A Grid Service Framework for Metadata Management in Self-e-Learning Networks*

George Samaras, Kyriakos Karenos, and Eleni Christodoulou

Department of Computer Science, University of Cyprus
{cssamara,cs98kk2,cseleni}@ucy.ac.cy

Abstract. Metadata management is critical for Grid systems. More specifically, semantically meaningful resource descriptions constitute a highly beneficial extension to Grid environments that started to gain significant attention. In this work we contribute to the effort of enhancing current Grid technologies to support semantic descriptors for resources – termed also the *Semantic Grid.* We use a *Self e-Learning Network (SeLeNe)* as the testbed application and propose a set of services that are applicable in such a case in alignment to the *Open Grid Services Architecture (OGSA).* We concentrate on providing services for the utilization of Learning Objects' (LO)[1] metadata, the basic of which, however, are generic enough to be utilized by other Grid-based systems that need to make use of semantic descriptions. Different service placement scenarios produce a number of possible architectural alternatives.

1 Introduction

Grid Technology has found uses in a wide area of applications that usually address large scale, process and data intensive problems. Our effort is to bring data-centric services adjusted to the Grid environment and to expand its functionality in the area of resource sharing using e-Learning as the testbed application. As we elaborate in section 2, we consider metadata management (viewed as semantically meaningful resource descriptions of learning material), crucial especially as the Grid expands to be supplemented with capabilities towards supporting (and incorporating) technologies from the Semantic Web, termed the "Semantic Grid" [1] under the guidelines of the Global Grid Forum (GGF) [2].

Our work derives from our IST project SeLeNe: The SeLeNe Project is aiming to elaborate new educational metaphors and tools in order to facilitate the formation of learning communities who require world-wide discovery and assimilation of knowledge. To realize this vision, SeLeNe is relying on semantic metadata describing educational material. SeLeNe offers advanced services for the discovery, sharing, and collaborative creation of learning resources, facilitating a syndicated and personalised access to such resources.

---

* This work has been supported by the E.U. Project "SeLeNe: Self e-Learning Networks", IST-2001-39045.

[1] A LO is generally defined as an artifact in *digital* form utilized during the learning process.

M. Dikaiakos (Ed.): AxGrids 2004, LNCS 3165, pp. 260–269, 2004.

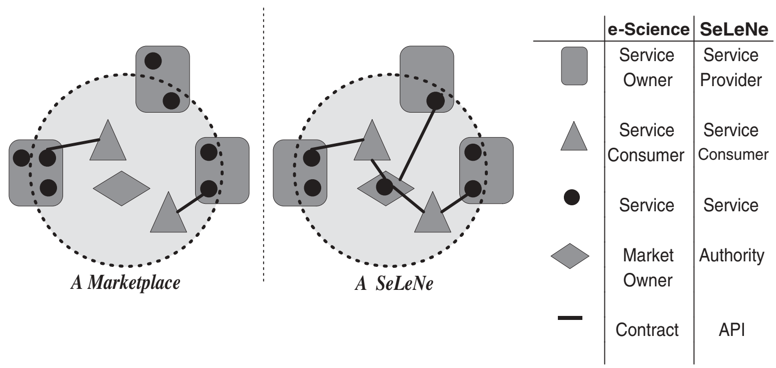


Fig. 1. SeLeNe and the Semantic Grid

Service-based educational systems open new ways in the usability of the Grid as their primary requirements include the provision of adequate services for sharing, syndicating heterogeneous resources and relevant content discovery. Efforts are already under way: In [3] an attempt is made to provide an infrastructure for future eScience. Of our interest in this work, is the adoption of a service-based perspective to meet the needs of a global and flexible collaborative system for educational purposes. The Grid is described as a collection of service providers and service consumers brought together in a "marketplace", initiated and managed by a "marketplace owner". We parallelize the "marketplace" to a SeLeNe (Fig.1) and the "marketplace owners" to the system point of entry, which provides reliability in accessing the system. These we later refer to as Authority sites.

