Open Science Grid

Ruth Pordes,
Fermilab

Open Science Grid

Goals.

Services & Principals.

Organization & Activities,

Intertwining with the EGEE and LCG.

In the Short Term.

http://opensciencegrid.org/documents/index.html

The Open Science Grid Consortium

Will Build a National Production Grid Infrastructure in the United States, Open to any Science. The Consortium will build on and

Evolve Existing Infrastructures: Grid3+, shared laboratory & experiment university facilities, RunII SamGrid, Campus Grids, all in close collaboration with Grid technology providers.

US LHC has a key role in contributing to & leadership in the roadmap: presenting resources to the common Grid infrastructure; developing distributed data and processing services to a common environment; running applications on the shared infrastructure; education and outreach programs.

The Open Science Grid Vision

To take a significant step forward in cooperative development and Grid operation across Virtual Organizations on the national and international scale.

To build a flexible real-time framework of services and Grid infrastructure using an engineered end-to-end approach.

To accommodate multiple Grid systems and technologies while ensuring coherent operation.

To support science working environments for individuals and diverse analysis teams, as well as large scale production Grid operations for 100s - 1000s of tasks with a wide range of demands for computing, data and network resources.

The Open Science Grid Consortium Participants

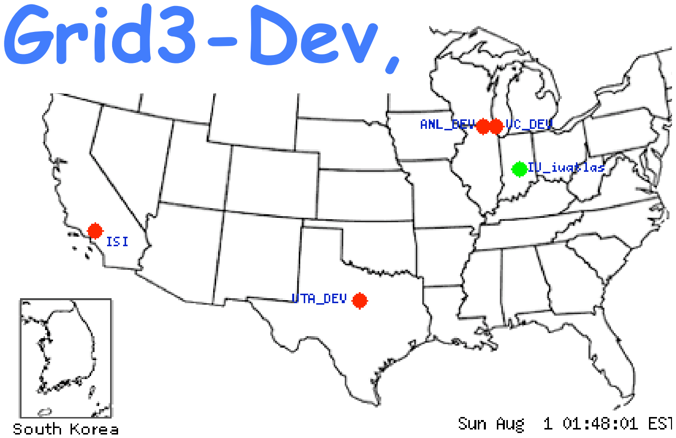
Include a broad base across many facilities, experiments, and grid projects: Fermilab, BNL, SLAC, LBNL, JLAB, Run II, BaBar, BNL, STAR, PHENIX, LIGO, SDSS, BTeV, PPDG, iVDGL, GriPhyN; as well as DOE & NSF sponsors.

Include Grid technology providers: Condor, Globus, NMI, VDT, Virginia Tech…

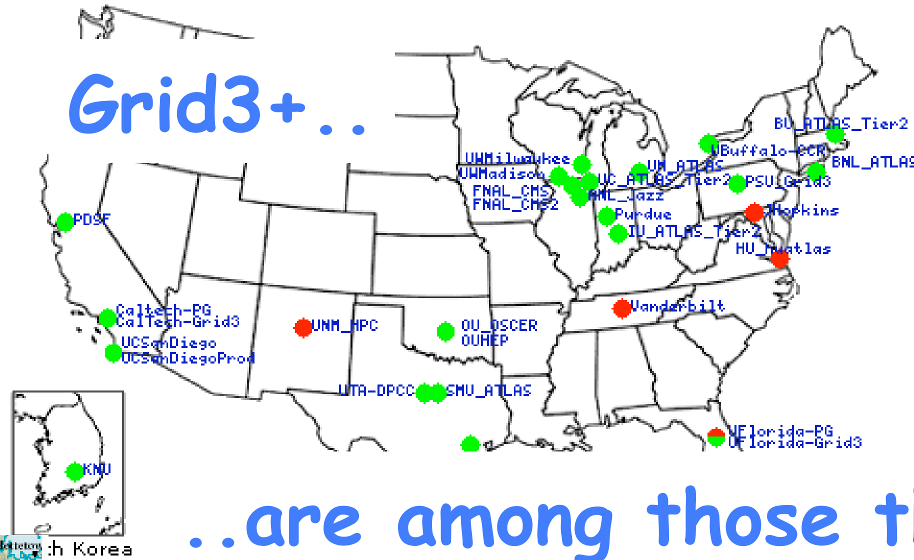
Organized as a Collaborative Partnership with community goals & contributions, deliverables and self-organized governance - like an Experiment.

