Open Science Grid

Document Name	**A Blueprint for the Open Science Grid**
Version	**1.2**
Date last updated	March 2009
OSG Activity	Blueprint

The Blueprint for the Open Science Grid records the guiding Principles and builds an evolving road map for the design, development and operation of Architecture and Services in the OSG infrastructure. The Blueprint in its mature form will provide a basis for planning a coherent and composite technical program of work, which will be utilized in the course of various iterations of different OSG Activities. The document is being prepared through consensus of the participants in the Blueprint Activity, and subject to review by a Review-Circle. The document is available from http://www.opensciencegrid.org/documents/ .

---

1 Introduction....3
2 Definitions....4
2.1 The Open Science Grid....5
3 Principles, Best Practice and Requirements....6
3.1 Principles....6
3.2 Best Practice....7
3.3 Requirements....8
3.3.1 Resource Providers & Sites....8
3.3.2 Virtual Organizations and Dynamic Workspaces....9
4 Discussions....9
4.1 Namespaces....9
4.2 Naming Conventions for Data....10
4.3 Data Management....11
4.3.1 Storage Elements....12
4.3.2 Storage Resource Managers....12
4.3.3 File Catalogs/Logical Namespace Managers and Metadata Services....12
4.3.4 Replica Catalogs....13
4.3.5 Higher-Level Data Management Services....13
4.4 Storage Management....13
4.5 Federation Services....13
4.6 Architecture of a Service....13
5 Architectural Decomposition....14
5.1 Basic OSG Components....16
5.2 Symmetry & Recursion relating Users, Resources, and VOs....16
5.3 Job and Data Management....18
5.4 Interfacing the Facilities....19

5.5 Areas of Responsibility 19
6 References 20
7 Appendix: LIGO Data Publishing and Access Scenario .... **Error! Bookmark not defined.**

V1.2		Add OSG definition of Cloud, Add principle for security components expecting to establish trust from the requestor.	
V1.01		Add new definitions for Community (Cyber)-Infrastructure	

V1.0		Stable since 2004 to 2008	
V0.9	12/24/04	Clean up	RP
V0.8	10/24/04	Input from October Blueprint	HN, RP
V0.7	10/24/04	Split the document – no text changes, only deletes	RP
V0.6	9/12/04		Post Blueprint meeting
V0.5	9/6/04		Prepare for Blueprint meeting
V0.4.2	8/15/04		Distributed to Joint Committees
V0.4.1	8/08/04	Additions	Ian Fisk, Rob Gardner, Ruth Pordes
V0.4.0	8/01/04	Comments and changes	Wyatt, Ruth
V0.3.5	7/30/04	Comments and changes	Abhishek Rana, Jerome Lauret, Conrad Steenberg, Frank Wuerthwein
V0.3.0	7/19/04	Tidy up	
V0.0.0	7/15/04	Blueprint face to face	Distributed to Review Circle who attended phone call

1 Introduction

The Open Science Grid Consortium will build a sustained production national infrastructure of shared resources, benefiting a broad set of scientific applications. The organization and framework for the consortium is described on the web site at `http://www.opensciencegrid.org`. This Blueprint for the Open Science Grid provides the guiding principles and roadmap for the building and operation of the infrastructure and will provide a basis for planning a coherent technical program of work. The Blueprint does not provide the actual plan or decisions on technologies for implementation. The Blueprint does include the broad outlines or principles of an architecture to support the technical goals. The Open Science Grid Consortium will work through a set of self-organized Activities. As work progresses, these activities will be integrated through a Technical Coordination Group.

The OSG infrastructure is being built and deployed through the set of Activities, each of which involves some or all of the participants in the Consortium. Within each Activity there are a dynamic and evolving set of participants, applications, services, and resource providers. Contributions are subject to ongoing negotiation with the associated activity, and are not statically defined at the start.

The Open Science Grid infrastructure relies on many diverse projects (research, development, design, operations) and groups who may be participants in the Open Science Grid Consortium but whose projects are outside the boundary of the organization's framework itself. This Blueprint takes account of this structure and in general refers to the documents of these projects rather than duplicating the information.

The Blueprint is guided by overarching principles to make the infrastructure – both conceptually and in practice – as flexible and functionally simple as possible, to build from the bottom up a system which can accommodate the widest practical range of users of current Grid technologies, in a context which maximizes the future convergence of those users to greater commonality in technology choices. The infrastructure spans multiple Grids. The production quality, scale and internal consistency of the infrastructure, its broad scope and the diversity of its client communities lead to additional requirements and principles in support of sustained and robust operations.

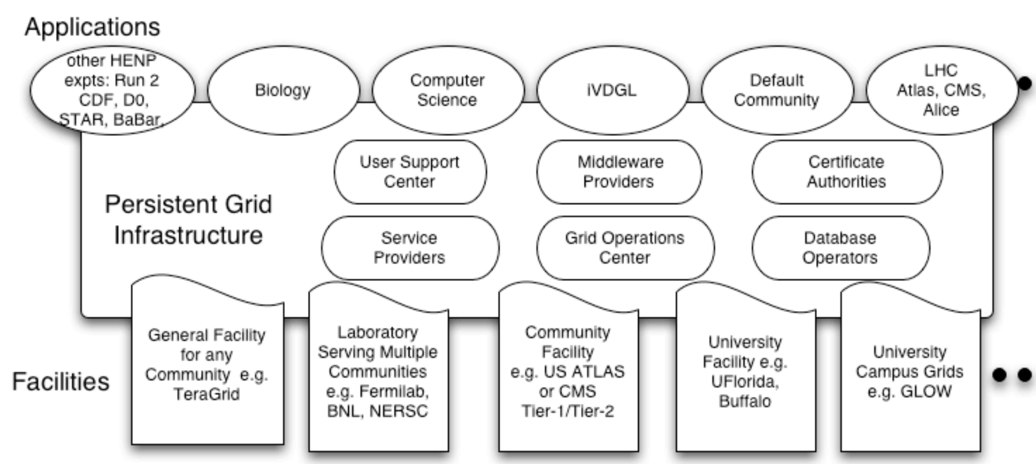


Figure 1: The Open Science Grid

2 Definitions

The basic terms are defined within the scope of the Open Science Grid. An attempt has been made to define a useful set of simple definitions upon which the end to end infrastructure can be built. Definitions that follow dictionary definitions and standard usage are not repeated here.

