The Open Science Grid: Collaborative Science on a Shared High-Throughput Distributed Facility

Ruth Pordes[1], Mine Altunay[1], Paul Avery[2], Alina Bejan[3], Kent Blackburn[4], Alan Blatecky[5], Robert Gardner[3], Sebastien Goasguen[6], Bill Kramer[7], Miron Livny[8], John McGee[5], Doug Olson[7], Rob Quick[9], Alain Roy[8], Torre Wenaus[10], Mike Wilde[3], Frank Würthwein[11]

1 The Open Science Grid (OSG)

The goal of the OSG is to transform processing and data intensive science through a cross-domain self-managed nationally distributed cyber-infrastructure that brings together campus, laboratory and scientific communities and supports the needs of collaborative research at all scales.

To meet this goal the OSG maintains and extends a robust secure heterogeneous facility, distributes and supports a growing common software toolkit, and enables a growing set of research and educational communities. The OSG provides access to resources and services to further the goals of a diverse set of important science projects, while at the same time solving unique technical and sociological challenges.

1.1 Introduction

The Open Science Grid (OSG)[1] is a unique consortium providing an innovative open distributed cyber-infrastructure engaging scientific research communities across the United States. The consortium is a multi-disciplinary collaboration of scientists, researchers, IT providers, software developers, educators and administrators. The members cooperate to run a diverse mix of data and computationally intensive applications. The goals of the consortium are executed through a project, jointly funded by the Department of Energy Office of Science (DOE/SC) SciDAC-II program and the National Science Foundation (NSF)[2].

---

[1] Fermi National Accelerator Laboratory, [2]University of Florida, [3]University of Chicago, [4]California Institute of Technology, [5]Renaissance Computing Institute (RENCI), [6]Clemson University, [7]Lawrence Berkeley National Laboratory, [8]University of Wisconsin-Madison, [9]Indiana University, [10]Brookhaven National Laboratory, [11] University of California, San Diego

Science communities contributing to and supported by the Open Science Grid include the thousand person, peta-scale particle physics experiments at the Large Hadron Collider[3] in Switzerland (represented by the US ATLAS and US CMS collaborations in the US), the Laser Interferometer Gravitational Wave Observatory collaboration (LIGO)[4], the Condor[5] and Globus[6] computer science groups, and the STAR[7], CDF and D0[8] US accelerator based experiments. Organizations contributing to the OSG storage and computing resources include the DOE multi-purpose facility at NERSC, the Fermilab and Brookhaven National Laboratories, more than twenty "Tier-2" university facilities that are organized parts of the above scientific collaborations, and more than twenty other university groups.

This paper explores the unique nature and accomplishments of, and challenges faced by, the OSG in its' multi-dimensional environment (Figure 1). In particular, the:

- Governance of the organization and distributed execution of an agreed upon program of work;
- Operation and extension of a heterogeneous production computational facility;
- Goals and implementations of common software, security and other services;
- Economic model;
- Mix and complexity of the applications; and
- Engagement and education of new communities, educators and students.

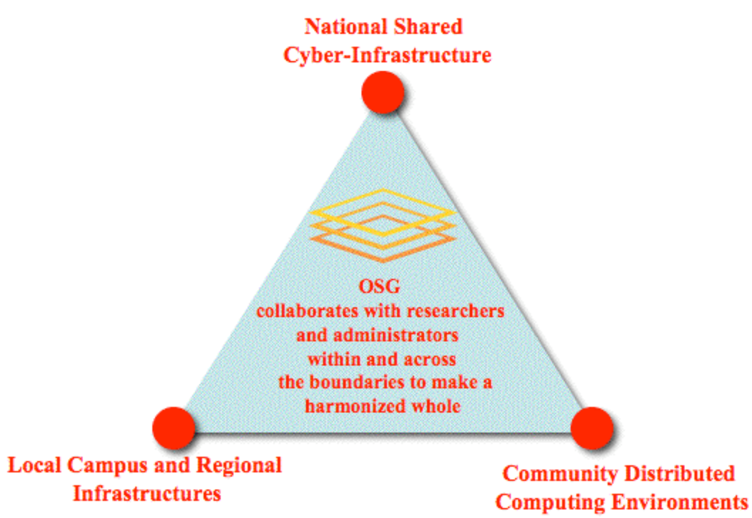

Figure 1: The Environment of the Open Science Grid Consortium

The OSG supports collaborative research at all scales, from the single-PI and small group to very large global communities. The OSG supports research in a diversity of scientific fields. The OSG facility provides access to storage and processing resources in university department clusters, campus-wide and regional cyber-infrastructures, and high-throughput and high-performance computing facilities at the international as well as the national and local level. The OSG contributes to the DOE/SC and NSF objectives to provide the scientific community access to world-class computational facilities[9] and to democratize collaborative research and education through the provision of shared cyber-infrastructure[10].

1.2 Background

It was recognized early in the planning stage of the experiments at the Large Hadron Collider (LHC) accelerator in Switzerland that the computational scale and complexity of the data analysis could only be met through collaboration of the worldwide physics community. This led to a globally distributed computing and data storage system implementation involving more than thirty countries, that supports round-the-clock data distribution rates of more than a Gigabyte per second and enables transparent access to more than a hundred thousand processors. It was further realized that many other sciences – e.g. biology, fusion, climate modeling, geology and, indeed, the social sciences - had rapidly increasing needs for data collection, access, modeling and analysis, and would soon reach the tera- and peta-scale. Government agencies used these challenges as an opportunity to engage computer and domain scientists in joint projects to stimulate timely solutions to the domain scientists' needs, encourage the validation of computer science innovations in distributed computing technologies and methods, and to foster a broad transformation in collaborative research.

