University of Zurich
Department of Informatics

Gabi Dreo Rodosek
Heinz-Gerd Hegering
Burkhard Stiller

Dynamic Virtual Organizations as Enablers for Managed Invisible Grids

TECHNICAL REPORT — No. 2005.09

November 2005

University of Zürich
Department of Informatics (IFI)
Winterthurerstrasse 190, CH—8057 Zürich, Switzerland

Gabi Dreo Rodosek, Heinz-Gerd Hegering, Burkhard Stiller: Dynamic Virtual Organizations as Enablers for Managed Invisible Grids
Technical Report No. 2005.09, November 2005
Communication Systems Research Group
Department of Informatics (IFI)
University of Zürich
Winterthurerstrasse 190, CH—8057 Zürich, Switzerland
URL: http://www.ifi.unizh.ch/csg/

Dynamic Virtual Organizations as Enablers for Managed Invisible Grids

Gabi Dreo Rodosek[1], Heinz-Gerd Hegering[2], Burkhard Stiller[3,4]

[1] Information Systems Lab IIS, University of Federal Armed Forces Munich, Werner-Heisenberg-Weg 39, D-85577 Neubiberg
[2] Leibniz Supercomputing Center, Barer Straße 21, D-80333 Munich, Germany
[3] Department of Informatics IFI, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
[4] Computer Engineering and Networks Laboratory TIK, ETH Zürich, Gloriastrasse 35, CH-8092 Zürich, Switzerland

E-Mail: gabi.dreo@unibw.de, hegering@lrz.de, stiller@ifi.unizh.ch

Abstract

As of today, grids provide the technology, applications, and platforms for a seamless access to resources, services, and content in a fully decentralized world of distributed information, computing power, and information technology business. Grid computing is about resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organization. Increasingly, advanced grid approaches address knowledge grids.

Therefore, the goal is to create a transparent grid infrastructure, to make the grid invisible and to overcome today's rather static approaches. Thus, the Dynamic Virtual Organizations (DVOs) approach provides the right incentives to act as an enabler for invisible grids. Firstly, this paper studies DVOs and its grid service management requirements. Secondly, a concept for service management of DVOs is presented according to the identified key requirements.

Keywords: *Grid, Invisible Grid, Dynamic Virtual Organization (DVO), Service Management.*

1 Introduction

The use of distributed resources to solve processing-intense or knowledge-based tasks has started since underlying network technology has grown to provide sufficient bandwidth across longer distances. Thus, a variety of distributed systems targeted at an efficient, reliable, robust, and efficient provisioning of resources, services, or knowledge. Besides different types of distributed middleware developments, more recently service-oriented architectures (SOA), and peer-to-peer (P2P) systems, grids have been developed to address a coordinated resource sharing across domains.
Grids are determined by a set of 3 main characteristics [8]:

- Grids "coordinate resources that are not subject to centralized control",
- Grids use "standard, open, general-purpose protocols and interfaces", and
- Grids allow "to deliver non-trivial qualities of service".

However, these three characteristics have evolved in two steps over time and the traditional view of a grid — typically termed a *computational grid* — has developed toward a *service grid*, which provides application domain-specific services, such as simulation libraries for weather, simulation packages for bio and neuro sciences, or automotive development software, all of which assume an underlying computational grid. Thus, the second step research is going to develop at this stage, is a *knowledge grid*, in which neither the supercomputing nor the application-specific service show a prevailing nature, but rather the scenario-dependent application of a grid over a mandatory set of aggregated services and computing. Additionally, the support of commercial knowledge grids and the offer of mobile grids requires enhancements on the mechanism level of such grids. Those include a.o. pricing, accounting, security, and mobility. As soon as from an application point of view such grid services become transparently usable from a variety of users and customers, the *invisible grid* is reached.
For instance, a crisis and emergency response scenario, as the most recent terror attacks unfortunately outline, does require the fast and ad-hoc aggregation of information — thus, highly relevant knowledge in a certain area —, which will synchronize on-the-spot sensor data with different data base content, which will simulate tunnel construction data with chemical data of explosives, and which will see decentralized and centralized control of information and decisions. Additionally, the heterogeneity of technology of devices, networking technology, and services is complemented by mobility aspects of sensors, data collecting devices, and decision points.
This paper argues on the technology side that *Dynamic Virtual Organizations (DVO)* can be enablers for invisible grids in such a way that service management principles required for their commercial operation are fully integrated. As there exist several definitions of the term *Virtual Organization (VO)*, those are investigated and adapted to the needs for service management integration. OGSA determines a VO as a group of individuals inside or outside a real organization [10], whereas in [9] a VO is defined as a set of resources and services. To define a VO as a group of individuals, results from the overall goal that individuals need to achieve a common goal. If the provision of (virtual) services is the main goal, it is more appropriate to refer to a VO as a set of resources and services. Additional definitions of VOs, however, not suitable for invisible grids, can be found, *e.g.*, in [7].
Thus, the new definition of a VO for invisible grids and its service management-related use reads as follows: *A VO is defined as a set of virtual resources and virtual services that can be used by individuals to achieve a common goal.* The set of *virtual* resources and *virtual* services of a VO need to be mapped in an operational manner to *real* resources and *real* services as provided by real organizations. A dynamic VO exists for a limited period of time and can be set-up in a dynamic way, according to a common application goal that needs to be achieved, either by individuals or the provisioning of (virtual) services. Thus, the integrated services management has to consider virtual and real resources as well as virtual and real services, respectively, at the same time.

The management of data transfers, the provision and supervision of services utilized, and the maintenance of required resources determines the crucial aspect for any physically involved party, mainly network operators, service providers, resource providers, and even large companies. Obviously, the number of existing management approaches and systems available today is large and they are restricted to each of those three domains separately. In turn, the management of DVOs and the underlying grid determines the key problems addressed within this paper.

The paper argues on the economic side, that only based on an integrated grid services management approach as outlined above, a commercial use and exploitation of invisible grid services will be possible. This is due to the fact that besides mapping functions for real and virtual services and resources, respectively, the integrated management will allow for an efficient, quality-driven, policy-based, and Service Level Agreement-based (SLA) approach, which follows a cost-based optimization of grid service provisioning in a multi-provider domain. Thus, accounting for resources and services, DVO operation, and SLA management are the key requirements for a commercial successful invisible grid.

The remainder of the paper is organized as follows. While Section 2 briefly overviews related work, Section 3 outlines the key dimensions of grid management. Based on those, the major requirements on grid management are presented in Section 4, which are utilized to propose the new concept of DVO management in Section 5. Finally, Section 6 summarizes and draws conclusions.

2 Related Work

Since the concept of DVOs as well as its management is quite new, the overview of the related work will cover a more broader spectrum. DVOs and their management is part of grid management which can be divided into two areas:

- Overall architectures including management tasks and
- Dedicated mechanisms for grid management.

