WSMO Deliverable

D25.1 v0.1
WSMO and Grid

WSMO Working Draft – November 26, 2004

Authors:
Ioan Toma, Dumitru Roman and Kashif Iqbal

Editors:
Ioan Toma

This version:
http://www.wsmo.org/2004/d25/d25.1/v0.1/20041126/

Latest version:
http://www.wsmo.org/2004/d25/d25.1/v0.1/

Contents

1 Introduction

Two of the most important areas in Computer Science that gathered a lot of interest these days are: Web Services and Grid Computing. Web Services try to solve the problem of enterprise application integration and automatization of business processes by making use of the web as a global infrastructure for distributed computation [Alonso et al., 2003]. They offer a new level of automatization in eWork and eCommerce, where fully open and flexible cooperation can be achieved, on-the-fly, with low programming costs. Grid computing, on the other hand, tries to provide the computational power and data management infrastructure necessary to support the collaboration of people, together with data, tools and computational resources [Foster and Kesselman, 1999]. The most common problems that Grid address are computationally hard and data intensive problems in science and engineering.

A closer look at Web Services and Grid computing shows that these two areas have a lot in common. A resource on the Grid can be view and described as service. Basically both Web Services and Grid computing deals with services and both architectures have the same underling design principles provided by Service Oriented Architecture. Latest directions in Grid and Web Services [Czajkowski et al., 2004] try to provide a unified framework that can deals with both Grid and Web Service requirements. What is missing, in Web Services and Grid areas, is a proper support for machine processable semantics and therefore human intervention is needed to actually discover, combine, and execute services. Semantic Web Services promise to solved this problem. They will provide a fully mechanized web infrastructure for computers interactions [Fensel and Bussler, 2002].

One of the major initiatives that aims to realize the Semantic Web Services vision is the *Web Service Modeling Ontology* (WSMO) [Roman et al., 2004]. WSMO provides the conceptual framework for semantically describing web services and their specific properties. The *Web Service Modeling Language* (WSML) [1] is a formal language for annotating web services with semantic information, using concepts provided by WSMO conceptual framework.

The aim of this document is to investigate how the conceptual model provided by WSMO can be used for Grid computing applications and what variant of WSML is more suitable for Grid requirements. Our focus in this version of the deliverable is to present the current effort in Grid and to identify the ”semantic” requirements of Grid computing applications. We provide as well some initial ideas of how WSMO/WSML can be used in Grid. Future versions will be mainly focused in this last aspect.

The document is structured as follows: Section 2 provides an overview of current efforts in Grid. Section 3 enlists the requirements of Grid community that can be fulfilled by Semantic Web and Semantic Web Service technologies, more precis WSMO. Section 4 tries to identify how WSMO/WSML can be used for Grid requirements that we identified in the previous section. Finally, in Section 5 we conclude the document and present our future work.

[1] www.wsmo.org/wsml

2 Overview of Grid

Grid computing has emerge a as new distributed computing infrastructure technology for advanced science and engineering [Foster and Kesselman, 1999]. The Grid technologies facilitates the global sharing of not just information, but also of tangible assets (computational and data storage resources). Popular technology like the Web or E-mail provide basic communication mechanisms that allow communities to span over states, countries and continents and to work together. But what if they want their data, computers, software and other resources into a single virtual office? These are actually current requirements by a range of collaborative problem-solving and resource brokering strategies emerging in industry, science, and engineering.

Grid seeks to make this possible by providing the protocols, services and software development kits needed to enable flexible, controlled resource sharing on a large scale. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. At the heart of Grid is the concept of **virtual organization** (VO) [Foster et al., 2001]. A virtual organization is a dynamic collection of individuals, institutions and resources bundled together in order to share resources as they tackle common goals. The following are examples of VOs [Foster et al., 2001]: the application service providers, storage service providers, computation providers, and consultants engaged by automobile manufacturer to perform scenario evaluation during planning for a new factory; a crisis management team and the databases and simulation systems that they use to plan a response to an emergency situation; and members of a large, international, multiyear high energy physics collaboration.

The Grid can be viewed as an integrated computational and collaborative environment that preforms different operations on user request. From the user perspective Grid provides a set a services [Baker et al., 2002]:

1. *Computational services* provides support for executing application jobs on distributed computational resources, individually or collectively.
2. *Data services* provide access to distributed sources of data.
3. *Application services* are concern with application management and access to remote applications.
4. *Information services* provide information about computational,data and application services.
5. *Knowledge services* provide support for knowledge management.