It is envisioned that the SeLeNe service layering will enable generic services to be provided which will support the high level Application-specific services. OGSA GridService [4] can be adapted to the requirements of an e-learning environment as it provides a process-oriented model on which our data-oriented model services will be based. However, the various services – as described in the OGSA Layers – need not be deployed within every physical site. Each node may create and offer different services to the system, included in a predefined set. It is apparent, also, that services may require the collaboration of many sites (e.g. a query service) functioning in a distributed manner.

An educational environment such as the one envisioned by SeLeNe, however, exceeds the requirement of a standard client-server Grid model. Firstly, information sharing must be extended to the semantic level. The semantic extension of the SeLeNe-offered services will aim to address the diversity among consumers and producers of LO descriptions (in addition to services) in terms of ontological contexts. These requirements would require high coupling among services and the ability for the combination of these services towards the completion of specific e-learning tasks. In addition, the need for personalization - which requires each participating learner or site to be viewed as individual - and for collaboration - which requires for a global view of the SeLeNe components' interaction – impose a model that should functionally allow for the handling of both cases.

Although we do view the problem through Grid lenses, we identify the need to incorporate techniques from both Grid and P2P technologies. Efforts have already been initiated towards the incorporation of P2P capabilities to the OGSA framework by the GGF community. Although currently efforts are still at an early draft stage, one can clearly see the practical need for P2P-usable OGSA [5].

To this end, the most relevant work to SeLeNe is done within the SWAP project [19], which combines P2P and Semantic Web technologies to support knowledge sharing. Web Services technologies [20] provide an excellent infrastructure on which SeLeNe services can be build. However we also consider other alternatives, especially in the light of P2P/Grid requirements mentioned earlier. The JXTA project framework [21] offers a purely Java-based services core and concentrates on a P2P-oriented model. On the other hand the Globus project [17] provides a range of basic services for the construction of Grid-based systems. These technologies have been studied extensively as part of a number of previous works [19, 16, 18]. Herein, we provide the definition of the required services and the architectural model that would suit the user requirements, assigning much less weight on the possible future implementation alternatives.

It is important to note that some vital services need to be available at all times (e.g. registration, mediated querying etc) as well as the fact that we need to provide some method for information integration. Therefore we propose that "authority" sites should be present that will be more reliable and may acquire the role of mediator (e.g. to interconnect related sites by clustering) or coordinator (e.g. to support ontology mappings when and if necessary).

2 An OGSA-Guided, Metadata-Centric Architecture

It has already been mentioned that SeLeNe is concentrating on the management of LO metadata having in mind. Based on the OGSA service layering we construct, in this section, a corresponding layered set of services required for a SeLeNe[2].

Management and manipulation of LO metadata is at least as important and critical as LO management itself. In an educational system, content descriptions are crucial in order to provide a uniform method for the discovery of LOs relevant to the user's queries and for combining multiple such descriptions to realize specific tasks that lead, eventually, to supporting the learning objectives. Additional requirements such as personalization support, change notification and automatic/semi-automatic metadata generation are only indications of the demanding nature of metadata management. When addressing large data set services in OGSA layering, metadata handling is usually present at the Resource and Collective layers. In our case, as metadata is the actual shared resource and due to the mentioned requirements we believe that it is required to provide metadata services covering all layers. For example there exists the need for descriptions of LOs to be accessed, manipulated and stored in an RDF repository

[2] User Requirement Analysis for Self e-learning Networks is available in [8] as further work in the SeLeNe project.

(i.e. the Repository's API). This is suitable to be included to the Fabric layer services since a number of different storage alternatives may be present. On the other hand there exist high-level services that will support Trails and Personalization of LO descriptions (i.e. adaptation of the learning material based on specific user profiling and paths followed in the LO space during the learning process) that need to be placed at the Application layer.

