

OSG Docdb xxx

August 10th 2009

DRAFT Version 1

|  |  |
| --- | --- |
| Document Name | Requirements and Principles for the Future of OSG (OSG’) |
|  | DRAFT |

# Introduction

This describes requirements and principles to which we are planning for the future of the Open Science Grid. To date these have been gathered through individual conversations with lead stakeholders, input solicited during the Council face to face meeting in March 2009 as well as the phone meetings in May and June 2009, input gathered from the “at-large VOs” by the VO group, OSG project blueprint meetings in May and July, and ongoing discussions within the OSG Executive Team.

# Goals

*OSG will provide the LHC and LIGO communities with a common virtual facility (operations services, security and software) operating in the US as an integral part of their global computing systems.*

*OSG will provide a common, shared distributed computing infrastructure which can be used, subject to policy, by all members of the Consortium – physics and non-physics, campus and regional, laboratory and national communities.*

*OSG will continue as a grass-roots hands-on collaboratory with contributions from member organizations integrated with activities of the directly funded staff to meet the mission and goals of the Consortium.*

*OSG will play a leadership role in the US National Cyberinfrastructure by supporting the nations largest global scientific communities and expanding the reach to all the nations university campus and other communities that come to the table.*

*OSG aims to have recognized responsibilities in partnership with the NSF XD program of projects.*

*OSG aims to have recognized responsibilities as part of the next round of the DOE SciDAC program of projects.*

## Principles of Operation:

We will continue to promulgate our principles and overarching guidelines:

* Apply and audit software and processes to sound system-engineering principles.
* Design and implementation to be stakeholder, user community, driven.
* Provide support for community specific distributed systems - user community end-to-end transparency and control.
* Provide services, software and support personnel for campus and region-wide shared infrastructures.
* Be flexible to changes in user needs.
* Provide a framework for research and scientific communities to build their own common user environment across heterogeneous resources.
* Cater both to research groups that require very large amounts of computational resources for long periods of time and to individuals who seek to use high-end computation to reduce the time required for running their applications to seconds or minutes so that they can rapidly and interactively explore their research questions.
* Cater to both researchers whose computations require very little data movement and to researchers who are performing very data-intensive computations.
* Include an Integration Testbed in parallel with the production infrastructure that is used for testing new services and software at scale and by the user applications.
* Rely on partner and “Satellite” projects for software developments and extensions; and to partner to provide significant new capabilities.

## Key Directions (not in priority order)

* Program of work driven by current and anticipated stakeholders.
* Solidify the concept and practice of OSG Satellites.
* Effort in core project ~same # of FTEs as current project (~33FTEs)
* Strategic plans informed by sponsors (NSF MPS, NSF OCI, DOE ASCR, DOE HEP, DOE NP, NIH?).
* Increased role and responsibilities on and across the university campuses.
* Core mission to support of the US LHC Tier-3s.
* Core mission to provide US operational services for the WLCG.
* Support for and working with Community Grids is a core activity, including the LIGO Data Grid.
* Plan for a 5-year program of work.
* Well understood and agreed to relationships, cooperation and interfaces with XD projects.

### Partnering with XD

OSG aims to approach XD[[1]](#footnote--1) proponents to discuss and agree on activities and interfaces that would have OSG having responsibility and authority for specific areas of the programs. Software, operations, support activities on the campuses is one such area.

## Core Mission

**Continue operations of the production infrastructure:**

* Improve the usability and reduce the effort to operate at all levels.
* Maintain a usable, secure, robust, distributed facility open to all contributors (resources and applications) in the scientific, research, and education domains. (No commercial use).
* Continue support for the OSG Virtual Data Toolkit for the OSG Consortium and other projects, including the evolution of WLCG and EGI.
* Solidify presence and usability of shared cyberinfrastructure within the US University campuses and transparent use of local and wide area cyberinfrastructure nationally.

**Extend:**

* The capabilities and capacities offered for the stakeholders at the table, including those represented on the OSG Consortium Council, VOs represented by the “at-large” Council member, and users through the Campus, Engagement, Education and Outreach activities.
* Strengthen organizational, operational and technical gateways between OSG and other peer infrastructures including the TeraGrid and XD, and EGI and NGIs in Europe, Campus and regional grids in the US and off-shore, and commercial cloud resource providers.
* Workforce training for the future…TBA.

# Continue Operations of the Production Infrastructure

* This section adds some detail to requirements and directions of the core mission based on input to date.
* Add support for Alice and LHCb based on requests from the experiment and discussions with DOE and NSF.

## Software

* Sustain support for the VDT. Complete support for native packaging. Simplify VDT/OSG installation, update and configuration. Make the integrated set of components more robust. Accommodate heterogeneity of needs of difference communities.
* Client Toolkits are very important
* Evolve to use web services (LIGO, ATLAS)
* Need data encryption and data security capabilities (SBGrid). 2 types of data from biomedical people - unpublished and likelihood of others getting the data is not crucial. Other data is regulated by HIPPA and has regulations for use. Clients with encryption would be good.
* Application level portals and gateway software.
* Integration with industry could be useful (SBGrid).
* Better define and publish boundaries for software regarding what OSG supports and what it doesn't support (SLAC). Determine how OSG decides when it must support software components themselves. OSG could have a role in sustaining software used by multiple communities. Continue contributions to external software developments for needs of the operational infrastructure (Fermilab).
* Full support for use of Virtual Machines (STAR).
* Better error code reporting and translation through the s/w stack with standard / homogeneous error format of logs (STAR, ATLAS?)
* Better end-to-end support for data management and storage use.
* Well-understood and simple data and storage management services.
* Include more components for network management and allocation.
* Track evolution and ensure interoperability with EGEE/EGI (CMS).
* Gradually adopt components of software sustainability paper. (<https://osg-docdb.opensciencegrid.org:440/cgi-bin/ShowDocument?docid=866>)

