Grid Interoperability: Joining Grid Information Systems

Martin Flechl[1] **and Laurence Field**[2]

[1] Department of Nuclear and Particle Physics, Uppsala University
[2] CERN IT-GD

E-mail: martin.flechl@tsl.uu.se, Laurence.Field@cern.ch

Abstract. A grid is defined as being "coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations". Over recent years a number of grid projects, many of which have a strong regional presence, have emerged to help coordinate institutions and enable grids. Today, we face a situation where a number of grid projects exist, most of which are using slightly different middleware. Grid interoperation is trying to bridge these differences and enable Virtual Organizations to access resources at the institutions independent of their grid project affiliation. Grid interoperation is usually a bilateral activity between two grid infrastructures. Recently within the Open Grid Forum, the Grid Interoperability Now (GIN) Community Group is trying to build upon these bilateral activities. The GIN group is a focal point where all the infrastructures can come together to share ideas and experiences on grid interoperation. It is hoped that each bilateral activity will bring us one step closer to the overall goal of a uniform grid landscape. A fundamental aspect of a grid is the information system, which is used to find available grid services. As different grids use different information systems, interoperation between these systems is crucial for grid interoperability. This paper describes the work carried out to overcome these differences between a number of grid projects and the experiences gained. It focuses on the different techniques used and highlights the important areas for future standardization.

1. Introduction

The traditional model of parallel computing used in scientific institutions consists of a site with a batch system and a storage system. Local users have access to these local resources in accordance with the local site policy. A user may have access to several sites but would have to authenticate himself at least once per site (or even once per service) following different security mechanisms, adhere to different site policies, and use different interfaces.

Grid computing is a model of distributed parallel computing which eliminates those barriers: There is a common security mechanism deployed across all sites and services, the Globus Security Infrastructure (GSI) which is a public/private key infrastructure based on X.509 certificates. All participating sites agree on a common policy for external users, and common interfaces exist to access services across all sites. Grid computing is clearly distinguished from other models of distributed computing by its concept of Virtual Organizations [1]. A Virtual Organization (VO) is a group of users from multiple institutions who collaborate to achieve a specific goal. The ultimate goal of grid computing is to provide a network of batch systems which, as far as the user is concerned, acts like a single supercomputer, offering resources which are easily accessible in a way similar to electric power grids.

After Foster's and Kesselman's definition of the grid architecture [2] and the development of the Globus Toolkit (GT, an open source software bundle for building computing grids following these definitions in the late 1990s) [3], a large number of grid projects based on GT emerged within a few years. Globus is not a deployment-ready grid solution but a toolkit for building other grid middleware. Each grid project started developing their own middleware which in spite of their common base, GT, lead to a variety of different solutions. Due to the regional character of the different projects this resulted in the current existence of largely isolated grid islands. It is the purpose of the interoperability project GIN to build bridges between those islands. Currently the focus is on the following grid infrastructure projects:

- EGEE[1], a project funded by the European Commission with the aim to provide a grid infrastructure for scientists, in particular the LHC experiments at CERN,
- NDGF[2], a joint grid project by Denmark, Finland, Norway and Sweden providing services for researchers in the Nordic countries,
- TeraGrid[3], a collaboration of nine large computing centers in the USA,
- NAREGI[4], the national Japanese grid project,
- OSG[5], a grid infrastructure for scientists from the USA,
- PRAGMA[6], a collaboration between universities and institutes around the pacific ocean from Asia, Australia and the Americas,
- DEISA[7], a collaboration of national supercomputing centers in Europe,
- NGS[8], a project with the aim to provide a Grid infrastructure to researchers in the UK,
- APAC[9], the national Australian Grid project providing a computing infrastructure for research communities.