Table 1. SeLeNe Services

<table>
<tr><th></th><th>Service Name</th><th>Core</th><th>OGSA Layer</th></tr>
<tr><td rowspan="7">App. Specific</td><td>Presentation</td><td></td><td rowspan="5">Application</td></tr>
<tr><td>Collaboration</td><td></td></tr>
<tr><td>Trails & Personalization</td><td></td></tr>
<tr><td>User Registration</td><td></td></tr>
<tr><td>LO Registration</td><td></td></tr>
<tr><td>View</td><td></td><td rowspan="5">Collective</td></tr>
<tr><td>ECA</td><td></td></tr>
<tr><td rowspan="8">Generic</td><td>Query</td><td>X</td></tr>
<tr><td>Update</td><td></td></tr>
<tr><td>Syndication</td><td></td></tr>
<tr><td>Locate</td><td></td><td rowspan="2">Resource</td></tr>
<tr><td>Information</td><td>X</td></tr>
<tr><td>Sign-on</td><td></td><td rowspan="2">Connectivity</td></tr>
<tr><td>Communication</td><td>X</td></tr>
<tr><td>Access</td><td>X</td><td>Fabric</td></tr>
</table>

2.1 Service Classification

It is understandable that not all services can be deployed at each and every SeLeNe-participating site. However, we feel that it is a requirement that there should be an as-small-as-possible set of specific services that each SeLeNe site will assume present in all other SeLeNe sites. The basic reason for this is to make sure that at least communication and discovery of available services will be possible as soon as a single SeLeNe site is identified as an entry point. These services, we can call *Core Services*. Additional *Appended Services* will be present in order to complete the larger percentage of SeLeNe functionality.

One can clearly see that the proposed interfaces in OGSA (GridService, Notification, Registry, Factory, HandleMap [6, 4]) are, to a major degree, process-centric. In SeLeNe, however, RDF metadata is the actual resource and for this reason we should provide additional or adapted interfaces to meet a, to a large extent, data-oriented system.

What we are envisioning is that the set of proposed services will be possible to be deployed in alignment to the OGSA guidelines (and possibly over widely used grid technologies such as Web Services and Globus) but also extended to provide additional functionality that is missing from today's Grids but required by an e-learning network (i.e. P2P support and expanded semantic (RDF) metadata usage). In this sense, as proposed in section 3.2, existing infrastructure can be utilized to mediate the underlying service functionality described below but targeting to support RDF descriptions as the requested resource (i.e. instead of computation cycles, storage, large data objects etc.). As argued next, generic RDF services can then be adopted by other grid systems. Besides characterizing services as being either core or appended, one other important distinctive factor for offered services is whether a service is generic or application specific (i.e. SeLeNe specific). *Generic* services will reside at the "hourglass neck" of the OGSA layers. These services will be usable for other applications or systems that require or make use of RDF. Examples of generic services include RDF view creation and change notification. On the other hand, *application specific* services concentrate on the specifics within the SeLeNe with respect to the e-learning requirements such as trail management and personalization services. In the following subsections we will describe the high level functionality for each of the proposed services.

2.2 Core Services

Access Service. This service is located at the lower layer of the Grid Architecture (Fabric). This service provides the direct access API to the local RDF Repository. It includes access methods for local requests as well as appropriate manipulation of locally stored descriptions (i.e. insert, delete and update of the repository content) irrespective of its low-level implementation. The actual storage repository can be realized over a number of implementation alternatives such as Sesame RDF Repository [9], Jena toolkit [11] and the ICS-FORTH RDFSuite [10].

Communication Service. This service provides the basic communication mechanisms for exchanging data. Current protocols may be used on which communication links can be established (such as TCP/IP) but we should also consider creating a simple "SeLeNe specific" communication service (i.e. for the exchange of specific types of messages e.g. task request submission.) Possible example technologies that can support this "SeLeNe specific" communication service are SOAP [12] and RPC techniques (e.g. Java RMI), however the message content and structure is not part of our current investigations. RPC is generally more appropriate for more formalized and concrete (e.g. local) communications and can be used in a local SeLeNe (e.g. installed at a single institution). On the other hand SOAP addresses incompatibility problems among multiple communicating and possibly remote groups.

Information Service. The Information service provides the capabilities of acquiring descriptive information on some SeLeNe site. Informally, it will be able to answer questions of the form: "what does this node understand in terms of metadata?" It provides the profile of the site (not the user). Put in another way, it provides metadata on metadata and more specifically the Namespaces used and the RDF Schema(s) for that specific site. The Information service is built on top of the Access service. It does not raise any new research issues for us.

