relevant approaches in the area of the Semantic Web [4]. For example, On2broker [9] (previously named Ontobroker) is a tool-environment that supports semantics-based query handling for semi-structured information from the world wide web (WWW). Ontologies are used to make explicit the semantics of web pages. Using extensions of HTML tags, web documents are annotated so that semantic query access and inference services can be provided. A similar approach of ontology-based annotation of web documents is also taken in SHOE [12].

Last but not least, an area of relevant work are taxonomic database systems. Here, the goal is to support multiple overlapping classifications in centralized, non-integrated database systems. For example, the Prometheus [22] database system for taxonomy provides an extended object-oriented data model called POOM to support scientific classification of organisms into multiple, overlapping classification hierarchies. That way, organisms can be simultaneously classified according to different criteria.

4 Evaluation Results

Our evaluation of the previously presented data integration approaches according to the ASME criteria leads to the following results (see Table 1):

- First, concerning multidatabase languages like SchemaSQL [15] or FRAQL [23], users are not abstracted from underlying data sources. Since multidatabase languages do not provide transparency, users have to directly locate and access the component database systems from which data is requested. Therefore, applying multidatabase languages is quite demanding since all tasks of data integration and reconciliation are solely put on the user. Additionally, multidatabase languages provide no explicit queryable data semantics to ensure semantically correct interpretations of schemas and data. Moreover, multidatabase languages focus on data integration in homogeneous environments, usually data from relational databases.

- Second, there are integration approaches that abstract from the local data sources by providing a conceptual layer on top of which data integration can be performed, like Unity [17] or KIND [11]. These approaches principally take a single global schema approach that forces all users to adapt to a single global schema. First, data of each source is separately and conceptually described. Then, these descriptions are used as a basis for global schema creation and query processing. Other approaches in this category, like InfoQuilt [25], require the user to model his domain of interest. Then, it is up to the user to identify relevant data sources and to create wrappers. Thus, users are required to deal with data sources directly without being shielded from underlying technical details and heterogeneity.

- Third, object-oriented virtual integration approaches like TSIMMIS [8] or Garlic [5] enable users to model object-oriented views of desired data. However, they lack support for explicit, queryable semantics on all available data. Moreover, users are insufficiently abstracted from underlying data sources and thus have to cope with low-level heterogeneities.

- Fourth, ontology-based integration approaches are considered. In single-ontology approaches like SIMS [2], an ontological commitment is needed in that respect that a user has to accept once and for all one global ontology. That is, a fixed way to perceive a particular domain is predetermined. On the other hand, additional mapping and similarity detection efforts between the involved ontologies are needed in multi-ontology approaches like OBSERVER [19]. Here, integration of different ontological domain models is required before integrated data access is possible.

- Fifth, approaches from the Semantic Web like On2broker [9] are limited to data from the WWW. Here, data is just annotated to represent explicit semantics for later retrieval. That way, information retrieval is possible that takes into account semantics as provided by the annotations. However, data is not provided in a structurally homogeneous, integrated form.

- Last, but not least, taxonomic database systems like Prometheus [22] do not consider integration but provision of multiple, overlapping taxonomies for single centralized database systems. That way, different ways to categorize a domain of interest can be supported. However, taxonomic database systems generally lack support for explicit data semantics.

5 Conclusions

In this paper, we have evaluated existing data integration approaches for their compliance with the ASME criteria (Abstraction, Selection, Modeling, and Explicit seman-

	Integratable data source types	Full abstraction of user from data sources	User-specific data source selection for integration	User-specific data modeling for integration	Explicit, queryable semantics
Multidatabase languages and declarative integration languages					
MSQL+	relational databases	no	yes	yes	no
SQL/M	relational databases	no	yes	yes	no
SchemaSQL	relational databases	no	yes	yes	no
FRAQL	object-relational databases	no	yes	yes	no
Approaches with conceptual-level abstraction from data sources					
InfoQuilt	databases and Web sources	yes	no	partially	no
Unity	relational databases	yes	no	no	no
KIND	relational, object-oriented, semistructured	yes	no	no	no
INDUS	relational databases, flat files	no	yes	yes	no
Object-oriented virtual integration approaches					
TSIMMIS	structured, semi-structured, unstructured	no	yes	yes	no
Garlic	databases, files, multimedia data sources	no	yes	yes	no
ViewSystem	relational databases, information retrieval system data, files	no	yes	yes	no
Ontology-based integration approaches					
SIMS	structured information sources	yes	no	no	no
OBSERVER	databases, files, Web sources	yes	no	no	partially
KRAFT	databases	yes	no	no	no
COIN	databases and semi-structured Web sources	no	yes	no	no
Semantic Web approaches					
On2broker	semi-structured Web documents (HTML, XML, RDF)	yes	no	no	no
SHOE	HTML files	yes	no	no	no
Taxonomic database systems					
Prometheus	n/a (not applicable)	n/a	n/a	n/a	no

Table 1. Characterization of Selected Integration Approaches According to the ASME Criteria

tics). In retrospect to the findings of our evaluation, three types of patterns can be recognized in Table 1:

- There are approaches (e.g., all presented multidatabase language and object-oriented virtual integration approaches) that support user-specific selection of data sources and user-specific data modeling for integration. However, these approaches do not abstract the user from technical-level heterogeneities.

