Open Science Grid

Document Name	**Open Science Grid Deployment, Spring 2005**
Version	**0.9.4**
Date last updated	May 24th 2005

Abstract. We present the plan for the initial deployment of the Open Science Grid (OSG) in the Spring of 2005. We refer to this set of functionality as OSG Release 0.2. The Grid3 sites will transition to OSG which will be a persistent, evolving grid. This implies that OSG embraces and extends the current Grid3 infrastructure in its initial deployment. This initial Deployment document outlines the deployment process as well as defines the Release functionality. We expect subsequent deployment activities to build off this base deployment, continue to support backward compatibility and add services driven by the needs of the stakeholders.

Version		
0.9.4	5/26/05	Update service definitions
0.9.3	5/23/ 2005	Update based on provisioning plans
0.9.2	2/14/2005	Plans for deployment

1 The Open Science Grid

The initial deployment of Open Science Grid strives to build upon the Grid3 success and relax several of the restrictions imposed on that collaboration. In particular, the Open Science Grid will contain services of various release levels, resources which service only a subset of all OSG VOs, and an increased membership pool.

Open Science Grid will include the continuing sites of Grid3, together with additional existing DOE and University facilities and campus Grids. The initial deployment of Open Science Grid will support an application mix from the US LHC collaborations, LIGO, SDSS, CDF, D0, BaBar and STAR experiments, GADU, SnB from ACDC, FMRI at Dartmouth, other bio-informatics applications such as BLAST, as well as computer science demonstrator applications. Open Science Grid governance documents describe the process by which Organizations join. That process is being ramped up slowly at first to debug it and avoid swamping the system. This document describes the technical requirements and milestones for OSG deployment.

The Open Science Grid environment will be an evolution of the Grid3 common grid infrastructure. The main goals of the evolution are:

- To continue and improve operation of a reliable production Grid for users while incorporating new resources.
- To take the next steps towards a widely used production Grid for our science Stakeholders.
- To further demonstrate interoperability between the OSG and LCG infrastructures, and demonstrate initial interoperation with the TeraGrid.

Open Science Grid is a “marketplace” for resource providers and consumers to barter for exchanges of value. As such, it defines a minimal set of common services and interfaces and allows for participants to form any combination of partnerships consistent with the charter. Members of the OSG consortium may provide software to implement services which conform with these requirements. Effort is taken to collect useful, consistent sets of this software into OSG maintained sets, however there is no requirement that participants run common software – rather they must provide conformant service interfaces. Thus an OSG “Release” refers to a set of functionality and agreed interfaces rather than the collection of software which provides exemplars of this functionality and interfaces. This statement of principle is tempered with the practical observation that many service definitions and interfaces are not yet stable enough for bulletproof specification, so members are encouraged to use common software where possible. One of the desirable features of the “marketplace” model is that the consumers will determine what level of conformity they demand and which features are most desirable.

The initial deployment will include a number of additional or extended services to those on Grid3. The Deployment is now being provisioned with the following elements:

- OSG Governance documents defining:
  - What is required to “be in” OSG.
  - What is the process for joining or leaving OSG.
  - What are the allowed uses of the OSG.
- Documentation for OSG installation, configuration and operation.

- Processing resources accessed through an OSG 0.2 Release Compute Element interface.
- Monitoring sufficient for operation of the OSG infrastructure and accounting of the number of jobs being run.
- Storage resources accessed through an SRM v1.1 interface and supporting GridFTP for data transfer.
- An Operations infrastructure sufficient to support the system and the users at least at the level of the Grid3 experience, including documented and operational support mechanisms.
- A Security Incident Response and Handling system.
- An OSG-wide Service Catalog.
- Support for applications beyond those on Grid3 including: US LHC analysis applications, D0, GRASE, Star and BaBar simulation jobs

The following principles of Grid3 will be maintained:
- Experiments must be able to effectively interoperate and run their applications on non-dedicated resources.
- Applications must be able to install themselves dynamically, thereby imposing minimum requirements on grid facility managers.
- The grid architecture consists of facilities (e.g., execution and storage sites), services (e.g., a prototype operations center, an information system for resource publication and discovery, and so on), and applications.