- **User** – A person who makes a request of the Open Science Grid infrastructure.

- **Resource Owner** – has permanent specific control, rights and responsibilities for a Resource associated with ownership.

- **Agent** – A software component in OSG that operates on behalf of a User or Resource Owner or another Agent.

- **Consumer** – A User or Agent who makes use of an available Resource or Agent or Service.

- **Provider** – Makes a Resource or Agent or Service available for access and use.

- **Ownership** – A state of having absolute or well-defined partial rights and responsibilities for a Resource depending on the type of control. OSG considers two such types: actual Ownership and Ownership by virtue of a Contract/Lease. A Lessee is a limited Owner of the Resource for the duration of the Contract/Lease.

- **Service** – A method for accessing a Resource or Agent.

- **Site** – A named collection of Services, Providers and Resources for administrative purposes. A Facility is a collection of Sites under a single administrative domain.

- **Virtual Organization** – A dynamic collection of Users, Resources and Services for sharing of Resources (Globus definition). A VO is party to contracts between Resource Providers & VOs which govern resource usage & policies. A subVO is a sub-set of the Users and Services within a VO which operates under the contracts of the parent

- **Virtual Site** is a set of sites that agree to use the same policies in order to act as an administrative unit. Sites and Facilities negotiate a common administrative context to form a "virtual" site or facility.

- **Dynamic Workspace** – A persistent, extensible, managed collection of objects and tools hosted on a grid that provides a runtime, real time environment and services for running applications.

- **Policy** – A statement of well-defined requirements, conditions or preferences put forth by a Provider and/or Consumer that is utilized to formulate decisions leading to actions and/or operations within the infrastructure.

- **Contract** – Agreement between Consumer(s) and/or VO(s) and/or Provider(s) expressed through Policies. Simplest contract is a consumer-provider match based on their policies.

- **Delegation** – An entrustment of decision-making authority during transfer of request for work or offer of resources from a User or Agent to another Agent or Provider, or vice versa. The latter is provided with a well-defined scope of responsibility and privilege at each such layer of transfer of request or offer.

- **Economy** – Set of benefits made and costs accrued as seen by Consumers and Providers, including an understanding of matching benefits between the parties.

- **Security** – Control of and reaction to intentional unacceptable use of any part of the infrastructure.

- **Grid** – A named set of Services, Providers, Resources, and Policies, overlapping and/or including other Grids operating as a coherent infrastructure in support to the contracting Virtual Organizations. Providers may delegate their contracts with the participating VOs to the Grid administration.

- **Community (Cyber-)Infrastructure** - A set of services and software that has been established by a community to meet the needs of its members. The management of the distributed infrastructure is the responsibility of the community, and the resources are all, or nearly all, owned by the VO and members.

- **Gateway** – A set of software and services that transforms and maps information, jobs, data, between Grids. Aka Bridge.

- Cloud – A set of Services, Providers, Resources and Policies providing a single point of access for all the computing needs of consumers. The resources are not necessarily owned by the consumer, but may be leased or otherwise "accessed.

- "Identity" Federation is a set of one or more Organizations and a set of zero or more Certificate Authorities that are Trusted. A Federation provides information about Individuals and the Organizations (e.g. to a CA).

Referenced Definitions:

- **Namespace** - `http://en.wikipedia.org/wiki/Namespace`

- **Resource** - Item 2 and 5 at `http://dictionary.reference.com/search?q=resource`

Notes:

There are approximate pairs of definitions that correspond to each other: User/Owner and Consumer/Provider. These pairs are not perfectly symmetric as User strictly refers to a person while Owner generally refers to an institution. There is some symmetry at the agent level such that both members of a pair delegate to engage in contracts in order to achieve their 'economic objective' within their expressed policies.

2.1 The Open Science Grid

The Open Science Grid (OSG) is the grid under the governance of the Open Science Grid Consortium operated as a sustained and production infrastructure for the benefit of the Users.

Other grids may operate under the governance of the OSG Consortium, for example the grid that validates the infrastructure before it becomes the OSG. The Open Science Grid includes facility, campus, and community grids that participate in the Consortium; The Open Science Grid interacts with grids external to the Consortium through federation and partnerships.

The Open Science Grid VO is open to those Users and VOs that have contracts with the OSG.

3 Principles, Best Practice and Requirements

Principles are basic rules and guidelines that govern (guide and influence) the fundamental aspects of the model, methods and architecture.

Best Practices are guidelines to be adhered to, as much as is possible, in practice. They are guided by the availability and use of existing components and technologies.

Requirements are formal statements that provide goals and constraints on the designs and implementations. Requirements affect functional aspects of the architecture, and can be presented through a set of Use Cases. The goal is to have a minimal set of requirements for participation in the OSG infrastructure.

The Principles, Best Practices and Requirements are not necessarily targeted for initial deployments of OSG. They are directed towards the long-term goals and requirements for the final infrastructure.

3.1 Principles

Principles are intended to apply to end-to-end use cases as well as the common infrastructure. For example, they are meant to be applied to the error handling, monitoring, information, security and management infrastructures, as well as the services and applications.

The OSG infrastructure must always include a phased deployment, with the phase in production having a clear operations model adequate to the provision of production-quality service. The new components and services required will be engineered, developed, integrated and progressively deployed in a way compatible with the continued evolution of OSG, in a series of carefully planned steps.