Thus, those areas have been investigated and selected approaches are discussed. In turn, the set of remaining issues and unsolved questions are listed.

2.1 Related Architectures

The Open Grid Services Architecture (OGSA) [10] determines for the provisioning of services in a grid environment the key challenges to achieve various Quality-of-Service levels when running applications on top of different platforms. The architecture proposed is based on concepts and technologies known from different grid and Web services communities. While the underlying architecture defines a uniformly exposed service semantics and applies standard mechanisms for creating, naming, and discovering transient grid service instances it provides for the location transparency and multiple protocol bindings for service instances as well. OGSA specifies interfaces, mechanisms required for the creation and composition of distributed systems, including various management functionality, such as lifetime, change, and notification. Finally, grid service bindings have been designed to support a reliable service invocation and the two key security tasks of authentication and authorization. To implement a service-oriented architecture, the set of resource sharing mechanisms have been design to complement grid services as such.

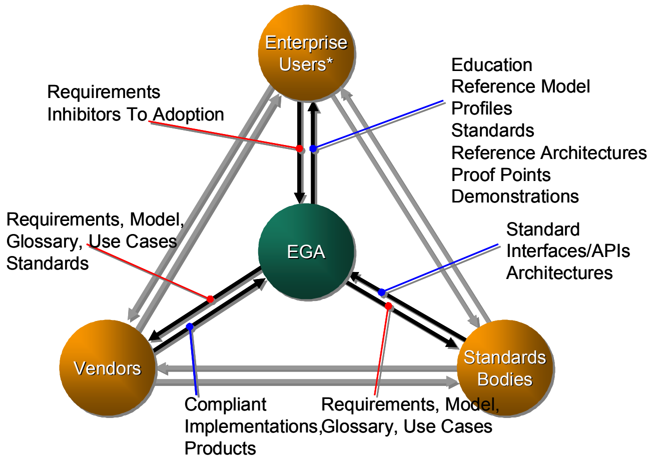

Figure 1: Enterprise Grid Alliance Reference Model [21]

In addition, the Enterprise Grid Alliance (EGA) Reference Model [21] shows a number of requirements addressing the user, providers (vendors), and standards bodies as depicted in Figure 1. The EGA Reference Model defines a framework for classifying grid resources and services together with their relationships and dependencies in a conceptual component architectural setting. The model covers contexts for requirements, solutions, and comparisons. Since it is vendor neutral and technology as well implementation agnostic, it does not assume a particular technology roadmap. Due to the set of commercial enterprise community-centric use cases consistent and relevant requirements for all enterprise grid stakeholders have been covered.

In addition, the current view on the management of grids in EGA [21] shows that for each grid component management as well as monitoring functionality are defined. This is complemented by enterprise-specific specifications of policies as well as accounting and billing interfaces, each of which seems to be in an early development state.

An important aspect of DVO management covers the functionality of services and service management. Due to the inherent complexity of the service management, most approaches develop a specific model to address particular problems within specific scenarios. Thus, a first approach toward a generic service model has been proposed in [5]. The Munich Network Management (MNM) service model, as visualized in Figure 2, has been defined as a result of a top-down oriented, systematic methodology. The service model has been defined with respect to the set of requirements such as the generic service definition, a separation of service definitions and service implementation, and determining management as an integral part of the service. The basis of the analysis is the inherent characteristic of every service that involves two players:

- One offering the service (service provider) and
- One requesting the service (customer).

According to those players identified, the model distinguishes between the

- Customer aspect of a service,
- Side-independent aspect of a service, and
- Provider aspect of a service.

For more technical information and detail, the reader may refer to [13]. Additionally, an extension of the proposed MNM service model is the proposed framework for Information Technology (IT) service management as discussed in [5]. Finally, the initiatives of Web Service Management, especially Management using Web Services (MUWS) [18] and Management of Web Services (MOWS) need to be mentioned in the context of grid service management.

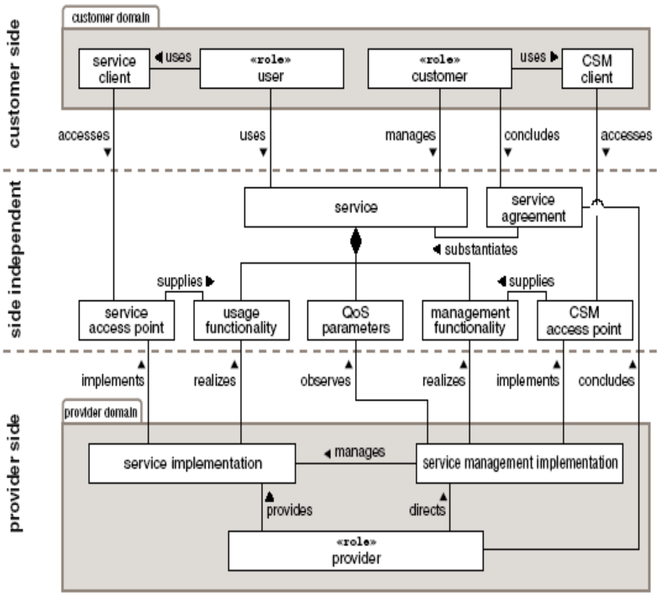

Figure 2: MNM Service Model [13]

2.2 Related Mechanisms

Further related work covers a number of dedicated grid management tasks, such as [2], where the Global Discovery Service (GDS) for the configuration and deployment of grid infrastructure and advanced monitoring services has been implemented with the JMX technology. Those described services have been designed to be scalable and lightweight and have been implemented within JIMS, the grid infrastructure monitoring system from the CrossGrid EU IST Project.

Due to the large scale of grid environments, [12] argues that the provisioning of a grid management mechanisms to enable grid computing is required to adapt to various application requirements in a flexible and automated way. Thus, [12] proposes an “Active Grid Architecture” and middleware for rapid and autonomic grid services creation, deployment, activation, and management.

Finally, for this selection, the GEON project’s task is on building the next generation cyberInfrastructure for the geosciences community, where the infrastructure is determined by a services-based infrastructure providing decentralized data federation capabilities across heterogeneous databases [22]. Based on the standardized software stack across all sites, while allowing controlled local customization a set of requirements being difficult to satisfy with existing cluster tools have been identified and extensions to the Rocks cluster distribution are described to satisfy several key goals.

The Globus Toolkit [9], [10] is an open source software toolkit used for building grids. It is being developed by the Globus Alliance and many others all over the world. The toolkit includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability. It is packaged as a set of components that can be used either independently or together to develop applications. Every organization has unique modes of operation, and collaboration between multiple organizations is hindered by incompatibility of resources such as data archives, computers, and networks. The Globus Toolkit was conceived to remove obstacles preventing seamless collaboration. Its core services, interfaces, and protocols allow users to access remote resources as if they were located within their own machine room while simultaneously preserving local control over who can use resources and when.