Query. The Query Service is of great importance: we need to define a powerful query language that will allow for the extraction of results from multiple, local RDF repositories. The Query Service should be distributed and should allow for search message routing in order to forward sub-queries to sites that can provide answers. It may also need to call the Syndication service to translate queries expressed against one RDF taxonomy to sub-queries expressed against different local taxonomies. It then passes a subquery to the Access service supported by a particular peer, expressed in terms of that peer's local RDF Schema. Another issue is the exploitation of the semantic meaning of our data to relate users of similar interests. A good, super-peer based technique is provided in [13] where a clustering technique is used to mediate heterogeneous schemas. Authority sites can become responsible for keeping semantically meaningful indexes about other neighboring sites.

2.3 Appended Services

Sign-On. A site is able to register to the SeLeNe in order to advertise its content and services. Also, in this way, it should be able to make its presence known to other sites. Sign-on allows for the update of the indexes of neighbors as well as the directly connected authority site(s).

Locate. This service relates to the OGSI GridService and makes requested service lookup possible. As soon as a site is connected, it should be able to discover where there are services that will be used, along with any required parameters that these services will need. We assume for now standard registry techniques depending on the architectural deployment of SeLeNe. A distributed cataloging scheme could suffice in this case (e.g. UDDI [15].) Semantic service descriptions is an issue not addressed within SeLeNe for now although it does pose an interesting future research issue for the evolution and expansion of the proposed set of services.

Syndication. The Syndication service is responsible for the translation between different RDF schemas. This is accomplished by using the user-supplied mappings between heterogeneous schemas. This implies both data-to-data and query-to-query translations. Syndication issues are also of high importance.

Update. The Update Service is used to appropriately transfer updates to descriptions expressed in diverse schemas. By analogy to the Query service, this service will take an update request for Peer 1 expressed in some RDF_Schema_2 and translate it into the equivalent update expressed in terms of RDF_Schema_1 by using the Syndication service. The Update Service would then request for the invocation of the appropriate operation of the Access service at Peer 1 to enact the actual update on its local RDF repository.

Event-Condition-Action (ECA). LO descriptions are gradually updated and enhanced due to the ongoing learning process. Users should be able to register their interest to receive changes when they occur that are relevant to metadata that are of their interest. This feature should be provided by the ECA Service, which will propagate updates and notifications to registered sites.

View. The View Service provides the functionality of creating personalized views by structuring (and re-structuring) virtual resource descriptions among the SeLeNe LO descriptions' space. By this way we allow for the user to actually built-up her own virtual learning environment which she can navigate and expand. The View Service will can be realized over RVL that is able to, additionally, allow the definition of virtual schemas and thus amplifies the personalization capabilities of the SeLeNe system.

LO Registration. This service provides the API for submitting a new LO by providing its description to the SeLeNe. Storing LO descriptions is handled by the use of the Access service. The registration process makes use of the Syndication service and allows the registration of both atomic and composite LOs.

User Registration. The user will be registering to a SeLeNe in order to create and later use her profile and thus acquire a personalized view of the system. User descriptions are also stored using the Access service. Issues of costing are not considered at this moment as we focus mainly on the personalization/profile creation aspect of the User.

Trails & Personalization. The Trails & Personalization Service is related to a specific user or group of users. It is concentrated on the educational characteristics of the user and provides the API to extract user-profiling information. It is proposed that this service should run as a user-side agent when possible while trails could be formed and managed by message exchanging of the participating person or group agent or agents.

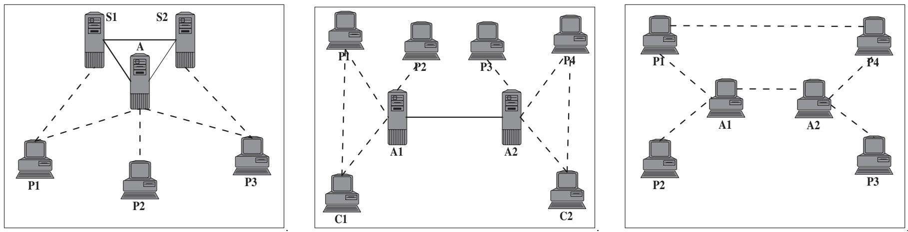
Collaboration. A Collaboration Service should allow the communication between users and groups of users and it is proposed that this is mediated by a central authority site. At least two sites should request the creation of a collaboration session and others may be added later. Collaboration services may include already available systems such as Blackboards, Message Boards, CVS (for collaborative code writing) or e-mail and instant messaging services. The SeLeNe Collaboration Service lies above these services in order to provide connections to other SeLeNe services.