Policy should be the main determinant of effective utilization of the resources. This implies that without governing policy there would be full utilization of the resources.

The OSG architecture will follow the principles of symmetry and recursion.

Services should work toward minimizing their impact on the hosting resource, while fulfilling their functions. (Any tradeoff between benefit and impact will constraint their design).

Services are expected to protect themselves from malicious input and inappropriate use.

All services should support the ability to function and operate in the local environment when disconnected from the OSG environment. This implies the local environment has control over its local namespace.

OSG will provide baseline services and a reference implementation. Use of other services will be allowed.

The OSG infrastructure will be built incrementally. The roadmap must allow for technology shifts and changes.

Users are not required to interact directly with resource providers. Users and consumers will interact with the infrastructure and services.

The requirements for participating in the OSG infrastructure should promote inclusive participation both horizontally (across a wide variety of scientific disciplines) and vertically (from small organizations like high schools to large ones like National Laboratories).

VOs that require services beyond the baseline set should not encounter unnecessary deployment barriers for the same.

3.2 Best Practice

The OSG architecture is Virtual Organization based. Most services are instantiated within the context of a VO. The OSG baseline services and reference implementation can support operations within and shared across multiple VOs.

Services may be shared across multipleVOs. It is the responsibility of the Service and Resource Providers to manage the interacting policies and resources.

Resource providers should provide the same interface to local use of the resource as they do to use by the distributed services.

Every service will maintain state sufficient to explain expected errors. There shall be methods to extract this state. There shall be a method to determine whether or not the service is up and useable, rather than in a compromised or failed state.

The OSG infrastructure will support development and execution of applications in a local context, without an active connection to the distributed services.

The infrastructure will support multiple versions of services and environments, and also support incremental upgrades.

The OSG infrastructure should have minimal impact on a Site. Services that must run with superuser privileges will be minimized.

System reliability and recovery from failure should guarantee that user’s exposure to infrastructure failure is minimal.

Resource provider service policies should, by default, support access to the resource. The principle ‘services should protect themselves’ thus implies that services should additionally have the ability to instantaneously deny access when deemed necessary.

Allocation and Use of a Resource or Service are treated separately.

Services manage state and ensure their state is accurate and consistent.

3.3 Requirements

Published information from resource providers, sites and services must be accurate.

All services must be (recursively) discoverable by the OSG discovery service. Registration implies name, contact identifier and other specific information.

Users, resources and service providers must accept the OSG Acceptable Use Policy. Services which receive delegated credentials additionally agree to be honest stewards.

A User must be a member of at least one participating organization (at least for the time being).

A service must be offered to at least one VO.

The minimal requirements for participating in the OSG infrastructure are: the ability to advertise services in the common infrastructure; to accept use of one or more resource by applications running on the infrastructure; and to abide by the security requirements; and to interact with the OSG services as needed for successful participation.

A minimal requirement on a Site is to provide some resources for OSG services and transient storage space for any job input and output. The amounts required for useful participation will evolve.

VOs, Sites and service providers will need to cooperate in order to permit the tracing of each transaction to a responsible user. (May not be the original user but a VO administrative user for example).

Policy of a resource provider takes precedence over the policy of a site which takes precedence over the policy of a VO which takes precedence over the Workspace (or sub-VO) policy. (In situations where the natural site policy is in conflict with that of the VO that owns it, a negotiation for consensus happens outside of the scope of OSG)

A VO (or sub-VO) can sublet a resource to another VO (or sub-VO). This transfers allocation of the resource & appropriate privileges, subject to policy and contracts, to another VO, but does not transfer Ownership.

A top-level Discovery Service, the main functions of which will be: (a) given a kind of service, return a list of service instance references; (b) given a service instance name, return a service instance reference. The Discovery Service will operate hierarchically.

3.3.1 Resource Providers & Sites

Sites can act as an administrative unit for: contracts with VOs; resource management and allocation; and providing services shared between VOs.

Sites may support a subset of the infrastructure, services and types of resource. A site should advertise its capacities and capabilities.

Sites must provide at least the well-defined set of OSG minimum services.

Sites need to be able to trace the responsible User when accessed.

Sites may deny access to a particular User and/or a VO based on security as well as contract and policy constraints.Permanent and durable storage space is provided by agreements between a VO and one or more Sites.

3.3.2 Virtual Organizations and Dynamic Workspaces

Sub-VOs operate under the context (contracts and policies) of the parent VO.

The execution environment is the responsibility of and within the scope of the VO and/or the Dynamic Workspace.

A VO must support use of VO based Dynamic Workspaces to the level of single transactions.

Validation of the infrastructure is the responsibility of the VO for their particular applications.

Resources and services can be shared by, and transferred between, VOs and Dynamic Workspaces.

VOs may have latency as well as performance requirements.

4 Discussions

These discussions have not yet resulted in well understood principles or requirements. Subsidiary documents may be available for more information:

4.1.1 Community (Cyber-)Infrastructure

Community (Cyber-)Infrastructures provide a set of services and software which has been established by a community to meet the needs of its members. The management of the distributed infrastructure is the responsibility of the community, and the resources are all, or nearly all, owned by the VO and members.

OSG provides services or resources to the Community (Cyber-)Infrastructure, or vice versa, based on agreements:

- Operational security
- Operations services such as trouble ticketing, monitoring, resource validation, accounting.
- Software from the common software stack.
- Brokering access to resources not owned by the Community or within OSG.
- Application services such as resource selection, workload management, data replication and archiving.

OSG and Communities collaborate to:

- Facilitate appropriate adoption of common and externally supported software and services to increase the cost effectiveness across the organizations.

- Make (more parochial) software designs and implementations more general, robust and effective.
- Facilitate sociological understanding and acceptance of the principles of the OSG, use of the contributions offered by OSG, and through this increase the combined value of OSG and the Community.

