Managing semantic Grid metadata in S-OGSA

Paolo Missier, Pinar Alper, Oscar Corcho, Ioannis Kotsiopoulos,
Ian Dunlop, Wei Xing, Sean Bechhofer and Carole Goble

School of Computer Science
The University of Manchester
Manchester, UK

Abstract

Grid resources such as data, services, and equipment, are increasingly being annotated with descriptive metadata that facilitates their discovery and their use in the context of Virtual Organizations (VO). Making such growing body of metadata explicit and available to Grid services is key to the success of the VO paradigm. In this paper we present a model and management architecture for *Semantic Bindings*, i.e., first-class Grid entities that encapsulate metadata on the Grid and make it available through predictable access patterns. The model is at the core of the S-OGSA reference architecture for the Semantic Grid.

1 Introduction

One of the prominent problems in large-scale distributed computing is the creation and management of *Virtual Organizations (VO)*, i.e., uniform views over pools of distributed computing and storage resources that are controlled by multiple, often autonomous organizations. The ability to address large-scale computational problems that require ad hoc configurations of resources, depends largely on the rapid and effective creation of VOs.

The Grid middleware infrastructure [4] provides a foundation for VO management by enabling, in principle, the discovery and sharing of various types of distributed hardware and software resources, e.g. computing, storage, data sets, scientific equipment. Forming a new VO, however, requires a consistent description of the available resources, and relies upon a shared understanding of their function and properties, such as access policies and other usage constraints. This makes descriptive metadata that annotates Grid resources an essential element of VO formation. Furthermore, in order for shared annotations to be of practical use, it is important that their intepretation be unambiguous. This can be achieved by providing a common reference framework for the interpretation of metadata: a shared vocabulary, a taxonomy of concepts, or, commonly, a full-fledged ontology.

In current Grids, however, metadata tends to be implicit, often embedded in the applications or in the middleware, and not associated to any interpretation framework. This results in latent knowledge that is prone to syntactic changes, not interoperable, and dependent upon extensive human effort for deployment configuration and maintenance. With the term *Semantic Grid*, we denote a Grid architecture in which such latent knowledge is made explicit and expressed in some uniform way, and where the metadata is based on a common interpretation.

Various mechanisms and patterns can be employed to maintain the association between a Grid entity, its metadata, and the reference framework for the intepretation of the metadata. In this paper, we describe a modelling approach whereby the standard OGSA framework is extended in order to accommodate a principled description of semantic metadata. The model is driven by the principle that such associations *should themselves be Grid entities*. In practice, annotations that describe the content of a file, or the purpose of a service, are themselves Grid resources. We have coined the term *Semantic Binding* (SB for short) to denote this new type of resource. After describing our proposed semantic extensions to OGSA, we focus on the stateful properties of SBs, and describe the functionality of a *Semantic Binding Service* dedicated to the management of SBs.

2 S-OGSA and Semantic Bindings

In previous work on Semantic Grid [3], we have introduced S-OGSA, an extension to OGSA that includes Semantic Bindings and related management services. Briefly, the proposal consists of three parts:

- a new set of services, called *Semantics-provisioning services*, that create metadata by annotating existing Grid services or resources;
- a conceptual model, where a SB is defined as an association between one or more Grid entities (GE), some metadata content that annotates the GE, and one or more Knowledge entities, KE. A KE is any reference framework that provides a semantic interpretation for the metadata –typically, an RDFS or OWL DL ontology;
- a new set of Grid services, called *Semantic Aware Grid Service* (SAGS), that are able to interpret and exploit the metadata found in Semantic Bindings, in order to enhance their own functionality.

