Open Science Grid

Document Name	**OSG Information Services: a Plan for OSG 0.6.0 – 0.8.0**
Version	**Draft 0.9.2**
Date last updated	8/15/2006
Authors	Gabriele Garzoglio, Ruth Pordes, Shaowen Wang
OSG document #	

Change History

8/10/2006 - 8/15/2006	Gabriele Garzoglio, Ruth Pordes, Shaowen Wang	Put things together

About this document

This is a draft white paper describing the gap analysis and a plan for preparing OSG information services to be ready in OSG 0.6.0 and 0.8.0 releases.

1. Introduction

The purpose of this white paper is to outline a plan for getting information services ready in the upcoming Open Science Grid (OSG) 0.6.0 – 0.8.0 releases. The plan addresses the following four major activities:

- Service-Oriented Architecture (SOA) information services examine the evolution of existing OSG information services. Emerging production-quality SOA information services such as MDS4 [1] are analyzed as well.
- Service discovery interfaces and definition address service discovery on OSG, which also takes the interoperability with EGEE into preliminary consideration.
- Definition of the GLUE Schema [2] v1.3 is a joint activity among several large-scale Grids with OSG included to enhance the GLUE Schema 1.2. Specifically, this activity reflects the interest of OSG to continue to promote the interoperability with peer Grids.
- Information services interoperability with peer Grids such as EGEE, NorduGrid, and TeraGrid is necessary for OSG to achieve interoperation with these Grids.

In the reminder of this paper, section 2 describes existing OSG information services. Section 3 presents SOA information services and a plan for evaluating such services with a specific focus placed on MDS4. Following the SOA topic, section 4 further elaborates about service definition and discovery interfaces and designs several tasks to iteratively test such interfaces. Section 5 focuses on the plan and specifications for the GLUE Schema 1.3. Section 6 addresses more general issues on OSG interoperability with peer

Grids including the EGEE, NorduGrid, and TeraGrid. Section 7 summarizes the key recommendations of this planning paper.

2. Existing information services

Existing OSG information services mainly include BDII (Berkeley Database Information Index) [1], GIP (Generic Information Provider) [4], GridCat [5], and MonALISA [6]. In addition, the CEMon [7] information service is being tested on the OSG integration testbed, and is scheduled to be released in OSG v0.6.0.

- The BDII provides a production quality LDAP server solution to publish OSG information using the GIP (see below) based on the GLUE Schema. LDAP query can be used for applications and users to access BDII information. OSG has maintained BDII instances for both the production and integration infrastructures.
- CEMon, an information system developed within the gLite project [8], interfaces with GIP to gather computing element characteristics, which can be represented in LDAP, XML, old and new ClassAd [9] formats. CEMon provides web services interfaces to publish information. These interfaces will be used to implement an OSG resource selection service.
- The Open Science Grid (OSG) GIP (OSG-GIP) has been developed to enable LCG-OSG interoperability, OSG resource selection services, GridCat, etc. The GIP is a grid information service that aggregates static and dynamic resource information for use with LDAP-based information systems such as the BDII. The GIP produces information based on the GLUE Schema.
- GridCat is a Grid cataloging system for OSG. This catalog system contains Grid readiness and other useful information to support job submission and job scheduling for application users and grid scheduler developers.
- MonALISA has been developed as a service-oriented distributed monitoring system using JINI/JAVA and WSDL/SOAP technologies. One of the typical use cases of MonALISA is to provide plots of VO, job and site usage with selectable views and timelines.

In the timeline for OSG 0.6.0 – 0.8.0 releases, OSG information services will remain diverse to serve different purposes. This diversity may be accommodated in SOA.

3. SOA information services

OSG future information services need to support services adhering to a SOA. Though existing OSG information services such as Clarens and BDII/GIP have capabilities for publishing service information, it is necessary to evaluate emerging service-based information service technologies such as MDS4.