1.1 OSG Deployment Plan

The Open Science Grid deployment activity is anticipated to repeat on approximately yearly timescales as new collections of functionality improvements warrant a coordinated effort to evolve the OSG. As mentioned before, at any time, there will be "legacy" elements of OSG from the previous Release, those for the current generation common functionality, and specialty elements with limited scope or "warranty". The initial deployment of OSG, as the first of these release cycles, seeks to put in place the supporting framework for this cycle as well as the particular elements of this release.

The timetable used for this deployment is maintained on the Provisioning web page at http://osg.ivdgl.org/twiki/bin/view/Provisioning/WebHome#Plan_of_Work

1.2 "Flipping the Switch" between Grid3 and OSG

As part of the provisioning activity sites on Grid3 will transition to OSG. Once this is complete the evolution of Grid3 to OSG will be "done". In order for this to occur the current Grid3 production grid must be reconfigured, the Site Administrators, Operations Staff and VOs must be aware and understand the change.

2 OSG Release 0.2

The goal of an OSG Release is to define a set of services and functionality that can be used as a basis for OSG applications. It is expected that over time, the mix of services actually present on the OSG will not map exactly onto any one OSG Release, however, one should be able to sensibly talk about an OSG Release as representing a well-defined subset of the OSG. Services are expected to understand their compatibility with various OSG Releases and to advertise this in their service description.

OSG Releases depend on the Virtual Data Toolkit (VDT). Many of the components in the software release are packaged, distributed and supported as part of the VDT. OSG Release 2 depends on VDT 1.3.5 or later. VDT 1.3.5 includes Globus Toolkit version GT3.2. OSG Releases use Pacman v.3 for the packaging, installation and configuration of the common grid infrastructure. The recommended software packages are available through central pacman caches, together with example configuration scripts.

The Open Science Grid Release 0.2 is has the following elements:

Compute Elements (CE)
- Gatekeeper with GT2.4 GRAM compatible interface.
- A batch system to manage job execution within the CE
- Local storage available to applications through a Posix interface

Storage Elements (SE)
- An SRM v1.1 compatible interface.
- Support for GridFTP data transfer.

OSG Operations Organization
- A distributed support organization with a central operations organization, with Indiana iGOC as the central point of contact.
- The organization Support Centers coordinate with each other and the iGOC for OSG operations and support.

OSG Service Catalogs
- Clarens Discovery Service is the common Web Service registry for OSG
- Gridcat is the common registry for Compute Elements

OSG Monitoring and Accounting Service
- MonALISA is used for CE and SE accounting (provides backward compatibility with Grid3)

Organization Membership Service (VOMS)
- Each Virtual Organization is required to maintain information about the users that have access to OSG resources through their membership in the VO.
- The VDT VOMS service (rebuilt from the INFN source to match GT) is the reference VOMS service for OSG.
- Organizations are required to have a documented membership registration process

Other services will be deployed on OSG. In fact, we expect Organizations will build substantial systems atop this base, but this common set defines the minimum functionality of OSG Release 0.2. Each of these services has a detailed description of requirements later in this document.

Recognized limitations of the Open Science Grid Release 0.2 (it is anticipated these will be addressed in future Releases):

- Data management services (include replica services) are the responsibility of the VO. In the initial deployment of OSG we are unable to support common data management services.

- There will be no grid-wide scheduling services or brokering of the workload management.
- Policies will, in general, be manually described and implemented by the VOs and/or Sites. The exception is that the priority of job execution on the sites will be through the batch system schedulers.

3 Deployment Process and Organizational Structure

The organization infrastructure for Open Science Grid includes Technical Groups and Activities, both of which contribute to the overall program of work needed for deployment. The few broadly scoped Technical Groups coordinate and oversee a technical or programmatic area. Technical Groups such as Security and Support Centers are responsible for gathering the requirements and overseeing the work done in a particular technical domain. The many Activities provide the organizational framework for the contributions and tasks of the program of work. Equivalent to a "project," an Activity will, in general, live long enough to meet a set of pre-defined deliverables, then disband. Once an Activity is defined, it is expected that the stakeholders will provide the effort needed to successfully meet the deliverables, subject to the oversight of the appropriate Technical Group(s).