Between 1999 and 2006 in the US, the DOE SciDAC-1 program sponsored the Particle Physics Data Grid, the High-Performance Data Grid Toolkit, and DOE Science Grid, projects and the NSF sponsored two complementary ITR projects, the Grid Physics Network and the International Virtual Data Grid Laboratory, that included the physics collaborations mentioned above, astrophysics collaborations, and the Condor, Globus and LBNL computer science communities. These projects were led and implemented by collaborating teams of computer scientists and scientific researchers – a model that continues today.

In 2003 the leadership of these projects recognized a common need for an operational distributed infrastructure for all to use. As a result of this a grass roots effort, the "Grid3" common infrastructure[11], was built as a demonstration for SC03, consisting of about twenty five computation sites and running several different applications. The value and benefit provided by Grid3 were demonstrated sufficiently that the projects maintained it as an operational infrastructure for the following two years. In addition, the communities working on Grid3 continued to self-organize, forming the OSG consortium and committing to further evolve the virtual facility and the communities using it. OSG was funded for five years beginning in September 2006 to evolve the support for the physics stakeholders (see Figure 2), and to broaden and deepen engagement with, and education of, other sciences.

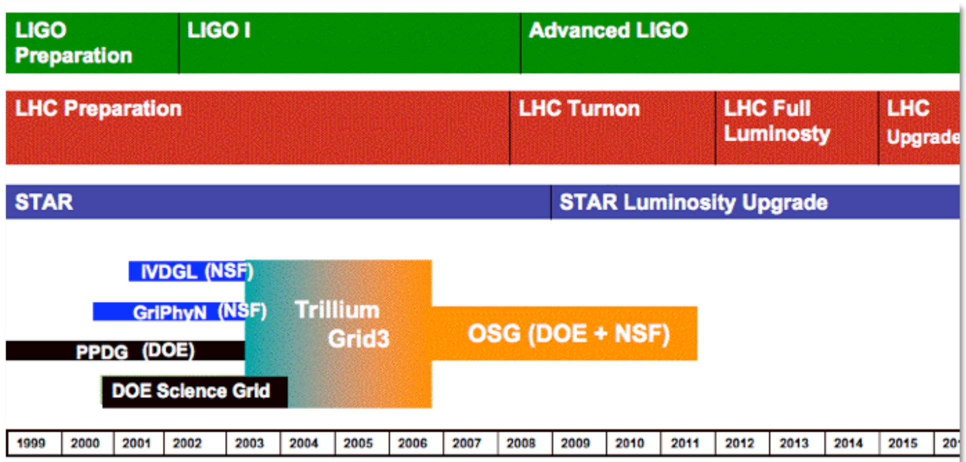

Figure 2: Timeline of the OSG Project and its Primary Scientific Stakeholders

2 Overview

2.1 *Virtual Organizations (VOs)*

VOs, autonomous and self-organized collaborative communities, are at the heart of the OSG model. A VO includes people working together, who share data, software, resources and research, to collectively achieve a common goal. The VO eco-system encompasses resources owned by the organization (permanently owned, temporarily leased or loaned) and common services used by the organization. VOs relate to, and are differentiated from, physical organizations through the delegation of physical identities – verification of the actual identity of the people, and the purchase and maintenance of physical hardware.

VOs range from "heavy-weight" long-lived, with strong governance and policies (e.g. the LHC communities), to "ad-hoc" self-organizing, groups of individuals who dynamically build on a few common ideas in much the same way as internet-based social networks (e.g. the Grid Interoperability Now, a VO that tests the interoperability between infrastructures). VOs engage with the OSG through agreements, policies and by-laws that describe the relationships between VOs and resource providers and define the expectations between the groups.

The OSG is itself a group of people with a common goal, sharing resources and services. The OSG is a (super) VO. Within the OSG-VO are sub-groups who form other VOs with related goals, software, and services: The Engage VO consists of members of OSG and new communities working together to enable effective use of the common infrastructure. Similarly the Education VO consists of OSG staff and students learning how to use the distributed facility. These sub-groups have all the characteristics of a VO, and demonstrate the iterative nature of the VO organizational concept.

2.2 The OSG Facility

The OSG Facility provides support, through the efforts of staff at nine separate institutions, for the round-the-clock operational infrastructure and grid-wide services, the deployment of the common software, system integration and testing, troubleshooting and fault diagnosis, security, customer support and the engagement of new communities. The OSG facility plays an important part in the World-Wide LHC Computing Grid by providing the distributed infrastructure in the US for the LHC experiments.

The processing and storage resources accessible through the facility are contributed by the members of the consortium. Today, the facility provides access to more than sixty separate processing resources (Figure 3), ranging from tens of thousands of cores at the NERSC high performance-computing center, to a small cluster at the NSF itself.