Grid computing is an innovative approach that leverages existing IT infrastructure to optimize cross organizational resource sharing in a controlled way. Grids can be further classified into *Computational*, *Data* and *Service* Grids. In addition to facilitate controlled resource sharing across geographically distributed VO, Grid middleware should enable new capabilities to be constructed dynamically and transparently from distributed services. Likewise to engineer new Grid applications it is desirable to be able to reuse existing components and information resources and to assemble and co-ordinate these components in a flexible manner. For this reason the Grid is moving away from a collection of protocols to a service-oriented approach: the Open Grid services Architecture (OGSA) [Foster et al., 2002]. This architecture uses the Web Services Description Language (WSDL) to describe Grid services. OGSA aims to unite Web services with Grid requirements and techniques.

3 Grid requirements

In this section we try to identify the main requirements of Grid services with main emphasis on "sematic" requirements. Many of the requirements of Grid services overlap with Web services requirements but there are requirements that are specific to Grid services only.

As explained in Section 2 currently Web Services and Grid efforts are converging. Early work in Grid was centered in defining a suite of *intergrid protocols* [Foster et al., 2001], but more recent Grid efforts shift to service-oriented approach [Foster et al., 2002]. Both communities make use of the same fundamental concept called *service*: *Web service* in WS community, *Grid service* in Grid community. A Web service is any piece of software that can be invoked over the Internet and uses a standardized XML messaging system. A Grid service is a service that is compliant to the Open Grid Service Infrastructure specification [Tuecke et al., 2003], and which exposes itself through a Web Services Description Language (WSDL) interface [Christensen et al., 2001]. Grid services actually extend Web services in three ways [Grimshaw and Tuecke, 2003]:

- Grid services use a two layer naming scheme. A service instances is identified by abstract name called Grid Service Handle (GSH), abstract name that is resolved into a Grid Service Reference (GSR).
- Grid services required a minimum set of functions and data elements.
- Grid services introduce explicit service creation and lifetime management.

Both, web services and grid services, need machine processable semantics that can be used to support automatic discovery, selection, mediation and composition into complex services. In consequence is important how Grid services are described because this has a major influence on how discovery, selection, mediation and composition are performed. Another important aspect in Grid applications is the huge amount of information. This information could be located in different locations, have different formats, interfaces, structures, etc. It is obvious that Grid have to consider the information integration problem. The current level of interoperability in Web Services and Grid areas must be developed forward by making data understandable and by providing automated support to integration process.

3.1 Conceptual model and language requirements

Major requirements that Grid services are a conceptual model and a language that can be use to "semantically" describe Grid services. Initial efforts in Grid to describe resource and services took a very simple approach. In Condor [Raman et al., 1998] for example requesters and providers advertise their characteristics and requirements in *classified advertisements* (ClassAds). Characteristics are expressed in a attribute/value style. An example is given below:

```
Disk = 10; //gigabytes
Memory = 512; //megabytes
```

The language allows as well to express constrains on the resource advertised/requested. An example is given below:

```
Constraint =
   !member(other.Owner, Untrusted)
      && Rank >= 10
        ?  true
```

```
:  Rank > 0
        ?  LoadAvg<0.3 && KeyboardIdle>15*60
:  DayTime < 8*60*60
        || DayTime > 18*60*60;
```

More recent efforts in Grid description language (eg. [Foster and Liu, 2003], [Solomon et al., 2000]) try to overcome these shortcomings. They address requirements like: resource sets description and matching, ability to describe requirements and preferences and support for query of descriptions. The problem with the previous mentioned approaches is that they work only at syntactic level. Some interesting approaches that make use of Sematic technologies to address Grid problems are described in [Tangmunarunkit et al., 2001] and [Wroe et al., 2003]. In [Tangmunarunkit et al., 2001] a semantic-enabled grid resource matchmaker that exploits ontologies, background knowledge, and rules for solving resource matching in the Grid is proposed. As part of myGrid project, [Wroe et al., 2003] propose to use DAML+OIL ontologies to describe bio informatics services and data.

The need for semantic technologies in Grid community is evident. We believe that WSMO/WSML extend with OGSA features (express explicitly resource and time issues) can provide the solution to ”semantic” requirements of Grid services.

3.2 Grid specific requirements

Although Grid Services and Web Services have many things in common, they still differ in many respects. Web Services have instances that are stateless and non-transient. In contrast, Grid Services can be either stateful or stateless, and can be either transient or non-transient. In [Goble and de Roure, 2002] some major characteristics of Grid services configuration were identified:

- *Dynamic and volatile*: Grid is a very dynamic environment. Grid services may be switched in and out as they become available or cease to be available.
- *Large*: The number of Grid service that are required by different applications at one moment in time could be huge.
- *Long-lived*: A simulation in Grid could take weeks. This implies a long term availability of Grid services.

A complete framework for Grid services must address the issues presented before. In consequence Grid application might involved large numbers of processes interacting in a coordinated fashion and must deal with process that appear and disappear.

