Executive Director Report
Open Science Grid
All Hands Meeting, 2010

Its been a productive year

Its been a productive week

Ruth Pordes

Very Interesting, Well Run All Hands Meeting

Kudos to Piotr,
his Program Committee,
Leaders of the Workshops,
Presenters
Attendees

The LHC took data and published results quickly

Usage of OSG rose and stayed high.

Average throughput 1.2 million CPU hours/ day 300 Terabytes data moved (WAN)

Non-LHC throughput ~50%;

Non Physics througput ~20%

~250 Scientific Publications in 2010

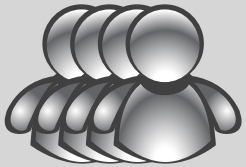
VO	# pubs
ATLAS	67
CDF	52
CDMS	2
CIGI	2
CMS	33
D0	20
Engage	20
GLOW	19
HCC	1
LIGO	6
Mini-Boone	3
MINOS	4
nanoHUB	3
NYSGRID	1
OSG	5
SBGRID	3
STAR	13

It's the Community that does it!

OSG's Community-Focused Architecture

Communities...
(LHC, LIGO, SBGrid, Campuses...)

... have **users**, application developers, site administrators, researchers, students etc:



... may use custom **services** in addition to OSG's:

Job management
Data management
...

...may provide diverse **resources** via one or more autonomous sites:

Compute Interface	Storage Interface
Site Fabric	System Status Accounting Security

OSG

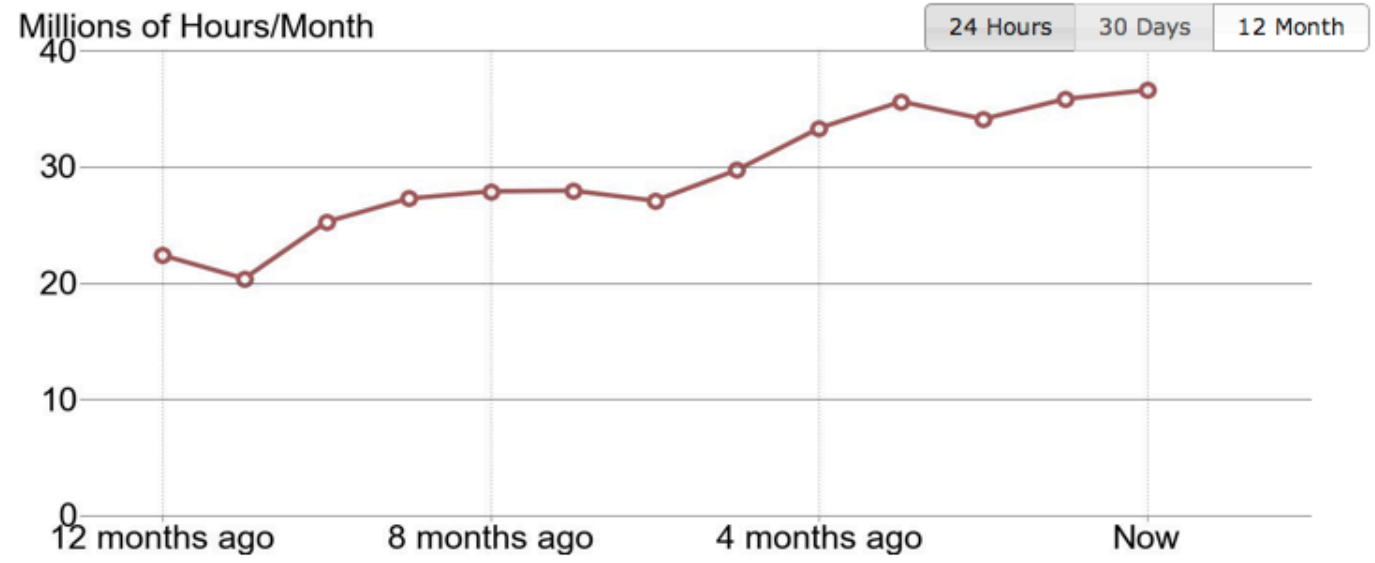
Consultation Services:
- Community
- Security
- Architecture
- Education/Training

Operational Services:
- Front-line support
- Security
- Info/monitoring
- WLCG Interaction
- IT Services

Software Services
- Software Distribution
- Software Evaluations
- Support

There is a sharing of software, operational services, and knowledge between the communities and OSG in each of these areas.