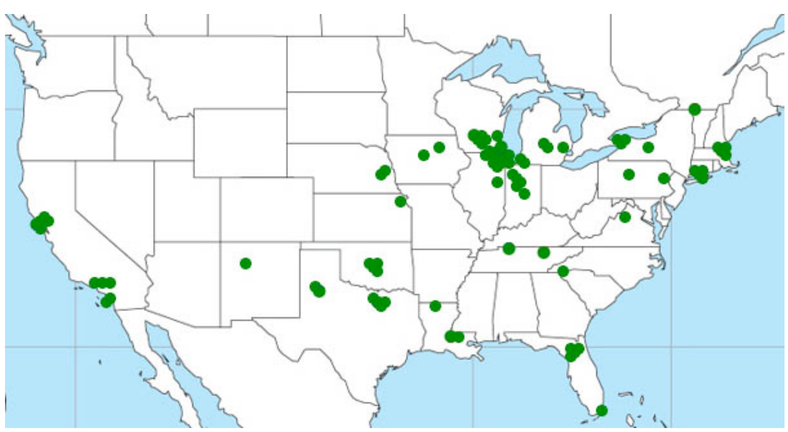

Figure 3: Location of OSG Resources in the US

More than forty-five thousand processing cores are typically accessible through the OSG facility - an increase of more than 20% over the past year. In many cases the owners (owner VOs) make heavy use of their processing resources, either locally or remotely through the OSG infrastructure. As members of the OSG consortium the owners commit to provide a percentage of the processors for use by other VOs (~5-10% when there is heavy owner use) and to share unused cycles with other OSG communities.

The OSG facility provides access to more than fifteen disk storage resources ranging from a few to 100 Terabytes in size, as well as access to more than five tape mass storage systems of more than twelve Petabytes total capacity. As with processing, the storage space is managed locally with a percentage available for use by other VOs than the owners.

2.3 The Economic model

Three policy models, operating in parallel, moderate the use of resources accessible to the OSG. The owners control and manage their resources, and are responsible for their security and operation. OSG acts as an agent between the owners and the VOs that use the resources through the OSG infrastructure.

First, the owners define Quality of Service policies for their resources to ensure they (and/or specific other VOs they agree to support) have prioritized use and/or deterministically allocated fractions of the available computational throughput. Second, the owners provide particular VOs allocations, typically for a defined duration in time, to meet a well-defined production throughput goal. Typically, the community needing such production approaches the OSG with their needs, and the OSG brokers between the owners and the VOs. Third, owners provide access to all or a subset of the OSG VOs such that any available resources can be used (so-called “opportunistic use”).

The total throughput provided by the OSG facility relies on the fact that communities receive and perceive benefit (over the short and long term) from contributing their resources according to these models. That is, they see the value in being able to use computational resources provided by other members of the consortium as outweighing the cost of allowing their resources to be used by others. For example, a VO may benefit because they can contribute resources that would be otherwise under utilized at a certain time of the year, and then be able to use more than their dedicated resource during surge periods at another point in time. Additionally, VOs see benefit from using common software and services as worth the cost of building or adapting their applications to run in the OSG software environment.

It should be noted that the policy models apply equally to resources allocated to the OSG VO itself. An example of this is the allocation at NERSC to the OSG VO. The OSG then determines the use, including sharing and policies, across all VOs in the consortium.

2.4 Software

The OSG software stack is key to making the OSG infrastructure work. The OSG Virtual Data Toolkit (VDT) is provided by the facility as a common software distribution for installation on resources for use by VOs, administrators, and users. The VDT includes the Condor and Globus technologies with additional modules for security, storage and data management, workflow and other higher level services, as well administrative software for testing, accounting and monitoring[12]. The members of the OSG consortium control the contents of the VDT and the schedule for new releases of the package.

The OSG supports multiple software versions and provides software that publishes what is available on each resource. This allows the users and/or applications to dispatch work to those resources that are able to execute it. Through installation of the VDT, users and administrators operate within a well-defined environment and set of services. The VDT currently includes the following software (roughly categorized):

- *Core Grid Infrastructure Software*: Condor and the Globus Toolkit.
- *Information Services* including information providers, LDAP repositories, accounting and monitoring services, resource validation scripts, and the Metronome build and test infrastructure.
- *Storage Service Implementations:* BeStMan from LBNL, dCache and XRootd.
- *Security tools and infrastructure:* X509 certificate management, services and utilities; resource "certificate to Unix ID" mapping tools.
- *Client Tools:* utilities for accessing OSG services; libraries to read/write data from/to grid-accessible storage sites; workflow tools.
- *Support Software:* Utilities used by many software packages, including Apache, Tomcat, Berkeley DB, MySQL, OpenLDAP, PHP, Squid web caching and miscellaneous VDT tools to help administrators, support staff and users.

The OSG supports distributions of the core middleware for other projects including the TeraGrid[13] in the US and the Enabling Grids for EScience (EGEE) [14] in Europe. Each project, including the OSG, augments the VDT with specific configuration scripts and utilities for its environment and users.

The OSG maintains a test-bed infrastructure in parallel with the production facility. The OSG staff perform regression and integrated system tests before each release of the software, in order to ensure its' quality. The number of components included in the VDT has been steadily increasing (Figure 4).