4 WSMO in Grid

Web Service Modeling Ontology is a conceptual model for describing services in the context of WWW and is a basis for constructing Semantic Web Service. Due to the fact that WSMO is based on ontologies, the base of the Semantic Web, it has a high potential to automatize service related tasks: discovery, selection, composition, mediation, execution, monitoring, etc.

Started far apart in applications and technology, grid service and web services converged (see Figure 1 for the evolution of them - grid services started with GT1, web services started with xml, soap and wsdl, and finally converged as WS-RF, an extension of web services that considers grid specific requirements), but still they are defined at a syntactical level, without any formal semantic that would make the suitable for automation. A technology is needed that will extend and enrich these service descriptions with new elements that will allow to automatize service related tasks. The conceptual model of WSMO, and WSML, the formal language that reflects the conceptual model, will allow different types of automatic proofs to be made in this language, thus providing a certain level of automation for service related tasks.

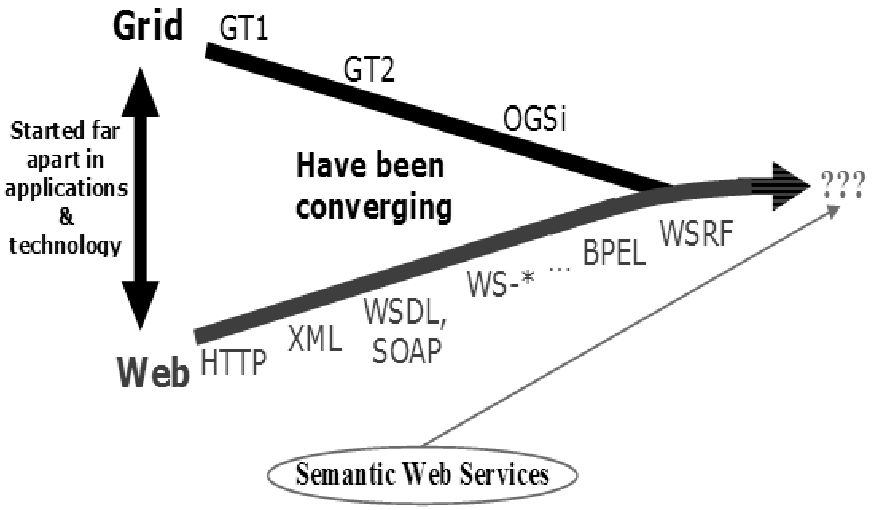

Figure 1: Convergence of Web Services and Grid Services

The OGSA framework, the conceptual model for grids, defines different types of services with specific capabilities that are needed for grid applications. However, OGSA doesn't provide a formal language for describing these services, thus being of little use in automatic performance of different service related tasks. Current languages used in grid, like GRAM (grid resource allocation manager), are based on XML and XML-schema, thus inheriting all its drawbacks (semi structured data format, no formal semantics, no reasoning support, etc); here is where OGSA could benefit from the conceptual model of WSMO and its associated language, WSML.

All the OGSA services, which are summarized below could employ WSMO for semantically describing thir properties:

- Infrastructure Services - OGSA leverages Web services architecture to implement a SOA. WSDL is used for service descriptions. SOAP is the communication protocol. Here WSMO/WSMX will be helpful. Initially this infrastructure was based on OGSI but now it will be based on WSRF.

- Execution Management Services - These services deal with the problems of task initiation and management. Grid resource broker uses the status information for each resource stored in monitoring and discovery service to discover a resource and initiate a job or set of jobs on that resource. After scheduling jobs, it also gathers information on the status of jobs.

- Data Services - Data services in the grid are responsible for efficient data access, data consistency, data persistency, data integration and data location management.

- Resource Management Services - These services allow the management of individual resource itself, management of resources in Grid (i.e. resource reservation, monitoring and control) and monitoring of Grid infrastructure which consists of resources as well like monitoring the registry service.

- Security Services - Security services provide controlled access to resources which can be in various administrative domains with different access and security policies. Grid Security Infrastructure (GSI) specifies the whole set of protocols and security architectures that are required for controlled resource sharing across the Grid.

- Self-Management Services - Self-management services include SLA, policies and service level manager model. SLA includes business and IT agreements between the provider and user of the service. Policies are used to govern the behavior of an SLM (Service Level Manager) and the manageable resources under its control. Service Level Manager Model provides the interface such that various human operators and SLM can work together without having knowledge about each other built in at design time.

- Information Services - Monitoring and Discovery Service in OGSA is an XML database called Xindice collect information from each of the resource in Grid. It Stores dynamic data for monitoring and discovery. It also acts a registry for grid services. XPath and XQuery languages are used to query data from information service.