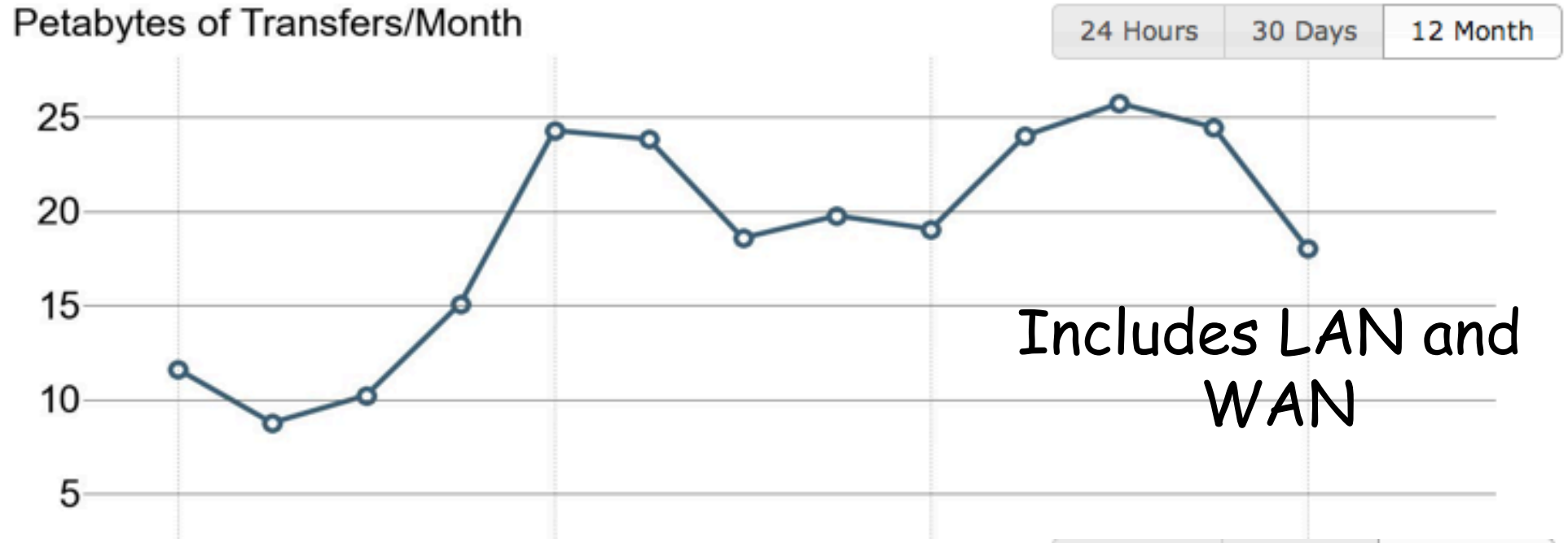
Jobs | CPU Hours | Transfers | TB Transferred | Status Map



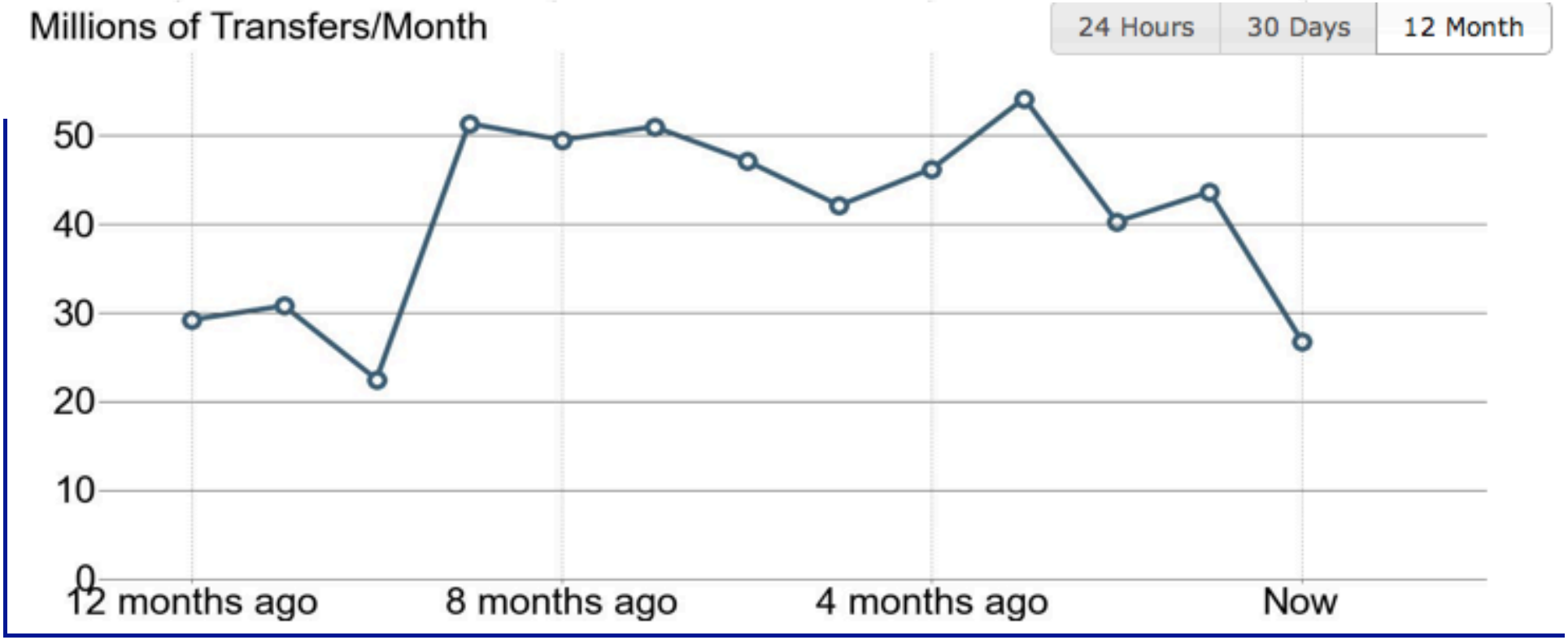
CPU hours spend on an OSG resource is reported to the central accounting system. The above graph shows the number of CPU hours per month. A total of 383,551,000 CPU hours were spent.