4.2 Namespaces

A namespace is a collection of names in which all names are unique within their semantic groups. Names in a distributed system can be organized in namespaces, which can be represented as directed graphs. The process of looking up a name is known as name resolution, and a knowledge of how and where to start resolution is generally referred to as closure mechanism.

An ideal namespace management scheme is expected to rely not on maintaining globally unique absolute names, but rather on schemes that exploit the relative uniqueness of names in the local namespaces.

OSG will consider various namespaces and their management. Namespaces may be defined by and potentially shared between any entity in a grid or VO. The scope will include grid and local namespaces.

Each Service potentially has namespace scope and responsibilities to manage. E.g. Physical – Device level; Logical – within the VO; User – which may be meta-data driven.

The “opensciencegrid” namespace is available for use by Services, Providers. VOs and Sites through a contract with the Open Science Grid consortium. Request for and review of such names is through the appropriate Technical Group.

4.3 Naming Conventions for Data

Over the last several years, a variety of physics projects have come to a loose consensus on how to name data objects in a grid environment. This naming convention consists of several levels of names, as shown in Figure 2.

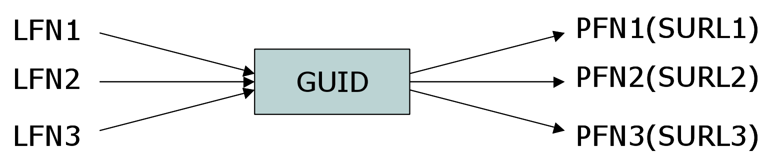


Figure 2: Shows a typical naming convention used by the physics community.

At the highest-level, there are *Logical File Names (LFNs)* that are unique within the defined namespace. A Logical File Namespace is generally defined and managed within the scope of a Virtual Organization. VOs may share LFN namespaces. Logical File Names are human readable and normally structured with the syntax of a UNIX file path and name.

The naming convention allows one or more logical names or aliases to map to a single Globally Unique Identifier (GUID) for a data item. GUIDs are generally considered to be automatically generated names that may have no particular significance to a human reader. There are multiple conventions for generation GUIDs, and several open source GUID generators are available. In general, these generators create a unique name based in part on the MAC address of the machine generating the GUID and on the current timestamp. In some systems, GUIDs are not exposed to the end user.

[Open issue: to what degree is global uniqueness is guaranteed by a particular GUID convention?]

[Open issue: should all OSG participants agree on a particular convention for generating GUIDs for interoperability?]

The naming convention allows one-to-many mappings between GUIDs and either Site URLs or physical file names, explained below.

Optionally, there may be a Storage Resource Manager (SRM) associated with each storage element or physical storage system. An SRM manages local storage and may assign and expose a logical name for each data object called a *Site URL* or *SURL*. Use of SURLs allows movement of data objects by the SRM to be opaque to the upper levels of the system.

Finally, a Storage Element itself assigns and manages physical names for each data object, for example, the physical path of a file on a file system. A *Physical File Name (PFN)* identifies both the storage resource and the location in which the data is stored. This name must allow the Storage Element and the name of the file as stored to be identified.

While Figure 2 shows a typical naming convention, the use of this convention is not universal. Some physics projects want to eliminate the separation between logical names and GUIDs and have a single unique logical identifier for each data object. Recently, the EGEE/ARDA project proposed the use of a single, unique, user-defined and mutable logical name associated with a GUID; in addition, there may be symbolic links associated with that logical name.

For the purposes of the OSG Blueprint document, we will assume the naming convention shown in Figure 2.

4.4 Data Management

Data (files) registered to a Grid are identified by a unique identifier within the GUID[1] namespace.

Data are stored in named containers, which can be nested and which are registered to the grid as Files. Files are given a unique identifier in the GUID namespace. OSG is agnostic on the question of the mutability of containers.

Several components are likely to be present in a typical Open Science Grid deployment. At the lowest level are physical storage systems or *Storage Elements*. Above these are optional *Storage Resource Managers* that optimize the use of storage elements. One or more catalogs exist at the next level of the hierarchy, possibly including a *logical namespace manager* (sometimes called a *file catalog)*, a *metadata catalog* and a *replica catalog*. Finally, we group together the collection of *higher-level data management services* that may include publication services, sophisticated

[1] UUIDs/GUIDs, ISO/IEC 11578:1996 http://www.iso.ch/cate/d2229.html, or DCE 1.1: Remote Procedure Call http://www.opengroup.org/publications/catalog/c706.htm

data movement and replication services, data validation services, and workflow managers. Next, we discuss each component in more detail.

4.4.1 Storage Elements

Storage Elements are physical sites where data are stored and accessed, for example, physical file systems, disk caches or hierarchical mass storage systems. Storage Elements manage storage and enforce authorization policies on who is allowed to create, delete and access physical files. They enforce local as well as Virtual Organization policies for the use of storage resources. They guarantee that physical names for data objects are valid and unique on the storage device(s), and they provide data access.

4.4.2 Storage Resource Managers

A Storage Resource Manager (SRM) is a component that may optionally be associated with a Storage Element. SRMs optimize the use of Storage Elements, for example, by providing file caching or staging of files from hierarchical storage. The use of SRMs is very likely in Open Science Grid deployments.

SRMs provide an additional level of logical names or Site URLs (SURLs) for data objects stored on the storage element. SURLs provide the SRM with the ability to move data objects on the storage element (for example, to facilitate efficient storage, retrieval, or free space management) without exposing that data movement to the upper levels of the data management hierarchy. In particular, the use of SURLs allows files to be moved at the storage element without requiring that the associated mappings in a replica catalog be updated.

4.4.3 File Catalogs/Logical Namespace Managers and Metadata Services

A Logical Namespace Manager maintains many-to-one associations between logical file names and GUIDs. It guarantees that newly-created logical file names have not been previously used in the system. It answers queries based on a logical file name or a GUID, and it enforces authorization policies about who is allowed to create, delete and access associations between logical names and GUIDs.