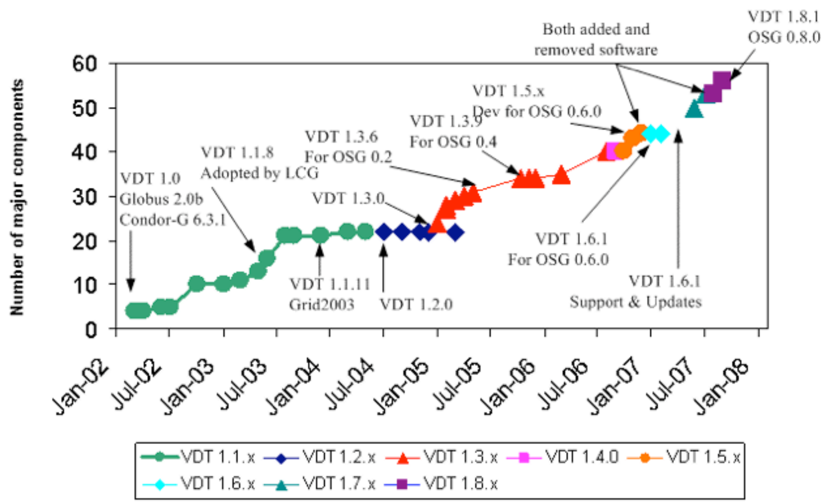

Figure 4: Releases and Growth in Number of Components of the VDT

2.5 Campus and Regional Facilities and the OSG

The OSG's mission includes advancing the use and state of cyber-infrastructure for university faculty, educators and students. The OSG works with, and supports the work of, campus and regional groups to help them build and operate their local resources as part of a distributed facility. The OSG provides interfaces between the OSG and campus/regional infrastructures, transforming adaptors and bridges between the different services (e.g. security and authentication domains), and technologies to allow data and applications to move and operate transparently across them.

There are currently five such infrastructures interfaced to the OSG at Clemson University, Fermilab, Purdue University, the University of Wisconsin at Madison and the New York State regional grid (NYSGrid). Each uses slightly different models and bridging technologies. Each shares, to a greater or lesser extent, the OSG software and operational services. For example: all the processing nodes at Fermilab run the OSG "worker node" software, whereas at Clemson the OSG software runs only on a gateway machine interfacing a Windows based clusters to the wide-area infrastructure.

2.6 Federating Grids at the National and International Level

The OSG provides interfaces and gateways between the OSG facility and other large-scale infrastructures using the same model as with campus and regional facilities, the difference between a large regional and a small national facility can be a matter for debate, in particular with the TeraGrid and EGEE.

OSG and EGEE are both part of the World Wide LHC Computing Grid (WLCG) Collaboration. The WLCG is a collective VO providing access to resources covered by formal agreements between countries participating in the LHC and providing common software and services to the four LHC collaborations. The OSG-EGEE interfaces and adaptors allow data and jobs to flow transparently between the two infrastructures through a single user interface. Users submit jobs that are run transparently on resources from either infrastructure. Data is moved and stored through common transport (GridFTP) and storage management interfaces (SRM). With the agreement of each resource owner, OSG publishes accounting and availability information to the WLCG.

The OSG is a TeraGrid Science Gateway and will provide for interoperation in both directions. Users can submit work to either infrastructure through their own gateways or portals. The OSG is currently testing a forwarder that will automatically, following defined policies, move jobs submitted by OSG VOs to run on an available TeraGrid resource.

3 Accomplishments

3.1 Increased Use and Effectiveness of the Infrastructure

The throughput of the OSG facility has increased steadily over the past year, with a maximum use being more than fifteen thousand CPU days per day (a factor of 2 more than at the start of 2007) and typically 15 VOs running jobs at any one time (See Figure 5, the peaks in summer 2007 are due to the LHC data analysis challenges and other physics users). Between 20% and 30% of the throughput is by "non-owners" of the resources used (the current assumption is that each resource has a single owner – we know this is not the case for several of the larger resources). In OSG, "opportunistic" use means the ability for one VO member to dynamically discover and use an available processing or storage resource which would have otherwise have been left idle. The physics communities typically use between 80-90% of the total cycles used on the OSG. We see significant throughput from the current round of Tevatron experiments. STAR and LIGO also met their scientific goals for use of the OSG over the past year.

More than thirty non-physics users are using the facility for production data simulation and analysis. The non-physics throughput ramped up over the month of March 2008 to around 1,500 cpu-days per day — all opportunistic use of available cycles.

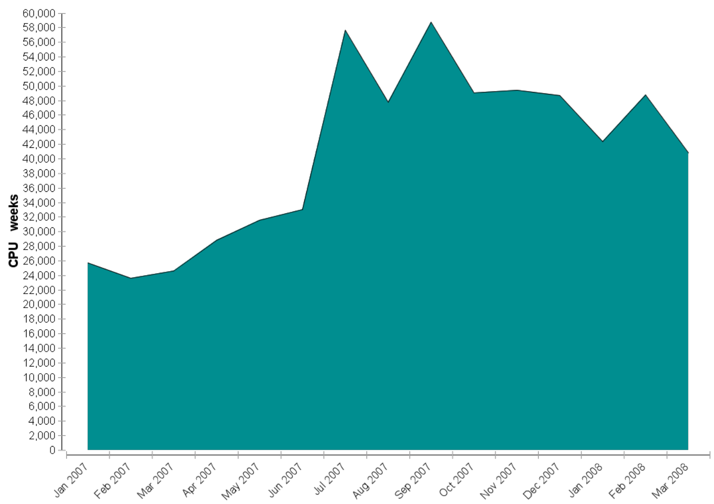

Figure 5: Total Reported OSG Usage, CPUWeeks per Week for the last 12 months

Several technology improvements over the past year have increased the usability and robustness experienced by the user: automated, local testing that regularly publishes the state and availability of the resource; a facility-wide resource selection service providing matchmaking, using Condor ClassAds, of the current state of the resources and the needs of a particular job; a client side matchmaking service that augments the facility provided information with selection criteria local to the specific user; managed storage resources accessible over the wide area infrastructure, where data needed by the applications can be independently moved and then accessed with higher performance and more robustness by the multitude of locally running jobs.

3.2 Science from the Tevatron

All the physics publications from the experiments at the Tevatron at Fermilab now depend on simulation running across multiple sites on the OSG[15]. The experiments' local resources are running at capacity to process the ongoing data acquired from the accelerator so are not available for the value added insight gained with additional computation.