MDS4 is the monitoring and discovery system that is built within the Globus Toolkit 4. MDS4 is distinguished from previous similar systems by its extensive use of interfaces and behaviors defined in the new WS-Resource Framework and WS-Notification specifications and its deep integration into essentially every component of the Globus

Toolkit 4 [10]. Peer large-scale Grids such as the TeraGrid has deployed and tested MDS4. A project named MIP (Modular Information Provider) [11] has developed information provider solutions to integrate MDS4 within OSG. However, further evaluation work on scalability and robustness needs to be performed to get MDS4 ready in OSG 0.6.0 – 0.8.0 releases. Such evaluation work needs to identify and evaluate application use cases, which will help understand how to effectively provide consistent and easy-to-use application or user interfaces for information services in OSG. Also, the evaluation work must leverage what has been done at peer Grids to minimize the cost and promote Grid interoperability.

A specific plan to evaluate SOA information services using MDS4 as an example is described as follows:

1. Understand the progress of peer Grids
2. Identify OSG application cases that will be interested to interact with MDS4
3. Design methods to test scalability and robustness
4. Improve information provider solutions based on the requirements of achieving scalability and robustness
5. Experiment with MDS4 on the integration testbed
6. Conduct experiments and provide feedbacks to MDS4 and VDT projects

Several iterations of these six steps may need to be executed for individual OSG 0.6.0 – 0.8.0 releases respectively.

4. Service definition and discovery interfaces

Assuming OSG is moving rapidly toward SOA in the timeline of 0.6.0 – 0.8.0 releases, it would be important to specify and publish service information in a coherent way as well as to provide effective and efficient interfaces to discover services. Existing service publishing and discovery technologies in OSG mainly include MDS4 and Clarens, both of which are part of VDT.

The Clarens [12] solution provides a registry for publishing and querying for Grid services. It has similar functionality to a UDDI registry in the aspect of supporting service registration and discovery. Clarens provides a lightweight API, xmlrpc accessibility, and a dynamic and distributed architecture. Current Clarens implementation depends on having access to a MonALISA server that accepts ApMon monitoring data publications.

MDS4 is naturally integrated with GT4 services. GT4 Web services support a minimal set of resource properties (an informal service name and a service startup time) and thus can be registered within MDS4 in a straightforward fashion. Specifically for GRAM and RFT, a larger set of service-specific information can be published [13]. MDS4 can support service definition and discovery using information provider solutions through its support for the GLUE Schema that has service components specified.

The following list of tasks is identified to develop service definition and discovery interfaces for OSG 0.6.0 – 0.8.0 releases.

1. Understand the progress of peer Grids on the adoption and applications of SOA information services and leverage their expertise
2. Develop example use cases to engage and promote the applications of SOA information services
3. Evaluate the performance of using existing information services for publishing service data and discovering services based on the use cases
4. Develop light-weight and user-friendly service discovery client tools and provide support for using the tools
5. Provide technical feedbacks and recommendations to respective information services projects and VDT

These tasks should be executed in an iterative way.

5. GLUE Schema 1.3 requirements

Several groups of representatives from the OSG, TeraGrid, EGEE, NorduGrid, and other relevant stakeholders have started collaborating to draft a document to define the GLUE Schema v1.3. The plan is for these groups to come up with their requirements by September 1, 2006, and to accomplish the document by November 15, 2006. This plan has been agreed upon by several representatives of participating groups.

The OSG requirements will be finalized at the parallel session on Information Services of the OSG consortium meeting during August 21 – 23, 2006. Preliminary discussions about such requirements are held weekly before the session. The requirements should be based on analyzing real use cases.

Plan

1. Proposals for specifying changes and gathering requirements – by September 1, 2006. Each project may create a proposal describing the requirements for proposed changes to the Schema 1.2. Each change should be specified with a description of the requirements it addresses. Several OSG members have already held a number of internal discussions on the use of the Glue Schema by a resource selection service and produced a document [14].
2. Proposals integration – by September 15, 2006. Sergio Andreozzi will compile the proposals from participating Grids into one coherent proposal.
3. Proposal reflection – by October 15, 2006. Each project will review all the proposed changes and provide feedbacks on each change to Sergio Andreozzi.
4. Proposal feedback compilation – by November 1, 2006. Sergio Andreozzi will incorporate the feedbacks from participating Grids into a single proposal.
5. Consensus on controversial issues – by November 15, 2006. A few meetings may need to be held to build consensus on controversial issues that might arise.
6. Schema implementation and deployment will work on adapting information provider solutions to the changes in the Schema and supporting the adoption of new information provider implementations. This work is required on both integration and production infrastructures.

