Towards a Semantic Grid Architecture

Ioannis Kotsiopoulos, Pinar Alper, Sean Bechhofer, Oscar Corcho, Carole Goble, Dean Kuo and Paolo Missier

The School of Computer Science
The University of Manchester,
Manchester, UK
<ioannis, penpecip, seanb, ocorcho, carole, dkuo, pmissier> @cs.man.ac.uk

María de los Santos Pérez Hernández

Universidad Politécnica de Madrid
Campus de Montegancedo,
Boadilla del Montem
Madrid, 28660
mperez@fi.upm.es

The Semantic Grid is an extension of the current Grid in which information and services are given well defined and explicitly represented meaning, better enabling computers and people to work in cooperation [1]. Existing Grid Services deal with knowledge in the form of metadata and its associated semantics in an implicit fashion, providing no way to share this knowledge with other Grid components. Semantic Grids not only share computational and data resources, but also explicitly share and process metadata and knowledge.

In the last few years, several projects have embraced this vision and there are already successful pioneering applications that combine the strengths of the Grid and of semantic technologies [2]. However, the Semantic Grid currently lacks a reference architecture, or a systematic approach for designing Semantic Grid components or applications.

We propose a Reference Semantic Grid Architecture (RSGA), extending the Open Grid Services Architecture, by (a) explicitly defining the mechanisms that will allow for the explicit use of semantics and (b) defining the associated knowledge to support a spectrum of service capabilities. We call the semantically enhanced OGSA "Semantic OGSA", or S-OGSA.

The OntoGrid S-OGSA architecture has a model of semantic resources, which extends the model of a Virtual Organisation and the OGSA model [3]; and two architectural pillars that support the model of *Semantic Provisioning Services* and *Semantic Grid Entities*.

(a) Semantic Resources - A Semantic Grid Model

A definition of the Semantic Resources that are supplied and consumed amongst the services extends the general model of Grid – such as the model defined in CIM or a VO such as the EarthScience VO [4] or the ^{my}Grid Information Model [5]. The model identifies two group of entities:

- **Grid Entities** (*G-Entities*) are anything that carries an identity on the Grid, including resources and services [6]
- **Semantic Entities** (*S-Entities*) are anything that encapsulates knowledge or is associated with other entities that encapsulate it. They are divided into two groups: Knowledge entities and Semantic Grid entities.
  - **Knowledge Entities** (*K-Entities*) are ontologies, rules, schemas, knowledge bases or even free text descriptions that encapsulate knowledge that can be shared.

- **Semantic Grid Entities** (*SG-Entities*) are Grid Entities that are associated with, or defined as a type of, Knowledge Entities. Grid entities will acquire and discard associations with knowledge entities through their lifetime. Grid entities could potentially be associated with multiple entities of different forms and capabilities, simultaneously. **Semantic Bindings** (*S-Bindings*) are the assertions that link a Grid Entity with a Knowledge Entity to transform it into a Semantic Grid Entity. Semantic Bindings are first class resources with an identity and their own metadata. A Grid Service that has a semantic binding is a Semantic Service; a Grid Resource that has a semantic binding is a Semantic Resource.

(b) Semantic Provisioning Services

They are services that provision semantic entities. These Semantic Services are themselves Grid Services. Following the aforementioned classification of semantic entities, two major classes of services are:

- **Knowledge provisioning services**: software components participating in a service-oriented architecture that can produce (and in some cases store) knowledge resources, and that can be used to manage knowledge resources. These supporting the creation, storage and access of different forms of knowledge resources. For example: ontology services (a major form of knowledge) and reasoning services.

- **Semantic Binding provisioning services**: software components participating in a service oriented architecture that can produce (and in some cases store) S-Binding resources, and that can be used to manage S-Binding resources. For example: semantic binding index services, for accessing and storing metadata associating Grid entities with knowledge entities; and annotation services for generating metadata from different types of information sources, like databases, files or provenance logs. S-Bindings are stateful, so they are subject to soft state processes; i.e. they will time out, get deleted or be removed.