Grid3 -> OSG, Spring '05

Grid3-Dev,

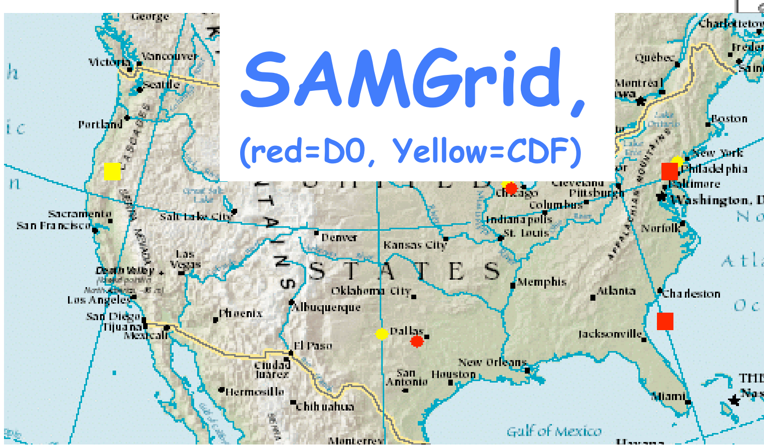


Grid3+..

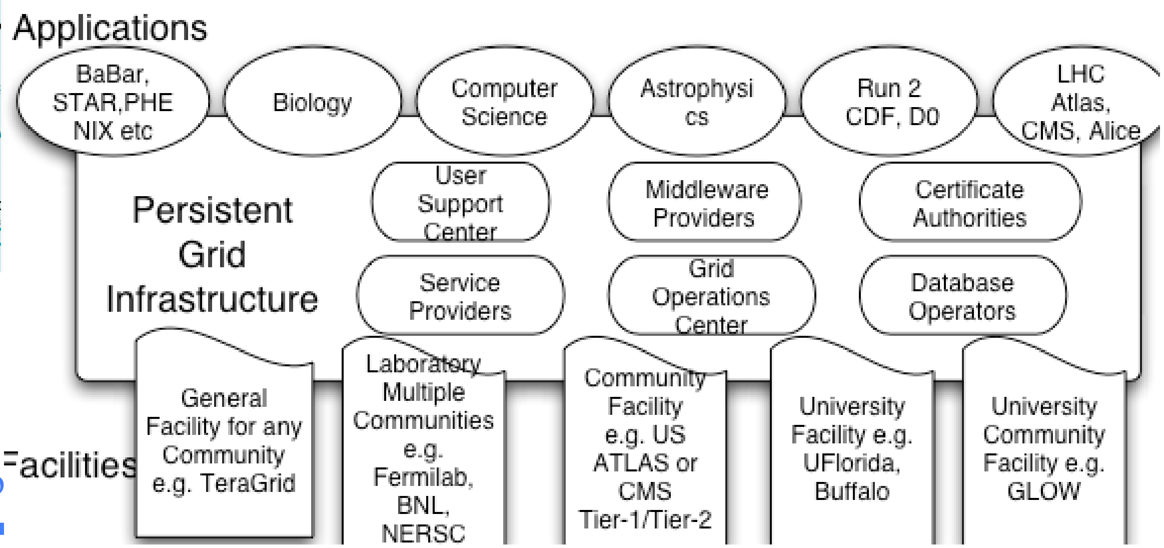


SAMGrid,

(red=D0, Yellow=CDF)