NorduGrid [17] is a grid research and development collaboration developing, maintaining. and supporting a free grid middleware, known as the Advance Resource Connector (ARC). The aim is to deliver a robust, scalable, portable, and fully featured solution for a global computational and data grid system. NorduGrid develops a set of tools and services, the so-called ARC middleware.

TeraGrid [25] determines an open scientific discovery infrastructure combining leadership class resources at 8 partner sites to create an integrated, persistent computational resource. Deployment of TeraGrid was completed in September 2004, bringing over 40 teraflops of computing power and nearly 2 petabyte of rotating storage, and specialized data analysis and visualization resources into production, interconnected at 10-30 Gbit/s via a dedicated national network.

The Agreement-based Grid Service Management (OGSI agreement) [4] defines a set of OGSI-compatible portTypes through management applications where services can negotiate with management services for the purpose of managing grid services and other applications and resources. These negotiations dynamically mediate between users and service providers within virtual organizations, related by potentially complex community relationships.

2.3 Open DVO Management Issues

The overview of the related work in terms of architectures and mechanisms has shown that the following issues have not been addressed so far:

- How is a DVO described, especially its offered services and resources, such that an integrated service management principle may become applicable?
- How can DVOs be automatically and dynamically set-up that the underlying real and virtual resources as well as service are cost-effectively manageable?
- What are the service management operations for DVOs?
- What are the dedicated management operations to join or leave a DVO, sometimes called membership operations?

- How are services and resources of a VO as well as a DVO advertised with respect to quality, security, and cost to enable an accounting and billing of grid services?
- Which policies and SLAs are essential for grid service provisioning and management to ensure that quality, security, and cost are met as specified by the application?
- How are trust relationships between VOs specified and maintained?
- How to support customer-tailored service and management policies for the specific needs of grid users?
- How to support flexible collaborations such as Computer Supported Cooperative Work (CSCW) or conferencing?
- How is a federated identity management supported in order to have directories for, *e.g.*, VOs, certificates, roles, or authorization?
- How to support workload management between VOs to distribute load?
- Which Management Information Bases (MIB) are to be defined for grid services or for supporting the management of DVOs, if at all?

Obviously, the grid paradigm imposes new management tasks. Therefore, dimensions of grid service management, their key requirements, and a DVO-based grid service management concept are a must and being studied in this paper.

3 Dimensions of Grid Service Management

Grids and especially their management as well as their service management determine a complex problem area with specific requirements of distributed systems and various abstraction levels, which need to be structured in a systematic and practically implementable manner. The objective of this section is in fact to recognize and define as detailed as possible which new challenges and requirements, respectively, are to be considered that grids pose on service management in general. Although all dimensions defined below cover overall tasks of grid management, the more detailed focussing aspect of the continuing discussion will be placed on service management challenges from the perspective of DVOs.

3.1 Generic Model of DVOs

Before going into major details of DVOs, it is necessary to define the generic model of a DVO, which will be utilized to derive the set of dimensions of grid management and in turn the key grid service management requirements. As visualized in Figure 3, a Virtual Organization (VO) — independent of any dynamicity — may consist of a set of virtual grid services and virtual grid resources. In the example presented, VO_A uses resources of the Real Organization B (RO_B) and services and resources of RO_A. It needs to be emphasized that only parts of real organizations may belong to a VO. Even the same part of a RO can be a part of several VOs.

Resources of ROs are the known managed objects from the traditional management, *e.g.*, network elements, end systems, or applications. IT services belonging to a real organization are also known managed objects from the area of IT service management. Thus, it needs to be determined, whether grid services and virtual services of a VO, respectively, can be managed in the same way as services of a real organization. A service of a VO needs to be specified with its functionality, the cost and quality provided, as well as its manageability. So far, these characteristics are the same as with services of real organizations. However, a virtual service is in most cases a more complex service that may consist of several services of one RO or even more ROs.

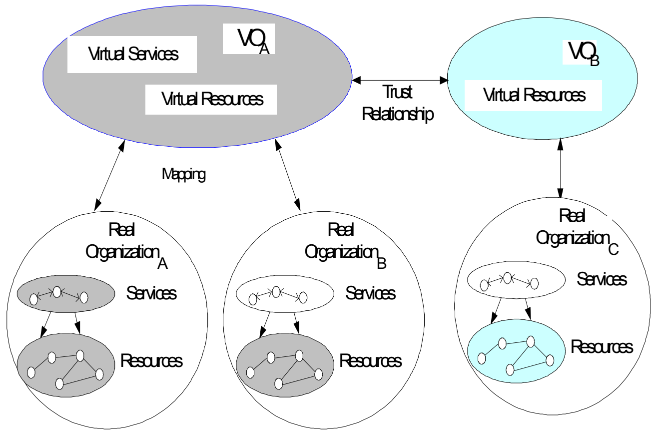

Figure 3: Generic DVO Model

In order to access services or resources of another RO it is necessary to have a trust relationship established between VOs, which is in turn mapped to all services and resources of ROs. Otherwise, explicit authentication and authorization procedures need to be performed each time a service or a resource of a VO or an RO, respectively, is accessed. Summarizing, participants and resources are governed, in addition to VO-specific policies, by rules and policies of those organizations they are members of. That means, a VO access has to be established and coordinated through a binary trust relationship between local users and their institution as well as between the VO and users.

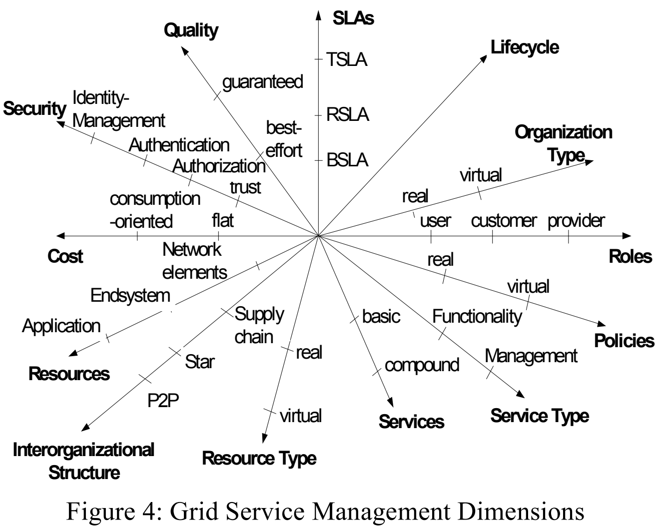

Figure 4: Grid Service Management Dimensions

3.2 Dimensions of Grid Service Management

As identified above, all dimensions of grid management developed here need to be structured systematically, to ensure that the complex area of grid service management can be highlighted appropriately. For that reason, Figure 4 outlines the 13 dimensions of grid service management.