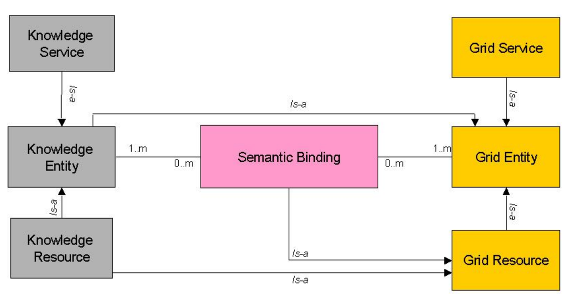

Fig. 1: Conceptual model for Semantic Bindings

As mentioned, the key property of SBs is that they are first-class Grid resources themselves, as shown in Figure 1. This makes it possible to incrementally incorporate semantics into Grid services, without disrupting existing OGSA-compliant services. As a result, we can envision a "mixed economy" of co-existing semantics-unaware services and SAGS, and we may also identify design patterns for the piecemeal migration of the former into the latter [7].

Consider, as an example, the Ontogrid-AuthZ service, a SAGS that manages XACML-compliant authorization requests and responses[1], providing the basis for access control in a VO. This service was developed in the context of the Ontogrid project,[2] to demonstrate how semantics on the Grid can be exploited to provide enhanced service functionality to applications. The service accepts a user access request and determines the requestor's eligibility based on declarative rules that define the user's role. Access control to specific resources is based on a set of access permissions and restrictions that are associated to each role. In this scenario, users are modelled as Grid resources (uniquely identified by their Distinguished Name, DN), and the information required to evaluate the rules is metadata associated to each DN, for instance the person's affiliation and their roles within their real organizations (i.e. a computing centre, an academic department, etc.). Rule evaluation results in a decision to grant or deny the user access to a certain pool of resources, and is based on the semantic interpretation of the metadata, using a modified version of the `KAoS` suite of ontologies [2], which includes descriptions of actors, groups, actions, resources, policy types, and more.

Here, Semantic Provisioning Services are the services that produce the metadata, i.e., they annotate the user's DN with affiliation and local role information. SAGS, in this case, are decision services that determine eligibility to access a service, by evaluating the rules based on the metadata and on the ontologies. This type of SAGS is known as a Policy Decision Point (PDPs) in the Common Open Policy Service (COPS) Protocol.[3] Finally, Semantic Bindings are Grid resources that encapsulate the metadata and maintain its association with the DN, as well as with references to the ontologies. They are made available to SAGS through the Semantic Binding Service, which manages SBs using that are stored persistently in some metadata repository.

In practice, KEs and SB are managed entities that are accessible through WSRF services, using WS-Addressing as the naming scheme. KEs are addressed using a combination of URIs and WS-Addressing endpoints, similar to the mechanism proposed in the WS-Naming specification.[4]

An implementation of the S-OGSA model is available from the Ontogrid project. Grid entities metadata is stored using an RDF model and managed by the Atlas distributed RDF store [8], and ontologies are accessed through the

[1] XACML is the eXtensible Access Control Markup Language, defined by the OASIS XACL Working Group, see http://www.oasis-open.org/

[2] The Ontogrid project: http://www.ontogrid.net

[3] COPS is defined in IETF RFC 2748, please see http://www.ietf.org/rfc/rfc2748.txt.

[4] Working draft July 5, 2005 available at https://forge.gridforum.org/projects/ogsa-naming-wg/.

WS-DAIOnt component [5]. A full implementation of the Semantic Binding Service is currently being developed.[5]

3 Semantic Binding lifetime

As mentioned, SAGS rely on Semantic Bindings to implement their behaviour. The SBs, however, are dynamic entities that are subject to change, and futhermore, the GEs that they describe, and the KEs that their content refers to, may also change in time. Some of these changes may cause a SB to become invalid, with the consequence that the SAGS that depend on them can no longer rely on their values. In particular, a Grid service may become unavailable, or a Grid resource may be destroyed. In the VO scenario, for example, users may change their affiliation, resulting in a new DN being issued to them. Although the user is indeed the same, the GE changes, and SB must be revalidated (indeed, the access rights may change because of the new affiliation). When this happens, the associated metadata becomes invalid. Similarly, the update of an SB itself in general has an impact on a SAGS, for instance when the roles or privileges associated to a VO participant change.