..are among those that will contribute to OSG…



Open Science Grid

Consortium Participants

Adams, David	Brookhaven National Laboratory	Finn, Lee Samuel	Penn State	Kim, Sang	National Science Foundation	Price, Larry	Argonne National Laboratory
Avery, Paul	University of Florida	Fisk, Ian	Fermilab	Knosp, Boyd	University of Iowa	Quick, Rob	Indiana University-Perdue
Aydt, Ruth	National Center for Supercomputing Applications (NCSA)	Foster, Ian	University of Chicago	Koranda, Scott	University of Wisconsin, Madison	Rana, Abhishek	University of California San Diego
Baker, Rich	Brookhaven National Laboratory	Fu, Yu	University of Florida	Kowalski, Andy	Jefferson Lab	Ranka, Sanjay	University of Florida
Bakken, Jon	Fermilab	Gaines, Irwin	Fermilab	Kramer, Bill	Lawrence Berkeley National Laboratory	Ravot, Sylvain	CERN
Ballintijn, Maarten	MIT	Gardner, Rob	University of Chicago	Lauret, Jerome	Brookhaven National Laboratory	Reed, Dan	University of North Carolina
Bardeen, Marge	Fermilab	Gibbard, Bruce	Brookhaven National Laboratory	Lazzarini, Albert	California Institute of Technology	Rodriguez, Jorge	University of Florida
Bauerdick, Lothar	Fermilab	Gieraltowski, Jerry	Argonne National Laboratory	Legrand, Iosif	CERN	Roland, Gunter	CERN
Berman, Fran	San Diego Supercomputing Center	Giles, Roscoe	Boston University	Livny, Miron	University of Wisconsin, Madison	Roy, Alain	University of Wisconsin, Madison
Boehnlein, Amber	Fermilab	Goldberg, Marv	National Science Foundation	Lorch, Markus	Virginia Tech	Rubi, Ernesto	Florida International University
Bourilkov, Dimitri	University of Florida	Gonzalez, Saul	DOE	Luehring, Fred	Indiana University	Schlagel, Thomas	Brookhaven National Laboratory
Bradley, Scott	Brookhaven National Laboratory	Goodman, Alyssa	Harvard University	Malon, David	Argonne National Laboratory	Schmidt, Jack	Fermilab
Branson, Jim	University of California San Diego	Gordon, Howard	Brookhaven National Laboratory	Mambretti, Joe	Northwestern University	Scott, MaryAnne	DOE
Bresnahan, Glenn	Boston University	Green, Dan	Fermilab	Markowitz, Pete	Florida International University	Seidel, Ed	Louisiana State University
Bunn, Julian	California Institute of Technology	Green, Mark	University at Buffalo	Melani, Matteo	SLAC	Shank, Jim	Boston University
Butler, Joel	Fermilab	Grundhoefer, Leigh	Indiana University	Merrit, Wyatt	Fermilab	Sheldon, Paul	Vanderbilt University
Cao, Junwei	MIT	Heavey, Anne	Fermilab	Moore, Reagan	San Diego Supercomputing Center	Shoshani, Arie	Lawrence Berkeley National Laboratory
Catlett, Charlie	Argonne National Laboratory	Helm, Mike	Energy Sciences Network	Mount, Richard	Stanford Linear Accelerator Center	Skow, Dane	Fermilab
Cavanaugh, Richard	University of Florida	Hinchcliffe, Ian	Lawrence Berkeley National Laboratory	Nefedova, Veronika	Argonne National Laboratory	Smith, Wesley	University of Wisconsin, Madison
Cook, Jeremy	University of Bergen in Norway	Hoeheim, Chuck	SLAC	Neilson, Ian	CERN	Snider, Rick	Fermilab
Cousins, Bob	UCLA	Humphrey, Mary	University of Virginia	Newman, Harvey	California Institute of Technology	St Denis, Rick	University of Glasgow
Cowles, Bob	SLAC	Huth, John	Harvard University	Odyniec, Grazyna	Lawrence Berkeley National Laboratory	Steenberg, Conrad	California Institute of Technology
Crawford, Glen	DOE	Ibarra, Julio	Florida International University	O'Fallon, John	DOE	Stevens, Rick	Argonne National Laboratory
Cvernak, Ann	Information Sciences Institute	Jacak, Barbara	Stony Brook University	Olson, Doug	Lawrence Berkeley National Laboratory	Stickland, David	CERN
Dasu, Srihara	University of Wisconsin, Madison	Jackson, Judy	Fermilab	Onel, Yasar	University of Iowa	Szalay, Alex	Johns Hopkins University
Demchenko, Yuri	University of Amsterdam	Jackson, Keith	Lawrence Berkeley National Laboratory	Pearson, Doug	Indiana University	Thies, Rick	Fermilab
Dinda, Peter	Northwestern University	Johnson, Eric	Florida International University	Pennington, Rob	National Center for Supercomputing Applications	Thomas, Michael	California Institute of Technology
Dobson, James	Dartmouth College	Johnston, Bill	Lawrence Berkeley National Laboratory	Perelmutov, Timur	Fermilab	Thomas, Tim	University of New Mexico
Ernst-Eschgarth, Michael	DESY	Kaletka, Mark	Fermilab	Petravick, Don	Fermilab	Thompson, Kevin	National Science Foundation
Evard, Remy	Argonne National Laboratory	Kelsey, David	Rutherford Appleton Laboratory	Philpott, Sandy	Jefferson Lab	Throwe, Thomas	Brookhaven National Laboratory
Welch, Von	National Center for Supercomputing Applications (NCSA)	Kennedy, Robert	Fermilab	Pinsky, Larry	University of Houston	Traumer, Mary	Georgia Institute of Technology
White, Vicky	Fermilab	Willis, Bill	Columbia University	Youssef, Saul	Boston University	Tuts, Michael	Columbia University
Whitmore, Jim	National Science Foundation	Womersley, John	Fermilab	Yu, Dantong	Brookhaven National Laboratory	Van Lingen, Frank	CERN
Wilde, Mike	University of Chicago	Wuerthwein, Frank	University of California San Diego	Yu, Jae	University of Texas Arlington	Walters, Adam	Fermilab