3.2.1 Dimension 1: Organization Type

The starting point of this discussion is the question about the *organization type*. Real organizations are already known from the area of network and systems management. Virtual organizations represent a new paradigm that is characterized by dynamics and temporal limited durations as well as the certain objectives or goals that have to be achieved by the VO in case the VO is a group of individuals (i.e., communities) or virtual services and/or resources, if the provision of those is in the center of discussion. Communities are groups of grid users who are sharing interests, *e.g.*, common research goals, common experiments, or common resources. They typically obey community-specific policies with respect to group membership, resource access, or accounting. Additionally, grid applications are typically distributed. Underlying resources and services are owned by different and independent organizations. Thus, a virtualization of resources and services is indispensable for a location-, system-, and organization-transparent access. VOs can be small or large, short- or long-lived, single- or multi-institutional, homogeneous or heterogeneous.

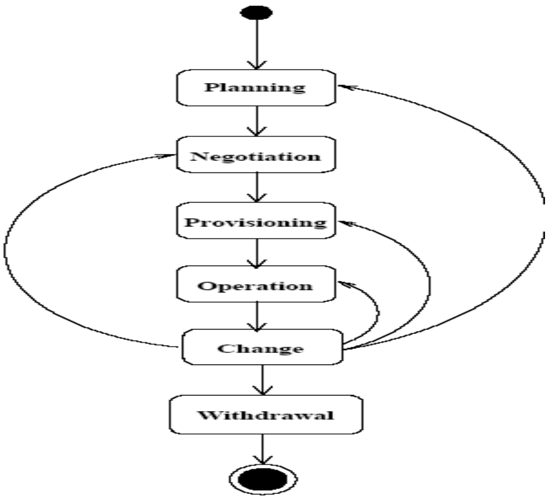

Figure 5: Lifecycle Phases [5]

3.2.2 Dimension 2: Lifecycle

All phases of the lifecycle of a VO can be defined as for example in [5], and as visualized in Figure 5. These phases consist of planning, negotiation, provisioning, operation, change, and withdrawal, and are defined from the service-oriented perspective. However, an alternative approach may be taken and the definition of a lifecycle for VOs is discussed in [14] and visualized in Figure 6.

3.2.3 Dimension 3: Service Level Agreements

Each organization needs to provide *Service Level Agreements (SLAs)* which cover the next dimension to discuss. In order to support application execution in the context of grids, the scheduling of grid resources is necessary. Grid resource scheduling is defined as the process of making scheduling decisions involving resources over multiple administrative domains. A grid resource broker must make resource selection decisions in an environment, where it has no control over the local resources, resources are distributed, and information about resources is limited or dated. The broker takes users' requirements as well as job descriptions and contacts resources that may fulfill these requirements to gather information on their current state, *e.g.*, workload. A decision is made, which resource(s) will be used to run a job. This step is followed by a negotiation with these resources. The negotiation is based on the framework provided by the Service Negotiation and Acquisition Protocol (SNAP) [24], whereby guarantees are obtained that user requirements will be fulfilled by the use of an SLA. SNAP comprises three main SLAs:

- Task Service Level Agreement (TSLA),
- Resource Service Level Agreement (RSLA), and
- Binding Service Level Agreement (BSLA).

While TSLAs are required, where the user provides a clear objective specification of his task requirements and any resource preferences, RSLAs relate to resource discovery and decision-making on the appropriate resources that meet the task requirements and securing the resources for utilization. Finally, BSLAs associate the task with the resources.

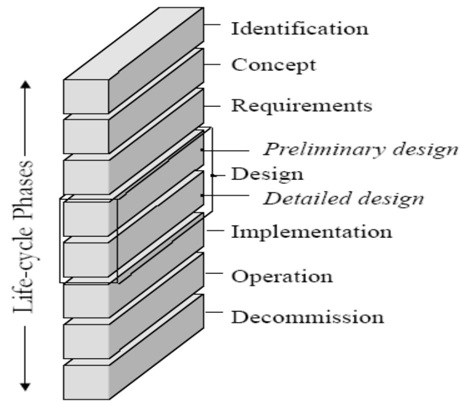

Figure 6: Phases of a VO Lifecycle [14]

3.2.4 Dimension 4: Quality

Each service needs to be provided with a *quality* of service, also if only a simple best-effort service may be required as pointed out with the next dimension. Quality of a service can be guaranteed either in terms of SLAs, this may cover deterministic as well as statistical guarantees, or it may be provided in a best-effort sense.

3.2.5 Dimension 5: Security

Efficient grid *security,* especially cross-site trust management, is a major prerequisite for VOs. Trust means confidence in or reliance in some quality or attribute of a person, institution, or object, or even the truthfulness of a statement. Problems of shared trust models for large heterogeneous VOs — involving people from multiple, international institutions — are still partly unsolved. For example, across administratively similar systems, (*e.g.*, within an organization) typically informal trust models exist. Across administratively separated systems (*e.g.*, across many similar organizations) in general formal trust models are needed. For heterogeneous systems being administered separately (*e.g.*, across multiple organization types) new formal trust models are required.

3.2.6 Dimension 6: Cost

Since services need to be provided with a specific quality, appropriate cost models need to be applied. Depending on the particular use and application area of grid services, the required accounting in support of those different cost types may either support consumption-oriented or flat-oriented schemes.

3.2.7 Dimension 7: Resources

The dimension of *resources* represents traditional managed objects of network and systems management, namely network elements, end systems, and applications.

3.2.8 Dimension 8: Inter-organizational Structure

Several *inter-organizational structures* of VOs exist, as discussed in [7], and shown in Figure 7. Depending on these structures the centralization or decentralization of management tasks as well as decisions has to be designed.

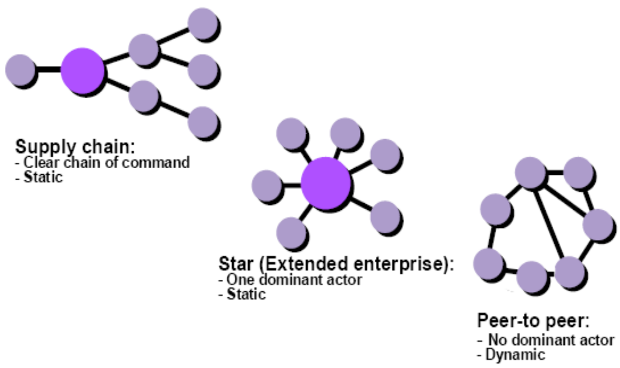

Figure 7: Sample Inter-organizational Structures [7]

3.2.9 Dimension 9: Resource Type

The dimension of *resource type* refers to either a real or a virtual resource. This distinction is mandatory, to ensure that the underlying mechanisms of scheduling the resource, accounting for the utilization of this resource, or maintaining this resource can be made available directly to the service or need to be offered in the virtualization case via an appropriate interface.

3.2.10 Dimension 10: Service

The dimension of *services* identifies the so-called basic services which do not consist of any further service, thus, forming a quasi atomic element. The so-called compound services may consist of several basic services and their composition may be based on application requirements or service necessities.