2. Grid Information Systems

The basic tasks of the grid information system are service discovery, service selection, service optimization, service monitoring and service accounting. It is a fundamental ingredient of every grid since jobs can only be pushed to adequate resources by using the information retrieved and indexed by the information system about the existence and characteristics of resources available at a certain moment. The Globus Toolkit version 2 which most grids are based upon contains the Metacomputing Directory Service (MDS) 2 as an information system component [3]. The MDS can be divided into three parts:

- The information provider, a script which obtains static information (from a configuration file) and dynamic information about local services, formats it into Lightweight Directory Access Protocol (LDAP) Data Interchange Format (LDIF) and returns this information.
- The GRIS (Grid Resource Information Services), which executes the information provider, obtains the LDIF and returns the result.
- The GIIS (Grid Information Index Services), which queries all the GRIS which have registered to it. A GIIS can register to another GIIS to build up a hierarchical structure.

[1] Enabling Grids for e-Science, `http://www.eu-egee.org/`
[2] Nordic DataGrid Facility, `http://www.ndgf.org`
[3] `http://www.teragrid.org`
[4] NAtional REsearch Grid Initiative, `http://www.naregi.org/index_e.html`
[5] Open Science Grid, `http://www.opensciencegrid.org`
[6] Pacific Rim Applications and Grid Middleware Assembly, `http://www.pragma-grid.net`
[7] Distributed European Infrastructure for Supercomputing Applications, `http://www.deisa.org/`
[8] National Grid Service, `http://www.grid-support.ac.uk/`
[9] Australian Partnership for Advanced Computing, `http://www.apac.edu.au/`

The MDS data model for the information service is based on the LDAP Protocol which consists of entries arranged in a hierarchical tree-like structure. Each entry contains a set of named attributes and values. The information schema defines the mandatory and optional attributes which are filled by entries of a certain type, as well as a type for the value of each attribute. A possible schema for an entry for a computing cluster could among other attributes require the attribute "SiteCPUs" to be filled by each computing cluster with its number of CPUs, value type integer, and allow to specify the operating system as a string. MDS does not define a complete schema and therefore several different schemas are used by the grid projects (see Table 1), namely the GLUE (Grid Laboratory for a Uniform Environment) schema used by most projects, the Nordugrid schema used by NDGF and NAREGI's schema which is based upon CIM (Common Interface Model) and implements NAREGI vendor extensions (NVE).

Table 1. Comparison of Grid Information Systems.

Project	Connection Protocol	Schema	Data Model
EGEE	LDAP	GLUE	LDIF
NDGF	LDAP	Nordugrid	LDIF
TeraGrid	WS-RF	GLUE	XML
NAREGI	WS	CIM+NVE	XML
OSG	LDAP	GLUE	LDIF
PRAGMA	WS	WebSIM	XML
DEISA	WS	GLUE	XML
NGS	LDAP	MDS	LDIF
APAC	LDAP	GLUE	LDIF

The hierarchical structure of the GIIS enables complete information retrieval by querying the top level GIIS. However, MDS has shown not to be a solution for large-scale production because it does not scale: Multiple client requests quickly lead to an overload of the top level GIIS [4, 5]. For this reason the large grid projects could not use the information system of the GT and had to provide their own solutions. Unfortunately, this was done largely in isolation and lead to many different solutions:

Within EGEE, the BDII (Berkeley Database Information Index) was introduced as a production quality replacement for the GIIS. The main difference is that the population of the cache and the query have been decoupled. With MDS, if the cache was out-of-date, the query would propagate down to the resource level. This mechanism was not fault tolerant and caused system failures when operated in a large-scale environment. Decoupling these processes additionally lead to a significant performance improvement. Each BDII consists of two LDAP databases, one for reading and one for writing to the database, which are periodically swapped.

The ARC middleware of the NDGF project uses the top level GIIS only to index the actual GRIS which are then queried in parallel, thus no caching is involved and the problem is avoided. OSG initially used MDS2 but is adopting the BDII to be ready for large-scale production. TeraGrid is currently testing MDS4, part of the most recent GT release, which is based upon web services. It has still to be shown to scale under production. This is currently being investigated. NAREGI's solution is completely different. At the site level there is an OGSA-DAI interface to a database. This database is populated by the CIM information provider. The combined entity is called a "cell domain".