- On the other hand, there are approaches (e.g., SIMS, OBSERVER, KRAFT, or On2broker) that abstract the user from the data sources and technical aspects; at the same time, these approaches do neither allow user-specific selection of data sources nor user-specific data modeling for integration. It thus seems that abstracting the user from technical-level issues and supporting user-specific data selection and modeling are conflicting goals.

- None of the presented approaches provides explicit, queryable semantics. OBSERVER is the only exception. However, OBSERVER only partially supports explicit, queryable semantics: Although OBSERVER internally uses a system based on Description Logics [3] for ontological knowledge, queries on data semantics from users against this Description Logics system are not supported. Ontologies are only used internally to rewrite queries based on inter-ontological relationships.

From Table 1 it can be seen that none of the presented approaches is able to fully attain the ASME criteria. While we argued in Sect. 2 that all the four criteria are essential for data integration, there is — to the best of our knowledge — no data integration approach that is able to achieve all of them. We thus conclude that an integration approach capable of (1) preserving user sovereignty by enabling the user to express his desired view for data integration, (2) shielding the user from technical-level heterogeneities, and (3) providing explicit, queryable data semantics is desirable.

References

[1] R. Ahmed, P. D. Smedt, W. Du, W. Kent, M. A. Ketabchi, W. Litwin, A. Rafii, and M.-C. Shan. The Pegasus Heterogeneous Multidatabase System. *IEEE Computer*, 24(12):19–27, 1991.

[2] Y. Arens, C. Y. Chee, C.-N. Hsu, and C. A. Knoblock. Retrieving and Integrating Data from Multiple Information Sources. *International Journal of Cooperative Information Systems (IJCIS)*, 2(2):127–158, 1993.

[3] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003.

[4] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.

[5] M. Carey, L. Haas, P. Schwarz, M. Arya, W. Cody, R. Fagin, M. Flickner, A. Luniewski, W. Niblack, D. Petkovic, J. Thomas, J. Williams, and E. Wimmers. Towards Heterogeneous Multimedia Information Systems: The Garlic Approach. In *5th International Workshop on Research Issues in Data Engineering-Distributed Object Management (RIDE-DOM 1995)*, pages 124–131, Taipei, Taiwan, March 6-7, 1995.

[6] J. A. R. Castillo, A. Silvescu, D. Caragea, J. Pathak, and V. G. Honavar. Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Information Sources – A Federated Ontology-Driven Query-Centric Approach. In *2003 IEEE International Conference on Information Reuse and Integration (IRI 2003)*, Las Vegas, Nevada, USA, October 27-29, 2003.

[7] R. G. G. Cattell and D. K. Barry, editors. *The Object Data Standard: ODMG 3.0*. Morgan Kaufmann, 2000.

[8] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, and J. Widom. The TSIMMIS Project: Integration of Heterogeneous Information Sources. In *16th Meeting of the Information Processing Society of Japan (IPSJ)*, pages 7–18, Tokyo, Japan, October, 1994.

[9] D. Fensel, J. Angele, S. Decker, M. Erdmann, H.-P. Schnurr, S. Staab, R. Studer, and A. Witt. On2broker: Semantic-based Access to Information Sources at the WWW. In P. D. Bra and J. J. Leggett, editors, *World Conference on the WWW and Internet (WebNet 99)*, pages 366–371, Honolulu, Hawai, USA, October 25-30, 1999. Association for the Advancement of Computing in Eduction (AACE).

[10] C. H. Goh, S. E. Madnick, and M. Siegel. Context Interchange: Overcoming the Challenges of Large-Scale Interoperable Database Systems in a Dynamic Environment. In *Third International Conference on Information and Knowledge Management (CIKM 1994)*, pages 337–346, Gaithersburg, Maryland, November 29 - December 2, 1994. ACM Press.

[11] A. Gupta, B. Ludscher, and M. E. Martone. Knowledge-Based Integration of Neuroscience Data

Sources. In O. Günther and H.-J. Lenz, editors, *12th International Conference on Scientific and Statistical Database Management (SSDBM 2000)*, pages 39–52, Berlin, Germany, July 26-28, 2000. IEEE Computer Society.

[12] J. Heflin and J. Hendler. Semantic Interoperability on the Web. In *Extreme Markup Languages 2000*, pages 111–120, Montreal, Quebec, Canada, August 6-9, 2000. Graphic Communications Association.