3.2.11 Dimension 11: Service Type

The dimension of the *service type* emphasizes that a service can provide either a functionality such as e-mail or that it can be a management service itself. This distinction is essential to differentiate the use of the service in a particular situation.

3.2.12 Dimension 12: Policies

Policies can be defined either for virtual or real organizations. If policies are defined for VOs, it is necessary to map them to policies of ROs and to avoid policy conflicts. The policy and authorization framework from EGEE [6] determines a viable approach in this direction (cf. Figure 8).

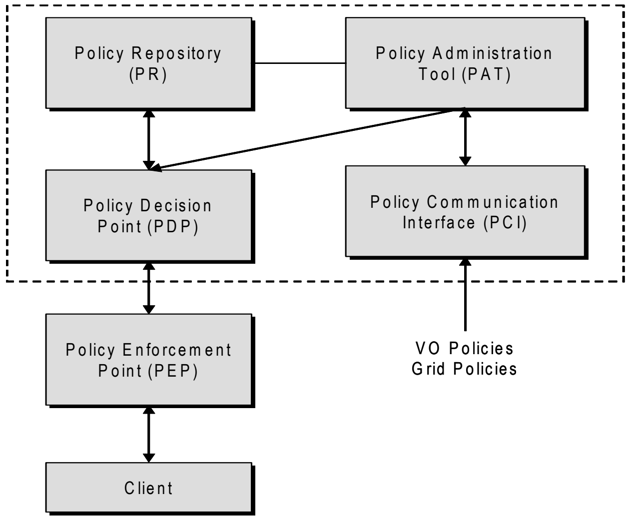

Figure 8: EGEE Policy and Authorization Framework [6]

3.2.13 Dimension 13: Roles

Finally, the dimension of *roles* enables a grid service management to distinguishes between

- A user (*e.g.*, student at a university) with a direct usage relationship,
- A customer (*e.g.*, university or a company), where a legal relationship is of relevance, and
- The provider (*e.g.*, service provider) being responsible for the provisioning of services.

Thus, this dimension determines many technical choices of mechanisms in support of service provisioning and service management.

4 Grid Management Requirements

The investigation of knowledge grids leads to a definition of their requirements on an integrated management. This management includes their communication-relevant as well as economically driven tasks, thus forming an approach of multiple management applications utilizing via an Application Programming Interface (API) the underlying platform of the knowledge grid as basic tools (cf. Figure 9) as an open and extensible platform. As this system visualizes an distributed approach, the location of those Management Applications as well as of the Knowledge Grid Platform my be fully decentralized. Based on the model developed above (including the information and functional model), the organizational and communication model are reflected in this section in terms of key requirements. In particular, all requirements taken from the perspective of Dynamic Virtual Organizations (DVO) onto the underlying technology, network operation, and service provisioning determine the focus. Finally, the overall effort for management must be balanced to the granularity of management tasks performed [19].

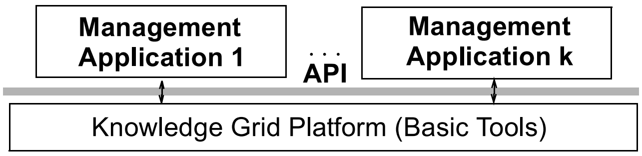

Figure 9: Integrated Knowledge Grid Management

Besides the organizational (in turn economic) and communication (in turn technical) requirements, the focus on invisible grid approaches foresees 2 orthogonal types of such requirements:

- Grid-specific (GS) requirements and
- Abstract-generic (AG) requirements.

Each of those are separated below and discussed in the following to ensure that the defined key functionality is met and most effectively managed.

4.1 Definition of Key Functionality

As defined in Figure 4, 13 grid dimensions determine relevant management tasks. Out of these, the following 6 dimensions addressing key mapping tasks are extracted and additionally 2 overall functions are derived, to form the respective management function:

- Policy: Mapping and management of policies
- SLA: Automated SLA management and negotiation
- Quality: Selectable QoS mechanisms and methods
- Security: Configurable security mechanisms and federated identity management mechanisms
- Resources: Automated mapping of resources and workload scheduling of resources
- Services: Automated provisioning of compound services
- Overall functions:
  - DVOs: Automated management of operations and, *e.g.*, specific join and leave actions
  - Accounting: Definition of accountable units and resource- as well as service-specific data bases

Depending on the grid dimension of the “Organization Type” and the “Resource Type”, real as well as virtual constituents may exist, thus, instantiating those functions above on the physical level or in the abstract level of the Virtual Resource or Virtual Organization. Furthermore, the dependency on the particular “Role”, such as a user, customer, or provider, the specific instance of the management functionality may differ, however, its existence is mandatory.
Moreover, each of those functions offered may show an effect onto the underlying “Cost”, thus, in a consumption-oriented model, where usage of the grid service will be taken into consideration, or in a flat rate manner, where simply the potential use of the grid itself will result in a cost to be paid for. Finally, depending on the proposed knowledge grid management architecture the “Service Type” by itself may consist of grid functionality or management functionality.

4.2 Technical Requirements

Based on the integrated management model and the key functionality defined, the following list of requirements defines the technical ones and their metrics, which include the classification above. It has to be noted, that the dynamicity and flexibility achieved by applying the DVO concept in a knowledge grid will result in a complex management architecture. This architecture needs to deal with temporary contexts, intermediate status information, and unique as well as secure identification of resources and roles. Thus, the particular requirement defined is followed by its type in brackets (GS or AG), its metric applied, and dimensions affected by:

- Enabling effective use of resources in terms of resource virtualization (AG): Specification of resources in use and their accountable units, following the resource and resource type dimension.
- Definition of grid resources (GS): Specification of resources in use and their accountable units, following the quality, resource, resource type, and organization type dimension.
- Enabling efficient use of resources (AG): Definition of resource allocation methods and their accountable units, following the service and service type dimension.
- Definition of grid-based SLAs (GS): Specification of mandatory SLA attributes and service-related parameters, following the SLA and quality dimension.
- Automated DVO SLA negotiation (GS): Specification of a negotiation protocol, following the quality, SLA, and policy dimensions.
- Secured DVO SLA negotiation (GS): Specification of a negotiation protocol, following the security, quality, and SLA dimensions.
- Federated identity management (GS): Specification of an identity management protocol and related system components, following the security, inter-organizational structure, and roles dimensions.
- Definition of accountable units (GS): Specification of resource-independent units reflecting physical resources in use, following the quality and policy dimension.
- Deterministic manageability (AG): Specification of management algorithms, which allow for an unambiguous management result, following the policy dimension.
- Provisioning of flexible, easy-to-use, and simple management interfaces (AG): Definition of an API for the knowledge grid platform, following the roles dimension.
- Running and controlling jobs in a secure and trusted environment (GS): Specification of job allocation mechanisms, following the resources, quality, security, SLA, services, and policies dimension.
- Ensuring fault tolerance (AG): Determination of redundancy, following the resource type, service type, and SLA dimension.
- Definition of DVOs (GS): Specification of DVO operation functions and mechanisms, following the organization type, service type, resource type, and quality dimension.