A major success of the model of the collaborative approach was D0's re-processing a complete dataset in early 2007, necessitated by recalibration of the detector components. D0 requested processing of a thirty-terabyte data set (500 million events) over the period of six months, needing successful execution of more than a million CPU hours (see Figure 6). As a result of the voluntary contributions of other consortium members, over 50% of the events were processed using opportunistically available resources on the OSG. This was an important demonstration of our ability to share resources while retaining most of the use for the resource owners. D0 used more than a dozen sites, and moved over seventy Terabytes of data to accomplish the reprocessing. The benefit to the experiment is further demonstrated by their continued and increasing use of the OSG for simulations since last year's reprocessing goal was achieved.

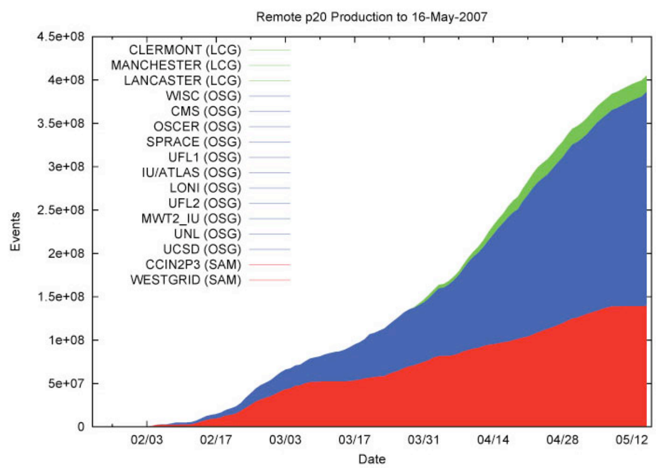

Figure 6: D0 Event Re-Processing, Spring 2007

3.3 Broader Scientific Results

Many of the non-physics communities run production applications across the facility as members of one of the OSG managed VOs – typically the Engage or OSG-VO itself. The OSG engagement staff provide embedded help to new communities to adapt their applications and make an effective interface to the OSG tailored to their particular science or research. The OSG has developed methods and tools that are reused for each new user community. The OSG provides added value to these communities by offering a means to smooth over peaks in load and usage (see Figure 7). Typically researchers perform a series of computations, followed by a hiatus for analysis of the results and development of new algorithms and hypotheses.

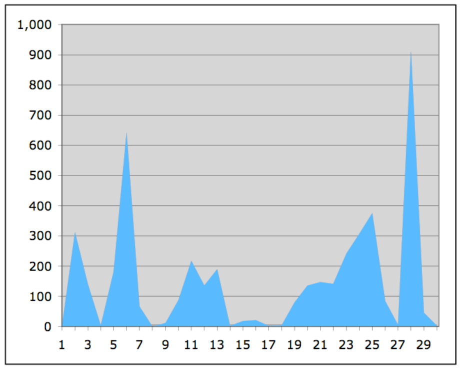

Figure 7: Engage VO CPUWallclockdays/day on OSG for February showing the cyclical nature of the use.

Three examples are illustrative of the production running on the OSG that are resulting in scientific results: Brian Etherton, at the University of North Carolina (UNC), regularly uses OSG resources to run the Weather Research Forecasting (WRF) application (see Figure 8). The OSG job submission tools allow effective running of parameter sweeps and ensembles containing multiple runs of the models. This has led to significantly improved accuracy of the results compared to individual models alone[16].

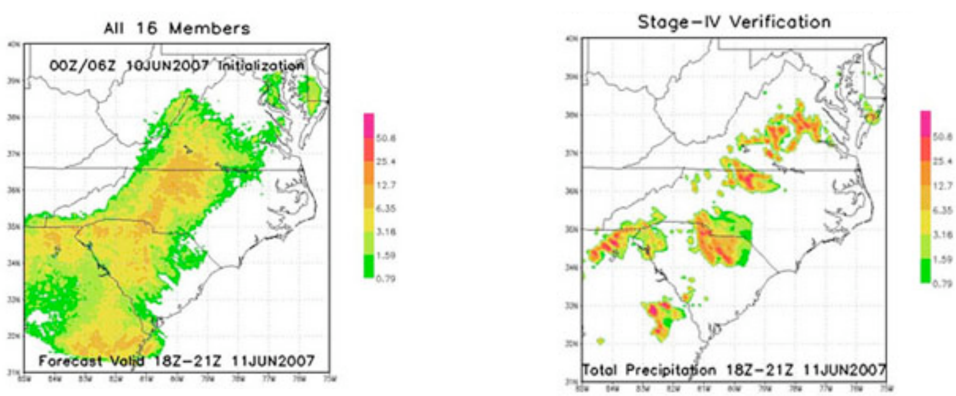

Figure 8: Ensemble of 16 different climate models to more accurately predict stormy "convective precipitation" in the Carolinas (left). Results are compared with those based on precipitation radars and gauges (right). Image Courtesy UNC

Catherine Blake, an assistant professor at the University of North Carolina's School of Information and Library Science, researches methods for retrieving, analyzing and finding relationships in published research texts from multiple disciplines. The project, called "Claim Jumping Through Scientific Literature"[17] requires a large amount of compute time in its first phase. With only a few days of work with OSG staff needed to adapt the application, Blake's research team successfully ran more than 15,000 hours of workload in just over a day across multiple OSG resources. Similarly, the Chemistry group at the University of Buffalo, running as part of the NYSGrid VO, was able to adapt their applications from running on a local cluster to using more than 10 sites on the broader infrastructure, with very little overhead. Andrew Schultz[18] is modeling the higher-order virial coefficients of water models by numerically evaluating Mayer integrals, via importance sampling. Up to 1 trillion configurations are sometimes needed to achieve the desired level of precision. In a few weeks Andrew has used more than more than 21,000 CPU-Days on the OSG. Other applications currently being run in production mode are a Rosetta protein folding application from RENCI and a, non-parallelized, version of the molecular dynamics program, CHARMM.