[13] M. Kaul, K. Drosten, and E. J. Neuhold. ViewSystem: Integrating Heterogeneous Information Bases by Object-Oriented Views. In *Sixth International Conference on Data Engineering (ICDE 1990)*, pages 2–10, Los Angeles, California, USA, February 5-9, 1990. IEEE Computer Society.

[14] W. Kim, I. Choi, S. K. Gala, and M. Scheevel. On Resolving Schematic Heterogeneity in Multidatabase Systems. *Distributed and Parallel Databases*, 1(3):251–279, 1993.

[15] L. V. S. Lakshmanan, F. Sadri, and I. N. Subramanian. SchemaSQL - A Language for Interoperability in Relational Multi-Database Systems. In T. M. Vijayaraman, A. P. Buchmann, C. Mohan, and N. L. Sarda, editors, *22nd International Conference on Very Large Data Bases (VLDB 1996)*, pages 239–250, Bombay, India, September 3-6, 1996. Morgan Kaufmann.

[16] T. Landers and R. L. Rosenberg. An Overview of MULTIBASE. In H.-J. Schneider, editor, *Second International Symposium on Distributed Data Bases (DDB 1982)*, pages 153–184, Berlin, Germany, September 1-3, 1982. North-Holland.

[17] R. Lawrence and K. Barker. Integrating Relational Database Schemas Using a Standardized Dictionary. In *2001 ACM Symposium on Applied Computing (SAC 2001)*, pages 225–230, Las Vegas, Nevada, USA, March 11-14, 2001. ACM Press.

[18] W. Litwin, A. Abdellatif, A. Zeroual, and B. Nicolas. MSQL: A Multidatabase Language. *Information Sciences*, 49(1-3):59–101, 1989.

[19] E. Mena, V. Kashyap, A. P. Sheth, and A. Illarramendi. OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies. In *First IFCIS International Conference on Cooperative Information Systems (CoopIS 1996)*, pages 14–25, Brussels, Belgium, June 19-21, 1996. IEEE Computer Society.

[20] P. Missier and M. Rusinkiewicz. Extending a Multidatabase Manipulation Language to Resolve Schema and Data Conflicts. In R. Meersman and L. Mark, editors, *Sixth IFIP TC-2 Working Conference on Data Semantics (DS-6)*, volume 74 of *IFIP Conference Proceedings*, pages 93–115, Stone Mountain, Atlanta, Georgia, USA, May 30 - June 2, 1995. Chapman & Hall.

[21] A. D. Preece, K.-y. Hui, W. A. Gray, P. Marti, T. J. M. Bench-Capon, D. M. Jones, and Z. Cui. The KRAFT Architecture for Knowledge Fusion and Transformation. *Knowledge Based Systems*, 13(2-3):113–120, 2000.

[22] C. Raguenaud, J. B. Kennedy, and P. J. Barclay. The Prometheus Taxonomic Database. In N. G. Bourbakis, editor, *1st IEEE International Symposium on Bioinformatics and Biomedical Engineering (BIBE 2000)*, pages 63–70, Arlilngton, Virginia, USA, November 8-10, 2000. IEEE Computer Society.

[23] K.-U. Sattler, S. Conrad, and G. Saake. Adding Conflict Resolution Features to a Query Language for Database Federations. In M. Roantree, W. Hasselbring, and S. Conrad, editors, *Third Workshop on Engineering Federated Information Systems (EFIS 2000)*, pages 41–52, Dublin, Ireland, June 19-20, 2000. IOS Press / infix.

[24] P. Scheuermann, A. K. Elmagarmid, H. Garcia-Molina, F. Manola, D. McLeod, A. Rosenthal, and M. Templeton. Report on the Workshop on Heterogenous Database Systems held at Northwestern University, Evanston, Illinois, December 11-13, 1989. *SIGMOD Record*, 19(4):23–31, 1990.

[25] A. Sheth, S. Thacker, and S. Patel. Complex Relationships and Knowledge Discovery Support in the InfoQuilt System. *The VLDB Journal*, 12(1):2–27, 2002.

[26] M. Templeton, D. Brill, A. Hwang, I. Kameny, and E. Lund. An Overview of the Mermaid System - A Frontend to Heterogeneous Databases. In J. M. Walker, editor, *Sixteenth Annual Electronics and Aerospace Conference and Exposition - Technology Shaping the Future (IEEE EASCON 1983)*, pages 387–402, Washington, D.C., USA, September 19-21, 1983. IEEE Aerospace and Electronic Systems Society.

[27] A. Tomasic, L. Raschid, and P. Valduriez. Scaling Heterogeneous Databases and the Design of Disco. In *16th International Conference on Distributed Computing Systems (ICDCS 1996)*, pages 449–457, Hong Kong, China, May 27-30, 1996. IEEE Computer Society.