## Operational Support, Security and Support

* Critical operations center services – problem recording, triaging, ticketing and ownership; gateways and publishing to the WLCG services – information, accounting, security, reliability & availability reporting etc. (ATLAS, CMS)
* Provide frameworks and exemplars for general utility to the communities within the operations framework to provide general benefit and value under the umbrella.
* Lower the barrier to use, improve usability of all services and installations.
* Increase automation, alarming. Improve problem management, change and configuration management processes.
* Reduce the impact of site-to-site variability and lower the barrier to use of data across the sites (CDF).
* Extended targeted operational support for US LHC Tier 3s and small sites in general.
* Active monitoring and notification of software bugs in s/w e.g. openssl and security vulnerabilities.
* Support for Virtual Machines across all operational services impacted (STAR).
* Improve usability of end-to-end security infrastructure – identity token acquisition and management.
* Support for MPI (CIGI, LIGO?, SBGrid)
* Increase the resources accessible for sharing (CDF, D0)
* Help grow sustainable and usable (lower the barrier to entry) regional grids.
* Campus grid implementation information or assistance. Information on how smaller OSG grid installations can be effective among larger partners. (GPN, GROW)

## Community Specific

* Support for more dynamic and flexible VOs and groups (SBGrid, Council).
* Sharing and use of storage by communities that don't own it using reservation, allocation (SBGrid, LIGO, Council)
* Strengthen relationships with Community Grids for Security Incident Response and alerts.
* Strengthen relationships on workflow and community capabilities, condor etc.
* Develop policies as to how a regional grid interacts with the OSG (NYSGrid)

# Extend

This section adds some detail to requirements and directions of the core mission based on input to date.

## Future capabilities and Capacities

* Understand role, impact, requirements, interfaces to and integration with commercial clouds, if any (Council).
* Develop appropriate interfaces, integration with TeraGrid.
* Virtual Machine technologies as supported resources and environments.
* Extend capabilities and capacities for managing and using data and (dynamically allocated) networks across the shared distributed infrastructure.
* Workflow – understand usability and scope of Pegasus and DAGMAN (LIGO).
* Commonly adopted standards (LIGO, ATLAS)

## Training

* Integrate Training with Community training. (SBGrid, LIGO)
* Training and education is an area in which the OSG has very little effort – benefit from collaboration with TeraGrid (Council)

## National and International aspects

* Work with specific international communities based on activities of the communities involved (LIGO - VIRGO, STAR- Korea, China, Council – China, US LHC – South America)

# Acknowledgements and References

(note: some documents have controlled access.)

National Workforce Development:

<http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=864>

Software Hardening and Infrastructure Sustainability:

<http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=866>

Original Concept of OSG Satellites: Solicitation for Preproposals, [https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=207&version=1&filename=OSG Pre-Proposal Solicitation.pdf](https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=207&version=1&filename=OSG%20Pre-Proposal%20Solicitation.pdf)

OSG and National CI - input to MPS request <https://osg-docdb.opensciencegrid.org:440/cgi-bin/ShowDocument?docid=879>

OSG Council meeting minutes: <https://twiki.grid.iu.edu/bin/view/Council/May-12-2009>

<https://twiki.grid.iu.edu/bin/view/Council/June-09-2009>

“At –Large” VO Stakeholder needs

<https://twiki.grid.iu.edu/bin/view/VirtualOrganizations/Stakeholder_PlansNeedsRequirements>

US ATLAS input for OSG year 4 plan: <https://twiki.grid.iu.edu/twiki/pub/Management/Year4Planning/1.0.6_ATLAS_Requirements.doc>

US CMS input for OSG year 4 plan <https://twiki.grid.iu.edu/twiki/pub/Management/Year4Planning/1.0.5_CMS_Requirements.doc>

Input from STAR for OSG Future [https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=880&version=1&filename=OSG -STARInput.doc](https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=880&version=1&filename=OSG%20-STARInput.doc)

Communications from members of the Executive Board, ATLAS, CMS, Fermilab, LIGO, SBGrid, SLAC, STAR

1. From the NSF solicitation: *“*The primary goal of the next phase of the TeraGrid is to enable major advances in science and engineering research, in the integration of research and education, and in broadening participation in science and engineering by under-represented groups, by providing researchers and educators with usable access to extreme-scale digital resources, beyond those typically available on a typical campus, together with the interfaces, consulting support and training necessary to facilitate their use. For this reason, we refer to the next phase of the TeraGrid as “eXtreme Digital”, “XD.”

   For many researchers and educators, it is likely that the digital resources that they draw upon will include both XD resources and computing systems and data collections on their home campuses, as well as digital data from observing systems or archives distributed around the world. XD should facilitate this. In addition, it is anticipated that an important external resource for some researchers will be the DataNet partnership (see NSF 07-601). One of the design goals of the architecture of XD should be ease of integration of XD and DataNet services.

   As noted above, it is anticipated that in the lifetime of XD, many researchers and educators will undertake projects that draw upon a mixture of XD resources, computing systems and data collections on their home campuses, and digital data from observing systems or archives distributed around the world. Drawing on lessons learned during the operation of the TeraGrid, it is important that the design of the architecture of XD facilitate, to the extent possible, the ability of users to “attach” local or other remote resources, such as a computing service or a data collection, to their XD environment in such a way that it can be used as if it were logically a part of the XD environment. For example, a researcher may wish to use this capability to execute a complex scientific workflow that combines the use of XD resources and local campus resources, or a user may wish to design a task that can, at different times, be executed on local campus resources and XD resources”. [↑](#footnote-ref--1)