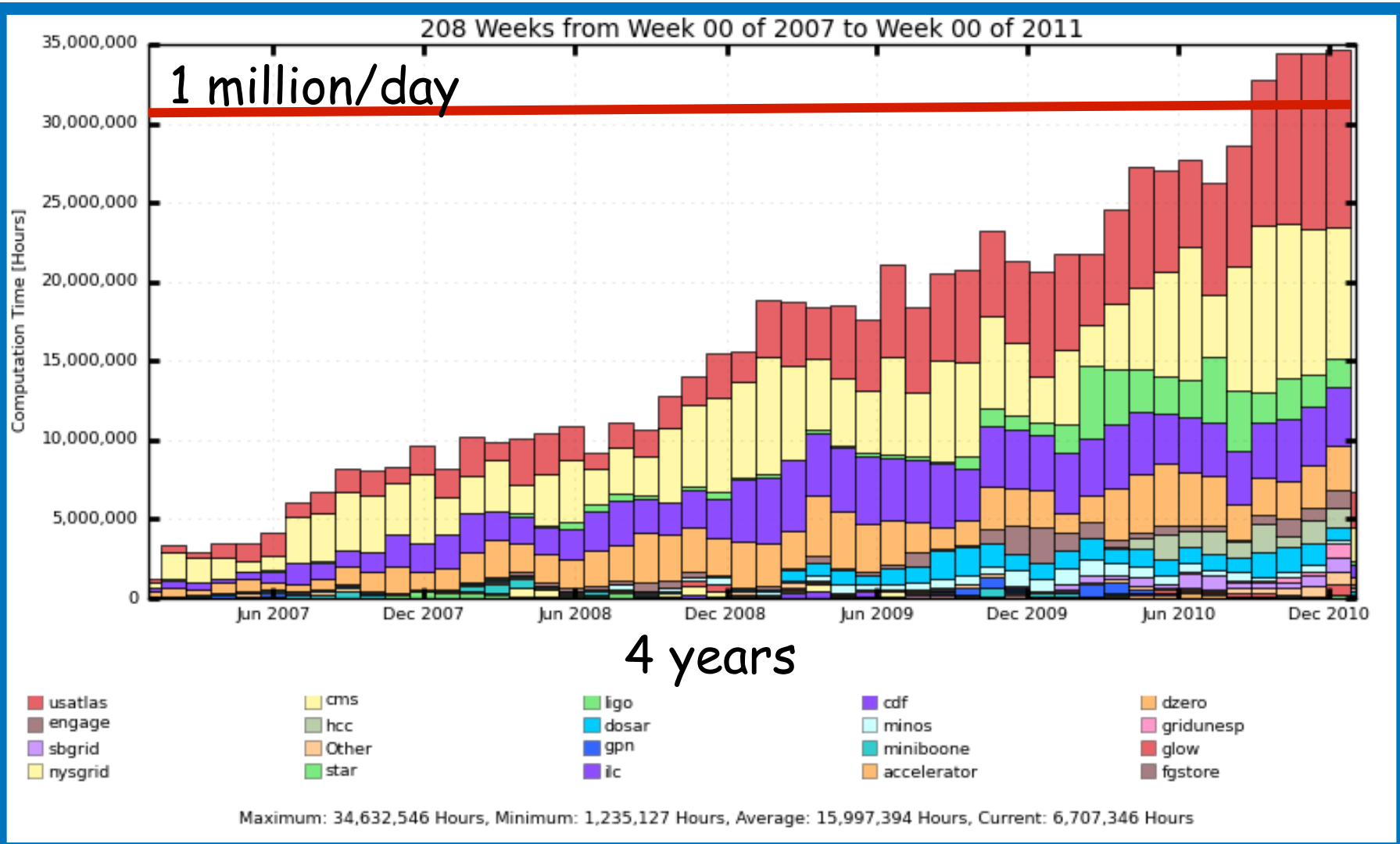
OSG delivered across 87 sites

In the last 24 Hours	
627,000	Jobs
1,149,000	CPU Hours
1,472,000	Transfers
611	TB Transferred
In the last 30 Days	
14,932,000	Jobs
35,951,000	CPU Hours
42,805,000	Transfers
19,758	TB Transferred
In the last Year	
176,079,000	Jobs
383,551,000	CPU Hours
534,493,000	Transfers
261,395	TB Transferred

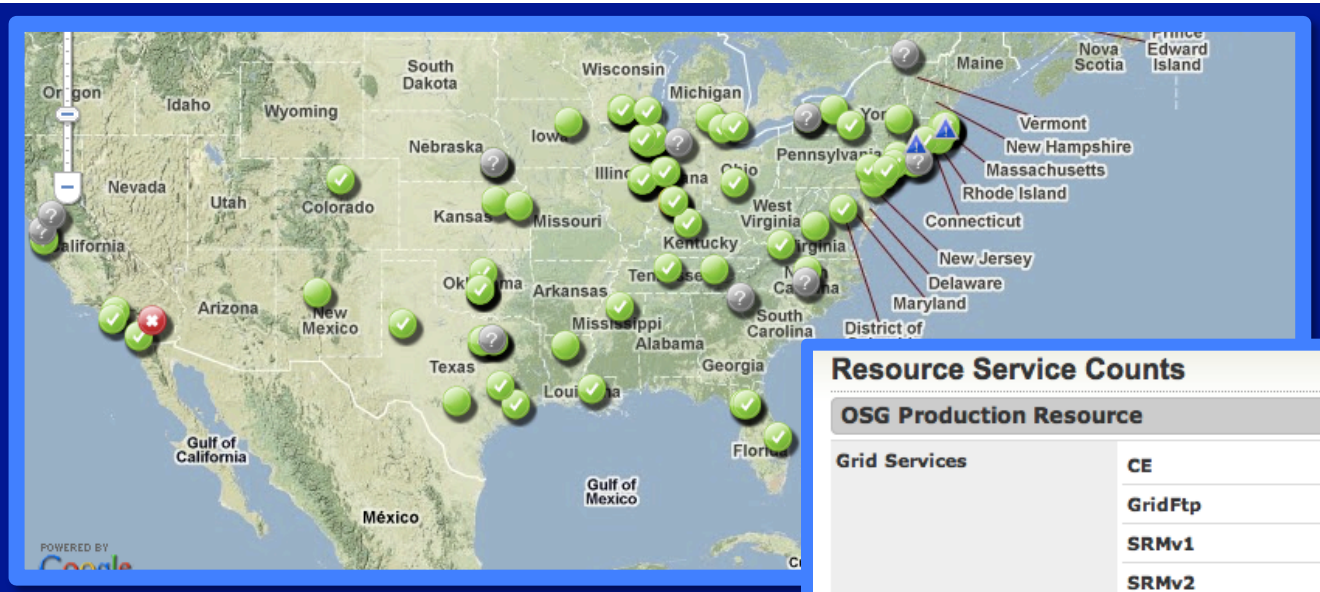


Includes LAN and WAN





13 main VOs
8 are multi-science



Resource Service Counts

OSG Production Resource		
Grid Services	**CE**	126
	GridFtp	48
	SRMv1	0
	SRMv2	62
	Submit Node	5
OSG Integration Test Bed Resource		
Grid Services	**CE**	25
	GridFtp	6
	SRMv1	1
	SRMv2	7
	Submit Node	0

Available Processing
95 OSG 1.2.X
3 OSG 1.0.X

Welcome New Entrants

Alice production
Belle-2 (EGI/OSG)
CSIU*
Dayabay

Many US LHC Tier-3s.