Technology Roadmap & Principles.

Open Science Grid

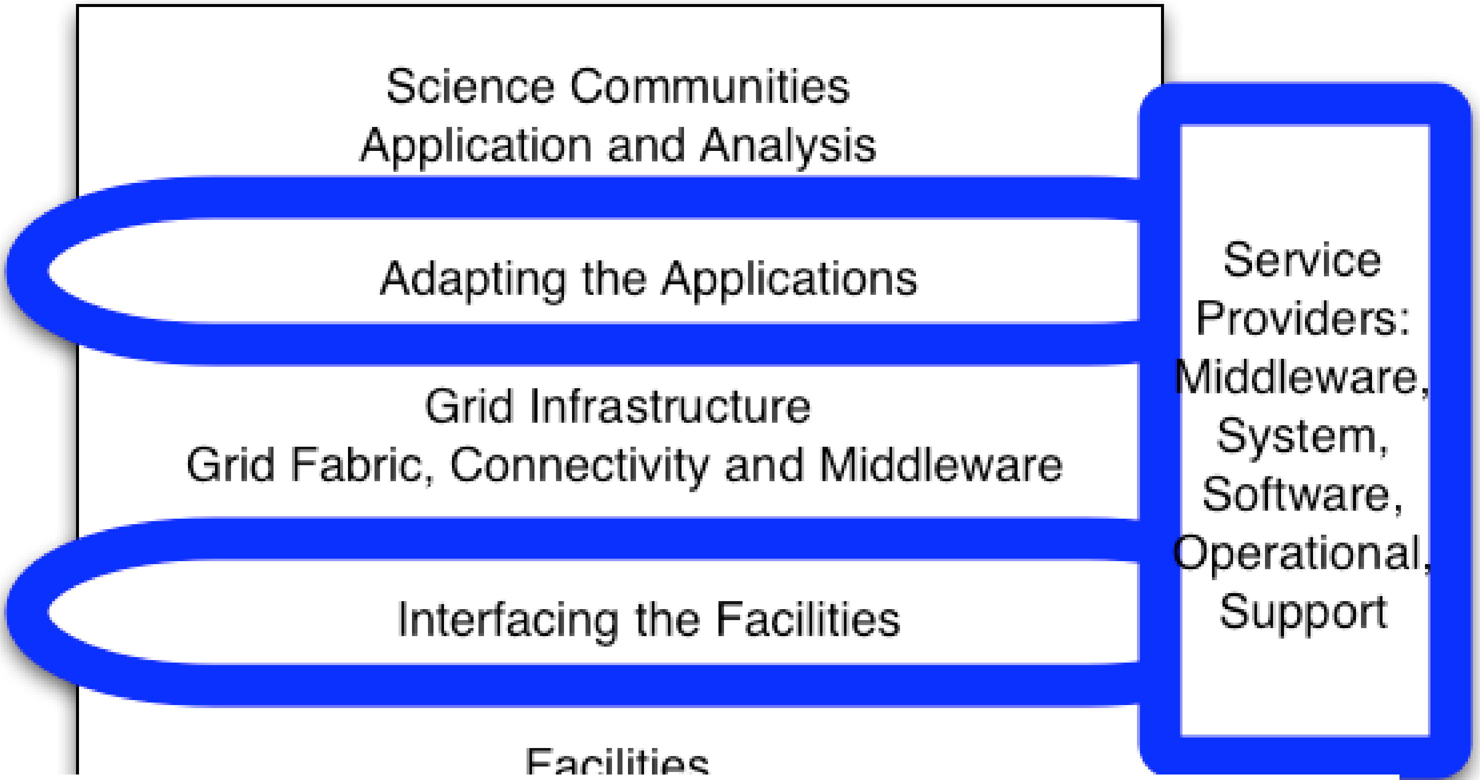
OSG Technical Approach

Build on experience with existing and emerging distributed system technologies, and prototypes; not already present in simple Grids:

Use grid-wide view of resources (site facilities and networks) to ensure consistent use and support policies. Provide a grid-wide real-time monitoring and tracking system to ensure good throughput of jobs & performance.