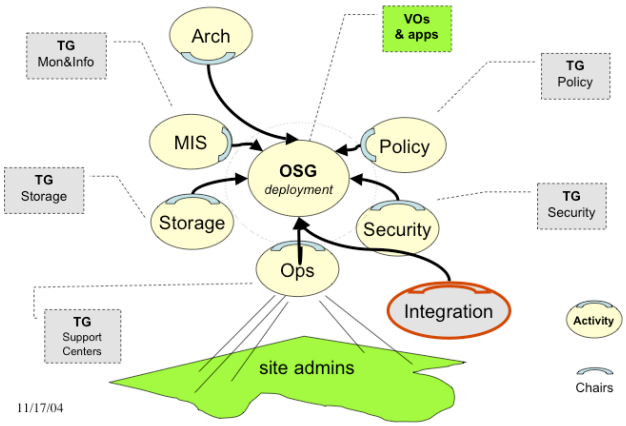
3.1 The Deployment Activity

The Deployment Activity defines and coordinates the program of work based on the requirements of the stakeholder contributors. The current Deployment Activity is responsible for delivery of the first deployment of the Open Science Grid. The OSG Deployment Activity includes:

- The co-chairs of the Activities contributing to the initial deployment of the Open Science Grid.
- Liaisons from each of the VOs contributing applications to run on the Open Science Grid.
- Liaisons from each of the VOs providing Facility and Campus Grid infrastructures which will interface to and integrate with the Open Science Grid.
- Liaisons to the partner grid infrastructures – the LCG, EGEE and TeraGrid.

The scope of the deployment activity includes the deployment coordination role and so recommendations and guidance from the blueprint group are reconciled with the deliverables and schedules. Maintaining technical coherence and making the strategic technical decisions and planning (e.g., which services to deploy when) are included in the responsibilities.

A graphical view of the relationship of Technical Groups and Activities to the Deployment Activity is shown below:



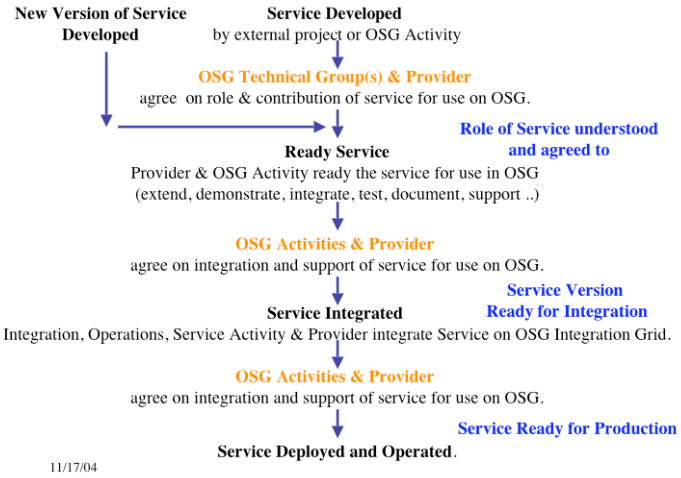
3.2 The Integration Activity

The Integration Activity has a unique role. It provides the coordination of and support for validation and testing of services, technologies, and applications before they are deployed into the stable production infrastructure. Services and applications are initially tested and used in parochial testbeds or grids outside of OSG. Once they are to be deployed on OSG they must be tested in an integrated environment to ensure they are robust, operate at the scales needed, can be operated and supported in a production environment, and provide the necessary documentations and tests. The Integration Activity is responsible for the integration and validation of the services to be deployed, and for testing consistency, performance, and robustness of the common infrastructure.

For example, if we want to add an implementation of or new interface to a Storage Element or Service: we expect that the service comes with a way to install it, a description of what the expected functionality is, and a method to validate the installation to confirm that the installation was a success. We expect that the service comes with a test harness that exercises all the functionality. We want to be able to say that it passed all the intended functionality at the intended scale: of the order of 10 files, 1000 files, of the order of 10 users, 1000 users,, whatever is agreed to between the developers and proponents of a service and the applications that will use it.