3.4 Contributions to the LHC experiments and WLCG

The ATLAS and CMS experiments depend critically on the OSG as their production distributed facility in the US. In 2006 and 2007, ATLAS and CMS made significant use of the OSG for their "data challenges" - data distribution and processing tests to test the capacity and capabilities of their systems. They further used the OSG for the production event simulations and analyses needed for the development and testing of the physics algorithms and, soon, to compare to the results obtained from data from the accelerator. The OSG provides software, operational support, and grid-wide monitoring, accounting and other services for the US LHC collaborations. Also, the OSG tests that new software and services interoperate smoothly with the European infrastructures and don't break the collaborations' codes.

The development and testing of the LHC infrastructure pushes the envelope in several dimensions: the sustained rate of data distribution and storage; the number of jobs submitted to the infrastructure per unit time; and the number of collaborators working on a single scientific mission. Figure 9 shows average data throughput of over 2 GBytes/Sec for several hours, a typical test, and Figure 10 shows the processing distribution by country of more than 5,000 CPUYears for ATLAS and CMS over the past year, to which the US has contributed more than 25%.

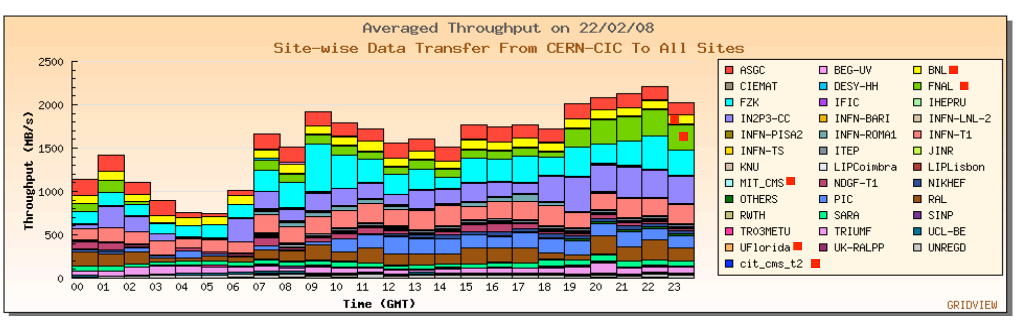

Figure 9: Data Transfer Tests on 22/2/08 from CERN to sites world wide (red dot indicates a US site).

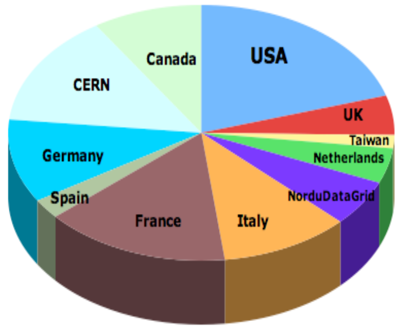

Figure 10: WLCG ATLAS and CMS processing for the past year. OSG is responsible for the USA portion.

The LHC has the largest scientific collaboration and computational distributed system in the world. The expertise, experiences, and insights the OSG is realizing from supporting their scientific applications informs and guides reuse for the other OSG communities.

4 Challenges in Sustaining a Virtual Facility for Collaborative Science

4.1 Organization and Governance

With the start of the project the OSG had to reorganize from a collaborative community into a managed program of work across more than fifteen sites, with planned deliverables and milestones, while maintaining the commitment and engagement of the broader group of consortium members. Maintaining the right balance of top-down, bottom-up, self-direction and expected deliverables is a work in progress. There remain challenges that are best addressed face to face – at least initially. Effective methods to coordinate over 50 Virtual Organizations and resource providers require a merger of traditional methods (all hands meetings, tele-conferences, etc.) and collaborative computing methods.

4.2 Maintaining and Evolving the Software and Services

The primary goal of the OSG software effort is to build, integrate, test, distribute, and support a set of common software for OSG administrators and users. The software stack must be easy to install and configure even though it depends on a large variety of complex software. Currently the VDT includes more than fifty separate components and is built nightly for more than thirteen Linux variants, with the client software built also for MacOSX and AIX.

Configuration issues are a significant challenge. Many VDT components use the same commonly available utilities such as Apache, Tomcat and libraries such as perl, and their use must be compatible. The challenges of the software include: coordinating the large matrix of components that must be verified against multiple OS versions and that must be built and tested; reducing the storage and memory footprints by ensuring components share libraries and utilities as much as possible; and balancing the benefits of including software useful to some against supporting its installation and administration by all.

Many independent groups use the OSG around-the-clock for production work. We thus face the challenge of installing new software, and new versions of software, into production without disrupting the ongoing operations and use of the facility. This can be done incrementally across the many resources, and indeed the OSG supports a mix of software versions. The installed version at any particular resource is advertised. Version skew adds to the testing and integration effort.

The full end-to-end set of software components and services can interact and fail in many complex ways. The OSG has well defined software validation and system integration infrastructures and processes for releasing new versions of the software. Full version releases of the VDT can take weeks to complete as they include testing by VOs, testing of interoperability with EGEE and the WLCG, reliability testing and documentation. The contents of the software stack are driven by the consortium members – including uses, administrators, operators, technology providers and OSG staff - and security requirements (see Figure 11). The OSG continually balances pressure for new and enhanced capabilities against the risk of disruption in service and the overhead of release preparation of an increasingly complex end-to-end set of software.