and

SURAgrid

A Regional Science Community

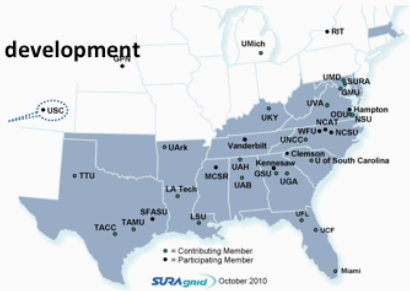
www.suragrid.org

March 2011 SURAgrid establishes an OSG VO

- **Facilitate Collaboration in the context of national research cyberinfrastructure**
- **Improved user support**
- **Easily extend resource sharing**
- **Participate in strategic and technology development**

History

- SURAgrid evolved from the NMI Testbed Grid in 2003
- 35 sites
- 18 resource providers
- 11 applications – storm surge modeling, genome alignment, gene database search, teaching and outreach



* Computational Science at Indiana University

Next steps for the Americas Partnerships

GridUNESP (Sao Paolo)

- Continues as significant opportunistic resource provider. 8 sites.
- Training school reuses much of OSG summer school material
- Readied for local users – next step Submission Factory and job submissions.

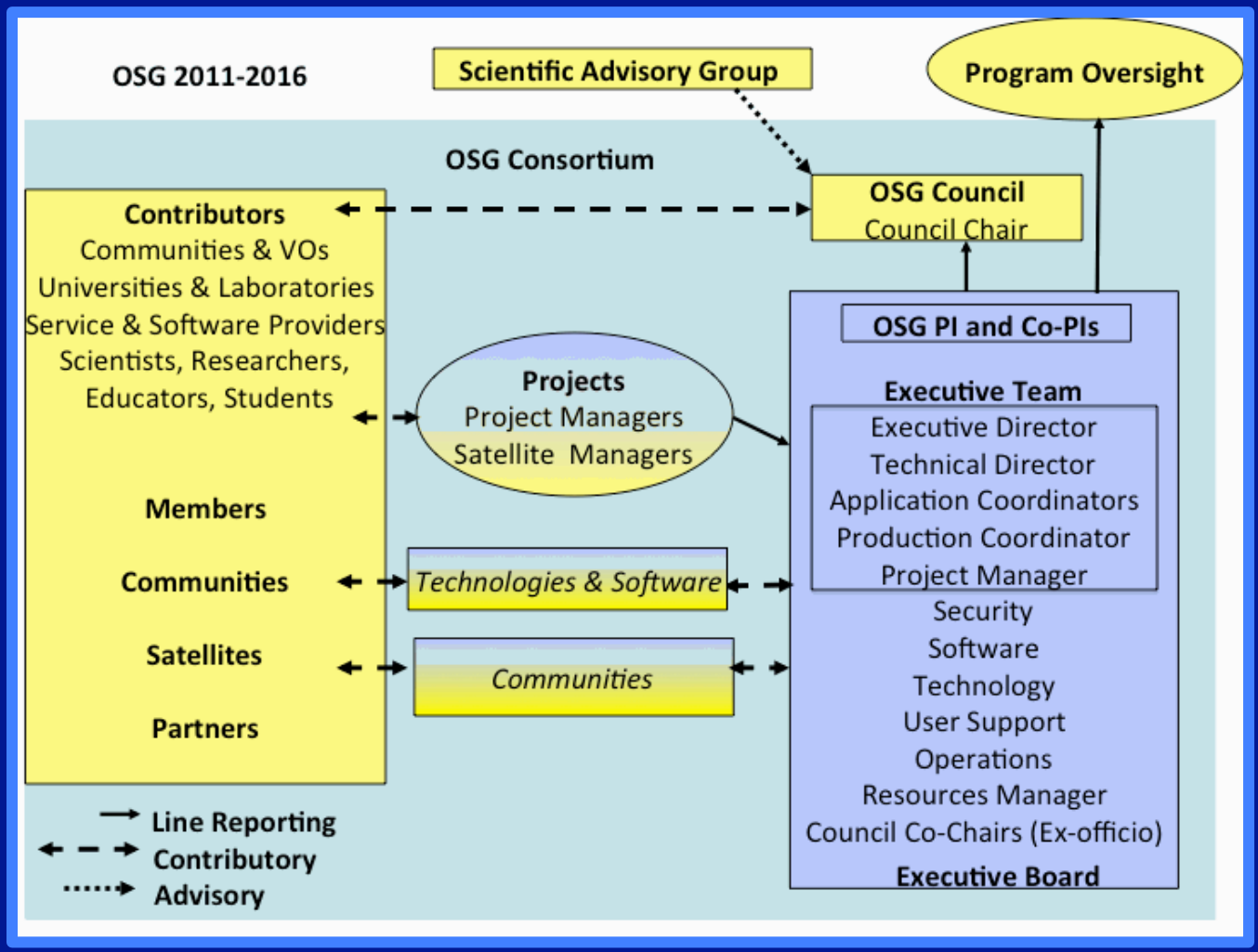
Colombia - Welcome to the lead Profesor Harold Castro, Universidad de Los Andes.