Finally, a change in the Knowledge entities that the SB refers to should prompt a re-evaluation of the SBs, too. Consider for example the case of a collection of roles defined in a VO ontology; when roles are deleted or moved within a taxonomy, the SBs that describe the role of an individual may have to be updated as well. The problem of assessing the impact of ontology evolution on an existing knowledge base has been addressed in the literature [10]. The consequences of evolution are summarized in [6] as the problem of re-aligning the data and the ontology. S-OGSA is concerned with the detection of ontology change events, and with the notification of those events to interested parties, so that appropriate realignment of the SB content –not discussed in this paper, may take place.

In order to deal with these changes in a principled way, S-OGSA defines SBs as stateful Grid entities with a defined lifetime; the state diagram associated to an SB includes a set of fundamental states and state transitions, as well as the external events that cause the transitions. The specification of SB lifetime extends WS- ResourceLifetime, a part of the WS-Resource Framework family of specifications that standardizes the way that resources are destroyed, and defines resource properties for the inspection and monitoring of a resource lifetime. While WS-ResourceLifetime is focused exclusively on resource destruction, we extend it to include any life-changing event that may affect the validity and updates of an SB. Furthermore, the basic state machine presented here can be extended with sub-states, as shown later.

[5] All of these components are available from the Ontogrid CVS repository, please see http://www.ontogrid.net for details.

3.1 Semantic Binding state machine

The state transition diagram is shown in Figure 2. When it is first created, a Semantic Binding SB is in the *Valid* state. We denote with GE_{SB} and KE_{SB}, respectively, the set of Grid entities and Knowledge entities that are part of the association, and with $content_{SB}$ the metadata payload within SB.

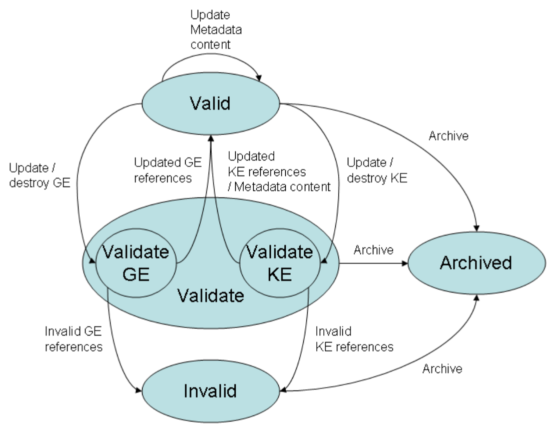

Fig. 2: State transition diagram for a generic Semantic Binding

State transition events are of the following types: changes in the Grid entities, denoted by $GE_{SB} \rightarrow GE'_{SB}$; changes in the Knowledge entities, i.e., $KE_{SB} \rightarrow KE'_{SB}$; or updates to the SB content: $content_{SB} \rightarrow content'_{SB}$. Note the Grid and Knowledge entities can also be destroyed: $GE_{SB} \rightarrow \emptyset$, $KE_{SB} \rightarrow \emptyset$. These transitions are listed in the second column of Table 3.1. In addition to these external events, a content expiration date can also be associated to an SB, so that it is automatically considered stale upon expiration. In the table, this is indicated as the event $obsolete(SB)$.

For a *Valid* SB, these events cause its transition to either one of two possible *Validate* states, *Validate GE* and *Validate KE*. These are interim states in which the SB may be invalid, and is awaiting re-validation. A re-validation process, either manual or automated, is any procedure that updates any or all of GE_{SB}, KE_{SB}, or $content_{SB}$, and which results in a decision as to whether the updated entities represent a new valid combination. For a *Validate GE* SB, such procedure determines whether the existing metadata can be associated to the new Grid entities, and provides an update to the references in SB to GE'_{SB}. For example, following a change in the identity of a Grid resource which had some VO-related metadata profile, the procedure determines whether the same profile can be associated to the new identity for the resource. For a *Validate KE* SB, the problem is to determine whether the new ontology can still be used to interpret the old metadata. As mentioned in the previous section, various