- Automated DVO operation (GS): Specification of a DVO operation scheme, following the organization type, roles, and policy dimensions.
- Secured DVO set-up (GS): Specification of a negotiation protocol, following the security, organization type, roles, and policy dimensions.
- Enabling an overall multi-provider capability and a support of collaborative environments being essential for distributed grids (GS): Specification of trusted associations between roles and definition of exchange data across those boundaries, following the security, quality, policy, and roles dimensions.

Although this list is already very detailed and long, it may not be fully complete and exhaustive, since technology-specific characteristics, *e.g.*, depending on the specific underlying grid architecture as outlined in Section 2.1 may be required to be added.

4.3 Economic Requirements

Based on the integrated management model and the key functionality defined, the following list of requirements defines the economic ones and their metrics, which include the classification above. Besides the dynamicity of DVOs and all roles involved, the overall economic goal of a commercial grid needs to be maintained. Thus, all requirements are defined as performed above:

- Viable business models (GS): Specification of clear grid services/roles relationships by determining value addition to be achieved, following the roles, service types, services, and cost dimensions.
- Implementable pricing models (GS): Definition of technically feasible and economically viable pricing models based on defined accountable units, following the service type, services, cost, policy, and quality dimensions.
- Suitable SLA to pricing parameter mappings (GS): Specification of existing and accounted for technical parameters of SLAs with their corresponding behavior to pricing parameters; following the SLA, quality, service, and cost dimension.
- Availability of a settlement scheme (AG): Determination of role-based settlements schemes, which may omit the use of any pricing and, thus, accounting to take place due to an economically balanced service exchange, following the roles, service types, and cost dimensions.
- Availability of trust management mechanisms (GS): Specification of dynamic set-up/re-use of trust associations between domains, following the service type, security, policy, roles, and cost dimensions.
- Management Efficiency (AG): Specification of the ratio of management granularity and effort for providing it, following the cost and almost all other dimensions.

4.4 Additional Requirements

Finally, a small number of overall and additional requirements for an integrated grid management may be considered, which have not been particularly addressed above. Thus, they are listed for completeness reasons only:

- Mobility-enabled grids (GS): Specification of effects of mobility protocols and mobile access links.
- Application-specific grids (AG): Determination of basic functionality in support of dedicated application fields.
- Proprietary grids with semi-open interfaces (GS): Tying down of selected interfaces for dedicated grid platforms.
- Embedding into existing network management systems (AG): Tying down of dedicated grid platforms and selected functionality with existing management capacity.

5 DVO Grid Service Management Concept

In order to satisfy the above requirements we propose for the management of dynamic virtual organizations an extension to the OGSA framework.The basic idea behind this approach is to leverage OGSA's concept of resource virtualization, manageability capabilities, discovery and query mechanisms and standardization efforts for protocols and schemes. Since OGSA's major purpose is the support of resource sharing across multiple administrative domains, OGSA provides a context for associating users, requests, resources, policies, and agreements across organizational boundaries. Thus, it is a logical steps to also base the management of DVOs on the (possibly extended) OGSA framework, especially when taking into account that OGSA already provides for global name spaces, for metadata services, for site autonomy, and for resource usage data.

5.1 Managing Gaps of DVOs in Grids

VOs are not a grid-specific concept and as such challenges in managing DVOs and their respective lifecycles are not new. It is well known that the effectiveness of a DVO management mainly depends on the availability of an adequate information base regarding potential VO partners and their "level of preparedness" for a VO involvement. What is new in Grid environments, however, are potentially short lifecycles of DVOs, their formation frequency, and a strong requirement for automated DVO management processes, Thus, the need for a mechanisms support of these tasks in an integrated manner becomes the driving force.

Investigating the VO formation and management processes of current grid projects, following conclusions are drawn:

- All grid projects typically assume the prior existence of prepared communities from where VOs are recruited. This corresponds to the VO breeding approach pursued in ECOLEAD [3].
- Nearly all grid projects restrict their VO management efforts to setting up adequate security mechanisms (*e.g.*, DEISA [16], LCG [15]) rather than addressing the whole VO formation process.
- VO management services typically cover site authorization and access rights management only, compare with, *e.g.*, ICENI [23], VOMS [1]

- There may be different types of DVOs depending on their mission. For example, DEISA uses a project-oriented peer-to-peer approach, LCG uses a hierarchical hub-and-spoke structure, a process-oriented supply chain topology can be found in GMES [11].

Consequently, there is a tremendous gap between those requirements identified above and the reality exemplified by those grid projects. Especially missing is an integrated approach for managing the complete lifecycle of a DVO and its services. An adequate DVO management concept has at least to allow for:

- The formation of DVOs includes the rapid identification of a set of partners best fitting an opportunity (*e.g.*, physics experiments or business cooperations) and quickly configure them into a collaborative network to exploit this opportunity. Finding right partners and establishing necessary conditions for starting the collaboration, however, this not trivial and generally suffers from information lacks and "unpreparedness" of potential partners.
- The definition and the agreement regarding DVOs' mission, strategies, objectives, policies, and business plans includes a description of both service and management requirements of the DVO on all levels of the management pyramid.
- The operational activities of DVOs need to be monitored, controlled, and evaluated during the operation phase against the DVO's mission. Deviations from goals and objectives or any other disturbances may lead to requests for change, *e.g.*, DVO process re-engineering.

5.2 An OGSA-Based Management Architecture

For overcoming these gaps, an OGSA-based management architecture is proposed and shown in Figure 10, separating the management of Grid resources (*e.g.*, rebooting a host or setting VLANs on a network switch) from the management of the resources on the Grid (*e.g.*, resource reservation, monitoring and control), the management of the OGSA infrastructure, which is itself composed of resources (*e.g.*, monitoring a registry service), and the management of DVOs constituting the Grid [2].

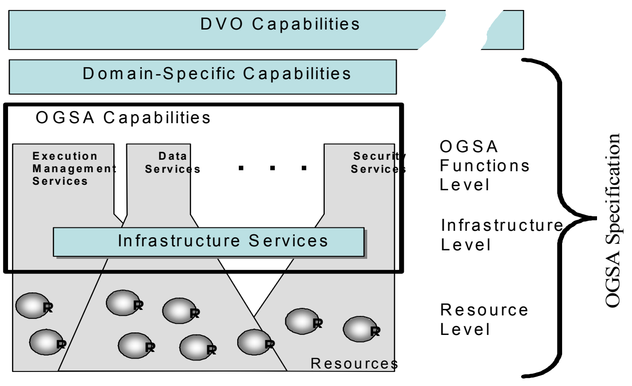

Figure 10: DVO Management Concept

While at the resource level resources are managed directly through their manageability interface (*e.g.*, through SNMP, CIM/WBEM), the infrastructure level provides a base manageability model representing resources as Web Service resources to be monitored and controlled through standard Web Services means. The OGSA functions level provides both, a functional interface and a manageability interface for the management of the respective capability. The resource management capability at this level comprises typical activities of distributed resource management and IT systems management and may be policy-based. The resource management functionality include apart from resource reservation policy management and security management also tasks of VO management.

