

1. NAME OF INITIATIVE: The Open Science Grid

List of major collaborating institutions (including non-US partners).

The Open Science Grid Consortium is a roadmap to build an open national computing production grid infrastructure for science in the U.S.. Fermilab is making significant contributions to the program of and leadership in the Consortium. Participants in OSG come from the DOE and NSF funded laboratories and universities. The Consortium will build up a broad base of participation – currently from High Energy and Nuclear Physics, participants in iVDGL Grid3, PPDG, and Computer Science and Information Technology groups.

2. SCIENTIFIC JUSTIFICATION:

Physics goals. How does it fit into the global physics goals for the entire field.

Over the next few years Fermilab will make an increasing fraction of its resources available on the Open Science Grid. US CMS, Tier-1 and Tier-2 centers and the software and computing program, are contributing to all aspects of the program and are aligning their developments, deployments and milestones to contribute to the Open Science Grid. This is being done in close collaboration with similar contributions from US ATLAS. The Tevatron Experiments are already contributing their remote computing model expertise and infrastructure to the Consortium Technical Groups and Activities. CDF and D0 will adiabatically adapt their compute facilities to participate in and be accessible from the common OSG infrastructure. An important aspect of Fermilab's contributions to the OSG are working with the LHC Computing Grid projects to ensure a usable, supportable sustained infrastructure to support the distributed collaborations across Europe and the U.S.

Other experiment collaborations within the Fermilab sphere of support, e.g. BTeV, SDSS, DES, are also initial participants in the OSG. Fermilab's Education and Public Affairs offices are also participants, aligning their activities in the area of Computing with the broader vision of the Open Science Grid. Additionally, The existing multi-community Physics Grid Projects – PPDG, GriPhyN, iVDGL, SRM – are aligning their programs of work to contribute to the Open Science Grid roadmap. The Open Science Grid is a strategic direction for the Fermilab Computing Division. Through its participation Fermilab is active in the wider facilitation and communication activities – and aims to effect a broad base of participation as well as timely and successful building and deployment of the infrastructure. The ultimate goal of the OSG is to develop an effective and broadly available common grid infrastructure through a self-organized collaborative framework. The goal is to ensure a low cost of entry to contribute resources and adapt applications to run on this grid.

The goals require a significant level of investment and attention over the next five to ten years. An open national production grid infrastructure for science will require widespread adoption of this distributed computing paradigm. The technologies and deployments of the “disruptive technology” Grid paradigm (capability, performance and robustness) are transitioning from early adopter and research activities to consolidation and market adoption – which requires significant investment and engagement to reach the final phase of ubiquitous use and availability.

The OSG infrastructure will be built and deployed through a set of Activities, each of which involve some or all of the participants in the Consortium.

In particular a series of Activities OSG-0, OSG-1 etc will build iterative releases of the infrastructure which will be usable by and supported for a set of applications defined within the Activity. The Joint Grid2003 project has already demonstrated the feasibility and effectiveness of the approach, and the resulting operating infrastructure, Grid3, will evolve into the Open Science Grid. The US LHC and Run 2 data simulation, processing and analysis applications will provide significant use of the infrastructure and proof of the approach.

3. VALIDATIONS FOR SCIENTIFIC JUSTIFICATION:

Examples of recommendations and supporting statements from the committees, panels, and the community at large.

Not applicable at this time.

4. DESIRED SCHEDULE:

List major milestones (month & year) such as design complete, construction start, construction complete, etc.

Open Science Grid-0 will be launched in Spring 2005. Participants in this milestone include the major stakeholders above. The scale of OSG-0 is expected to be several thousand CPUs, 100TBs managed disk store, a mix of 10-20 applications. The performance metrics for OSG-0 will advance from those of Grid2003 by providing a robust framework for strategic and tactical storage, managed transfer of data between sites, a heterogeneous middleware base, dynamic participation in the infrastructure, and an evolution in the security model.

Subsequent milestones for OSG will increase the resources available on the Grid to several tens of thousands of CPUs and petabytes of managed disk storage over the next 3 years, with a specific readiness goal of support for the simultaneous processing of LHC and Run 2 data in 2007. Our HEP community has a further goal of enabling a broad community collaborative model for science (as described in the DAWN ITR proposal for example) to provide open, easy to use, managed resources for the individual and small group of physics researchers in the US LHC era and beyond.

While a large fraction of the effort and resources necessary to build the Open Science Grid will come from the realignment and contributions of existing facilities, experiments and projects, the broader vision and goals of the Consortium need additional focused work to achieve. In particular, Fermilab's contributions to the system engineering and deployment, facilitation and coordination, communication and dissemination, outreach and community involvement,

negligible. The broadly stated needs for the Consortium are stated in the white paper from August 2003. We encourage participation and contributions from nearby and far universities.

Projected growth of OSG throughout the decade:

Milestone + Date	#CPUs	# Sites	Managed Storage	Application Groups
Grid2003 1/04	2500	25	Not part of common infrastructure	10
OSG-0 2/05	4000	30	MSS: 100 TB Disk: 200 TB	15
OSG-1 2006	10,000	40	MSS: 1PB; Disk: 1PB	20
OSG 2007	20,000	80	MSS: 2 PB; Disk: 2 PB	25
OSG 2008	30,000	100	MSS: 4PB; Disk: 4 PB	30

5. ROUGH ESTIMATE OF COST RANGES:

Whatever the best information available (eg. \$M +/-30~50%, \$150~250M, etc.). Total cost range including non-DOE funding (if any other funding sources are assumed and if known, state from where and how much. Also indicate remaining R&D cost to go.

In order to migrate the experiment-specific grid services to the Open Science Grid some core servers and services owned and operated by the federation, rather than simply contributed by one of the partners, will have to be put in place and operated as a robust round the clock service (**\$2M/year**).

The initial phase work of federating the LHC resources and migrating both CMS and ATLAS applications to a common U.S. Grid is estimated to require approximately **\$4 M in the first year** in development costs and subsequently between **\$6M and \$7.5M per year** in ongoing integration and operations costs.

In parallel, opportunities to consolidate services and integrate non-LHC experiments and resources should be studied and demonstrated (**\$750K/year**). The additional funding necessary to fully migrate a particular experiment or application area in phases two and beyond will then be better understood and detailed proposals for this will have to be developed.

6. DESIRED NEAR TERM R&D:

Major activities needed to be completed before start construction.

Not applicable.

7. BRIEF DESCRIPTION OF LABORATORY’S ANTICIPATED ROLE:

Expected unique capabilities to be provided by lab. Rough estimate of human resources from lab (#FTE in what type labor).

To assure success there must be a constant effort of approximately 15 FTEs applied to the following activities: Management (4 FTEs), Engineering (4 FTEs), International Coordination (4 FTEs) and Education and Outreach (3 FTEs). **(\$2M/year total).**