Other differences between the information systems of the different grid projects concern the method to query the database and the representation of the schema (LDIF and XML are most

common). Addressing all the project-specific differences of the information systems is the most important step towards interoperability since a grid project cannot use the resources of another project without at least knowing about existence and characteristics of these resources.

3. Grid Interoperability Now: Status and Outlook

In their influential paper [1], Ian Foster et al. claimed that it is a "fundamental concern [of grid computing] to ensure that [resource] sharing relationships can be initiated among arbitrary parties, accommodating new participants dynamically, across different platforms, languages, and programming environments". Today, however, these possibilities are limited to the artifical regional boundaries constituted by the different grid infrastructures. An important long-term goal for grid computing is to overcome these boundaries by providing and establishing universal protocols and formats, just like the success of the World Wide Web would have been impossible without HTTP and HTML. Such a solution, however, cannot simply be imposed on the different grid architectures but can only be achieved by starting with bilateral and multilateral interoperability initiatives.

The purpose of these initiatives is not only to provide a proof-of-concept for interoperability and potential standards for the future, but also to provide a production-ready solution for the present. Such a solution is essential for current international VOs which fall within different regional grid architectures, like the Virtial Organizations (VO) of the LHC experiments at CERN, the most important driving force in scientific grid computing. This use case requires at least EGEE, OSG and NDGF to act as one infrastucture. Without such interoperability, these VOs would have to separate their entire work by region.

A successful example of bilateral interoperability activities is the collaboration between OSG and LCG[10] [6], the primary production enviromnent of EGEE. Driven by the demand of the LHC VOs to be able to seamlessly use both grid infrastructures the interoperability activities started at the end of 2004. The main difference between the two grid projects was that OSG used the Grid3 schema. To solve this problem, OSG also moved to the GLUE schema after a new version of the GLUE schema (1.2) was introduced, including new attributes required by OSG. This shows that present interoperability activities may lead to universal standards in the future. After a few more adjustments concerning the information system and setting up a BDII representing the OSG sites, interoperability was achieved and today jobs submitted via an EGEE user interface may be processed at a US site without the user even noticing. At the same time interoperability activities between EGEE and NDGF were already in an advanced state [7] and the goal of EGEE, OSG and NDGF acting as one infrastructure within reach.

Following this sucess, an ad-hoc meeting at Super Computing 2005 in Seattle between representatives of different grid projects lead to the start-up of the "Grid Interoperability Now!" (GIN) activity. The purpose of GIN is both to support and learn from existing bilateral interoperability efforts as well as to initiate new projects in this context, with the aim to unite these bilateral initiatives to a multi-grid interoperability platform. The focus is on finding working solutions within short terms but it is also intended that the experience gained will contribute to future standardization efforts. The GIN activities focus on the following corner stones of interoperability:

- Security Model: For authentication, almost all grid projects use the GSI which is based on X.509 certificates, already an international standard. The proposed authorization solution is the VO Management System (VOMS) which is not used by all grid projects yet.
- Computing Interface: Many different types of job submission interfaces are used and there is no agreement on any standard. A candidate for the job description is JDL (Job Description Language).

[10] LHC Computing Grid, `http://lcg.web.cern.ch/lcg/`

- Storage Interface: The Storage Resource Management (SRM) as storage interface along with GridFTP for file transfers is widely deployed, but several different versions are used.
- Information system: The differences concerning schemas, representations, query languages and software systems have already been addressed in Section 2. The natural candidate for a common schema is GLUE, either in LDAP or XML format.

4. Interoperability of Grid Information Systems

Jobs submission from the interface of one grid to computing elements of another grid requires information about the resources of the second grid to be available within the first one. Thus the first step towards grid interoperability is to join the information systems of different grids. In the following, the work carried out to achieve interoperability between different information systems is described and conclusions from the gained experiences are presented.