The management of DVOs is proposed to be placed on the OGSA standard, extended by an additional DVO Capabilities Level for the DVO lifecycle management. Capabilities at the DVO management level need to include lifecycle management services (*e.g.*, formation, configuration and re-formation, decommissioning), operational services (*e.g.*, suspend, resume), monitoring services, security services, and auditing and statistical services. Services expose their capabilities through respective functional interfaces. In accordance with [13] there is also a management interface through which all capabilities are managed, *e.g.*, monitoring of specific registries or monitoring of workflow managers.

5.3 DVO Operation and Use Case

To enable the grid service management concept for DVOs to be applied to the invisible grid, firstly, the operation of such DVOs is defined and outlined, and secondly, a use case is discussed. For that reason, the operation of DVOs is structured in 7 phases:

1. *Initialization* and specification of service offering(s)
2. *External service requests* based on trust relationship settings and own service offer
3. Optional automated *services discovery* to enable potential service composition
4. *SLA negotiation* for service(s) offered and *policy settings*, which is being complemented with selected service accounting definitions to be applied
5. Productive *operation* of the DVO
6. Optional *membership operations* on customers, users, and providers, based on service-specific actions
7. *Tear-down* of the DVO explicitly or implicitly

The grid service management architecture being applied leads to the discussion of a use case, *e.g.*, based on 2 DVOs D_1 and D_2 being initialized based on 3 providers P_1, P_2, and P_3 and 2 users U_1 and U_2. These DVOs utilize real resources offered from the 3 providers. While it is assumed that P_1 to P_3 are of a real organization type, D_1 and D_2 have to be initialized. Within phase 1 their service offerings are proposed, in the example, D_1 offers service S_1 and D_2 offers service S_2. As DVO D_1 requires an additional service S_3 to complete his service S_1, thus forming a simple grid service, he needs to position an external service request in phase 2 to a provider, say P_3, he may know about already. In this case, the trust relationship has been established beforehand. Alternatively, in phase 3, D_1 may have started a service discovery for finding a provider for his particular request of S_3. As a result of phase 2 or 3, the final SLA needs to be negotiated in phase 4 to make sure that the service quality is met. This is complemented by setting the right policies, *e.g.*, on the duration of the resource usage, and the selected accounting principle, *e.g.*, on a usage-based scheme. At this stage, the DVO D_1 goes into operation, as determined by phase 5. Depending on the application field, D_1 may be required to include additional resources during run-time, thus, he performs in phase 6 membership operations, *e.g.*, inviting P_1 to join the DVO. Finally, as the service provisioning has been successful or the task of an application has been fulfilled, within phase 7 the DVO may start the tear-down process and all relationships are opened, resources are released, and accounting processes are finalized.

Within these phase the service management processes will be performed and actions required are undertaken. Thus, the multi-party approach between all roles and organization types has to be applied. As a dedicated use case originating from the e-health environment, Figure 11 shows in an updated manner relationships between multiple roles, organizations, and the formation of the DVO [26]. In this case, the management of roaming agreements, contracts, and of SLA established between all roles will form the key commercial success for the DVO presented.

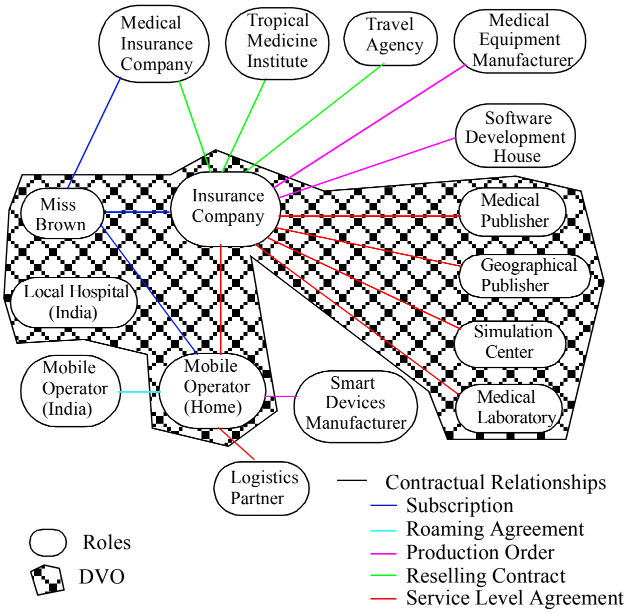

Figure 11: Organizational Alignment of Business Scenario Players [26]

5.4 Grid Service Management Consequences

The consequences for Grid management have already been implicitly stated. In order to emphasize them explicitly, examples of the following management challenges need to be addressed:

- Development of virtualization concepts
- Dynamicity of organizational structure and services
- Change of interaction schemes
- Discovery of services, resources and capabilities in order to form efficient VOs
- Automated VO formation workflows
- Service user-driven: Web services as facts/applications and as management mechanism
- The necessity to develop a DVO MIB.

5.5 Major Management Support Functions

Based on those investigations above, the following set of major functionality has been derived to achieve a service management approach for knowledge grids. Note that those functions are partly reflected from the service provided and partly from the management infrastructure required:

- Mapping of policies between roles, service types, and resource types, specifically between real and virtual resources and services, respectively
- Explicit mapping of resources from virtual onto real ones
- Automated SLA management and negotiation based on cost and policy definitions
- Provisioning of configurable security functions
- Dynamic and static establishment of trust relationships between roles and their respective service or resource
- Definition of generic accounting tasks and general accounting units, depending on resources and services as well as their metering, aggregation, and accounting functions
- Mapping of accounting units for virtual resources onto real resources
- Maintenance of distributed accounting data bases for real and virtual resources and services, respectively
- Dynamic management of VOs, such as for create, teardown, join, leave, and update
- Service composition support in a manual, semi-automated, or automated manner
- Determination of decentralized or centralized management mechanisms depending on the inter-organizational structures in place
- Specification of management information on a per service, service type, resource, and resource type basis

The grid service management approach will be successful only, if the management architecture as well as the service support integrate these functions in an open manner, where standardized interfaces, Management Information Base (MIB) details, and well-defined parameter settings according to those 13 dimensions of Section 3.2 determine the key success factor.

6 Summary and Conclusions

The complex task of managing grid services for next generation invisible grids has been investigated throughout this paper. Based on an overview of related grid architectures and grid mechanisms, the set of key open issues for the particular case of Dynamic Virtual Organizations-based service management have been derived. Due to the large number of influencing factors, a set of 13 key grid service management dimensions has been defined. Those dimensions formed the basis for the specification of grid service management requirements, which have been exploited to outline a grid service management concept for DVOs offering knowledge and invisible grid services.