3.3 Service Deployment Process

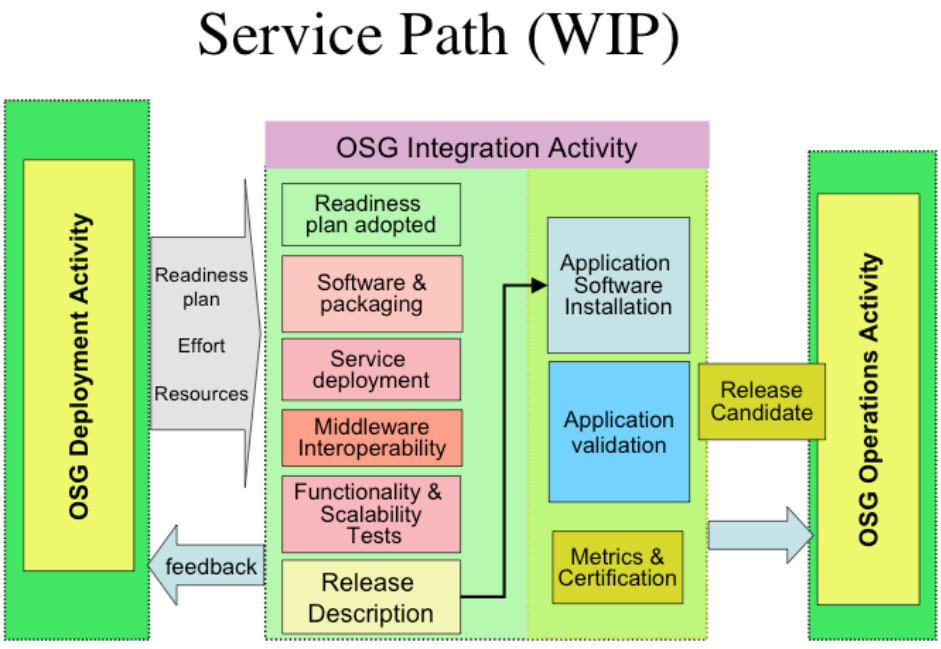
Services and VO based applications are proposed and deployed on OSG through a lightweight process and series of steps designed to provide for robustness and maintainability of the infrastructure and prevent denial of service and performance breakdown of the grid.



A short Service Readiness Plan is presented for each service to be deployed. A template for a readiness plan is included in Appendix 2.

3.4 Provisioning Activity

Once the Integration Activity has identified a set of software and services that can be deployed and supported on OSG it recommends the version of the pacman cache as "ready for deployment". The Deployment Activity reviews and accepts the recommendation.
A Provisioning Activity is created to take the software cache from Integration to Operations, to review and revise documentation, and oversee the first installations of the Release.



4 OSG Release 0.2 Services

This section provides the definitions for OSG Release 0.2 standard services.

4.1 Compute Elements (CE)

Requirements for OSG Release 0.2 Compute Elements (CE):

Interfaces:

- Gatekeeper with GT2.4 GRAM compatible interface
- Grid3-info.conf file containing Grid3 schema element assignments
- Accounting interface to support information to MonALISA through the VO-modules.

Functional Capabilities:

- Job queuing system such that jobs are either accepted and queued, or rejected if busy
- Must be able to accept (though not necessarily interpret) VOMS extended proxies
- Must accept validation jobs from OSG operations VO (MIS) to be listed as "operational" in OSG Gridcat catalog.
- Local Storage
    - $APP points to the path to local application storage space available to VO applications.
    - $DATA points to local data storage space accessible to each worker node within the CE.
    - $TMP points to local temporary caching space accessible to each worker node within the CE.
    - $WNTMP points to local temporary caching space accessible to the individual worker node.
- VOs are expected to implement policies to prevent overload of the headnodes.

Execution Environment:

- grid3-info.conf file must be executable to establish Grid3 schema element environment variables for a grid job.

