

# Earth Cube Technical Solution Paper – the Open Science Grid Example

Miron Livny<sup>1</sup>, Brooklin Gore<sup>1</sup> and Terry Millar<sup>2</sup>

<sup>1</sup>Morgridge Institute for Research, Center for High Throughput Computing, <sup>2</sup>Provost's Office  
University of Wisconsin-Madison, Madison, WI 53706

This is one of four Technical Solution papers that, along with a Design Approach paper, is a collection of papers submitted to Earth Cube by a large group of University of Wisconsin-Madison related researchers and educators at the University of Wisconsin-Madison spanning many colleges, centers, departments, and institutional partners including three Technical Solution papers from the Space Science and Engineering Center and a Design Approach paper.

The most important and perhaps difficult challenge that the Earth Cube initiative will face is governance across diverse disciplinary, social, and political cultures relevant to the success of its goals. Success in this endeavor will be difficult initially and must be built on trust, mutual understanding, and a shared vision that this can be a non-zero sum enterprise.<sup>1</sup> It is important that neither the wheel nor the flat tire be reinvented too often in this process. Although there will be new and unique challenges, social science research and lessons from previous experience will be paramount. There are valuable lessons that can be learned from other successful CI/Science collaborations, such as the Open Science Grid (OSG).

The mission/vision of the Open Science Grid (OSG) is to advance science through open distributed computing. The OSG is a multi-disciplinary partnership to federate local, regional, community and national cyberinfrastructures to meet the needs of research and academic communities at all scales. Funded jointly by the NSF and DOE, OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project – see <http://www.opensciencegrid.org/>. A map showing sites primarily in North America follows:



**Figure 1. Open Science Grid Sites in North America**

This technical solution paper presents aspects of the Open Science Grid (OSG). Although the OSG consortium involves high-energy physics rather than the geosciences, it faced many similar challenges and has had many

<sup>1</sup> See Nonzero: The Logic of Human Destiny by Robert Wright, Pantheon Books, 2000

similar structural features. Also, some aspects of the OSG CyberInfrastructure (CI) could be directly relevant to Earth Cube.

Funded jointly by the NSF and DOE, OSG is a consortium of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project – see <http://www.opensciencegrid.org/>.

The OSG came about primarily because of the CI needs of the high energy physics community at the Large Hadron Collider located at CERN in Geneva, Switzerland. The ‘community’ is actually a large collection of smaller communities that vary by disciplinary emphasis, country of origin, affiliated organization, etc. These various communities and the challenges they faced have a number of similarities to Earth Cube:

- The physics community was ready to take on the CI challenge;
- There was an existing infrastructure and knowledge base on which OSG was built, but there was much to do in order to build such an integrated framework;
- The relevant technologies have continued to evolve at a pace that has allowed for substantial convergence and integration of the systems that now constitute OSG;
- The community came together in a set of processes marked by distinct events and face-to-face and online dialog about all or parts of the required CI;
- The challenge was not easy or rapidly addressed. However, there has been a convergence to a common framework over time.

In addition, the success of OSG depended on:

- Physicists and engineers who had visionary knowledge of the necessary fields and research;
- Users who had a strong grasp of both the high energy physics community’s scientific and CI needs;
- Cutting edge CI architects, builders and other technologists;
- Experts in knowledge management/information systems who contributed to discussions related to turning user and data requirements into CI functionality;
- Individuals who had experience with the governance of multi-user infrastructure and had engaged the community in creating, building, maintaining, and modifying facilities; and
- Postdocs and graduate Students who had training in high-energy physics, engineering, and/or computer sciences and related fields with an interest and ability to participate in discussions related to high energy physics CI.

Thus we believe that the Earth Cube initiative can take some lessons from the OSG experience.

### **User Requirements**

Developing communities of shared resources required a framework of mutual trust, whereas maximizing throughput required dependable access to as much processing and storage capacity as possible. The inherent stress between these requirements underpins the challenges that the Distributed High Throughput Community (DHTC) community faced in developing frameworks and tools that translated the potential of distributed computing into high throughput capabilities. The OSG addressed these challenges by following a framework that is based on four underlying principles that will be very relevant to Earth Cube:

- **Resource Diversity:** Maximizing throughput required flexibility to accept many types of resources and the integration of multiple layers of software and services;
- **Dependability:** Throughput had to be tolerant to faults since the scale and distributed nature of high throughput environments meant some service or resource would always be unavailable;
- **Autonomy:** Enabled users and resource providers from different domains and organizations to pool and share resources while preserving their local autonomy to set policies and select technologies;
- **Mutual Trust:** The formulation and delivery of a common goal through sharing required a web of trust relationships that crosses the boundaries of organizations as well as software tools.