- Annual schools with OSG tutors over last few years. Many small sites have installed s/w stack.
- One site supports both glite CE and OSG CE managing common set of worker nodes! Will become part of OSG documentation.
- Ready to Engage Users.

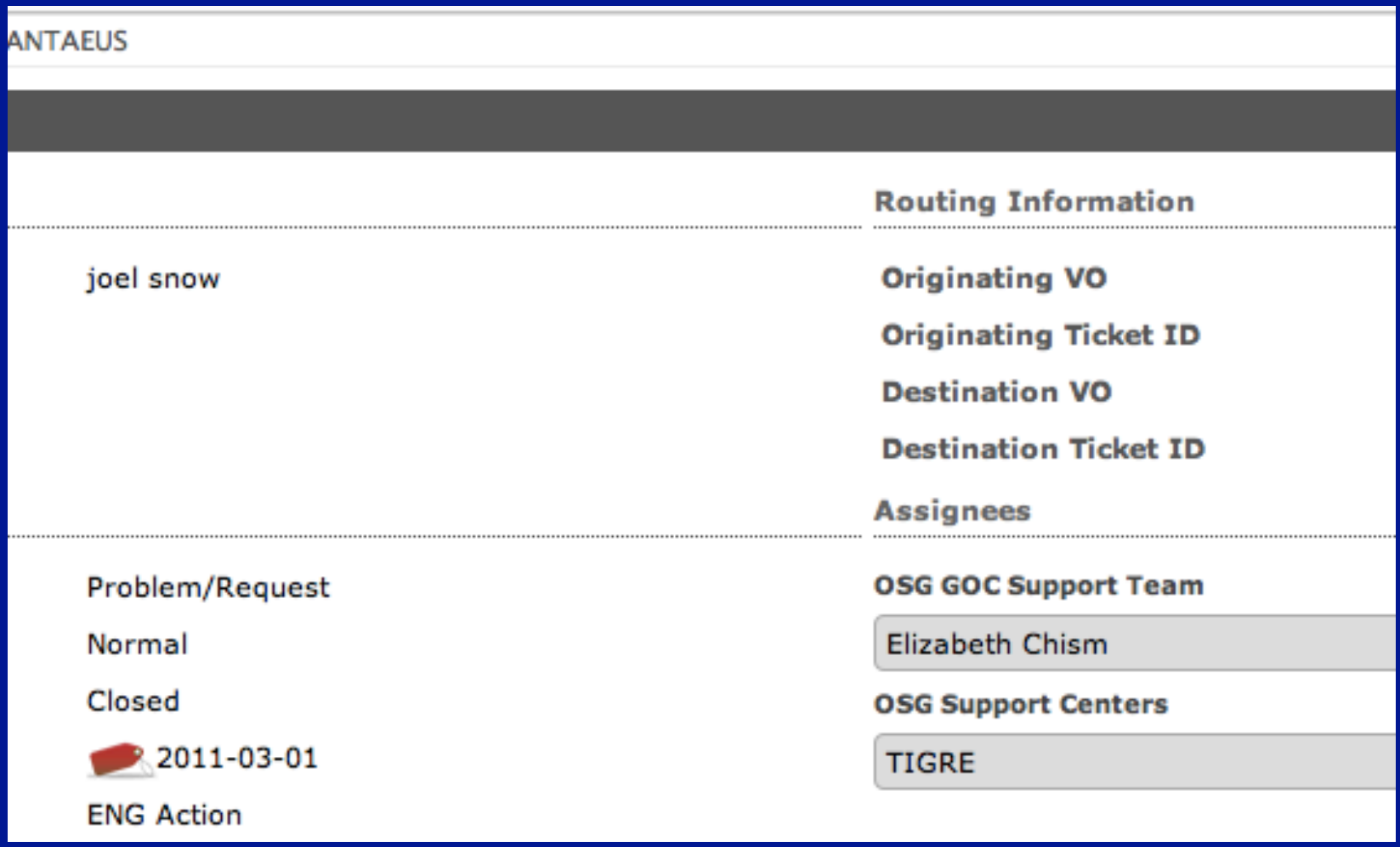
Jose and Horst continue to be active contacts in OSG.

Led by Tim Cartwright

Reminder of the Organization



OSG Operations handled its 10,000th ticket



Led by Rob Quick

Stability of Operations Services

Availability Metrics

Between Feb 5, 2011 00:00:00 UTC and Mar 7, 2011 00:00:00 UTC

Resource	Service	Availability	Reliability
GRATIA-OSG			OSG Production Resource Group
GRATIA-OSG-ITB	Gratia Collector	99.56%	99.56%
GRATIA-OSG-PROD	Gratia Collector	98.58%	98.58%
GRATIA-OSG-TRANSFER	Gratia Collector	99.42%	99.42%
GOC_BDII			OSG Production Resource Group
GOC_BDII_1	OSG BDII	99.98%	99.98%
	WLCG Interoperability BDII	99.63%	99.63%
GOC_BDII_2	OSG BDII	100%	100%
	WLCG Interoperability BDII	100%	100%
GOC_OIM			OSG Production Resource Group
GOC_OIM	OIM	99.95%	99.95%
GOC_RSV_Collector			OSG Production Resource Group
			99.7%
			on Resource Group
			99.72%
			99.95%
			on Resource Group
			99.77%
			99.95%
MyOSG			OSG Production Resource Group
MyOSG_1	MyOSG	99.7%	99.7%
MyOSG_2	MyOSG	99.91%	99.91%
OSG_Display			OSG Production Resource Group
OSG_Display_1	OSG_Display	97.87%	97.87%
OSG_TWiki			OSG Production Resource Group
OSG_TWiki	Twiki Server	99.81%	99.81%

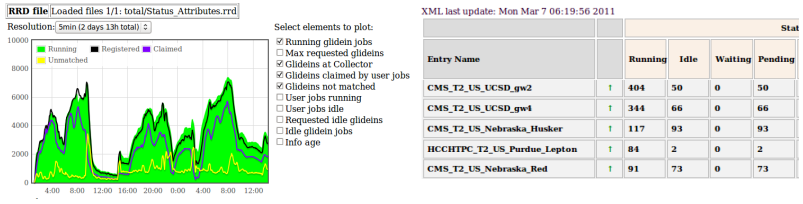
Information Services

OSG BDII	99.98%	99.98%
WLCG Interoperability BDII	99.63%	99.63%
OSG BDII	100%	100%
WLCG Interoperability BDII	100%	100%

Addition of OSG Pilot/Glidein Factory Operations Service

What We Do

- Site debugging
  - We have a set of monitoring tools to ensure glideins are running as expected
  - If we see something is wrong we open GOC tickets and work closely with the site to debug



XML last update: Mon Mar 7 06:19:56 2011

		Stat			
Entry Name		Running	Idle	Waiting	Pending
CMS_T2_US_UCSD_gw2	↑	404	50	0	50
CMS_T2_US_UCSD_gw4	↑	344	66	0	66
CMS_T2_US_Nebraska_Husker	↑	117	93	0	93
HCCHTPC_T2_US_Purdue_Lepton	↑	84	2	0	2
CMS_T2_US_Nebraska_Red	↑	91	73	0	73

UCSD Factory Statistics

- Currently active VO Frontends:
  - CMS Analysis
  - HCC (Nebraska)
  - SBGrid / NeBioGrid
  - CDF
  - GlueX
  - UCSD Campus Grid (under Engage)
  - GLOW (Wisconsin)
  - nanoHUB

Software Infrastructure

Delivering increased % of software in native packaging – RPMs. Improved configuration management, availability probe infrastructure along the way.

Updated, releasing, testing, supporting new versions and new components regularly. Storage Discovery Tools, Hadoop, Xrootd, Bestman2, "CREAM coming", GT5.

Incorporate changes for "High Throughput Parallel Computing"

PANDA based automated test harness on Integration Testbed.

Led by Alain Roy

Documentation appreciates & needs Community Contributions

- We reviewed, tested and improved 200 Twiki documents since April 2010, that are 70% of all relevant documents. We thank all of you, who have contributed to this effort!
- We received good feedback from Twiki users regarding individual documents (example: Compute Element Installation Guide), but documents are still hard to find due to lack of good navigation.
- The documents were integrated into a navigational tree, which is being implemented and tested on twiki-docteam - Access to documents will get easier soon!
- The Content Management Project is a *Community Project* and the support from the *OSG Community* could still improve! Resource Administrators, Software Developers and Grid Users should take

Led by Robert Engel

Project Statistics 02/11

Document Area	Documents	Released	Progress [%]
Compute Element	78	32	41
General	51	47	92
Storage	40	36	90
Security	35	35	100
Tier3	26	18	69
Operations	25	8	32
Virtual Organization	22	3	14
User	20	19	95
Integration	4	4	100
All	301	202	70

Protection & ID Management

Maintained our security protections. Daily communication with WLCG, TeraGrid, etc

Working with CILOGON to enable Campus based Shibboleth identities on OSG resources.