Recommendations for use of OSG Release 0.2 CE's

- Applications should use the Grid Monitor interface and submit jobs using Condor-G to mitigate the performance and scaling issues with the GT2 gatekeeper interfaces on the headnode of each site.
- An alternative may be provided by the emerging Condor-C interface to job execution.
- CE's wishing to utilize extended VOMS attributes for dynamic mapping and/or role based access control are advised to use the OSG provided tools:
    - PRIMA modules connect the Globus gatekeeper with OGSA Authorization services (eg. GUMS)
    - GUMS provides dynamic mapping of DNs to local UIDs
- It is recommended that $TMP and $DATA be a minimum of 50 GB, and $APP 10GB in size.

4.1.1 Resource Allocation Policy

For resource allocation policy will be enforced through batch queue priorities as for Grid3. Additional information relating to policy such as http://griodine.uchicago.edu/%7Ecldumitr/data/up.status.latest.up.nicer.txt will be made available.

It is strongly recommended that CE's provide at least 2 priority levels. They should apply a lower priority for monitoring and validation jobs such that they do not negatively impact the running of the application jobs.

4.2 Storage Elements (SE)

Grid accessible Storage Services on the OSG are defined through an SRM interface and publication of information through the GLUE SE Schema.

Requirements for OSG Release 0.2 Storage Elements (SE):

Interfaces:

- A SRM v1.1 compatible interface
- Publish of the storage information through the GLUE SE schema to Gridcat.

Functional Capabilities:

- Must accept validation jobs from OSG operations VO (MIS) to be listed as "operational" in OSG Gridcat catalog.

We are currently readying 2 implementations of Service Services – the srm/dCache and lbnl/drm. Testing of the interoperability of these services is part of the deployment activity and any differences will be noted as part of the operations documentation.

4.2.1 Storage Element Types

Permanent Storage:
Permanent storage will be provided by Storage Elements with SRM interfaces. The meaning of "permanent" is negotiated through direct agreement between the Site and the VO. BNL, Fermilab and LBNL already have such installations. The work to be done includes providing the information about the services, incorporating them into the monitoring frameworks, and agreements between the storage resource providers and the VOs who will use them.

Transient Managed Storage:
The primary motivation is simply to provide application and site administrators with a rational way to reserve and allocate storage resources to the Grid3/OSG user community at the VO level of granularity. The current technologies expected to be deployed are:

1. DRM: An SRM implementation from LBNL. It can be deployed on top of a normal unix type file system (local or NFS). It is already part of the VDT as an optional install, and has been tested outside of its development environment to some degree.

2. dCache+SRM: This is a virtual file system type implementation with SRM.
3. NEST

4.3 Monitoring and Accounting Services

Grid3 included several parallel monitoring and information systems. OSG will continue to facilitate multiple tools for monitoring however, there is concern about load on resources due to multiple monitoring tools and inefficient use of multiple collection mechanism. Following a monitoring workshop in 2004, the monitoring and information Technical Group has worked towards definitions of interfaces for required and recommended information to support monitoring and accounting on OSG.

We have to balance needs from resource providers, VO specific support systems, and interfaces to partnering grids - EGEE, LCG, campus grids like GLOW , GRASE, FermiGRID, and TeraGrid.

MIS-CI is proposed as a common OSG resource information framework to which all monitoring systems would subscribe. A version exists and can optionally be used in OSG today.

For OSG 0.2, we have decided that service monitoring itself is optional, however, each of the available monitoring services may have individual requirements.

For accounting purposes, there will be a common MonALISA instance as a central repository for OSG. To receive "credit" for OSG production, a resource provider will need to report utilization data to this central accounting repository. There are detailed requirements being worked now to determine what that requires for CE's and SE's connecting to MonALISA in OSG 0.2. It is clear that a more general model is needed to account for a wider variety of services in the future. This is left as a work item for future Releases.