Guided by these principles, the OSG advanced the state of the art of DHTC technologies as the consortium implemented the concepts and integrated, deployed and operated the software tools from many projects at demanding scales and operational standards. Together with the user communities, the leadership team strove to develop methodologies that improved the cost effectiveness of the national CI and thus serve as a catalyst and partner in creating and evolving novel software technologies. Noteworthy examples are location insensitive access to very large data, overlay resource managers, and more facile single sign-on systems. OSG has also driven innovation in methods and software that provide “health status” and “catalogs of available services” for a national scale CI. Working within a framework of high-level principles amplified the impact of its work as it promoted sharing of ideas, experiences and tools across the DHTC community and facilitated the development of education materials.

### **Partnership and Evolution**

Over the first 5 years of the OSG, the involved communities found that bringing High Throughput Computing (HTC) capabilities to new communities was most effective and sustainable via campus and regional affiliations. The original model for campus based HTC preceding OSG was the Grid Laboratory of Wisconsin, followed by FermiGrid and NYSGrid. This step-by-step evolution showed that the shared HTC capabilities that are part of a national CI, can be successfully implemented at universities, national laboratories, and even at the state or regional level.

The Consortium is the overarching organizational framework for the OSG partnership and includes all contributing organizations. A Council is the governing body. The program of work of the Consortium has been managed and executed by an Executive Team (ET) and consists of a core Project, independent (collaborative) satellite projects, and the contributions of consortium members. The core OSG Project provides services needed by the Consortium to meet its mission. Satellite projects are independent projects that contribute to the OSG and where OSG was involved in the planning process and committed support for collaboration. The OSG provides an intellectual anchor for satellite projects as well as a laboratory for deployment and hardening of new technologies. The strength of the organization is in the diverse, engaged teams formed by project contributors and staff working on challenging common goals in the context of a shared framework of high-level principles.

The management of the OSG is distributed – form follows function. The ET leads the partnership, manages the program of work, and sets priorities. The responsibilities are distributed across an Executive Director at Fermi Lab, the PI and Technical Director at UW-Madison, Application Coordinators who provide a direct interface to the U.S. ATLAS and U.S. CMS communities respectively (both at LHC), an Executive Associate Director, and a Project Manager. The work of the management team is leveraged across all the constituencies, satellites and partnerships of the OSG “planetary system”. The core expertise and experience as a long time collaborating group of this distributed management team are a crucial component of the past, current and future success of the OSG.

The members of the OSG consortium are united in a commitment to promote the adoption and to advance the state of the art of DHTC – shared utilization of autonomous resources where all the elements are optimized for maximizing computational throughput. The U.S. LHC scientific program embraces OSG as a major strategic partner in developing, deploying and operating their novel and cost effective DHTC infrastructure. As daunting as this has been, the Earth Cube challenges are probably more substantial and of course different. However, this is the type of model that the Earth Cube initiative should consider and investigate as the involved communities do the initial work to form an integrated community to develop, test, implement, and maintain a fabric of cyber tools to advance the research and education mission of the earth sciences as it gives life to the NSF Earth Cube initial vision.

### **CI architecture design, development and integration**

We believe that a key technical challenge for Earth Cube’s Cyberinfrastructure is not necessarily in the data per se, but in the discovery of that data and in providing scalable services that present a common access methodology to the data. By focusing on a common data access service, we can simplify the application development process of integrating multiple, interesting datasets whose value combined is greater than the individual data elements alone. A key goal is to make it easy for Earth Science domain scientists and application developers to create applications that are rich with Earth Science data.

A similar approach to high throughput computing (HTC), used by OSG, has resonated not only with the science community in academia, but also the private and commercial sector. By providing a unified view of computing resources, it simplifies the task of scaling computing from the desktop to local resources, national resources and cloud resources.

A key tenant of OSG’s approach to HTC is matchmaking. Jobs with given requirements are matched to resources that meet those requirements. And OSG does so with the core belief that resources will be unreliable and that jobs will need to be restarted on other resources during their execution lifecycle. This matchmaking paradigm could apply aptly to the Earth Cube. In Earth Cube’s case we would be matching the needs of applications to the capabilities of Earth Science data. In doing so we could provide a common view of and interface to that data. This is similar to the way HTC provides a unified view to computing resources, regardless of location or type (e.g. Linux, Mac, Windows).

### **OSG Organization and Model of Operation**

The OSG Consortium builds and operates the OSG project. Consortium members contribute effort and resources to the common infrastructure, with the goal of giving scientists from many fields access to shared resources worldwide. See org chart for OSG Consortium (being revised) and the project.

The OSG model of operation is that of a distributed facility which provides access to computing and storage resources at various sites in the US and abroad.

- Resource owners register their resource with the OSG.
- Scientific researchers gain access to these resources by registering with one or more Virtual Organizations (VOs).
- The VO administrators Register their VOs with the OSG.