Design to an architecture that deals with complexity and global scalability with moderate effort.

Applications, Infrastructure, Facilities



Interfacing & Adapting to a Common Infrastucture &

Providing Production Quality Common Services

Initial OSG Services & Infrastructure

Compute Elements: Condor-G, GRAM, evolving to Condor-C, SAMGrid; interfacing to N>5 Batch systems

Storage Elements: GridFTP + SRM with multiple implementations.

Grid Middleware: evolve Virtual Data Toolkit.

Grid Security Infrastructure: X509; VOMS based administration, including implementation of AuthZ Callouts; + Site & VO specific.

Monitoring, Information and Accounting: Ganglia, MDS, MonaLisa, MDViewer,ACDC, GridCAT, + Site & VO specific.

Workflow Management: Chimera, DAGMAN, RunJob + VO specific

Data Management - Grid-wide Replica Location Service (RLS/RLI), SAM, + VO specific.

Grid Operational Support Infrastructure: evolve iGOC, Discovery Service, Very simple Policy infrastructure, + Site & VO specific

Applications & Partners

Applications known to date:

US LHC simulation, processing and analysis. Data Challenge + ongoing program.

Run II, RHIC, BaBar Applications - piloting use of the shared infrastructure for simulation and analysis

Cpu bound, non data-intensive, applications e.g. BTeV simulation, GADU, BLAST, MRI simulations.

Computer Science application demonstrators e.g. Virtual Data, RLS stress tests etc.

Partner Grids to date:

Campus Grids e.g. ACDC, GLOW, Florida, Dartmouth.

TeraGrid - understanding details for interfacing and integration.

The Open Science Grid

Open Science Grid is the Grid under the governance of the Open Science Grid Consortium operated as a sustained and production infrastructure for the benefit of the Users.

Other grids may operate under the governance of the OSG Consortium, for example the grid that validates the infrastructure before it becomes the OSG.

OSG infrastructure interfaces to facility, campus, and community grids that participate in the Consortium;

OSG federates with grids external to the Consortium through partnerships & agreements.

Open Science Grid VO is open to those Users and VOs that have contracts with the OSG.

Driving Principles for OSG

The infrastructure - both conceptually and in practice – should be as simple and flexible as possible,

to build from the bottom up a coherent, performant system which can accommodate the widest practical range of users of current Grid technologies,

and provides an infrastructure that maximizes the potential for eventual commonality in technology choice.

Principles, Best Practices and Service Decomposition captured in a Blueprint

Examples of Principles in the Blueprint

OSG will provide baseline services and a reference implementation. Use of other services will be allowed.

All services should support the ability to function and operate in the local environment when disconnected from the OSG environment.

Users are not required to interact directly with resource providers.

The requirements for participating in the OSG infrastructure should promote inclusive participation both horizontally (across a wide variety of scientific disciplines) and vertically (from small organizations like high schools to large ones like National Laboratories).

The OSG architecture will follow the principles of symmetry and recursion

The OSG infrastructure must always include a phased deployment, with the phase in production having a clear operations model adequate to the provision of production-quality service.

Examples of Best Practices in the Blueprint

The OSG architecture is VO based. Most services are instantiated within the context of a VO.