Tab. 1: Events, state transitions, and validation actions

Events	**State after event**	**States after validation**
$GE_{SB} \rightarrow GE'_{SB}$	*ValidateGE*	*Valid / Invalid*
$GE_{SB} \rightarrow \emptyset$	*ValidateGE*	*Invalid*
$KE_{SB} \rightarrow KE'_{SB}$	*ValidateKE*	*Valid / Invalid*
$KE_{SB} \rightarrow \emptyset$	*ValidateKE*	*Invalid*
$content_{SB} \rightarrow content'_{SB}$	*Valid*	N/A
$obsolete(SB)$	*Invalid*	*Valid / Invalid*
$content_{SB} \rightarrow \emptyset$	N/A	N/A
$archive(SB)$	*archived*	N/A
$destroy(SB)$	N/A	N/A

approached can be followed, which are outside the scope of our work. In both cases, the SB goes back to the *Valid* state in case of successful validation, and to *Invalid* otherwise.

The possible outcomes of the validation procedures are listed in the last column of Table 3.1. Note that, as a particular case, when the Grid or Knowledge entities are destroyed, the validation procedure is always assumed to fail, leading to a *Invalid* state. Finally, note that according to the table, an update to valid metadata, i.e., $content_{SB} \rightarrow content'_{SB}$ when the state is *Valid*, always results in new valid metadata. Finally, the *Archived* state indicates that a SB is still available for inspection, but it has been superseded by a more recent version.

3.2 myGrid example with extended state model

The basic model just described can be extended by introducing sub-states, resulting in finer-grain definition of the behaviour of specific types of metadata. The example presented here refers to service annotations in the myGrid project[6]. A service is a Grid entity that is annotated with metadata describing its function. Each annotation is a SB having the metadata as its content and the service reference as its GE. Knowledge entities consist of myGrid vocabularies and domain ontologies [11] that are used to express the annotations.

Figure 3 shows the extended state diagram for these SBs. The new sub-states within *Valid* allow a distinction to be made between metadata that has been reviewed by human experts (i.e., "Quality Assured"), and metadata that is awaiting QA. Note that in both cases, the metadata is indeed valid, in the sense that the annotation is plausible. The sub-states add explicit information regarding the *quality* of the annotation, which some applications may want to take into account. This is important for instance when annotations are automatically generated, as in [1], requiring experts' inspection prior to their release.

Along with the state diagram, the transition table shown in Figure 3.1 can be refined with specific state transition rules, as follows:

- $GE_{SB} \rightarrow \emptyset$ results in a transition to the *archived* state;

[6] myGrid: http://www.mygrid.org.uk

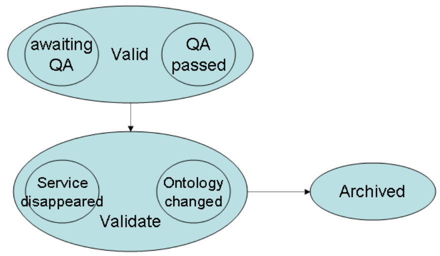

Fig. 3: Extended state diagram for myGrid Semantic Bindings states

- $KE_{SB} \rightarrow KE'_{SB}$ triggers the invocation of a change detection tool, e.g. [9], which analyzes the SB content and issues a report to the annotator;
- a transition to the *Awaiting QA* state triggers a notification to the annotator, to carry out the QA task.

3.3 State changes notification and Semantic Binding Service

From the previous discussion, it should be clear that SAGS that use any SBs, ought to be informed of any state change for those SBs. In S-OGSA, a notification mechanism based on the WS-Notification standard is available for this purpose. Specifically, a set of pre-defined topics are associated to changes of the form $GE_{SB} \rightarrow GE'_{SB}$, $KE_{SB} \rightarrow KE'_{SB}$, and $content_{SB} \rightarrow content'_{SB}$. Consumers who subscribe to those topics are notified of the changes. A dedicated service in particular, called *SB housekeeping service*, monitors all the SBs by subscribing to all topics, and is responsible for activating standard re-validation procedures.