As pointed out in the previous sections, the information systems of the different grid projects use different languages (schemas, data models and connection protocols) and it is thus the first and most essential step towards interoperability to establish a working communication (information exchange) between the projects. This can be divided into two separate tasks: Setting up a communication channel, and translating the communication. The first problem is solved by creating an environment within one grid project capable of querying the information system of the other project, and to perform the actual query; the second problem by translating one schema representation into the other (note that if both the schema and the representation are different then this requires two steps). Work has been undertaken to achieve information system interoperability in several bilateral projects, the current status is:

- OSG → EGEE: Successfully completed and used in production.
- NDGF → EGEE: Prototype completed, testing is ongoing.
- NAREGI → EGEE: Prototype completed.
- TeraGrid → EGEE: Prototype completed, essential information still missing.

The schema translators produced in these bilateral activities are used within the GIN framework with the aim to achieve truly multilateral information system interoperability. The architecture which has been used is shown in Figure 1.

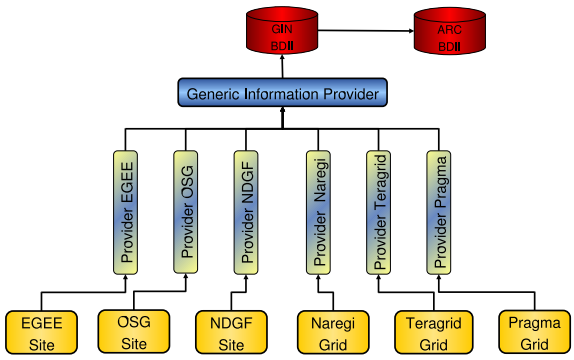

Figure 1. Architecture of the GIN-BDII. The Generic Information Provider (GIP) is used as a framework to support the use of pluggable information providers for the different grids. The GIN-BDII contains information in the GLUE schema and can be queried via ldapsearch. The information is also transferred to the ARC-BDII where it is translated into Nordugrid schema and provided to the NDGF sites.

Both EGEE and NDGF use LDIF to represent their information and populate an LDAP server which can be queried for this information. However, they use different schemas and interoperability of their information systems thus requires a translation from the Nordugrid to the GLUE schema (and vice versa). This is done via a script which performs the following steps:

(i) querying a web page for the most recent list of the NDGF GIIS'

(ii) creating parallel queries (ldapsearch) to all the NDGF GIIS' from the list

(iii) translating the result of these queries from Nordugrid schema LDIF to GLUE schema LDIF and writing this information to one temporary file per GIIS

(iv) adding values which do not exist in the Nordugrid schema from static files (e.g. GlueSiteLocation)

(v) returning all the information

This information is then used to populate LDAP databases on the GIN-BDII. Concerning the mapping between the GLUE and the Nordugrid schema (described in [8]), two main problems were identified: Firstly, the mapping is not 1:1. This is most obvious at the entity level where the GlueSite, GlueCluster, GlueSubCluster and GlueCE correspond to only two ARC entities, nordugrid-cluster and nordugrid-queue. The mapping currently chosen thus does not constitute a unique solution and only tests in a production environment will be able to identify problems and show if better alternatives exist. The second problem is that some entries only exist in one schema and not in the other, so data is missing and has to be either hard-coded or filled with default values. For the mapping Nordugrid → GLUE, this is mainly information specific for EGEE sites and irrelevant within NDGF, or at least not essential for job submission. Among the more important missing attributes is GlueSiteLocation (Format: City, Country) - the Nordugrid schema only offers a postal code - and GlueSiteLatitude/Longitute holding the geographic coordinates. This is solved by means of a static file holding this information which is read in by the script.

Currently, the GIN-BDII is running and can be queried via ldapsearch. The next step will be job submission tests once the EGEE → ARC interoperability setup is completed. EGEE and NAREGI also have ongoing bilateral interoperation activities. After initial meetings, a plan was suggested to create adaptors, translators and gateways to try and bridge the differences between the middleware. The information system was the first area which was worked upon and a mapping document between their different information schemas has been created by Yuji Saeki. Both EGEE and NAREGI have created translators between their systems.