The OSG Release 0.2 Accounting service has the following requirements:

Interfaces:

- Resources utilizing this service must report to MonALISA the data required to support the MDviewer plots from Grid3
    - For CE's this means:
        - Run VO_modules on the gatekeeper node
        - Run the MonALISA deamon on the gatekeeper node
        - Provide the local account to VO mapping file at (standard location)
    - For SE's this means:

Functional Capabilities:

4.4 OSG Service Catalogs

OSG strives for a heterogeneous, dynamic environment where services may come and go, provide differing levels of service, provide differing versions of interfaces, etc. The current state

of the art is not capable of providing such an environment, so we make a number of compromises and different attempts to progress toward that goal. Release 0.2 includes multiple service catalogs that have different features and capabilities. To provide for a common infrastructure, OSG designates a particular catalog as "primary" for each of the Release 0.2 services. For CE's and SE's the primary catalog is the Gridcat instance run by the OSG Operations Activity. For VOMS, the primary catalog is the VOMS list maintained on the OSG Operations website. The Clarens Discovery Service is designated as the primary catalog all Web Services based OSG services.

The Clarens Discovery Service is the closest approximation we have available to the OSG goal and so new OSG services are strongly encouraged to review how they might utilize this catalog or one like it.

Requirements on an OSG Release 0.2 Catalog Service:

Interfaces:
- Gridcat query API for Gridcat
- ASCII file import of vomes format file for VOMS catalog
- jClarens API for Clarens Discovery Service

Functional Capabilities:
- A catalog must have a documented procedure for accepting new entries. This procedure should strive to be automated, rapid, and low maintenance overhead.
- A catalog for an OSG standard service should record the result (pass, fail, indeterminate) of periodic attempts to validate the service.
- The catalog query interface should support retrieval and selection based on the most recent validation attempt result.

4.5 Virtual Organization Membership Services (VOMS)

OSG VOs are required to provide membership information for their VO. The functional requirements for an OSG Release 0.2 VOMS are:

Interfaces:
- An edg-mkgridmap compatible query interface to get list of member DNs
- (optional) A VOMS Admin interface to support obtaining VO attributes (eg. for creation of a VOMS extended grid proxy)
- (optional) Registry of the VOMS Admin service with the Clarens Discovery Service

Functional Capabilities:
- Must contain (links to) a backing store of contact information to facilitate contact with a member in case of problems.
- Must has a written member registration process describing how new members are authorized.

Within OSG, not all VOs will be given access to each resource. It will be the responsibility of the Resource Administrator to configure their resource correctly for the VOs they wish to support. Similarly, the VOMS administrator will be responsible for configuring the VOMS correctly to support all resources the VO wishes to use.

Recommendations for use of OSG Release 0.2 VOMS's

- VOMS admins are recommended to configure their VOMS to permit world read of their membership list to minimize maintenance overhead in accessing new resources.
- The OSG software package includes several tools to facilitate use of VOMS attributes for dynamic account mapping and role based access control. These are collectively know
- VOMRS is a tool provided to VO administrators for managing a VOMS database and is particularly useful for large VOs (>100 members).

5 OSG Operations

The Operational Model will be an extension of the Grid3 central operations model (run by the Indiana iGOC) to a distributed set of Support Centers. A full operations plan is being developed in parallel with the Deployment Activity.

5.1.1 Documentation and Publication

Information about each activity and the technologies deployed will need to be available from the Document Repository.

5.1.1.1 Security

The following documents and agreements are being developed for OSG

Incident Response Plan

Acceptable Use Policy – for users, and services (including resource providers).

5.1.2 Incident Response

Operations team will maintain the incident response mailing lists and support infrastructure for the adhoc incident response teams.

6 References

1. Open Science Grid web site http://www.opensciencegrid.org
2. Grid2003 web site http://www.ivdgl.org/grid2003
3. Samgrid web site http://projects.fnal.gov/samgrid
4. Grid2003 Plan http://www.ivdgl.org/grid3/documents/document_server/uploaded_documents/doc--707--Grid3_v21.pdf
5. Grid2003 Lessons Learned http://www.ivdgl.org/grid3/documents/document_server/uploaded_documents/doc--751--Grid2003%20Project%20Lessons-6.doc
6. Teragrid http://www.teragrid.org
7. LHC Computing Grid http://lcg.web.cern.ch/LCG/
8. EGEE http://www.cern.ch/egee/
9. Grid User Management System (GUMS) http://grid.racf.bnl.gov/GUMS/
10. Site Authorization Service (SAZ) http://www.fnal.gov/docs/products/saz/SAZ.htm