

Open Science Grid

Collaborative Science

On a National Cyberinfrastructure

part of the Worldwide Federated Grids

Ruth Pordes, OSG Executive Director, Fermilab

Globus World March 2nd 2010



Project 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.



“Local to Global”

Open Science Grid
Services and Infrastructure

End-to-End
Community
Systems

Contributions

Science
Community

Software
Provider

Resource
Provider

Executive
Board

Consortium

Project

Satellite
Project

Satellite
Project

OSG does not own processing, storage, networking and does not develop software.

\$ DOE

\$ NSF



We take the “P” and “M” seriously

Project Management Process, Planning and Goal Setting

Staff at 16 institutions & about 3, 4 FTE of effort by about 50 staff members organized into functional work areas (e.g. Operations, Security, Education, etc.) each of which is led by an Area Coordinator

Project year starts in October and ends in September, program planned one year at a time starting in June preceding the start of each project year. five sub-processes Strategic Planning and Goal Setting

WBS and Staffing Plan, Budget Allocation and Statements of Work , Execution Management , Closure and Feedback

Also mid-year corrections and updates to the plan as warranted by changes in staffing or needs of key stakeholders. OSG Council provides high-level input from the science stakeholders for the goal setting process.

June of each year, develop high level strategic initiatives to drive each of the functional area of work. Inputs from the original work proposal, plans of the key science stakeholders, evolution in technologies, feedback from the progress made and specific inputs from funding agency stakeholders. Area Coordinators develop high level goals describing their major work items; e.g. Recruit and develop three new science communities in achieving production

Reviewed to insure alignment; shared across the Area Coordinators then we proceed to WBS and staff planning.

3. WBS and Staffing Plan, Budget Allocation & SOWs

Transform these goals into a WBS with tasks and dates. Can move staff from one functional area to another (assuming a match in technical capability) as required by the work program. OSG staff retreat, provides peer review and ensure alignment of work

Budget plan reviewed and approved by the OSG Executive team and then concurred to by the OSG Council. The Statements of Work Overview of the work and responsibilities with high-level goals,

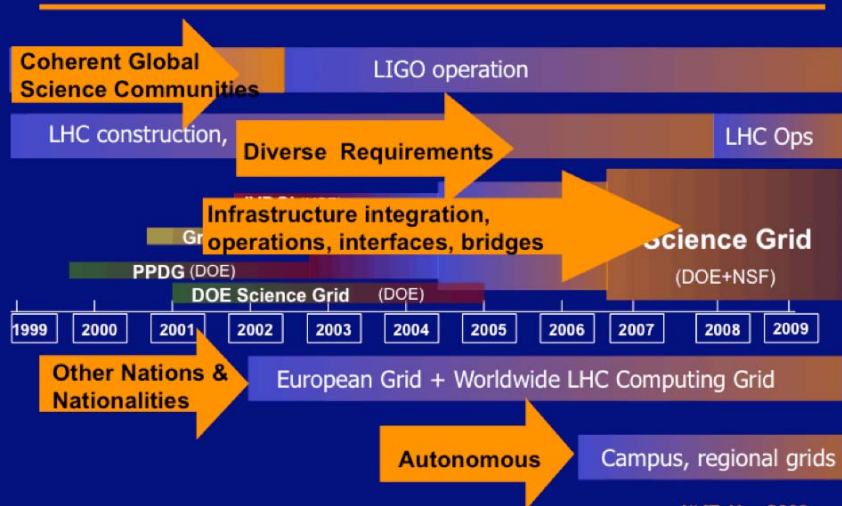
Specific tasks directly extracted from the WBS. Specific staffing commitments along with percent FTE and responsibilities. Specific budget allocations spanning NSF and DOE sources (directly extracted from the budget plan), and a set of common operating guidelines, terms and conditions, and principles that define the relationship between OSG and each participating institution.

The Statements of Work are then routed for approval

Status updates are solicited from the WBS owners on a monthly basis and, as needed, issues are discussed and managed. Each functional area presents their status every 6-8 weeks at a teleconference meeting.

St the end of September, a final WBS status update is undertaken. A set of metrics show what parts of the project were completed and which WBS items remain incomplete.

Provenance



NLIT, May 2008

And Bridging to Other Grids



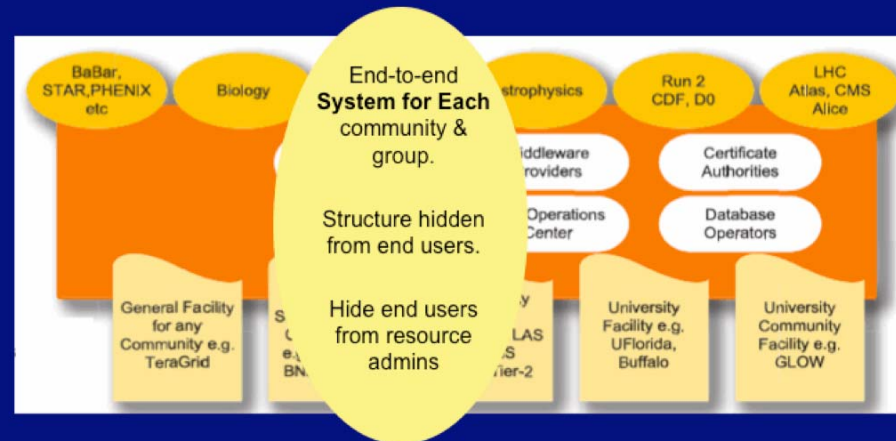
NLIT, May 2008

Acting as an Agent



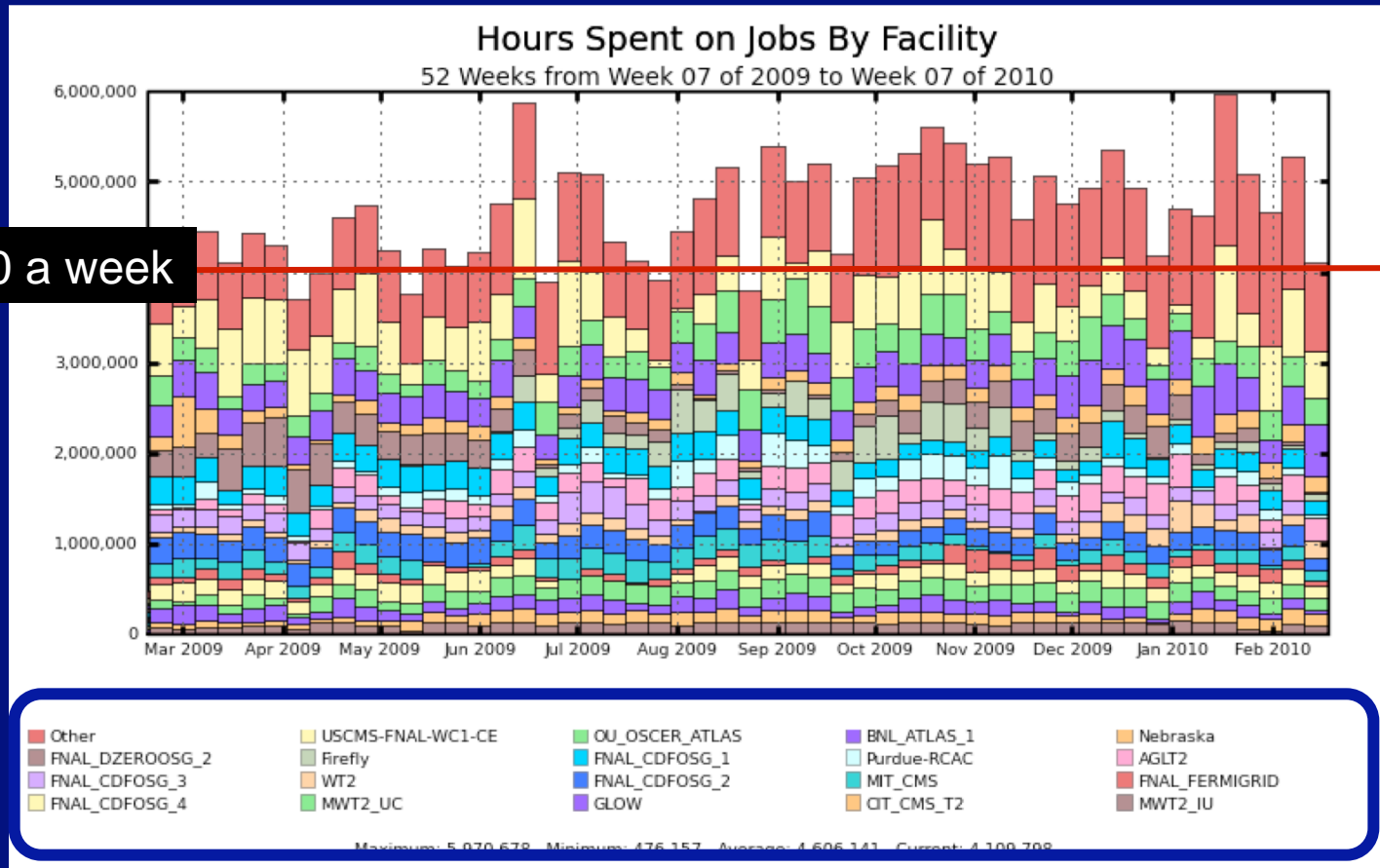
NLIT, May 2008

Providing End-to-end Integration





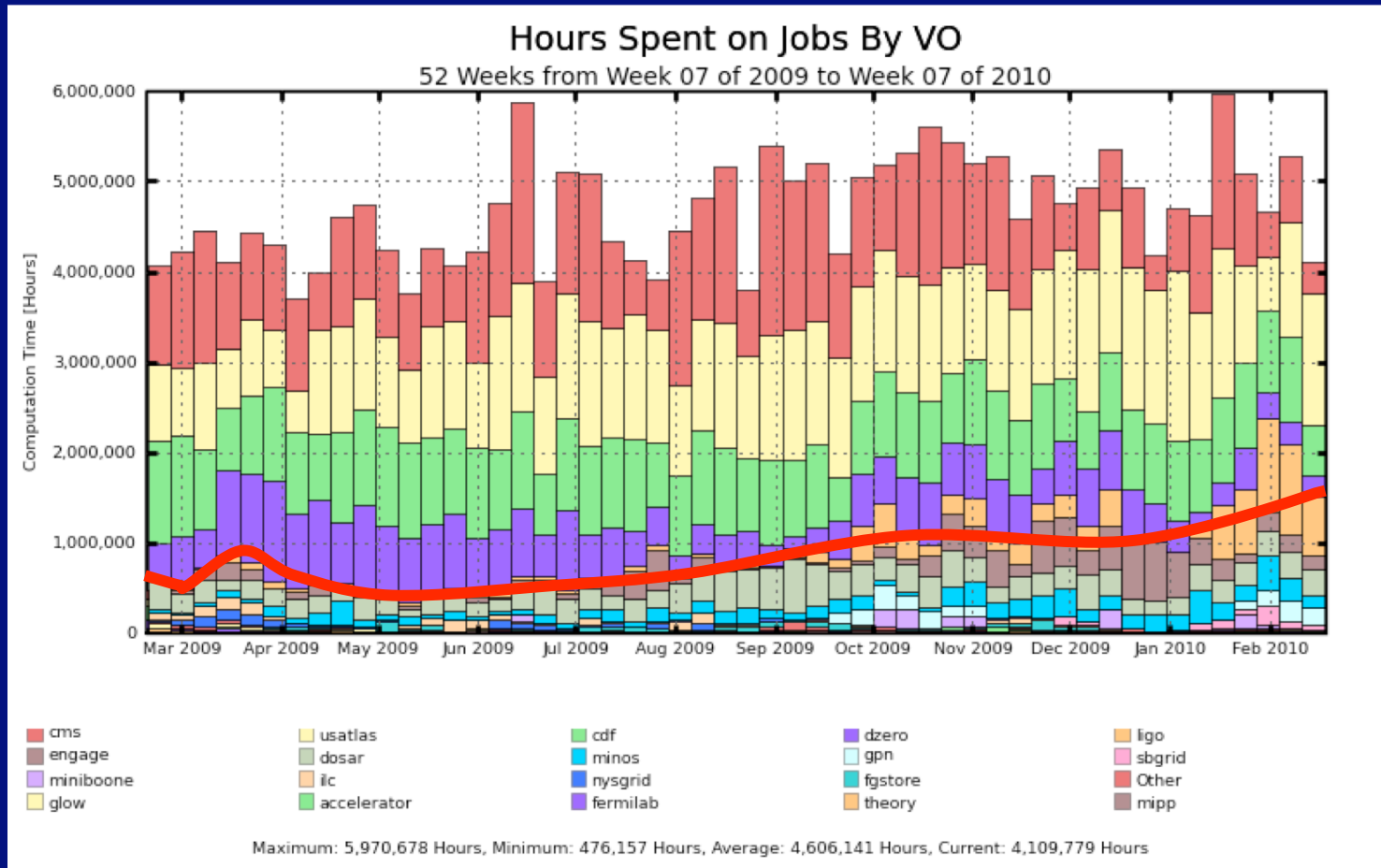
Use by Site



20 Sites represented explicitly. Largest contribution is “other”.
Peak doubled, average more than doubled, in 2 years;



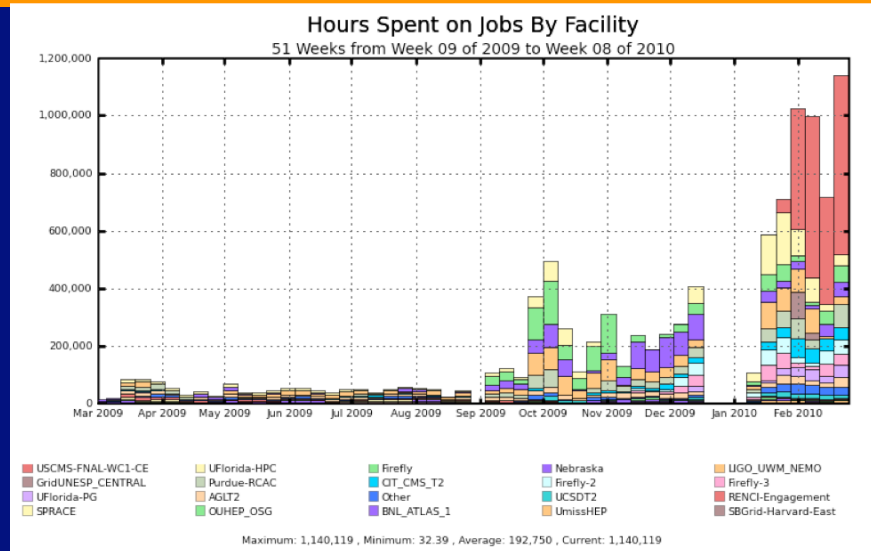
Use by Community



Curve shows “not-big 4” also ~doubled



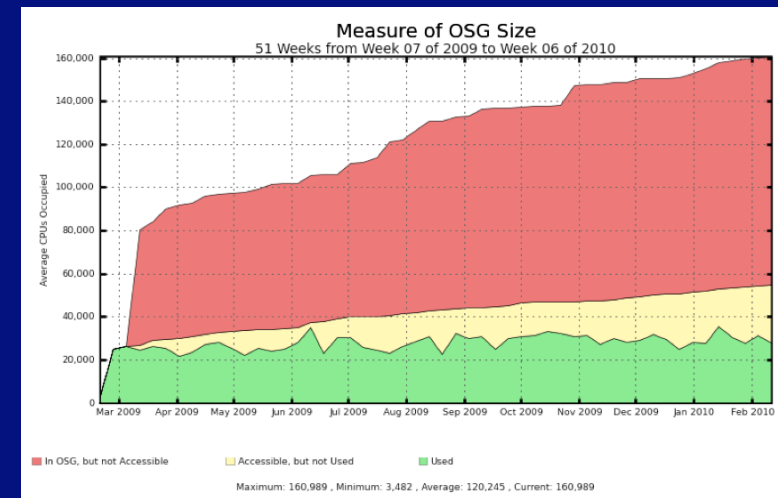
“Site Filler” – not there yet!! LIGO BOINC based Einstein@home



We determine the approximate size of the OSG by measuring how many compute hours per day are being reported, then dividing by 24. This establishes a good minimum bar for the size of the OSG, but is an under-estimate because not all resources are used every day.

For each site, we calculate the accessible cores by looking at the maximum usage over the past year. This assumes that the site has been utilized at its maximum capacity at some point in the last year.. This metric is based on actual usage.

The total cores of the OSG. This is the number of cores the site has in its batch system. Due to the fact that many sites have a batch system configuration which limits the number of grid jobs that can run, this metric is usually an over-estimate.





LHC Core Stakeholder

Global single "sign-on" system used by >1k physicists to

- o understand the experimental apparatus
- o develop & refine software to process the "raw" data
- o develop & refine software to analyze the refined data and derive scientific results

Tiered Structure

- o globally distributed archival storage: National Labs in US, Germany, Italy, France, Spain, UK, Taiwan

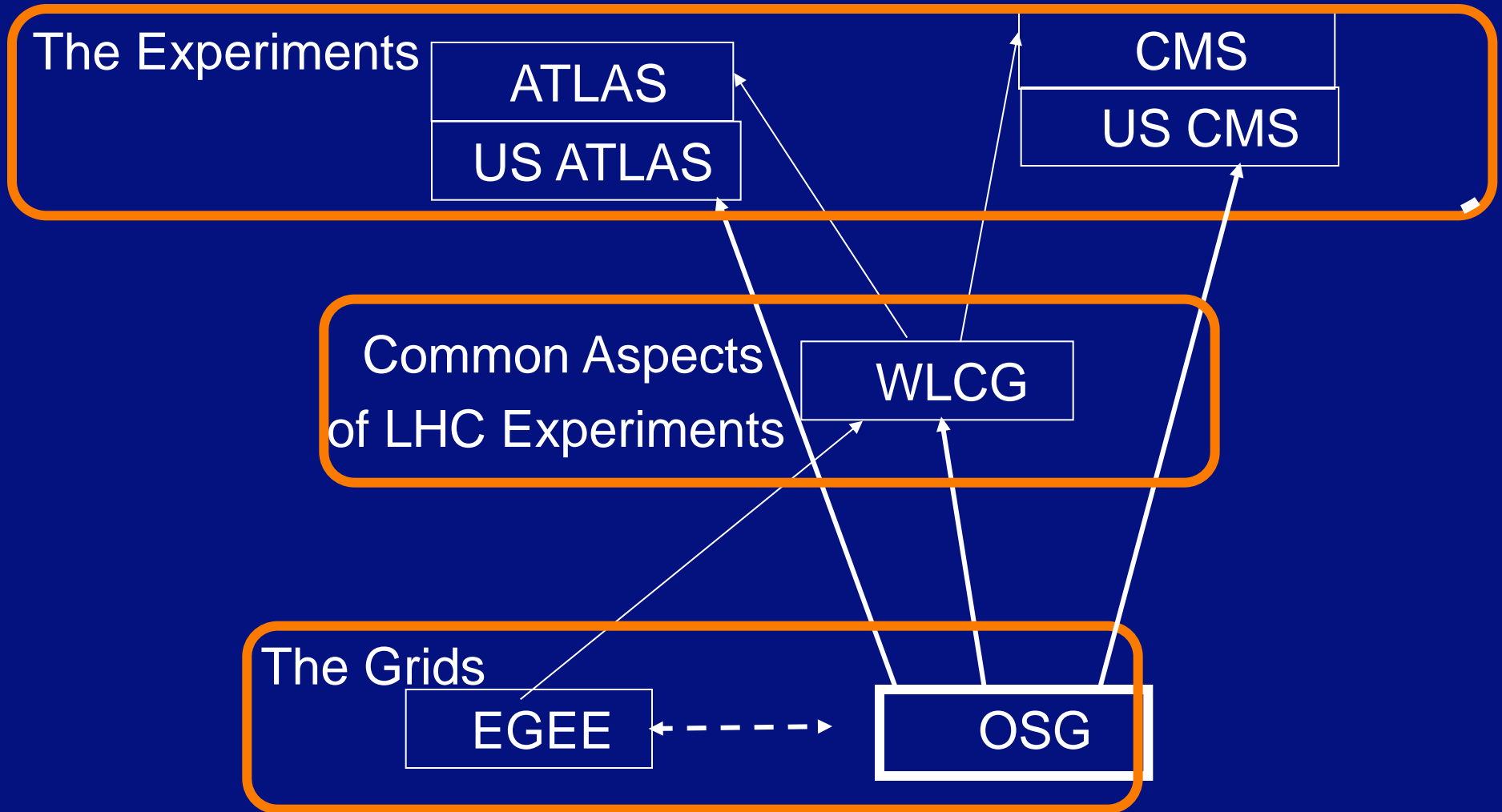
Live archive & primary processing

- o globally distributed analysis infrastructure: ~50 clusters of 0.1-10k cores and 0.1-1 Petabyte disk

- analysis to derive scientific results



Multi-Layered VOs





Engagement - our word for immersive, embedded help and training

Bring the power of the OSG to scientists, educators,
and campus beyond the physics domain...

and to use these experiences with new users and
domains to drive new requirements to evolve the
OSG.



Support and Sharing of Software and Knowledge

Virtual Data Toolkit – many different collections for sites,
users, managers, grids

OSG specific configurations and tools.

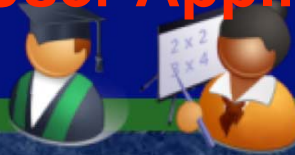
Only required service is Accounting!



Architecture of S/W

e.g.
OSG

Users, User Applications,



Community Wide Common Software and Support



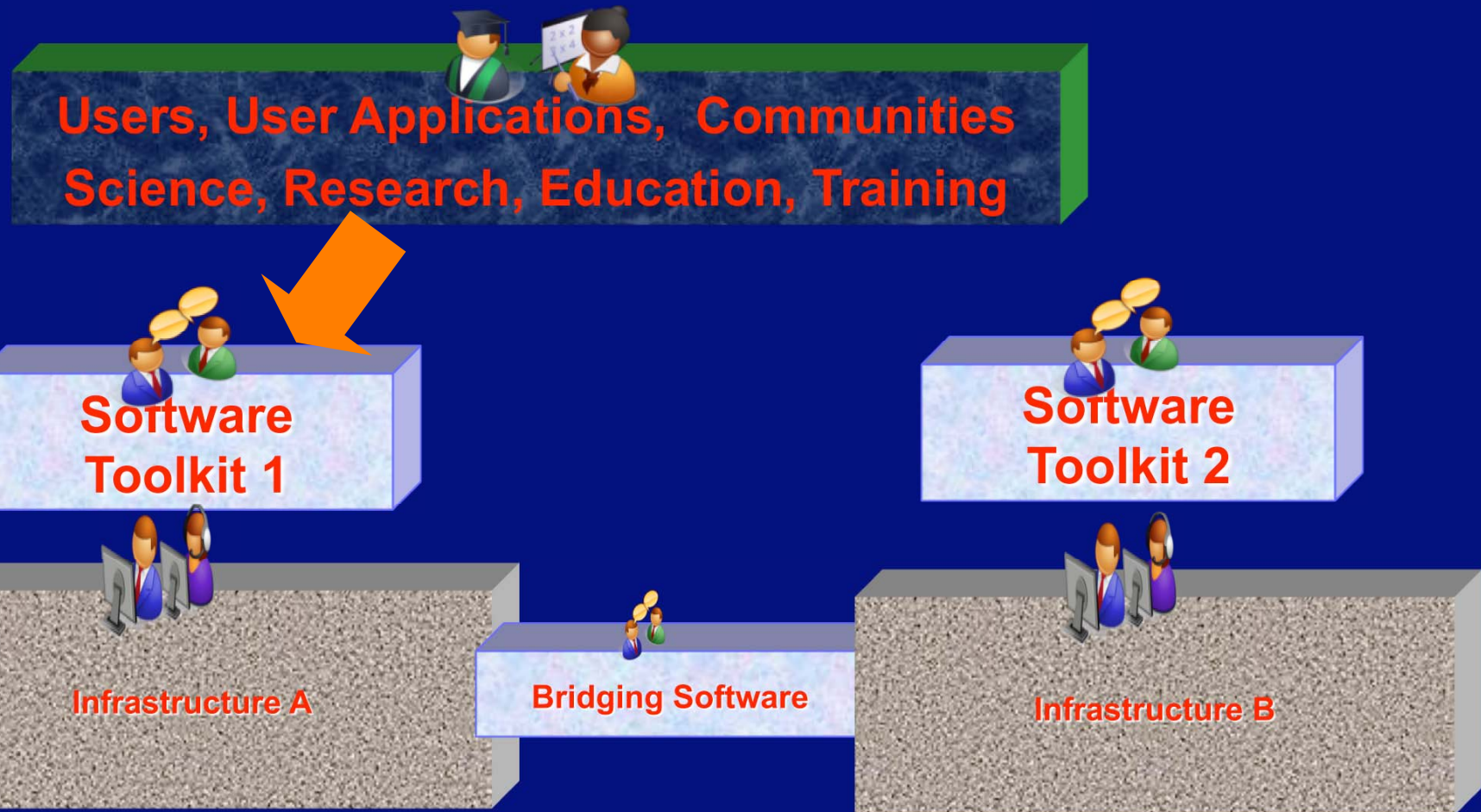
**Software, Services & People to help
(~60 modules)**



**Processing, Storage, Data and Administrators,
(at ~75 US Labs and Universities)**



Using Federated Infrastructures





OSG and Globus

Core use and support of GSI, GridFTP, GRAM 2,

Common/interoperable Authz call-out protocols.

Starting to integrate with scientific cloud frameworks.

Commonality of stakeholders and therefore
commonality of needs.

Every 6 weeks Collaboration phone-cons cover
issues, testing, new releases, futures... more on
this from later speakers.



OSG & TeraGrid(XD)?

Joint Statement of Agreed Upon Principles between OSG and TeraGrid, July 30th 2009

https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=882&version=1&filename=OSG_TG_SharedPrinciples_v4.pdf

OSG works with all XD teams equivalently.

Task Mission/Goals	TG Owner	OSG Owner	Stage
OSG & TG Joint Activity Tracking and Reporting	Tim Cockerill	Chander Sehgal	Active
iSGTW continuation and joint support plan; proposal for US based contributions	Elizabeth Leake	Judy Jackson	Planning
Resource Allocation Analysis and Recommendations	Kent Milfeld	Miron Livny, Chander Sehgal	Planning
Resource Accounting Inter-operation and Convergence Recommendations	Kent Milfeld, David Hart	Phillip Canal, Brian Bockelman	Planning
Explore how campus outreach and activities can be coordinated	Kay Hunt, Scott Lathrop	John McGee, Ruth Pordes	Active
SCEC Application to use both OSG & TG	Dan Katz	John McGee, Mats Rynge	Active
Joint Middleware Distributions (Client side)	Lee Liming, J.P. Navarro	Alain Roy, ?	Active
Security Incidence Response	Jim Marsteller, Randy Butler	Mine Altunay, Jim Barlow	Active
Workforce Development	John Towns, Scott Lathrop	Ruth Pordes, Miron Livny	Planning
Joint Student Activities	Scott Lathrop, Laura McGiness	David Ritchie, Jim Weichel	Active
Infrastructure Policy Group	J.P. Navarro, Phil Andrews	Ruth Pordes, Miron Livny	Active



Communications & Outreach

International Science Grid This Week (iSGTW).

International Grid Summer Schools – now one integrated into

Monthly OSG newsletter.

Grid schools in Colombia and collaboration with Sao
Paolo State Grid in Brazil.

**Goal to
integrate into
OSG program of
work and
increase
collaboration
with others.**

EGEE, EGI, EMI: 17 points of contact once count all the
working groups, security and operations, software
activities.



Happening “now”

LHC data taking and Analysis.

Increased usage by a number of communities.

Start to look to integrate commercial and scientific clouds, support “small-scale parallelism” applications, interface to Magellan/FutureGrid/

Try and improve Usability and lower the threshold to sustained use.