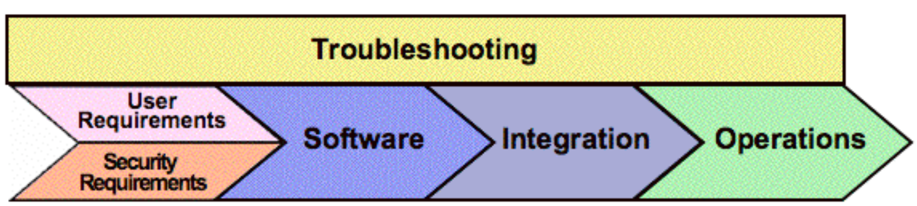

Figure 11: The OSG Software Process - from Requirements to Operations

4.3 Performance Limits and Resilience to Overload

We need to ensure the performance of each service on the OSG, as well as the end-to-end community systems, scales to the anticipated needs of the VOs. And, when a limit in performance is reached, the components of the system must to react gracefully to protect themselves and prevent breakdown and failures.

There is an effort within the OSG to test the limits of performance and understand the behavior of critical services and interfaces under performance overload. The goal of the effort is to put in place short-term mitigations and work with the software support groups to develop longer term improvements.

At the job submission client level the present performance goal is 10,000 jobs running simultaneously and 100,000 job runs per day from a single client installation, with a minimum success rate of 95%. This has been met in a variety of independent tests using realistic applications jobs. Additionally, one of the current large-scale storage management implementation in OSG (SRM as implemented in dCache [19]) has been shown to scale to more than 3,500 files handled per hour, an anticipated need of the LHC.

While we can measure the software performance limits in a test environment, this does not necessarily translate into a similar level of performance on the production infrastructure. It is an ongoing challenge to provide sufficient tools, configuration scripts, and documentation to resource administrators and VOs to help them configure the most performant system. We must provide robust and effective services, provide tools and know-how to enable solution of complex end-to-end problems, and provide sufficient information back to the operations and software teams to enable them to develop improvements. All this must become easy and robust.

4.4 Operational Security

As an open infrastructure, the OSG continuously assesses risk when putting in place security mechanisms and procedures. There are security incidents and alerts on average once a month. Each alert requires in depth investigation and analysis and then decisions on how and/or whether to continue operations.

OSG puts in place agreements and policies defining appropriate use and security expectations. However, OSG has little or no responsibility in the security internal to a site or resource, nor internal to the VOs and their management of users. Nonetheless, we try to ensure that legitimate usage of the infrastructure is robust and not impacted unnecessarily by security controls.

Timely communication between those affected and those at risk in a security incident is vital for implementing well-constructed mitigations and for remediation to be found. This frequently leads us into national and international issues of privacy and treaties. To meet these needs, the OSG provides a full time

Security Officer and dedicated security team. We pay attention to security and policy issues in all project and consortium activities. It needs significant and talented effort to define, decide and develop sufficient but not excessive best practices and tools.

4.5 Shared Data Storage and Management

The OSG aims to be a shared, common data grid, in the same model as for data processing, The software and policies to support shared data storage are not yet fully developed. Data specific issues including retention policies, space reservation, release and retrieval, as well as access control in a shared community environment, are complex problems to solve. The release of the OSG V1.0 software in May 2008 provides the first opportunity to test out the usability of and benefit from of dynamic data storage sharing by multiple VOs.

The OSG economic model includes the expectation that the owner VOs make 10-20% of the space available to other VOs, while retaining overall control and responsibility. The OSG provides software to publish the availability of the space, and plans to provide mechanisms to communicate the policies of use. The OSG will also provide common tools for VOs to make use of such storage across many sites.

4.6 A Mix of Parallel and Non-Parallel Applications

An increasing number of research-based simulation and analytic calculations require parallel processing spanning either a small number of processors (8-64, rapidly becoming “parallelism on the chip” to a large number of processors, 128-1024). The OSG is positioned to offer and support access to resources that provide effective throughput for such applications. The initial focus is on solidifying the infrastructure for sites offering MPI and that specialize in support of parallel applications. This includes ensuring that the OSG model of remote execution still works – no direct user login, dynamic installation of applications through the infrastructure, the use of well defined storage elements for anything but transient, per job data, and accurate monitoring and accounting of the resources used regardless of the degree of parallelism the application requires.

Currently three quite different sites offer MPI on the OSG – NERSC as a multi-user high performance facility, Purdue and NCSA with resources shared between TeraGrid and OSG, and the Wisconsin Campus Grid GLOW. We are currently understanding the specifics of each configuration and working out how to best abstract these into a more general common service. Issues include recompilation of the software, advertising policies that are appropriate for parallel usage, publication of the location of the MPI libraries and utilities, the maximum parallelism supported by the resource, and how to provide most effective throughput in a mixed environment of MPI and non-parallel jobs. Of the three applications currently being supported two are running production producing scientific results and working towards publication, and the third is in the testing phase.

4.7 Communities New to the OSG

For communities new to using cyber-infrastructure and distributed computing, the sociological and cultural change presents a significant hurdle. Expectations and working methodologies must be adjusted. Users and application administrators must learn to accommodate: lack of remote user login; lack of support for remote compilation (application binaries are typically installed using distributed computing tools); lack of ability to run local debuggers on a failing job. They must also accept that invariable some fraction of the resources will be unavailable at any one time.