The infrastructure will support multiple versions of services and environments, and also support incremental upgrades

The OSG infrastructure should have minimal impact on a Site. Services that must run with superuser privileges will be minimized

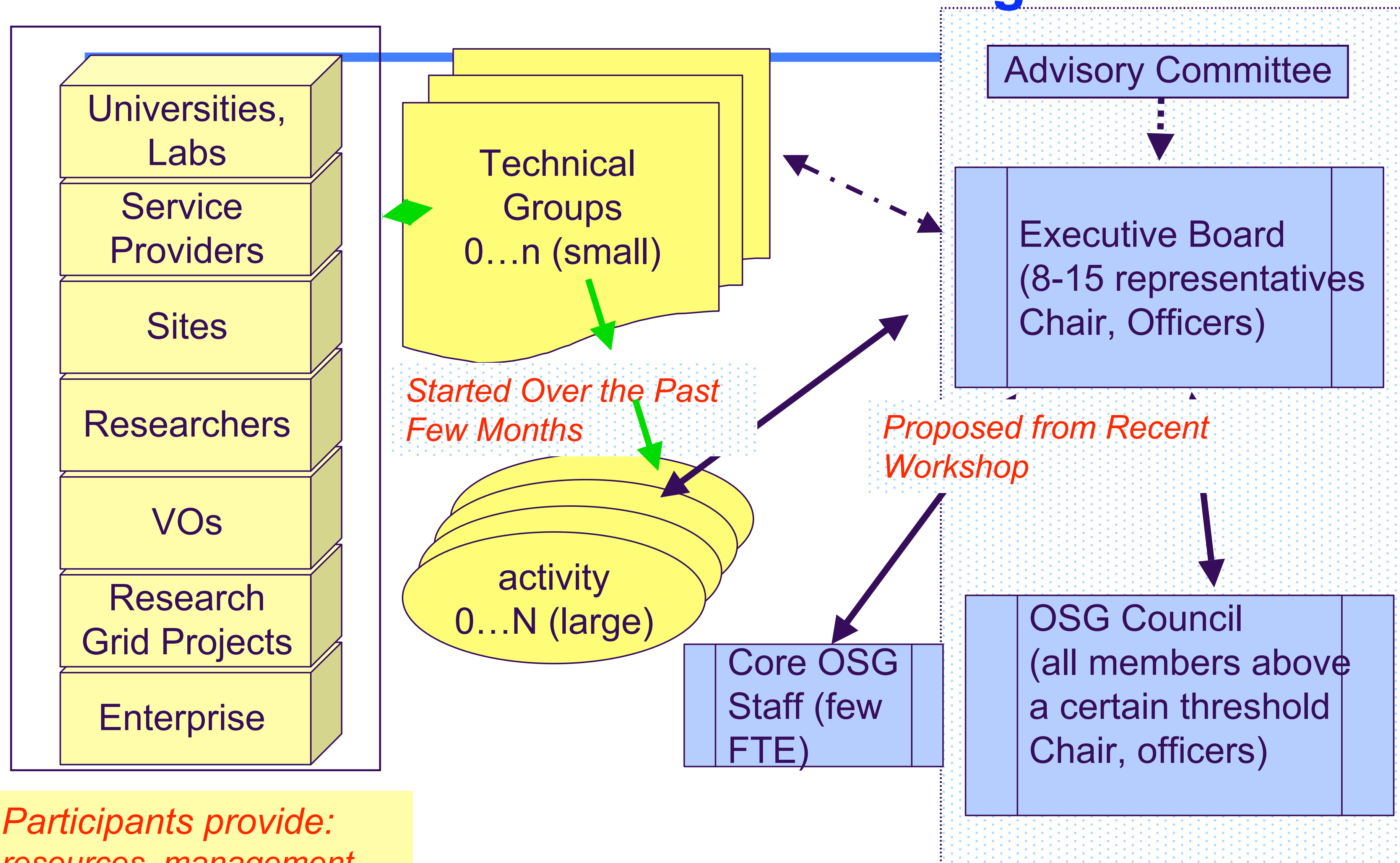
The OSG infrastructure will support development and execution of applications in a local context, without an active connection to the distributed services.

Services manage state and ensure their state is accurate and consistent.

OSG Blueprint is written in conjunction with and using as a basis the gLITE Architecture/Design documents- comparing & contrasting the details of the approaches, services and interfaces.

Organization & Activities

Details of Governance In Progress



Participants provide: resources, management, project steering groups

Open Science Grid

OSG Development Guidelines

Use and evolve existing Grid components and operational procedures. Minimize impact of the OSG framework, except as needed to ensure consistent operation. Agree among all partners on the wrapping and/or re-engineering of components as needed.

Involve all partners in the design, development and deployment, while meeting the needs of individual VOs. Our goal is to provide "common" flexible and easy to deploy tools that a VO can use to support their activities. OSG software may be (preferably)provided by member VOs, or a consortium of VOs; as agreed by the partners.