The discussion and developments of this work show that existing technology is still far away from an integrated service management concept and practical operations of those in a distributed system fail at this stage due to the lack of standardized management interfaces, standardized DVO operations, and standardized management information for grid services. Thus, the set of those 13 dimensions allow for a clear determination of influences, which have been taken up initially in the proposed DVO-based grid service management concept. Of course, at this stage the theoretical approach developed for managing DVOs and their invisible grids will need further attention with respect to services as well as service type specifications and their detailed utilization for composed services. Based on an at least semi-automated services management approach, invisible managed grid services will become possible eventually and their commercial use as well, mainly due to available and open service management schemes, which determine their key enabler.

Acknowledgements

The authors thank members of the Munich Network Management Team for helpful discussions and valuable comments on previous versions of this paper. Special thanks go to Michael Schiffers who provided extensive materials on topics and was always a critical discussion partner. The MNM team, form a group of researchers from the University of Munich, the Munich University of Technology, the University of Federal Armed Forces Munich, and the Leibniz Supercomputing Center of the Bavarian Academy of Sciences (http://www.mnm-team.org).

References

[1] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, A. Frohner, A. Gianoli, K. Lorentey, F. Spataro: *VOMS, an Authorization System for Virtual Organizations;* 1st European Across Grids Conference, Santiago de Compostela, Spain, http://infnforge.cnaf.infn.it/voms/VOMS-Santiago.pdf, February 2003.

[2] K. Balos, K. Zielinski: *JMX-based Grid Management Services;* 1st Workshop on Networks for Grid Applications (GridNets), San Jose, USA, October 29, 2004.

[3] L. M. Camarinha-Matos, I. Silveri, H. Afsarmanesh, A. Oliveira: *Towards a Framework for Creation of Dynamic Virtual Organizations;* Collaborative Networks and Their Breeding Environments (PRO-VE'05), Valencia, Spain, http://www.uninova.pt/~cam/ev/Prove2.PDF, September 2005.

[4] K. Czajkowski, A. Dan, J. Rofrano, S. Tuecke, M. Xu: *Agreement-based Grid Service Management;* http://www.globus.org/alliance/publications/papers/OGSI_Agreement_2003_06_12.pdf, Globus Alliance, 2003.

[5] G. Dreo Rodosek: *A Framework for IT Service Management;* Habilitation Thesis, University of Munich, 2002.

[6] EU Project: *EGEE — Enabling Grids for E-Science;* http://public.eu-egee.org, August 2005.

[7] EU Project: *VOMap — Roadmap Design for Collaborative Virtual Organizations in Dynamic Business Ecosystem;*, IST-2001-38379, May 2003.

[8] I. Foster: *What is the Grid? A Three Point Checklist;* GridToday, Vol. 1, No. 6, June 2002.

[9] I. Foster, C. Kesselman, J. M. Nick, S. Tuecke: *The physiology of the Grid: An open Grid services architecture for distributed systems integration;* http://www.globus.org/research/papers/-ogsa.pdf, January 2002.

[10] I. Foster, C. Kesselman (eds.): *The Grid 2: Blueprint for a New Computing Infrastructure;* Morgan Kaufmann Publishers, ISBN: 1-55860-933-4, 2005.

[11] L. Fusco, J. van Bemmelen: *Earth Observation Archives in Digital Library and Grid Infrastructures;* Data Science Journal, Volume 3, http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS388.pdf, December 2004.

[12] A. Galis, J.-P. Gelas, L. Lefèvre, K. Yang: *A Programmable Network Approach to Grid Management and Services;* International Conference on Computational Science 2003, LNCS 2658, Melbourne, Australia, June 2-4, 2003.

[13] M. Garschhamer, R. Hauck, H.-G. Hegering, B. Kempter, I. Radisic, H. Rölle, H. Schmidt, M. Langer, M. Nerb: *Towards Generic Management Concepts - A Service Model Based Approach;* 7th IFIP/IEEE International Symposium on Integrated Network Management (IM 2001), Seattle, Washington, U.S.A., May 2001.

[14] GERAM — *Generalized Enterprise Reference Architecture and Methodology;* Version 1.6.3, IFIP-IFAC Task Force on Architectures for Enterprise Integration, March 1999. Globus Toolkit; Globus Alliance, http://www.globus.org/toolkit/, August 2005.

[15] LCG Baseline Services Working Group: *LCG Baseline Services Group Report Version 1.0;* Geneva, Switzerland, http://lcg.web.cern.ch/LCG/peb/bs/BSReport-v1.0.pdf, June 2005.

[16] R. Niederberger, O. Mextorf: *The DEISA Project — Network Operation and Support First Experiences;* TERENA Networking Conference 2005, Poznan, Poland, June 2005.

[17] NorduGrid; http://www.nordugrid.org/, August 2005.

[18] Oasis: *Web Services Distributed Management — Management Using Web Services (MUWS 0.5);* http://www.oasis-open.org/committees/download.php/6234/cd-wsdm-muws-0.5.pdf, Committee Draft, April 2004.

[19] D. Oberle, S. Lamparter, A. Eberhart, S. Staab, S. Grimm, P. Hitzler, S Agarwal, R. Studer: *Semantic Management of Web Services using the Core Ontology of Services;* http://www.w3.org/2005/04/FSWS/Submissions/25/paper_html.html, 2005.

[20] B.J. Overeinder, N.J.E. Wijngaards, M. van Steen, and F.M.T. Brazier: *Multi-Agent Support for Internet-Scale Grid Management;* Symposium on AI and Grid Computing (AISB'02), London, UK, April 2002.

[21] D. Pearson: *Accelerating the Deployment of Grids in Enterprise Data Centers;* EGA EMEA Regional Steering Committee Presentation.

[22] F. Sacerdoti, S. Chandra, K. Bhatia: *Grid Systems Deployment and Management using Rocks;* Cluster 2004, San Diego, USA, September 20-23, 2004.

[23] A. Saleem, M. Krznaric, S. Newhouse, J. Darlington: *ICENI Virtual Organisation Management;* UK e-Science All Hands Meeting, Nottignham, UK, http://www.lesc.ic.ac.uk/iceni/downloads/materials/AHM2003/vom.pdf, 2003.

[24] Service Negotiation and Acquisition Protocol, http://www-unix.mcs.anl.gov/~schopf/ggf-sched/GGF5/sched-GRAAP.3.pdf

[25] TeraGrid; http://www.teragrid.org/, August 2005.

[26] M. Waldburger, B. Stiller: *Toward the Mobile Grid: Service Provisioning in a Mobile Dynamic Virtual Organization;* Internal Report, IFI, University of Zürich, Switzerland, August 2005.