There are also challenges to engaging new users - making the initial contact, catching them at the right time within their work cycle to ignite an engagement, and developing the trust and relationships necessary to drive lasting changes in their working methodologies. Patience is required when working with new VOs. Not only is there a learning curve on both sides, but there is also a lag time of 12 to 18 months before the first peer reviewed publications.

5 The Future

The OSG continues to follow a bottoms-up and top-down dual approach of pragmatism and attention to fundamental computer science concepts in providing high-throughput computational resources and services to collaborative science. The number of scientific communities using the OSG infrastructure continues to grow. This needs to be sustained in conjunction with a significant increase in the scale and performance of LHC collaboration use of the resources.

In the next three years the OSG will focus on improving the reliability, usability and effectiveness of the infrastructure while working with existing and new communities to support their science and research goals.

We will need to deploy and support technologies and processes to manage the OSG facility when the demand for resources exceeds the capacity. We will continue to be vigilant and active in our approaches to security, privacy and troubleshooting. We will continue to analyze our experiences to gain insight in order to evolve the technologies and organizational practices. Whether the current buzz words is “grid” “cloud” or “ecological”, at the technical level the concepts are similar and the methods and techniques the OSG is pioneering will continue to provide additional value and benefit.

6 Summary

We have described technical and research accomplishments and challenges facing the OSG. The benefit to collaborative research and the computational facilities that serve it are resulting in the engagement and expansion in the number of communities and organizations participating in the Consortium.

Our initial successes in integrating resources from campus infrastructures and high performance facilities show the model of federation works effectively without significantly increasing the operational support load on the OSG itself. This is allowing organizations to manage their own distributed facilities while providing access to and using the broader set of resources provided through the OSG.

The successful use of the OSG infrastructure by new communities demonstrates that the common infrastructure and software is reusable and accommodating to new applications and modes of use. The provision and support for a common software stack is proving its value in the adoption by an increasing number of communities, and through the requests from the stakeholders for additional capabilities and services. The next year will see a challenge to the OSG consortium as it supports the ramp up of LHC data taking while simultaneously meeting the science milestones of our other stakeholders and continuing to increase the use of the facility by an increasing number of collaborative research groups.

7 References

[1] The Open Science Grid, R, Pordes et al 2007 J. Phys.: Conf. Ser. 78 012057 (15pp) SciDAC 2007, doi: 10.1088/1742-6596/78/1/012057. OSG web site: http://www.opensciencegrid.org

[2] OSG is supported by the Department of Energy Office of Science SciDAC-2 program from the High Energy Physics, Nuclear Physics and Advanced Software and Computing Research programs, and the National Science Foundation Math and Physical Sciences, Office of CyberInfrastructure and Office of International Science and Engineering Directorates.

[3] LHC Technical design report. I. Bird, (ed.) et al. CERN-LHCC-2005-024, Jun 2005. 152pp http://weblib.cern.ch/abstract?CERN-LHCC-2005-024

[4] LIGO: http://www.ligo.caltech.edu/

[5] "Distributed Computing in Practice: The Condor Experience" Concurrency and Computation: Practice and Experience, Douglas Thain, Todd Tannenbaum, and Miron Livny, Vol. 17, No. 2-4, pages 323-356, February-April, 2005.

[6] Globus Toolkit Version 4: Software for Service-Oriented Systems. I. Foster. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2006.

[7] STAR detector overview, Published Mar. 1, 2003, Nucl. Instrum. Meth. A 499 (2003) 624 http://www.star.bnl.gov/

[8] An overview of Tevatron collider run II at Fermilab. Ronald S. Moore, FERMILAB-CONF-02-320-E, Dec 2002. 6pp. Tevatron: http://www-bdnew.fnal.gov/tevatron/

[9] DOE Office of Science Five-Year Plan, FY 2007-FY 2011 http://sc.doe.gov/Budget_and_Planning/Five-Year%20Plan/FYP%20Vol%20I%20-%20final%20version.pdf.

[10] NSF'S Cyberinfrastructure vision for 21st Century Discovery, Cyberinfrastructure Council, http://www.nsf.gov/od/oci/ci_v5.pdf; Report of Blue-Ribbon Advisory Panel on Cyberinfrastructure http://www.nsf.gov/publications/pub_summ.jsp?ods_key=cise051203.

[11] The Grid2003 Production Grid: Principles and Practice, Thirteenth IEEE International Symposium on High-Performance Distributed Computing (HPDC13), Honolulu, 2004

[12] A list of the software in the VDT is at http://vdt.cs.wisc.edu/

[13] TeraGrid: www.teragrid.org

[14] EGEE: http://www.eu-egee.org/

[15] CDF publications in 2008: http://www-cdf.fnal.gov/physics/preprints/2008.html. D0 publications http://www-d0.fnal.gov/www_buffer/pub/Run2_publications.html.

[16] B. Etherton, L. Breeger, "SHIPS Ensemble", AMS Tropical Conference, Orlando Florida, May 2008, available at http://personal.uncc.edu/betherto/papers.htm .

[17] Language Processing to Automate Information Synthesis: http://ils.unc.edu/~cablake/research.htm

[18] Prior publication: K. M. Benjamin, et al., "Virial coefficients of polarizable water: Applications to thermodynamic properties and molecular clustering", J. Phys. Chem. C 111 16021 (2007).

[19] Managed data storage and data access services for data grids. Jon Bakken et al. Computing in High-Energy Physics (CHEP '04), Interlaken, Switzerland, 27 Sep - 1 Oct 2004.