  Semantic bindings can be produced in two broad ways:

  - **Annotation**. The Grid resource is unchanged and annotated with semantics. For example, export data from a data service in XML annotated by RDF. This is the conventional approach of the web. It enables third party annotations on resources for which the annotator does not have write permission; we can reuse semantic web tooling and it minimises the impact of services and interfaces. This physically separates the resource from the knowledge entity. For messages, we can (a) annotate a message with knowledge entity (e.g. markup an XML doc with RDF/OWL) or (b) specify the knowledge entity's identity in the message. A service can fetch the knowledge entity from a service to retrieve the message's knowledge entity. This is one usage of metadata services in semantic Grid.
  - **Encoding.** Encode the data and resources with a semantics based approach, for example export data from a repository in RDF. This is appropriate for content managers and service providers who are able to directly generate RDF. There is a tighter physical binding of the resource and knowledge entity, dissolving the grid resource and the knowledge entity into one physical representation. This potentially impacts tooling and interfaces, though XML to RDF are available.

(c) A framework for Semantic Grid Entities

Semantic Grid entities are supported through a framework for evolving Grid entities to become semantically aware. Grid services should be capable of consuming and producing Semantic Grid Entities and processing them. Moreover, Grid resources should be able to be associated with Knowledge Entities through Semantic bindings.

Therefore, a specification for the S-OGSA infrastructure requires:

- A specification for S-WSRF, or more generally Semantic Stateful Services; that is a definition for the semantic bindings of stateful Web Services, expressed in WSRF or WS-I+, with knowledge entities.
- A specification for semantically enriched messages between Grid Services, that is messages that are semantically bound to knowledge entities.

A framework for new Grid entities to be designed as semantically aware and a migration framework for current Grid entities to become semantically aware have two potential evolutionary mechanisms:

- Semantically annotating existing entities that could facilitate dynamic discovery, dynamic composition or in general the development of "smarter" clients. This is re-factoring *external* to the component and should aim to minimize the impact on the component itself.
- Re-factoring existing services to become (Semantic Grid) Services capable of processing Semantic Grid Entities. This is re-factoring internal to the component, or creating it from scratch, and therefore has a profound impact on the component itself.

We explore the impact of adding semantics on current Grid services and type systems, arguing that a RSGA must be able to host a mixed economy of semantic and non-semantic entities. In particular we argue that Grid Services should be able to take advantage of knowledge on a best effort basis, in order to facilitate incremental acquisition of knowledge in a Grid, and to minimize the impact on Grid services that are not, and do not need to be, Semantic Grid Services. We use a case study of Virtual Organisation forming as a running example.

Acknowledgements

This work is supported by the EU FP6 OntoGrid project (STREP 511513) funded under the Grid-based Systems for solving complex problems.

References

[1] C.A. Goble, D. De Roure, N. R. Shadbolt and A.A. Fernandes *Enhancing Services and Applications with Knowledge and Semantics* in The Grid 2 Blueprint for a New Computing Infrastructure Second Edition eds. Ian Foster and Carl Kesselman, 2003, Morgan Kaufman, November 2003

[2] D. De Roure, Y. Gil, J. A Hendler *Guest Editors' Introduction: E-Science*. IEEE Intelligent Systems 19(1), January/February 2004: 24-25

[3] I. Foster, H. Kishimoto, A. Savva, D. Berry, A. Djaoui, A. Grimshaw, B. Horn, F. Maciel, F. Siebenlist, R. Subramaniam, J. Treadwell, J. Von Reich, *The Open Grid Services Architecture*, http://www.ggf.org/documents/GFD.30.pdf, 2005.

[4] L. Pouchard, L. Cinquini, B. Drach, et al., *An Ontology for Scientific Information in a Grid Environment: the Earth System Grid*, CCGrid 2003 (Symposium on Cluster Computing and the Grid), Tokyo, Japan, May 12-15, 2003.
[5] N. Sharman, N. Alpdemir, J. Ferris, M. Greenwood, P. Li and C. Wroe, *The ^{my}Grid Information Model*, Proceedings of All Hands Meeting, 2004.
[6] J. Treadwell, *GGF Open Grid Services Architecture Glossary of Terms*, http://www.gridforum.org/documents/GFD.44.pdf, 2005.