Strategic Accomplishments of the Open Science Grid 2006-2011
Open Science Grid Council

The objectives below are based on the original Charter of the Open Science Grid Consortium.[1]

Objective 1:

Build a persistent national grid infrastructure for large scale US science: the Open Science Grid.

Accomplishment:

The OSG based infrastructure has been operating continuously from November 2003 when it was instantiated for SC2003 as the "Grid3 infrastructure".

Lack:

Sustained funding for the infrastructure has not been attained.

Lessons Learned:

Important to have a mix of sponsorship and partnerships from a wide range of disciplines to help ensure sustainability. Additional strategies to engage researchers in campus-based contexts was not recognized earlier enough in the project.

Objective 2:

The Open Science Grid Consortium will provide a set of goals and an overall infrastructure within which the Grid resources of the different members can be operated coherently and compatibly.

Accomplishment:

OSG resources have been used by more than 15 different scientific and research groups during the period 2006-2011. There have been two full downtimes of the infrastructure during that time, due to certificate authority issues.

Lack:

The threshold to entry is regarded by small groups of researchers as too high.

Lessons Learned:

Spend more time understanding the different community needs and cultures and adapt developments to better match the needs. Would benefit from a standing effort (user support/troubleshooting team) to help small/new teams. Given the focus on stability and reliability of core baseline services, not enough attention was given to common, baseline end-user capabilties; those capabilities were developed within and tailored specifically for participating VOs.

Objective 3:

Computational and application scientists, working together to provide and support the set of facilities, services and infrastructure needed.

---

[1] http://osg-docdb.opensciencegrid.org/cgi-bin/RetrieveFile?docid=25&extension=pdf

Accomplishment:

OSG management and execution teams are all multi-disciplinary. The OSG Executive Team includes 3 engineers, 1 computer scientist, and 3 application scientists. The 5 areas coordinators are 3 computer scientists and 3 engineers.

Lack:

The balance of computer scientists' use of and contributions could be improved.
The balance of application scientists who are Area Team Leads could be improved.

Lessons Learned:

Within the OSG management team we learned how to leverage and coordinate with facility managers of the participating VOs who own the computing and stroage resources as well as the systems administrative effort which comprise the the OSG infrastructure.

Objective 4:

A structure of management and coordination bodies will oversee and coordinate the work of the Open Science Grid Consortium.

Accomplishment:

The OSG Consortium developed and agreed upon by-laws. The Consortium also agreed upon a project management plan and a cooperative agreement that have been adapted as needed.

Lack:

The co-engagement of the project Executive Team and Council members has not been as active as hoped. An Executive Board was envisioned to guide technical progress but proved ineffective and was abandoned.

Lessons Learned:

Objective 5:

The Open Science Grid will be open to all sciences that have a need for distributed large scale computing and data management, and can bring resources to be federated.

Accomplishment:

The OSG has been used by biology, mathematics, climate modelling, proteomics, structural biology, molecular dynamics, computer science, nuclear physics, astrophysics and high energy physics researchers.

Lack:

The contributions of non-physics communities and organizations was hoped to be broader and more active. There are a notable small group of contributors – Structural Biology Grid consortium and the University of Nebraska campus.

Lessons Learned:

As noted above, the importance of using the campus context to connect and coordinate with groups whose computing environments are evolving was not recognized until late in the project.

Objective 6:

An engineered production quality grid infrastructure will be built and operated in the US.

Accomplishment:

The usage of the infrastructure has steadily increased.
The Blueprint (principles and best practices) established in 2005 has been used.

Lack:

Much of the Blueprint is inconsistent and out of date. The updating has been uneven.

Lessons Learned:

We realized early on from failures of other large grid infrastructure projects that naiive top-down architectures for distributed grid was not a viable approach in the beginning, and that a baseline approach incorporating fundamentals distributed high throughput computing coupled to ideas of resource ownership and facility autonomy was the better approach. While this proved to be true, our approach could be improved by capturing and incrementally evolving a reference architecture and Bluprint vision roadmap so as to give Satellite development projects a better target environment.
Allocate standing staff effort to maintain the architecture and Blueprint throughout the project.
Perhaps Council contributions can help in an effort to keep the Blueprint up to date.

Objective 7:

Extended internationally through participation in the global grid infrastructure for science.

Accomplishment:

Not only is OSG is recognized as an important contributor to the Worldwide LHC Computing Grid, we are responsible for the driving the current model in which WLCG is the LHC experiments' organization connecting the experiments' computing organizations to the (multiple) underlying grid infrastructures. Addintionally OSG has federated infrastructure partners in Colombia and Brazil. Notably Colombia has created an independent national grid infrastructure (NGI) modelled after OSG's architecture and using OSG's software stack.

Lack:

Lessons Learned:

Good documentation and training materials, as well as in person visits and education might reduce effort overall. Given the lack of resources available to create specific training documentation, we developed a framework for re-use of production documentation as well as materials contributed by expert volunteer 'faculty' at OSG site administrator workshops.

Objective 8:

The grid will support managed access to large computing and storage resources (up to tens of thousands of CPUs and 100s of petabytes).

Accomplishment:

OSG today provides access to more than 50,000 job slots and over 30 petabytes of useable storage through a variety of bakend technologies and protocols. It remains a powerful organizational framework that connects thousands of users to these resources on a continuous basis.

Lack:

While the OSG provides access to tens of thousands of cores, 100 petabyte scale is yet to be reached in part due to variability in the computing models (data distribution and caching strategies) of the participating VOs. It is likely however to be reached in the next three to five years, given resource projections.

Lessons Learned:

The buy-in to sharing of resources is still constrained by the threshold of effort needed for buy-in. We need to pay more attention to showing the value of sharing and fast turnaround to initial results.

For small VOs to share they must support staff and infrastructure to answer and address problems from a diverse community with which they are not necessarily familiar. Funding such effort to improve the effectiveness of the whole could be considered.

Objective 9:

The Open Science Grid will create opportunities for educators and students to participate in building and exploiting this grid infrastructure and opportunities for developing and training a scientific and technical workforce.

Accomplishment:

OSG has held three summer schools where more than 75 students have been educated in distributed high throughput computing concepts and use. Additionally through site administator workshops more than 50 systems administrators have been trained in best practices for grid service deployment and operations, network tuning, troubleshooting and monitoring, and site security best practices. OSG sponsored two week-long workshops that assisted Colombia in launching their NGI, leveraging OSG-trained systems adminstators and other experts collaborating in the Consortium.

Lack:

A user-focused education program was slow to develop in part due to difficulties designing a program appropriately scoped for a consortium whose primary charter was to build a production scale grid infrastructure. As the sphere of expertise of OSG staff was on core middleware services rather than user interfaces and application integration (these were the responsibility of the VOs), the program remained challenged until Condor-specific technologies and “pilot/overlay” methods took root as a standard for large scale job submision.

Lessons Learned:

Education, training and outreach activities should in general be closely coupled to the core activities of a project; therefore as the educational mission included students and distributed computing and new science domains, more resources should be given to development of user interfaces and tools, as well as reusable, sustainable curricula and training materials. Having robust, common user frameworks with a quick application integration capabilities for new science domains is vital.

Objective 10:

The Open Science Grid Consortium will ensure that the U.S. plays a leading role in defining and operating the global grid infrastructure needed for large-scale collaborative and international scientific research.

Accomplishment:

OSG is recognized as an important contributor to the Worldwide LHC Computing Grid, supports international collaborations including LIGO, STAR, CDF, and D0.
OSG has federated infrastructure partners in Colombia and Brazil.

Lack:

Lessons Learned: