



The Open Science Grid

Miron Livny - OSG PI & Facility Coordinator, University of Wisconsin

Bill Kramer - OSG Council Chair, NERSC/LBNL

Ruth Pordes - OSG Executive Director, Fermilab

Open Science Grid

July 25th 2007



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.





Open Science Grid

Introduction



Why We are Here Today?

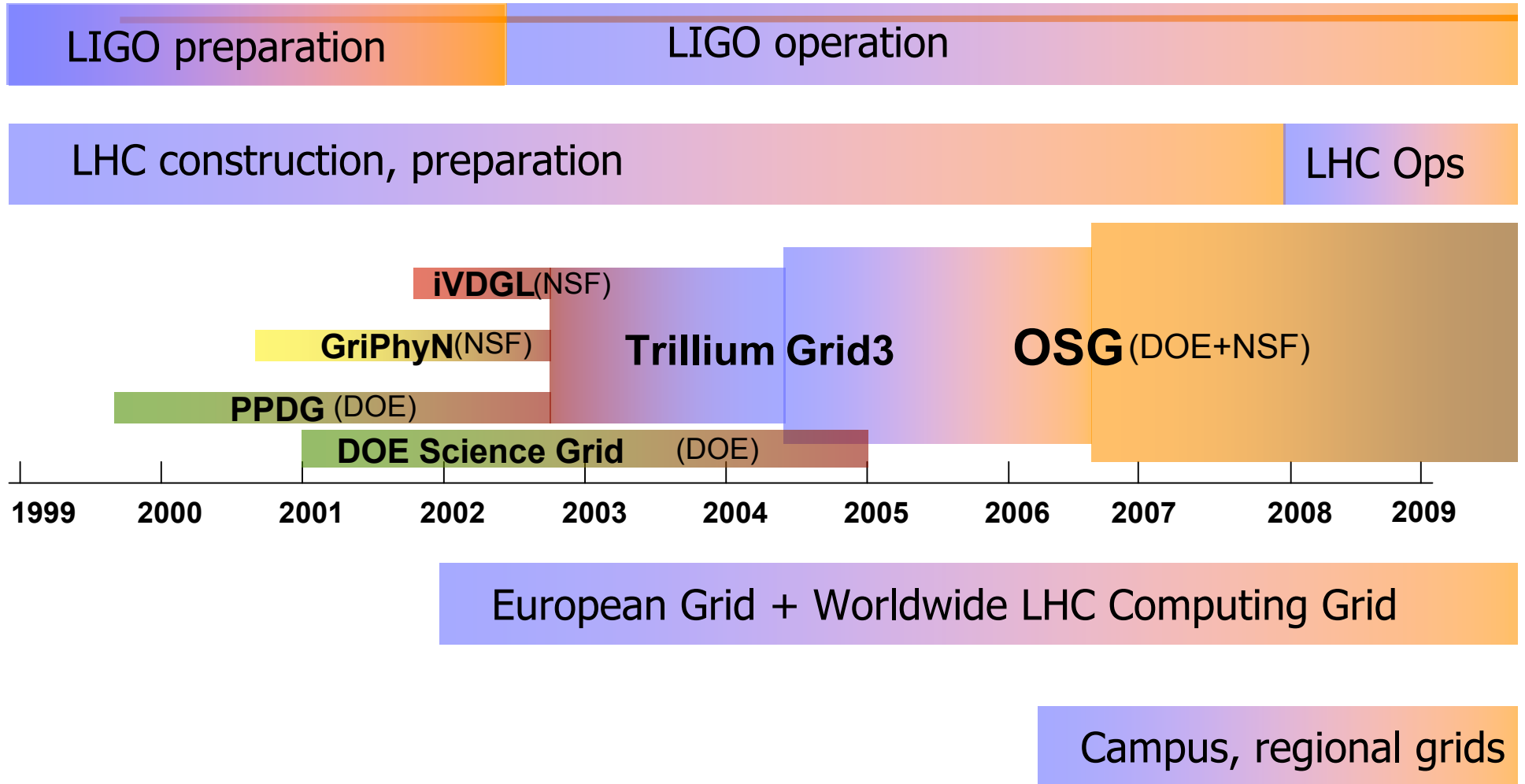
As part of a dialogue on the organization, management and operation of national cyber-infrastructures to brief the OCI on how the Open Science Grid (OSG) is organized, managed and operates.

- What can other national CIs learn from the OSG experience?
- What can OSG learn from the experience of other national CIs?
- How should we go about identifying and improving “best practices” in CI organization, management and operation?



Open Science Grid

The Evolution of the OSG





Open Science Grid

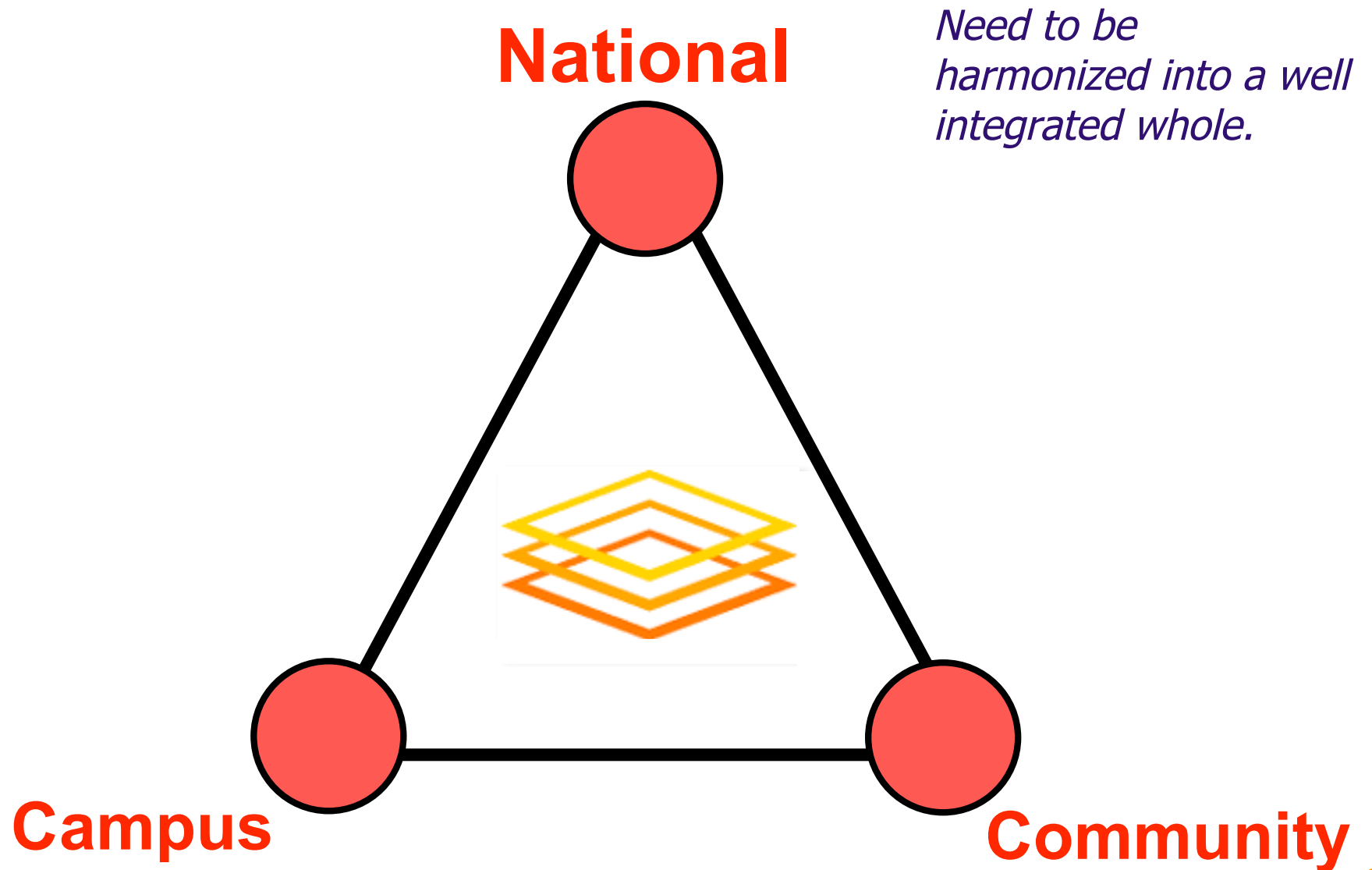
The Open Science Grid (OSG) vision

Transform processing and **data** intensive science through a cross-domain self-managed **national distributed** cyber-infrastructure that **brings together** campus and community infrastructure and facilitating the needs of **Virtual Organizations (VO)** at all scales



Open Science Grid

The Three Cornerstones





The Three Building Blocks

The OSG organization, management and operation is structured around three components:

- the **Consortium** – Bill Kramer,
- the **Project** – Ruth Pordes,
- the **Facility** – Miron Livny



Open Science Grid

The OSG Consortium



Background and Role

- I am the General Manager/Deputy Division Director of NERSC and Head of High Performance Computing at LBNL.
 - Spent time as the PI for the DOE Science Grid
 - Co-PI of the Petascale Data Storage Institute
- As OSG Council chair I am responsible for facilitating the interest and contributions of the stakeholders, making sure the needs of each participant are communicated to the Project and that the project delivers to those needs.



OSG is a Federation of Virtual Organizations and Sites

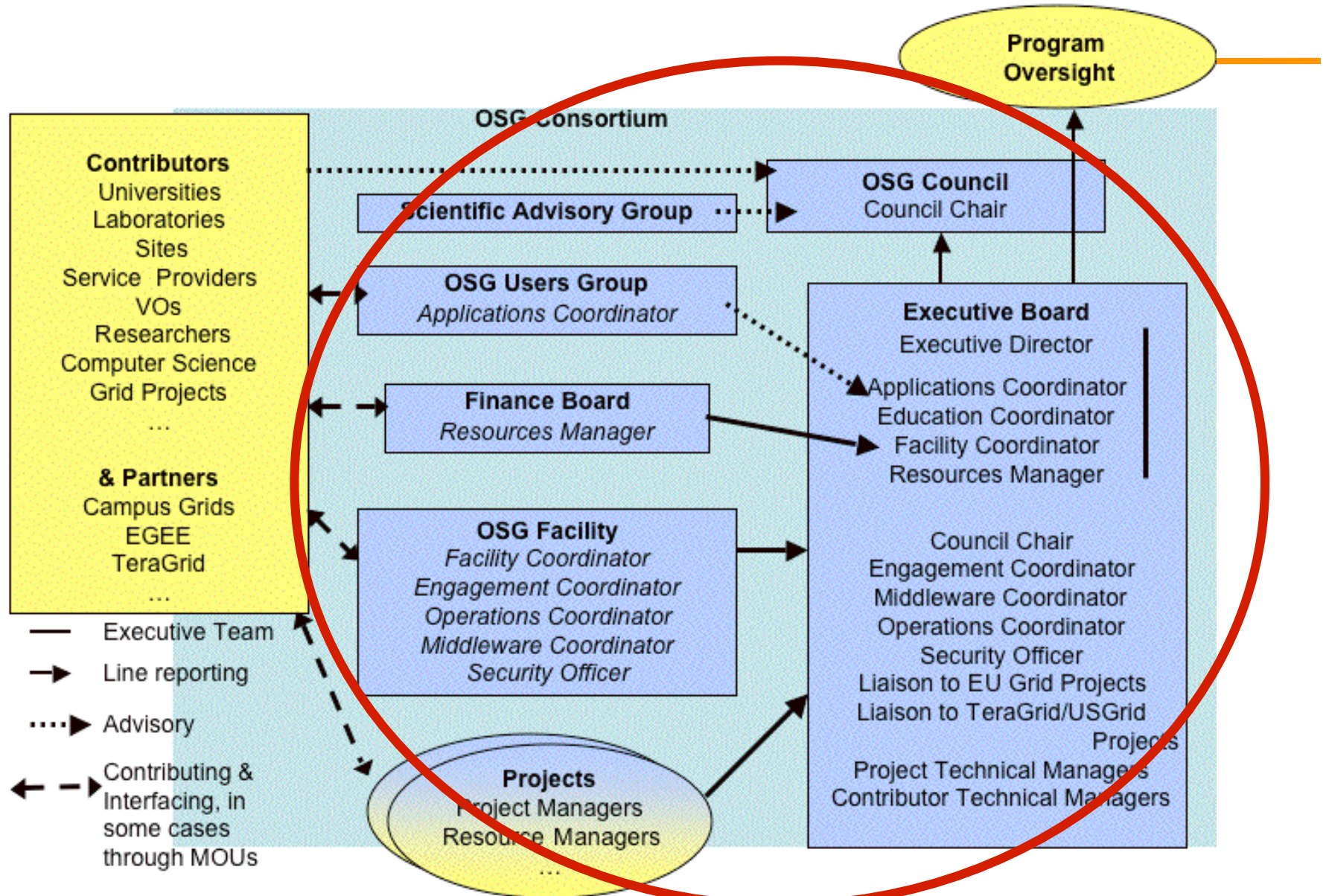
- OSG brings together many VOs for
 - opportunistic sharing of resources in a grid environment
 - allowing for more effective use of their collective resources
 - Allowing easier use of dedicated/allocated resources that are distributed
- A Virtual Organization (VO) is a collection of people (VO members),
 - A VO's member structure may include groups, subgroups and/or roles into which it divides its members according to their responsibilities and tasks, such that they are accorded appropriate levels of authorization.
 - In order to receive the appropriate authorization at another VO's site, a user's grid job must be able to present an authentication token along with a token indicating the desired computing privileges.
- A Site is a collection computing/storage resources (sites) and services (e.g., databases). and the terms "Site," "Computing Element" ("CE"), and/or "Storage Element" ("SE") to refer to the resources owned and operated by a VO or other organization.
 - Use of the resources at a site is determined by a combination of the site's local policies and the user VO's policies. VOs are responsible for contracting individually with each other for guaranteed access to resources.
 - Groups that provide software are also known as resource providers

The OSG Consortium and the OSG Project

- OSG is a large collaborative activity. Its implicit goal is to dramatically “lower the bar” for users and resource providers to join together
- The OSG Consortium comprises all institutions and projects that contribute to OSG.
 - The OSG Council is the governing body of the Consortium. The OSG Council provides the scientific coordination and oversight of OSG activities.
 - The OSG Council self selects a chair for two years to lead the council. Individuals can serve no more than two full consecutive terms as chair.
- The OSG Project is funded to provide staff for specific aspects of managing and sustaining the OSG.
 - The OSG Council elects an Executive Director to manage programmatic activities with the help of an Executive Team.
 - The Executive Director appoints an Executive Board to direct the OSG program of work, draw up policies and represent the OSG Consortium in dealing with other organizations and committees



Structure of the Consortium





Membership represents all aspects

- **60 Scientific Virtual Organizations:** LHC, STAR, LIGO, NanoHub etc.
- **27 Resource Providers:** DOE National Labs, University Facilities etc.
 - 10 Storage focused resources
- **>10 Software Providers (including External Projects):** Condor, Globus, Storage Resource Manager, Internet2, ESNET, CEDPS, Fermilab Accounting etc.
- **>4 Partners - Ex-Officio:** EGEE, TeraGrid, NWICG, TIGRE, APAC etc.



Examples OSG Council Issues over Past Year

- How to organize
 - Organization, Bylaws, Business Practice Review
- Quality Assessment of OSG Services
 - Review of OSG activities
 - Place to bring critical issues
- Allocation of resources
 - D0 and Nanohub – above and beyond resources
- Inclusion of new organizations
- Prioritization
 - Remedial vs new releases



Long Term Issues OSG Council Is Facing

- How to set expectations for providing opportunistic services
 - Surplus, probability, joint scheduling
 - What can people count on
 - Timing conflicts
- How to make the bar to participation low, but effective
 - For users to join and get access to resources – most VOs use multiple resources, some dedicated to them, some not
 - For sites to join – few sites provide resources only to OSG
 - Resource providers to offer quality software
- Sites and VO alignments/mis-alignments
- OSG does not develop nor does it control any SW it uses
- What does it mean to be a facility that has no physical resources
- How to evolve to a sustaining effort



Role of the Science Advisory Group

- The Scientific Advisory Group periodically reviews OSG activities and advises the OSG Council (and Executive Team) on matters related to OSG.
- The SAG members are appointed by the OSG Council Chair with input from the Council and the Executive Team.
- The comments of the SAG will be summarized and shared with the Council and others
- Questions considered by the last SAG Meeting
 - Is the OSG road map appropriate and will it likely result in substantial impact on improved science impact for our stakeholders?
 - Please make recommendations on how OSG should assess its impact on current and future scientific projects and recommend areas of NSF/DOE science that OSG should engage for wider impact.
 - Please provide comments on the steps OSG should take to define a sustaining model of service after this phase of the OSG project completes in 2011.



Current SAG

- Advisors
 - Dr. Barry Barish
 - Faculty at Cal Tech; LIGO large systems experience.
 - Dr. David Schwartz
 - Life science faculty at U of Wisconsin
 - Dr. Helene Armitage
 - VP for S/W development at IBM
 - Grid, Linux, AIX
 - Dr. Ari Patrinos
 - Formerly of DOE, now at Synthetic Genomics
- Outcomes of June Meeting



Open Science Grid

The OSG Project



Background

- Associate Head of the Fermilab Computing Division and US CMS Grid Services Coordinator
- >20 years working on data acquisition and offline systems at Fermilab, leading projects to develop common solutions.
- PPDG coordinator and PI, member of the iVDGL management team.
- Coined “Trillium” for grass-roots DOE-NSF project collaboration.



OSG Executive Director

- Elected by the Council.
- Day-to-Day responsibility for all aspects of the Project.
- Work closely with the PI:
 - The PI is responsible for the contractual obligations to the funding agencies, including the management of funds, reporting and contracts between the OSG and the funding agencies.
- Communicate with the Funding Agencies.
- Respond to the needs of the Consortium and Council.
- Attend to external communications and relationships.



OSG Project Resources

- \$30M over 5 years, jointly funded by DOE and NSF.
- Responsible to the Consortium for specific activities through funding of 33 FTEs of effort.
- The Project staff are distributed across 17 institutions.

	Year
NSF MPS	\$2.45M
NSF OCI (Year 1 only)	\$0.3M
NSF OISE (Year 1 only)	\$0.15M
DOE HEP	\$1.8M
DOE OASCR	\$1.3M
DOE NP	\$0.1M

	FTEs
Facility Management and operations	7.0
Security & troubleshooting	4.5
Software release & support	6.5
Engagement	2.0
Training & Education	2.0
Extensions	8.0
Executive Director and administration	3.0



Project Organized into Areas

Executive Team selected by Executive Director and endorsed by the Council.

Executive Team / Area Coordinators responsible for:

- Education and Training
- Extensions and Applications (Users)
 - Provide User Support Groups and Coordination.
- Resources
- Facility
 - Size of the Production Facility Today: 80 Processing Resources, 10 Storage resources.
 - Size of the Integration Grid: 31 Processing Resources, 2 Storage Resources.
 - 11 inactive (registered) Resources.
 - Facility Areas are part of the management structure (see Miron's talk)



Managing a Complex Distributed Project

- Area co-coordinators - (almost) all have other significant non-OSG responsibilities.
- Regular (1 hour) meetings:
 - Weekly meetings of Executive Team.
 - The technical areas: Facility, Operations, Security, Storage, Users group , Integration, Communications.
- Every six weeks meetings:
 - Executive Board - includes external projects we depend on.
 - Blueprint meeting discussions on architecture and longer term roadmap.
- OSG@work web and mail lists heavily used.



Deliverables and Milestones are Defined and Tracked

- Annual Project Plan (including WBS) gives details of deliverables and timeline for the year.
 - Deliverables driven by the science stakeholders.
 - Buy-in through “Science Milestones” deliverables - owned by the stakeholders and included in the plan.
 - Area Coordinators responsible for milestones and deliverables under their branch.
 - Well defined Software releases part of the plan.
- WBS updated by Area Coordinators quarterly. Missed milestones subject to discussion.
- Adjust plans based on experience, requests, feedback and problems.
- Change request process for project deliverables



Reliance on External Projects

- Deliverables depend on contributions not owned by the project:
 - Resources contributed by the Consortium members
 - Software contributed by external development projects.
- External Projects are members of the Executive Board.
 - Planning includes discussion and agreements between OSG and external projects.



Managing the Resources

- Resources Managers with responsibility for management and tracking of all funding issues.
 - Project Office to help me and resource managers.
- Annual Statement of Work (SOW) with each organization, signed by OSG and organization PI in conjunction with funding distributions.
 - Quarterly Reports expected from area coordinators and organization PIs.
 - Change request process defined and exercised - managed by the Resources Managers signed off by the Executive Director.



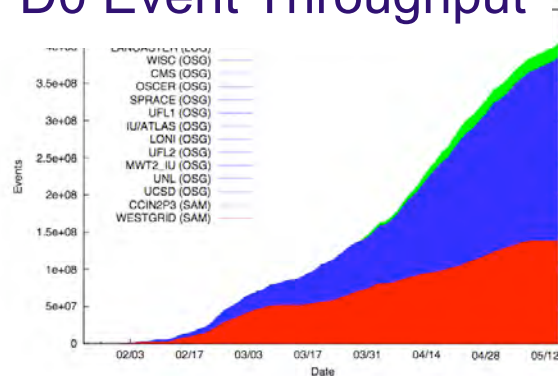
Open Science Grid

How we actually get something done!

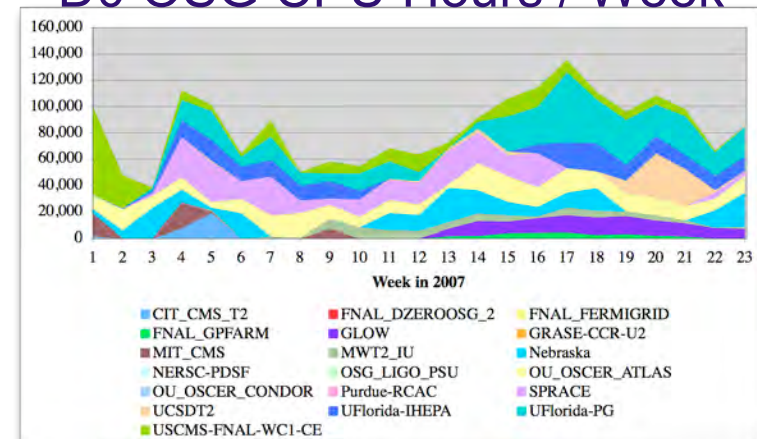
Results of D0 Reprocessing

- OSG responded to an unexpected request for a significant amount of resources- typical of discovery process.
- Through team work:
 - OSG Members providing processing cycles,
 - OSG and D0 worked together to troubleshoot problems
 - Daily attention to availability and throughput
- Reprocessing was finished in time for summer conferences:
 - 12 OSG sites contributed up to 1000 jobs a day, 2 M CPU hours
 - 286 million events, 286M Jobs on OSG
 - 48 TB Input data, 22 TB Output data transferred

D0 Event Throughput



D0 OSG CPU Hours / Week





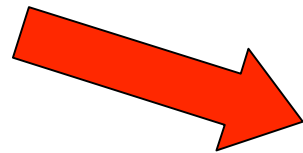
Training and Engagement

- Training program consists of 2-3 day schools of lectures and hands on training.
 - For administrators, users and software developers.
 - Education Coordinator follows up with alumni to help them use their knowledge.
- Engagement works and helps new Communities (User and Campus). directly for several months each.
- Campus Cyberinfrastructure (CI) Days joint with Educause, Internet2, TeraGrid, facilitate communication and infrastructure planning across CIO, Faculty, IT locally.

Becoming a Full OSG Citizen

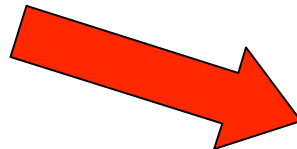
Join the OSGEDU VO:

Run small applications after learning how to use OSG from schools



Be part of the Engagement program and Engage VO:

Support within the Facility to bring applications to production on the distributed infrastructure



Be a standalone VO and a Member of the Consortium:

Ongoing use of OSG & participate in one or more activity groups.



Example of IceCube

- IceCube members attending International Summer School for Grid Computing.
- Single researchers included in OSG owned VOs and given help them use the infrastructure.
- When using the infrastructure in production will have an independent VO and be a regular OSG Community.



The Extensions

- It takes real effort to integrate and deploy new capabilities and technologies in the end-to-end system
- We work with development groups and the user communities to help with:
 - Scalability and system tests.
 - Ensure administration and monitoring work included
 - Interfacing to applications
 - Matching to OSG stakeholder needs



Security is part of the Project in the Facility

- Security scope must cover the OSG assets in the Facility and the VOs/Users.
 - Grid wide Services
 - Interfaces to and expectations of Resources & VOs.
- Communication is essential across many projects in many places key when risks and vulnerabilities identified.
- The Security Office and their team must respond quickly 24x7 to incidents.



Open Science Grid

The OSG Facility



Background and Role

Professor of Computer Sciences at the University of Wisconsin-Madison

- >30 years of research in distributed systems
- Leading the Condor project
- Co-PI (Computer Science) of PPDG and Senior Personal (leading the Virtual Data Toolkit – VDT effort) in GriPhyN and iVDGL.

Coordinator of the OSG facility

- Member of the OSG Executive Team
- Manage the Facility effort and project plan - ~ 2/3 of the OSG FTE count



Organization

The facility effort is organized in six groups:

- **Engagement** – Identify and support new groups
- **Integration** – Transition the OSG software stack to deployment
- **Operation** – Monitor activities and support VOs and sites
- **Security** – define, implement and monitor the security plan of the OSG
- **Software** – evolve, package and support the VDT
- **Troubleshooting** – work with sites and VOs to resolve “problems” in end-to-end functionality



The OSG Virtual Data Toolkit

- A collection of software initially delivered by GriPhyN, iVDGL, PPDG. The VDT doesn't write software, but gets it from providers
 - Condor, Globus, EGEE Components, Community Software (e.g. Fermilab accounting), open source utilities (Apache, MySQL etc).
 - Composition driven by stakeholders and controlled by the OSG project
 - An easy installation.
 - A support infrastructure.
 - Provides a middleware foundation for the software stack of several production Grids - including the OSG.



Why have the VDT?

- Everyone could download the software from the providers...
- But the VDT:
 - Figures out dependencies between software
 - Works with providers on bug fixes feature enhancements
 - Facilitates interoperability
 - Offers automatic configuration
 - Packages it
 - Tests everything on 15 platforms (and growing)



OSG Middleware Deployment

Domain science requirements.

**Condor, Globus,
Privilege,
EGEE, ...**

**OSG stakeholders and middleware
developer (joint) projects.**

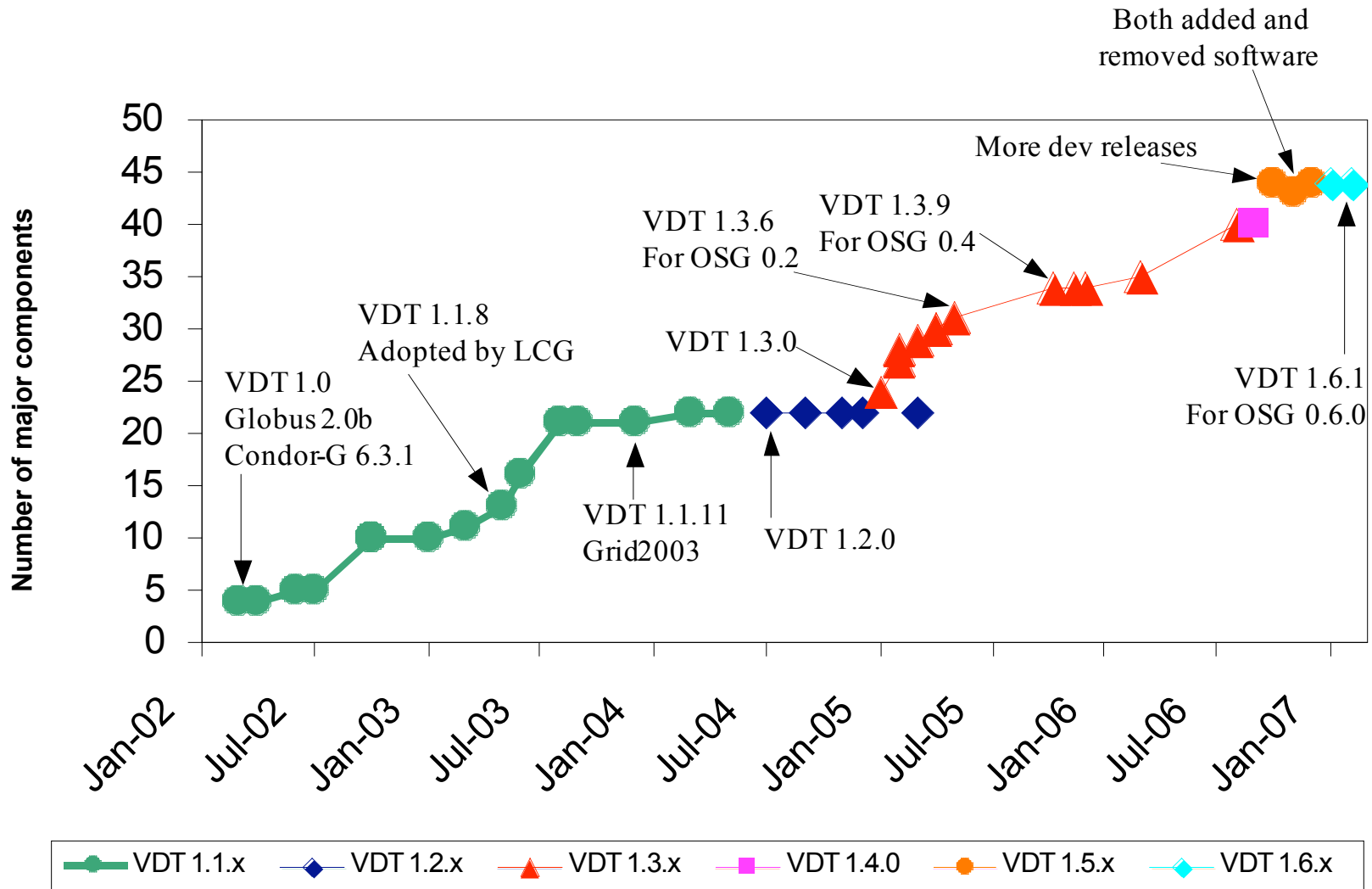
Test on “VO specific grid”

**Integrate into VDT Release.
Deploy on OSG integration grid
Interoperability testing**

**Provision in OSG release &
deploy on OSG sites.**



How much software?





Open Science Grid

It's the people who make it work!