Provide Technical Coordination (only) to the degree needed for consistent end-to-end and inter-VO operation, e.g. OSG Technical Coordinator reports to the Executive Group; Some global services and tools, and top-down design may be provided by small expert team, as needed

Technical Groups and Activities

Technical Groups address and coordinate a technical area.

- propose, endorse, form and oversee activities related to their given areas.
- liaise and collaborate with their peer organizations in the U.S. and world-wide,
- participate in relevant standards organizations.
- chairs participate in the Blueprint, Grid Integration and Deployment.

Activities are well-defined, scoped set of tasks contributing to the OSG.

- has deliverables and a plan.
- self-organized and operated.
- overseen & sponsored by one or more technical groups

Technical Groups & Activities - I

Security

- Representatives from DOE Laboratories, US LHC Regional Centers, Grid Operations Centers, TeraGrid, ESNET, University Partners (e.g. Dartmouth).
- Propose Security Model and Oversee Security Infrastructure
- Write Registration Policies (evolution of Grid3), Acceptable Use Policies etc.
- Collaborate with peers in other Grid organizations:
- Members of the EGEE JRA3 and LCG Joint Security Group

Incident Response Activity

- Write & Execute a Security Incident Response Plan for iVDGL and OSG.
- Interfacing also with LCG/EGEE & TeraGrid.

Technical Groups & Activities - II

Storage

- Coordinate activities related to data storage & management.
- Identify requirements and technology gaps.

Storage Service Challenges across OSG and in partnership with LCG

- Compatability across SRM implementations at BNL, Fermilab, LBNL, JLAB, UWisconsin.
- Global Replica Management Service tests.
- STAR sustained data transfer has been in place for ~2 years between BNL and LBNL.
- BNL/Fermilab, US ATLAS/US CMS now participating in sustained, robust data movement challenge with LCG.
- Evolve SAMGrid

Technical Groups & Activities - III

Monitoring and Information Services

- Extension of Grid3 Monitoring Group..
- Includes Accounting which will evolve from Grid3 accounting.
- Includes some testing/validation and publishing services. Test Harness or Grid Exerciser ?
- Execution Environment requirement and framework?

Policy

- Includes Authorization & Access Policies of Organizations (Sites, VOs) and Services (Resource Providers)

Open Science Grid

Support Centers (Infrastructure, Technology, Operations)

Representatives from iGOC, BNL, Fermilab, LBNL, SLAC, VDT

Define Operational Model.

- Cover "daily operations" "weekly operations" "monthly evolutions and changes" "yearly transitions".
- Ensure Architecture & Services Provide for Operational Needs (Scalability, robustness, management & administrative interfaces, maintainability etc.)

Communicate and Publish Information

- Coordinate helpdesk and trouble ticket infrastructures.

Participate in LCG & EGEE Operations Infrastructures.

Collaborate with TeraGrid & NMI Grid Center Grid Operation Infrastructures.

Physics Participants
Special Relationship to the LCG & EGEE

- Technical Groups interact and collaborate with peers in LCG & EGEE as much as possible.
- Blueprint ongoing attention to gLITE/ARDA architecture & design.
- Infrastructure attention to common interfaces between OSG & LCG/EGEE.
- Grid Technology based on common packaging/distribution of Virtual Data Toolkit.
- Site configurations to simultaneously present resources to OSG & LCG environments - access governed by local Site Policies.
- Applications data management and processing enabled across Federated Grid Infrastructures of OSG & LCG.
- Not to forget collaboration with GridPP through Experiments Data Handling and Analysis Projects.

Short Term Plans

Maintain Grid3 operations in parallel with extending Grid3 to OSG deployment in Spring '05:

- Add full Storage Elements
- Extend Authorization services
- Extend Data Management services
- Interface to sub-Grids.
- Extend monitoring, testing, accounting.
- Add VOs. Add OSG(-VO) Wide VO Services.
- Add Discovery Service.
- Service challenges & collaboration with the LCG

Continue with the Blueprint; Planning joint session with ARDA workshop in October; increase interaction with gLITE/ARDA.

Operations Workshop in Dec. Exchange information with LCG workshop on operations in Nov. Applications Meeting in February (probably)

Open Science Grid

the OSG Consortium goal is to build a production quality grid infrastructure to meet the long term data analysis needs of science in the US and to partner with other such grids in the US and across the globe.