Presentation. Based mainly on the user profile, the Presentation service should be able to produce graphical visualization of metadata. This could, for example, be a RDF graph. It could also be produced locally or via a web-based engine. Since visualization and presentation are highly related to the learning experience itself, there is no simplified methodology for it and will most possibly require much work.

3 Approaches to Service Placement

3.1 Architectural Models

In Figure 2 three models are shown: Continued lines represent direct connections between sites while discontinued lines represent possible connections established



(a) Centralized (b) Mediation-based (c) Autonomic

Fig. 2. Service Placement Approaches

due to service calls. Detailed interaction flows among site service calls with respect to the services proposed can also be found in [22].

One first approach is to take a look at the *Centralized* scheme. In such an architecture, a number of "fixed" service providers exist which are highly available and powerful enough to accommodate a large number of services. The centralisation has to do with the fact that the greater percentage computation and the totality of the RDF descriptions storage are found at a centralised location.

Provider servers are connected and together they provide a service provision cluster. Clients (or consumers) connect to the cluster via a specific entry point or an Authority. Metadata located at consumer sites need to be registered at any cluster server. In this sense, servers act as metadata repositories for LOs. Query and Integration/Mediation services are provided for metadata among the servers and replies are sent back to the requester. Since all tasks are handled within the group of servers, consumer sites are not actually aware of each other. This strategy is similar to a brokering system such as EducaNext/Universal [7].

In a *Mediation-based scheme*, consumers and producers (of both LOs and Services) are logically clustered around mediators/brokers that in our case will be taking the role of Authorities. This is also similar to the Consumer-Broker-Producer model (in terms of services) and also resembles the super peer scheme (in terms of content). The reason for this model to be named Mediation-based is due to the fact that its functionality is primarily facilitated by mediator machines, similar to "Brokers"/"Authorities." Authorities are affiliated with a number of "Providers" that become known to them. Sites may be both LO producers and providers but need to register their content to a broker which will provide the means for communication with other sites by creating logical communities. This last characteristic is highly desirable in SeLeNe. Edutella [16] is a mediation-based educational system built on the the JXTA infrastructure.

An *Autonomic* system is characterized by the fact that each site is autonomous in terms of service provision (i.e. each site may provide any number of services). In such cases, a core services requirement is the existence of a Service Discovery protocol (such as the previously described Discovery Service), which should be completely distributed. Metadata is maintained at each site and there is no centralization. Therefore, a distributed and possibly partially replicated

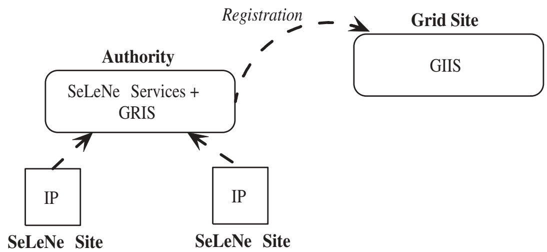


Fig. 3. An example of SeLeNe Services over Globus Information Services

metadata catalog should exist to address intermittent connectivity issues. One such autonomic (P2P) approach is found in the SWAP project. The core difference however is that SWAP is component-based, not service-based.

It is noted that extensive support for P2P environments will require a new global infrastructure [5]. Therefore, in addition to these efforts it is expected that the new version of the Globus Toolkit (GT3) [17] will adopt open protocol standards also applied in Web Services technologies. An improved OGSA specification in combination with GT3 support for standard technologies will bring this goal closer to realization.

3.2 Proposed Initial Globus Integration

The most relevant components to resource discovery and Grid services' information are the Globus *Information services* [14] or *Monitoring and Discovery Service (MDS)*. We omit description of this service due to space limitations.