6. Interoperable information services

In Grid architectures, information services function as a hub of information for users as well as other Grid services. Grid information services may vary from Grid to Grid in terms of the conventions and protocols used to present and access information. Therefore, information services are of central importance to achieve interoperability of Grid services between OSG and peer Grids such as EGEE, NorduGrid, and TeraGrid. Past and existing efforts to achieve interoperable information services in OSG are driven by the needs of individual VOs through real use cases. The goal of interoperable information services is to support VOs for their end-to-end interoperability use cases that access these information services. This section summarizes the status of current OSG interoperable information services and describes a plan.

Goal

To support VOs for their end-to-end interoperability use cases that access OSG common services.

Status

The current OSG interoperability activities have followed two main approaches:

1. VO-driven projects are sponsored by an individual VO with a specific need for interoperability. For example, a VO may wish to get its computing jobs executed on both OSG and EGEE. These projects are top-down, in the sense that they achieve end-to-end compliance for a given application or use case.
2. Service-driven projects are generally sponsored by OSG. They study general interoperability issues related to specific services (e.g. BDII) and address the compatibility of such services between OSG and EGEE. These projects are bottom-up, in the sense that they pursue the compliance of basic services.

Several methods have been investigated to achieve interoperable information services, which include

1. deployment of common information services (e.g., BDII and GIP for LCG and OSG) and standards (e.g., the GLUE Schema) among different Grids,
2. development and deployment of bridging information services (e.g., a translator service) between Grids,
3. the migration of the current Grid infrastructures to a Service Oriented Architecture.

Plan

This plan adopts a VO-driven approach. It is open-ended and its initial deliverables are focused on building a common framework to report and organize VO-driven activities for achieving interoperable information services.

1. A document summarizing requests for VO-driven interoperability projects will be developed to include a description for each project, key common services and the Grids involved. An initial draft of the document will be ready in September 2006.

2. A forum will be organized for new activities that address interoperable information services (e.g., leveraging tg-mig@opensciencegrid.org), and to provide a repository for status reports (e.g. http://osg.ivdgl.org/twiki/bin/view/Interoperability/).
3. Quarterly reports will be prepared to provide updates about documentation change and the outcome of ongoing and new interoperability projects. OSG common information services will be checked to assure successful end-to-end interoperability tests, which will support continuing compatibility with other interoperability projects.

In addition to information services, other services that have been considered to support Grid interoperation include Security (Authentication / Authorization / Accounting / Auditing), Data Storage (catalogues, storage interface, data transfer), and Job Management (submission, monitoring, resource selection, scheduling). Also, there are VO-specific services that may require interoperability validations, as their configuration might be different on different Grids.

7. Summary

References

[1] MDS4: http://www.globus.org/toolkit/mds/
[2] Glue Schema: http://glueschema.forge.cnaf.infn.it/
[3] BDII: http://lfield.home.cern.ch/lfield/cgi-bin/wiki.cgi?area=bdii&page=documentation, and http://scan.grid.iu.edu/ldap/index.html
[4] OSG GIP: http://grow.its.uiowa.edu/research/gip/
[5] GridCat: http://www.ivdgl.org/gridcat/home/index.html
[6] MonALISA: http://monalisa.cacr.caltech.edu/monalisa.htm
[7] CEMon: http://grid.pd.infn.it/cemon/field.php
[8] gLite: http://glite.web.cern.ch/glite
[9] ClassAd : http://www.cs.wisc.edu/condor/classad/
[10] GT4 : http://www.globus.org/toolkit/
[11] MIP: http://grow.its.uiowa.edu/research/mip/
[12] Clarens: http://clarens.sourceforge.net/
[13] GT4 resource properties: http://www.globus.org/toolkit/docs/development/4.1.0/toolkit-rp.html
[14] An OSG resource selection service: http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/GlueSchemaAndReSS