5 Conclusions and Future Work

In this version of the deliverable we have provided a first overview of how WSMO/WSML can be used in Grid environment settings. We have presented the requirements that Grid applications have and how WSMO and WSML can help in these regards. As described in Section 3, the main requirements of Grid application are: a conceptual model and a language to semantically described grid services and gird resources. We belief WSMO/WSML with some restrictions/extensions will provide the conceptual model and language for Grid requirements.

The future work will include a more detailed evaluation of the conceptual model and language required to semantically describe Grid services and resources. We will investigate in more depth how WSMO conceptual model and WSML languages can be used in Grid settings.

6 Acknowledgements

The work is funded by the European Commission under the projects DIP, Knowledge Web, InfraWebs, SEKT, SWWS, ASG and Esperonto; by Science Foundation Ireland under the DERI-Lion project; by the Vienna city government under the CoOperate programme and by the FIT-IT (Forschung, Innovation, Technologie - Informationstechnologie) under the projects RW^2 and TSC.

The editors would like to thank to all the members of the WSMO, WSML, and WSMX working groups for their advice and input into this document.

References

[Alonso et al., 2003] Alonso, G., Casati, F., Kuno, H., and Machiraju, V. (2003). *Web Services*. Springer.

[Baker et al., 2002] Baker, M., Buyya, R., and Laforenza, D. (2002). Grids and Grid technologies for wide-area distributed computing. In *International Journal of Software: Practice and Experience (SPE)*, volume 32, pages 173–203, USA. Wiley Press.

[Christensen et al., 2001] Christensen, E., Curbera, F., Meredith, G., and Weerawarana, S. (2001). Web services description language (wsdl) 1.1. http://www.w3.org/TR/wsdl.

[Czajkowski et al., 2004] Czajkowski, K., Ferguson, D., Foster, I., Frey, J., Graham, S., Sedukhin, I., Snelling, D., Tuecke, S., and Vambenepe, W. (2004). The WS-Resource Framework.

[Fensel and Bussler, 2002] Fensel, D. and Bussler, C. (2002). The web service modeling framework WSMF. *Electronic Commerce Research and Applications*, 1(2):113–137.

[Foster and Kesselman, 1999] Foster, I. and Kesselman, C. (1999). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann.

[Foster et al., 2002] Foster, I., Kesselman, C., Nick, J., and Tuecke, S. (2002). The physiology of the grid: An open grid services architecture for distributed systems integration.

[Foster et al., 2001] Foster, I., Kesselman, C., and Tuecke, S. (2001). The anatomy of the Grid: Enabling scalable virtual organizations. *Lecture Notes in Computer Science*, 2150:1–26.

[Foster and Liu, 2003] Foster, I. and Liu, C. (2003). A constraint language approach to grid resource selection. unpublished menuscripts. Technical report, University of Chicago. Available from `http://www.cs.uchicago.edu/files/tr_authentic/TR-2003-07.pdf`.

[Goble and de Roure, 2002] Goble, C. and de Roure, D. (2002). The Grid: An application of the Semantic Web. In *ACM SIGMOD*, volume 31, pages 65–70.

[Grimshaw and Tuecke, 2003] Grimshaw, A. and Tuecke, S. (2003). Grid Services extend Web Services. Available from `http://www.sys-con.com/webservices`.

[Raman et al., 1998] Raman, R., Livny, M., and Solomon, M. H. (1998). Matchmaking: Distributed resource management for high throughput computing. In *HPDC*, pages 140–147.

[Roman et al., 2004] Roman, D., Lausen, H., and Keller, U. (2004). Web Service Modeling Ontology – Standard (WSMO-Standard). Working draft, Digital Enterprise Research Insitute (DERI). Available from `http://www.wsmo.org/2004/d2/v1.0`.

[Solomon et al., 2000] Solomon, M., Raman, R., and Linvy, M. (2000). Resource management through multilateral matchmaking. In *Proceedings of the Ninth IEEE Symposium on High Performance Distributed Computing (HPDC9)*, number 5, page 290291, Pittsburgh, Pennsylvania.

[Tangmunarunkit et al., 2001] Tangmunarunkit, H., Decker, S., and Kesselman, C. (2001). Ontology-based Resource Matching in the Grid - the Grid meets the Semantic Web. *IEEE Intelligent Systems*, 16(2).

[Tuecke et al., 2003] Tuecke, S., Czajkowski, K., Foster, I., Frey, J., Graham, S., Kesselman, C., Maquire, T., Sandholm, T., Snelling, D., and Vanderbilt, P. (2003). Open grid service infrastructure 1.0.

[Wroe et al., 2003] Wroe, C., Stevens, R., Goble, C., Roberts, A., and Greenwood, M. (2003). A suite of DAML+OIL ontologies to describe bioinformatics web services and data. *International Journal of Cooperative Information Systems*, 1(12):597–624.