For many physics applications, a particular type of logical namespace manager is considered desirable: one that provides file-system like interfaces and semantics. In particular, this namespace manager provides associations between logical names for files and GUIDs. It groups files into directories and provides a hierarchy of directories. It provides a POSIX-like interface that allows creation, listing and deletion of directories as well as files. Examples of file-oriented logical namespace managers include the POOL file catalog, AliEn and the EGEE/ARDA file catalog.

Logical names for data objects are actually a particular type of metadata attribute associated with the data object. Thus, Logical Namespace Managers can be considered a specialized type of *Metadata Service*. They may contain additional metadata attributes, such as domain-specific metadata attributes and generic attributes such as creator, creation timestamp and checksum.

Alternatively, there may be separate Metadata Service(s) in the infrastructure that provide associations between metadata attributes and GUIDs for data objects. These Metadata Services provide the ability to perform attribute-based discovery of data objects.

4.4.4 Replica Catalogs

A replica catalog manages many-to-one mappings from the GUID namespace to either the SURL or physical file namespace. As a file is replicated or moved to other storage resources, the Replica Catalog Service maintains mappings to the physical file names of replicas. The replica catalog also allows users to associate attributes with a GUID or SURL/physical file name. Typical attributes include file size, access control information and a checksum associated with the file. The replica catalog responds to queries based on GUIDs, SURLs/physical names, or attributes. The catalog also enforces authorization policies regarding who is allowed to create, delete and access its mappings. The replica catalog implementation may be either centralized or distributed. A distributed implementation offers greater scalability and reliability, but raises some difficult consistency issues, for example, how to guarantee that a GUID registered at one catalog site has not been registered before or how to delete all mappings associated with a GUID if some of the replica catalog servers have temporarily failed.

The Virtual Organization sets policies that define what constitutes a valid replica. In most cases, replicas are defined to be files that contain identical data and have the same checksum.

4.4.5 Higher-Level Data Management Services

[Open issue: We need to discus what higher-level data management services, if any, should be included in the OSG Blueprint.]

4.5 Storage Management

VO's contract for storage space with storage resource providers (ranging from guaranteed to opportunistic use). A site providing storage for multiple VOs may manage the resources as a common service with common policies and operational procedures and dynamic mechanisms for resource sharing and allocation. These are transparent to the User (and the VO?)

4.6 Federation Services

The Open Science Grid will also provide an LFN Mapping Service to map an LFN structure between OSG and a grid it is federating with (e.g. Teragrid)

4.7 Architecture of a Service

Some of the OSG principles and best practices affect the architecture of each Service.

Services can enhance robustness through self-management and monitoring – for example, ensuring that if the service crashes it is automatically restarted.

The Replica Catalog Service must be distributed such that there are local catalogs well connected to the Storage Elements.

No service should present a single point of failure.

When a service fails, an attempt needs to be made to determine why. This can be done right after the initial restart-attempts; or in some case it will be important to do this analysis Before attempts to restart the Services (e.g. if that would trigger other failures).

4.8 Policy Infrastructure

Policy plays a central role in the OSG since it enables participation by and contributions from a diverse set of VOs and resource owners, across organizational boundaries. Both VOs and owners need to be able to express their policies This document will help define such contracts, what they will mean and what levels of obligation they imply:

All contracts with OSG are “best effort”. Contracts between members of OSG may be more formal.

Dynamic site and VO policies will be supported.

A framework needs to be in place that allows both application (soft enforcement), enforcement and reconciliation of dynamic policies.

Note: Enforcement is generally restrictive. It’s the basis for trust for both users and owners that their policies are adhered to. Policy reconciliation is enabling; it allows both users and owners to reach their *economic* objectives.

The challenges in policy enforcement are largely in reliable delegation and guarantees on maximum latencies for revocation. E.g., a site may change its policy and expect the new policy to be enforced within some time limit. How are contracts revoked that no longer satisfy the changed policy?

For policy reconciliation, the challenges lie in allowing policies that are sufficiently expressive. A very powerful policy reconciliation paradigm that OSG expects to employ is the concept of *matchmaking* between offers of and requests for resources. However, this is unlikely to be sufficient. In addition to simple matching, OSG infrastructure will most likely require a concept of preference, or ‘rank’, a notion of quota on aggregated resources, and the possibility to encode hierarchically structured policies.

An example of a hierarchically structured policy that needs to be supported is hierarchical fair share. A site invariably has policies that express preferences for some VOs over others. The VO in turn needs to express policies that reflect (some of) its organizational structure. It is thus a requirement for the end-to-end policy infrastructure to allow for policy expressions that may be depicted in form of a decision tree, or a directed acyclic graph.

5 Requirements on the Security Infrastructure

All participants in OSG have some security responsibility.

Security is integrated into the OSG activities and plans.

Sites determine what CAs to support, informed by the VOs they wish to support.

GSI is the authentication protocol for the near future so Users will have to have the ability to acquire Proxy credentials. While OSG will not have to operate this infrastructure, authentication methods available to the VO's users and acceptable for the resource requests they need to make must be developed.

A robust authentication challenge process must exist.

Trust between the requestor (for example, through providing an identity) and a security service shall be established (separately) before any functional response.

An OSG service will negotiate, as part of its contract with a VO, how to determine if a requesting identity is a member of the VO. For the default VO, some general OSG membership service will need to be defined. In the near term, this might be a union of all the specific VO membership services.

A baseline acceptable use policy is provided for use of OSG. VOs may have more restrictive AUPs that are invoked if users assert their VO membership authorization.

A Policy Enforcement Point control (or set of controls) needs to be implemented for services that are expected to enforce authorization decisions.

Services which accept delegated credentials must be auditable to resolve claims of challenged authentication and exposed risk. Action logs on grid services must be sufficient to determine which identity was associated with all processes and which AA tokens might be exposed in an incident.