- All members of the VO who have signed the acceptable use policy (AUP) are allowed to access OSG resources, subject to the policies of the resource owners.
- Each resource and each VO is supported by a designated, and in some cases shared, "Support Center (SC)," determined at registration time.

There is a collaborative wiki for OSG management activities. The Consortium Council governs the consortium. The OSG Consortium Governance Procedures and By-laws explain how the OSG Consortium works. The Executive Team manages the project. Within the OSG, work is organized into Technical Activities, often with joint projects between the OSG project and members of the consortium. Access to more detailed information is available here.

Each OSG Consortium member and partner organization sends a representative to the OSG Council. The OSG Council governs the OSG Consortium, ensuring that the OSG benefits the scientific mission of its stakeholders. The Executive Director and Executive Board direct the OSG program of work, write policy and represent the OSG Consortium in relations with other organizations and committees. Figure 2 is a diagram of the OSG organizational structure.

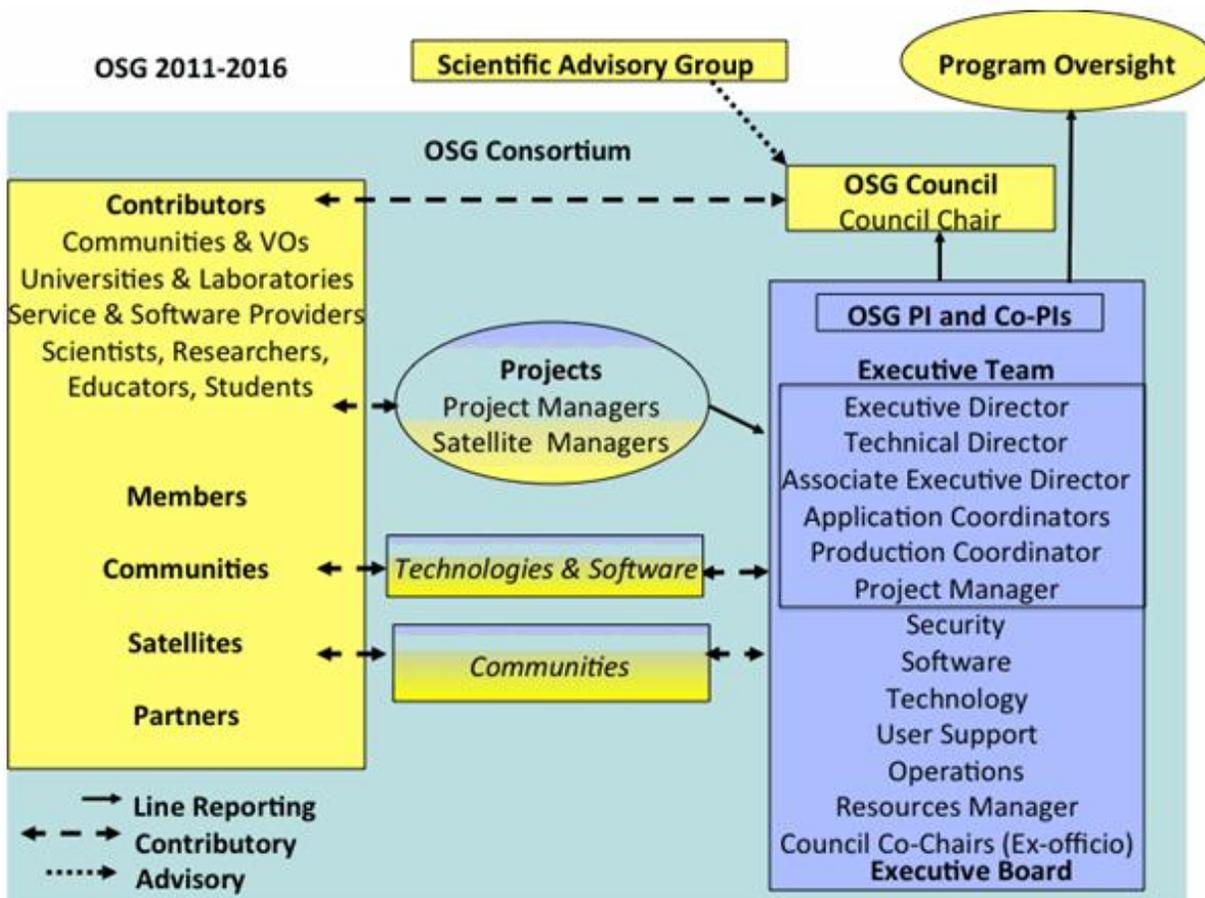


Figure 2. Open Science Grid Organizational Structure

### Future plans

The OSG experiences have many features that could be relevant to the Earth Cube Initiative, and there are key OSG leaders who are prepared to help make these experiences available and relevant to that initiative.