We conclude the section with a mention of the *Semantic Binding Service (SBS)*, a WSRF service for Semantic Binding management. Its functionality include (i) the creation and indexing of SBs given GEs, KEs, and a semantic metadata payload; (ii) retrieval of the semantic payload for a SB; (iii) inquiry on the SB state; (iv) archival or destruction of a SB; and (v) query on the collection of SBs, typically using the associated GEs as a key.

4 Ongoing work

In this paper we have described the main model and mechanisms for managing semantic metadata on the Grid, and their implementation as part of the Semantic Grid reference architecture, S-OGSA. The SBs and the other S-OGSA components for SB, lifetime management and notification will be released in the public domain in the near future. The architecture is being tested on a number of use cases provided by the Ontogrid project.[7]

[7] Prototypes for the use cases are available through the Ontogrid CVS server, and documentation is published on the Ontogrid portal, cited earlier.

Two main research questions are currently being investigated, namely (i) how can Grid services evolve into SAGS in a non-disruptive and cost-effective way. This is leading to the definition of architectural patterns that describe service evolution; and (ii) how SAGS are affected by missing metadata, i.e., when SBs are not available as expected, a common problem with metadata annotations.

References

1. Khalid Belhajjame, Suzanne M. Embury, Norman W. Paton, Robert Stevens, and Carole A. Goble. Automatic annotation of web services based on workflow definitions. In *International Semantic Web Conference*, pages 116–129, 2006.
2. J. Bradshaw, A. Uszok, R. Jeffers, N. Suri, P. Hayes, M. Burstein, A. Acquisti, B. Benyo, N. Breedy, M. Carvalho, D. Diller, M. Johnson, S. Kulkarni, J. Lott, M. Sierhuis, and R. van Hoof. Representation and reasoning for DAML-based policy and domain services in KAoS and nomads. In *Second International Joint Conference on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, 2003. ACM Press.
3. O. Corcho, P. Alper, I. Kotsiopoulos, P. Missier, S. Bechhofer, D. Kuo, and C. Goble. An overview of s-ogsa: a reference semantic grid architecture. *Journal of Web Semantics*, 4(2), 2006.
4. I Foster and C Kesselman. *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 1999.
5. Miguel Esteban Gutierrez, Asuncion Gomez-Perez, and Sean Bechhofer. Wsdaio: Ontology access provisioning in grid environments. In *Proceedings of the 1st CoreGRID Workshop on Knowledge and Data Management*, Poznan, Poland, September 2005.
6. M. Klein and F. N. Noy. A component-based framework for ontology evolution. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'03), Ontologies and Distributed Systems*, Acapulco, Mexico, 2003.
7. I. Kotsiopoulos, P. Missier, P. Alper, O. Corcho, S. Bechhofer, and C. Goble. Architectural patterns for the semantic grid. In D. Talia, A.Bilas, and M. Dikaiakos, editors, *CoreGRID Institute on Knowledge and Data Management, Poznan Workshop, September 2005*, volume XVIII of *CoreGRID Series, Knowledge and Data Management in GRIDs*. Springer, 2007.
8. I. Miliaraki, M. Koubarakis, M. Magiridou, and Z. Kaoudi. Semantic grid resource discovery using dhts in atlas. In D. Talia, A.Bilas, and M. Dikaiakos, editors, *CoreGRID Institute on Knowledge and Data Management, Poznan Workshop, September 2005*, volume XVIII of *CoreGRID Series, Knowledge and Data Management in GRIDs*. Springer, 2007.
9. P. Plessers and O. De Troyer. Ontology change detection using a version log. In *Proceedings of the 4th international Semantic Web Conference (ISWC05)*, Lecture notes in computer science, Galway, Ireland, November 2005. Springer.
10. L. Stojanovic. *Methods and Tools Ontology Evolution*. PhD thesis, Univ Karlsruhe (TH), 2004.
11. C. Wroe, R. Stevens, C.A. Goble, A. Roberts, and M. Greenwood. A suite of daml+oil ontologies to describe bioinformatics web services and data. *Special issue of the International Journal of Cooperative Information Systems*, 12(2):197–224, 2003.