The OSG will provide an Incident Response service.

A common method of restricted delegation will be provided

Intrusion Detection is an important part of the OSG security infrastructure.

The OSG will provide Recovery Procedures for those services it supports

There is a policy that governs the publication of data owned or stored by the OSG.

Identify modification, multiple identities and credentials. (TBA)

Where-ever possible the OSG security infrastructure will be compatible and in common with those of grids with which it federates. When needed, interfaces and mediation between the security infrastructures will be provided.

A user shall be able to use local resources with a local identity independent of any remote services or infrastructure.

Resources and services should accept only the minimum set of authorities necessary to provide the services to their users.

Services shall protect themselves with the semantics of the functions allowed.

All services should allow configuration parameter to ban proxies of greater than a certain lifetime.

(Condor-g can do proxy renewal now)

6 Architectural Decomposition

This section includes a set of sketches to explore the architectural decomposition and it will grow with interface and service definitions, and dependencies as we proceed.

6.1 Basic OSG Components

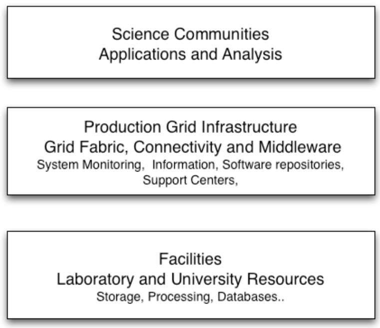


Figure 3: OSG Architecture

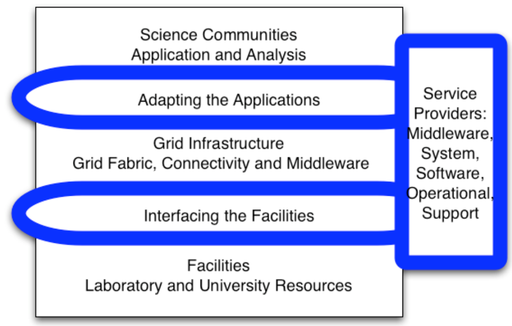


Figure 4: Missing Capabilities

6.2 Symmetry & Recursion relating Users, Resources, and VOs

OSG aims to federate across heterogeneous grid environments, large-scale distributed enterprises and communities. To facilitate this task, the OSG infrastructure views VOs as recursively-defined entities comprising of users, resources, and (sub-)VOs. The different ways a VO can be formed is shown in Figure 5. In this figure, users and resources organize themselves as VOs in order to enter into contracts resulting from negotiations based on their respective sets of policies. These contracts are manifested at the middleware level as matchmaking, and the related services are provided by the VOs.

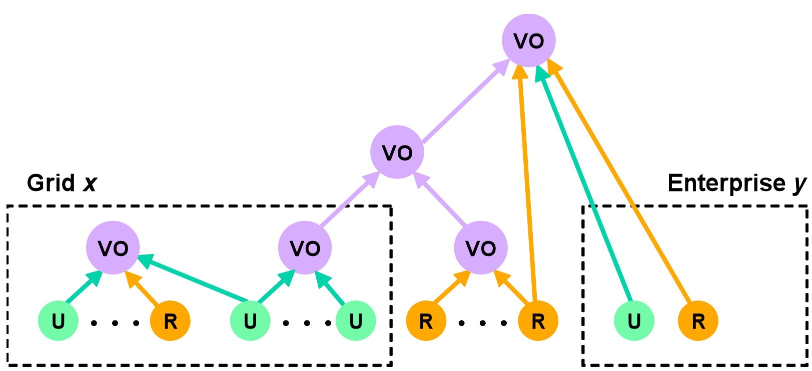


Figure 5: VO Hierarchy and Recursive VO Formation in OSG

VOs may choose to enter into sub-contracts in order to more effectively utilize their resources or better satisfy their users. For sake of simplicity, agents and services are not shown along with VOs in this figure. A VO can be solely a resource-provider or consumer or both. Figure 5b shows symmetry in this relationship by considering a typical flow of request from a user to a resource owner (via a resource provider). This figure takes into account Agents with delegated rights and policies, communicating and working together to establish end-to-end functionality. Users, Providers and Agents play roles of producers and consumers depending on the direction of workflow being considered. However, the conventional nomenclature for this role has been followed (the bold line in the figure) throughout this document. Policy representation and policy reconciliation generally implies delegation of responsibility in such a heterogeneous and dynamic environment. (This delegation may or may not include forwarding of identity and role of the user and/or resource. E.g., the cache management system of a VO generally can not be required to know which user requested what data movement as files in cache are used by more than one user. On the other hand, access to a user's quota does of course require a user's identity/role to be forwarded.)