The EGEE/TeraGrid case is different: Both projects use the same schema (GLUE), but in a different representation - LDAP for EGEE, and XML for TeraGrid. A similar script as for the EGEE/NDGF-case was set up which first submits a WS-RF query to the central TeraGrid index server. This query required setting up a node (which was named TERAGRID-BDII) with GT 4 installed. The result of the query is information about the TeraGrid resources in the GLUE schema format, which is then translated from the XML to the LDIF representation. Although both projects use the same schema the problem with mapping and missing data is not entirely eliminated since the GLUE specification [9] is not complete (most importantly, the relation between the entities is not defined). Different projects may interpret the schema slightly differently, and additionally decide not to use some of the optional attributes other projects use. However, the mapping of the essential attributes is straight-forward.

The interoperability activity between EGEE and OSG was already described in [6]. Translators in the direction EGEE → other projects exist in some cases (OSG, NDGF, NAREGI) or are planned for the near future. Thorough testing in a production environment has not been performed except in the case of OSG. However, the current activites prove that, at least in principle, bilateral information system interoperability between EGEE and other grid projects

described in this paper (NDGF, OSG, TeraGrid, NAREGI) is possible. In order to avoid $n \otimes n$-mapping between all grid projects we thus propose to use GLUE in LDIF representation as a go-between schema.

5. Conclusions

A fundamental aspect of every grid is its information system, which is used to find available grid services. This paper describes the work carried out to overcome the differences between the information systems of a number of grid projects. Most grid information systems have similar architectures. They all have a method to publish information and a method to query information. To create a translator between the information systems requires a component which queries one system, translates the information and publishes the result in the other information system. The difficult part is the translation. While it is straightforward to translate from one data model to another, it is more difficult to translate from one schema to another. If the information is not available in one system then it is impossible to insert it into the other.

The information schema is defined by the use cases it has to meet. The use cases of different grid projects are not necessarily the same. For grid projects, in order to agree on a set of common attributes, the "cross-grid" use case needs to be identified. Such a "cross-grid" use case is a task that must be performed the same way in all grids. A first example of such a use case is the GoogleEarth map showing all the participating sites and their grid infrastructure affiliation[11] which has been created using the information provided by the GIN-BDII.

This work has highlighted that using different information systems is not necessarily a problem. However, using different information schemas is the limiting factor for grid interoperation. The schema is therefore one of the main areas which needs standardization. Today, the grid paradigm is "a grid of grids", different grid federations working together to provide a seamless grid infrastructure. A truly federated grid will need a standardized or common information model.

6. Acknowledgments

The authors would like to thank Balazs Konya for providing the Nordugrid/Glue schema mapping [8] on which the corresponding translator is based, Yuji Saeki for the NAREGI-CIM/Glue schema mapping and Hitoshi Sato for implementing it (resulting in the NAREGI/EGEE-translator), and the GIN-Info group for their general support.

References

[1] Foster I 2001 *Intl. J. Supercomputer Applications* **3**
[2] Foster I 1999 *The Grid: Blueprint for a New Computing Infrastructure* (Morgan Kaufmann)
[3] Foster I and Kesselman C 1997 *Intl. J. Supercomputer Applications* **2**
[4] Field L and Schulz M W 2004 *Proc. of CHEP*
[5] Schulz M W, Field L, Nyczyk P and Novak J 2006 *Proc. of CHEP*
[6] Schulz M W et al 2006 *Proc. of CHEP*
[7] Gronager M et al 2006 *Proc. of CHEP*
[8] Konya B NorduGrid/ARC Information System `http://www.nordugrid.org/documents/arc_infosys.pdf`
[9] Andreozzi S et al GLUE Schema Specification Version 1.2 `http://glueschema.forge.cnaf.infn.it/uploads/Spec/GLUEInfoModel_1_2_final.pdf`

[11] The file `http://gridportal-ws01.hep.ph.ic.ac.uk/gin/gin-locations.kmz` can be downloaded and opened in GoogleEarth in order to see the map.