Although SeLeNe services are self-contained (providing registration, query, syndication, adaptation service), still, it is extremely difficult to claim the replacement or even direct integration of semantic resource descriptions with Globus MDS. One alternative could be implementing SeLeNe services as completely independent entities, i.e. as additional Grid Application, OGSI-compliant Services. Below we provide a possible set up, depicted in Fig. 3.

- SeLeNe sites act as Information Providers, (IPs) where Information are the descriptions available at the local repositories. It is assumed that Core SeLeNe services run on these sites including Information and Access services, essential for this functionality.
- The Grid Resource Information Server (GRIS) runs on Authority sites. SeLeNe IPs register resource descriptions to the Authorities. Note that Authorities can be providers themselves. Authorities, thus act as "gateways" to the rest of the Grid.
- GRISs Register with *any* available Grid Information Index Server (GIIS). In this way SeLeNe services are made accessible to external users by queering the GIIS.

4 Conclusion

The usage of semantic metadata resource descriptions can highly benefit Grid technology. In our work within the SeLeNe project we have proposed a set of core and appended services that allow for the query, access and syndication of heterogeneous RDF-based descriptions and propose the incorporation of such services to the current Grid Infrastructure. We use an educational e-learning application as a testbed and find that the usability of such a service set can be applied to multiple architectural models. We believe that semantic metadata for the Grid constitutes a critical extension towards the realization of the Semantic Grid vision.

References

1. The Semantic Grid Community Portal. Available at `www.semanticgrid.org/`
2. Global Grid Forum (GGF). Available at `www.gridforum.org`
3. D. De Roure, N.R. Jennings, N.R. Shadbolt. The Semantic Grid: A Future eScience Infrastructure. National e-Science Centre. Univ.of Edinburgh, *UK. UKeS 2002.*
4. I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The Physiology of the Grid: An open grid services architecture for distributed systems integration. *Technical report, Open Grid Service Infrastructure WG, GGF, Jun. 2002.*
5. The OGSAP2P Group. Available at `www.gridforum.org/4-GP/ogsap2p.htm`
6. Ian Foster, Carl Kesselman, Steven Tuecke. The Anatomy of the Grid Enabling Scalable Virtual Organizations. *International J. Supercomputer Applications, 2001.*
7. EducaNext (Universal) Project. Available at `www.educanext.org`
8. K. Keenoy, G. Papamarkos, A. Poulovassilis, M. Levene, D. Peterson, P. Wood, G. Loizou. Self e-Learning Networks - Functionality, User Requirements and Exploitation Scenarios, *SeLeNe Project Del. 2.2*, Aug. 2003.
9. Sesame Open Source RDF Schema-based Repository and Querying facility. Available at `sesame.aidministrator.nl`
10. ICS-FORTH RDFSuite. High-level Scalable Tools for the Semantic Web: Available at `http://139.91.183.30:9090/RDF/index.html`
11. Jena Semantic Web Toolkit. Available at `www.hpl.hp.com/semweb/jena.htm`
12. Simple Object Access Protocol (SOAP). Available at `www.w3.org/TR/SOAP`
13. Leonidas Galanis, Yuan Wang Shawn, R. Jeffery David J. DeWitt. Processing XML Containment Queries in a Large Peer-to-Peer System. *CAiSE 2003*
14. K. Czajkowski and S. Fitzgerald and I. Foster and C. Kesselman, Grid Information Services for Distributed Resource Sharing, *HPDC 2001.*
15. The Universal Description, Discovery and Integration Protocol. Available at `www.uddi.org`
16. Edutella Project. Available at `edutella.jxta.org`
17. Globus Toolkit. Available at `www.globus.org`
18. OntoWeb: Available at `www.ontoweb.org`
19. The SWAP System. Available at `swap.semanticweb.org/public/index.htm`
20. Web Services at W3C. Available at `www.w3.org/2002/ws`
21. Project JXTA. Available at `www.jxta.org`
22. The SeLeNe Consortium. An Architectural Framework and Deployment Choices for SeLeNe. *SeLeNe Project Del. 5.0. Dec. 2003.*