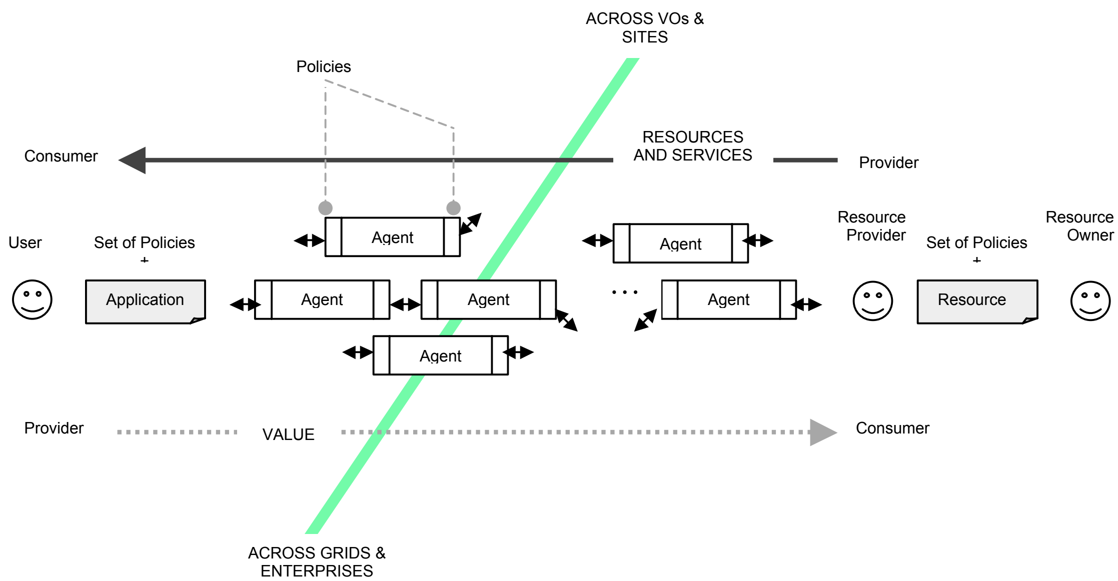
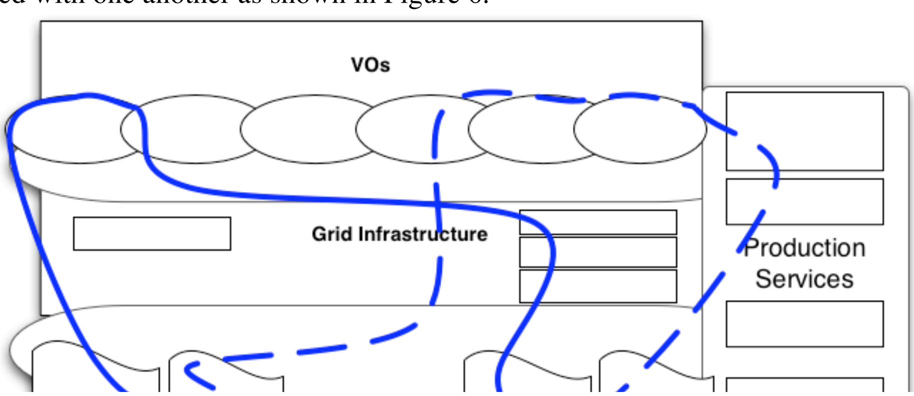


Figure 5b: Symmetry between Consumers and Providers

Each functional level in this model may have the capability to monitor its appropriate use. To make this relationship fault-tolerant, OSG may explore looking into error recovery and rollback mechanisms that would allow a workflow request to trace back by following only a limited number of steps.

6.2.1 Relationship between VOs, Grid Infrastructure, and Sites/Facilities

As a federation of grids, OSG infrastructure considers VOs and Sites to be dynamically associated with one another as shown in Figure 6.



This is made possible at an operational level by timed leases of Resources by Sites to the consumer VOs. Once party to a contract, a consumer VO takes the responsibility to dynamically deploy VO-specific services on Sites for the period of the lease. OSG will provide a persistent grid services layer and service specifications to guarantee interoperability, as well as reference implementations for those services. This includes both services provided by Sites as well as VOs. Both the Site and the consumer VO have the freedom to do monitoring and accounting in such an environment.

In the above mentioned symmetry in the OSG architecture it is important *where* a *decision* is made. Distributed systems fundamentally should allow components to have as little knowledge as suffices the need. Robustness, however, is dependent on effectual error-propagation and thus decision-making points. There is a trade-off involved since too many decision-making junctions in the workflow route may become an overhead.

6.3 Job and Data Management

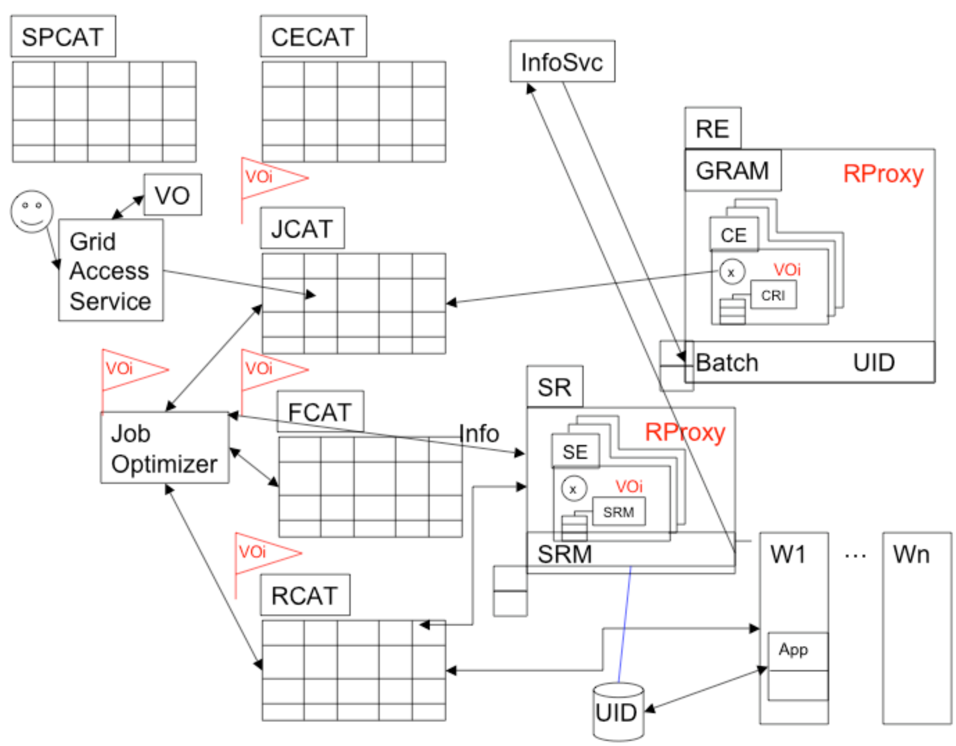


Figure 7: Job and Data Management Components

There is consensus to accept this architecture as a baseline model for review.

CE	Compute Element
CRI	Interface to Compute Resource (Condor-G)

GRAM	Interface to Compute Element
Job Optimiser	looks at list of files, decides which SR to use based on replica locations, to minimize data movement for instance. A VO can feed information from the SE (VO specific) to the Job Optimizer to tune this picture appropriately. Information also can flow from the SE to the SR.
RE	Resource
RProxy	Remote Proxy. Adapts from grid to local infrastructure (e.g. translate for private network)
SE	Storage Element
SR	Storage Resource
SRM	Interface to Storage Resource
CECAT	Compute Allocation Catalog
JCAT	Job Catalog. Has rules to control who is next in the compute center (RE). The RE can push out information that can influence the decision made by the JCAT who is going next... perhaps based on availability of RE's required services.
RCAT	Replica Catalog (a Reliable File Transfer Job Queue)
SPCAT	Space Allocation Catalog
WN	Worker Node
VOi	Possibility for VO-specific CE and/or SE implementations.

The architecture sketch depicted in Figure 7 is based on the notion that job & data management are conceptually symmetric, especially at the level of the job optimizer. In both cases, a VO leases resources from sites. It maintains catalogues of available and requested resources, and matches them based on policy driven optimization of workload throughput. This matching takes into account co-location of data and CPU as needed.

The architecture places minimal requirements on the sites. The responsibility for providing functionality is shifted to the VOs as much as possible. The latter is motivated by the notion that VOs are by definition internally cohesive wheras sites are distinct and may generally differ in a variety of ways.

6.4 Interfacing the Facilities

Facilities and sites are responsible for administering and supporting the services, resources and infrastructure within their administrative domains. These include storage, processing, network, and database services as well as the security, operations, and policy infrastructures.

Sites have services used by local or remote users not on the common grid infrastructures. Sites and facilities will support local grid infrastructures which will federate with or partially be made accessible to the Open Science Grid, local resources that will be shared with VOs accessing them through OSG. and local VOs that will want to use both the local resources as well as share those available through the OSG infrastructure.

These will be taken into account when defining each service (interfaces, capabilities, architecture) as well as in engineering the infrastructure.

6.5 Areas of Responsibility

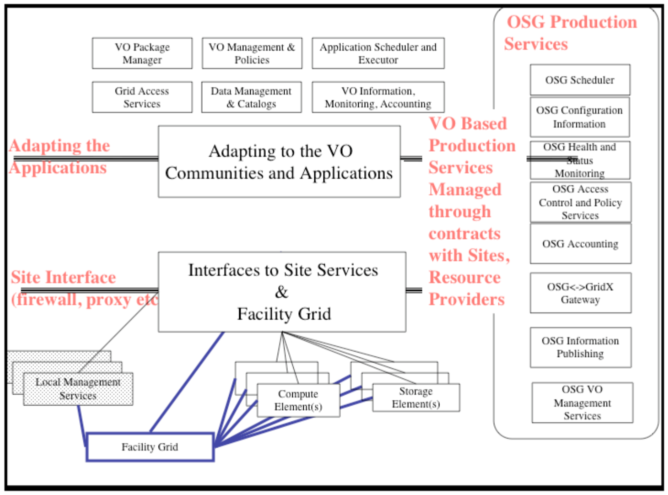


Figure 8: OSG Responsibilities

7 References

The Open Science Grid Consortium, http://www.opensciencegrid.org/.

Open Science Grid white paper, v2.3, August 10, 2003.

Open Science Grid Presentation to Fermilab Computing Division, May 14, 2003, http://cdinternal.fnal.gov/Org2003/BriefingMtg/2003Briefings/2003-05-CDbriefing.pdf.

R. Pordes, L. Bauerdick, V. White, *The Open Science Grid*, abstract submitted to CHEP 2004.

OSG Security Technical Group, http://www.opensciencegrid.org/techgroups/security/.

OSG Storage Technical Group, http://www.opensciencegrid.org/techgroups/storage/.

The Grid 2003 Project, http://www.ivdgl.org/grid2003/.

The Grid 2003 planning document, v21, GriPhyN 2003-33/Grid3 2003-3, Nov 5, 2003.

International Virtual Data Grid Laboratory (iVDGL), http://www.ivdgl.org/.

The Virtual Data Toolkit (VDT), http://www.cs.wisc.edu/vdt/.

LHC Computing Grid Project (LCG), http://lcg.web.cern.ch/LCG/.

The Particle Physics Data Grid (PPDG), http://www.ppdg.net/.

The Grid Physics Network Project, http://www.griphyn.org/.

I. Foster, and C. Kesselman (eds.), The Grid 2: Blueprint for a new Computing Infrastructure, Morgan Kaufmann, 2004.

H. Wang, S. Jha, M. Livny, P. McDaniel, *Security Policy Reconciliation in Distributed Computing Environments*, IEEE Fifth International Workshop on Policies for Distributed Systems and Networks (POLICY 2004), New York , June 2004.

I. Foster, N. Jennings, C. Kesselman, *Brain Meets Brawn: Why Grid and Agents Need Each Other*, The 3rd International Conference on Autonomous Agents & Multi-Agent Systems (AAMAS 2004), New York City, July 2004.

Grid Resource Allocation and Management (GRAM), `http://www.globus.org/gram/`.

A. Shoshani, A. Sim, and J. Gu, *Storage Resource Managers: Essential Components for the Grid*, Grid Resource Management: State of the Art and Future Trends, Kluwer Publishing, 2003.

B. Clifford Neuman, *Scale in Distributed Systems*, Readings in Distributed Computing Systems, IEEE Computer Society Press, 1994.

Tim Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web*, Scientific American, May 2001.

R. Raman, M. Livny, and M. Solomon, *Matchmaking: Distributed Resource Management for High Throughput Computing*, Proceedings of the Seventh IEEE International Symposium on High Performance Distributed Computing, Chicago, July 28-31, 1998.