Steps achieved in simplication of request, supply, use the ID certificates.

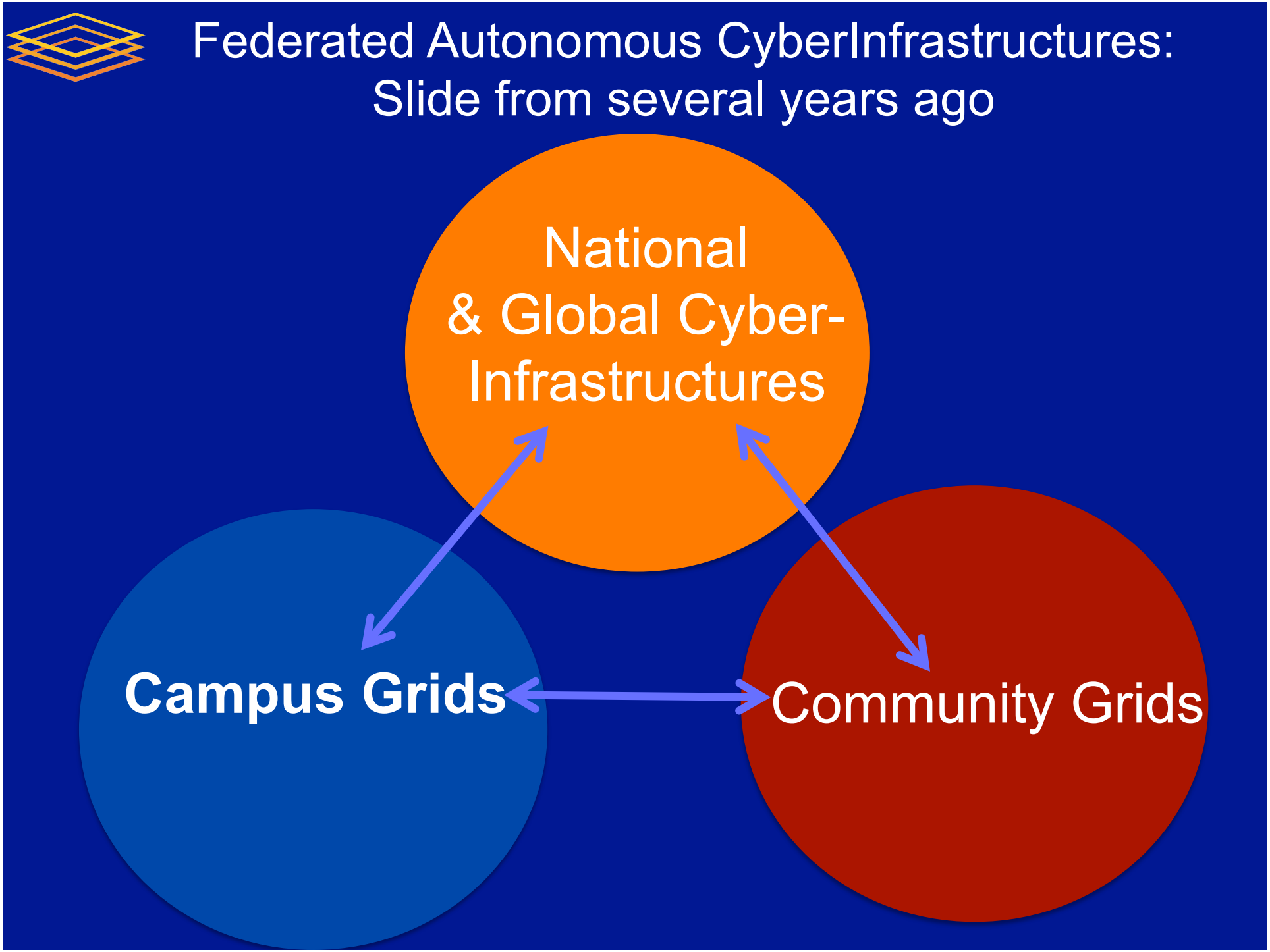
Continued with Security training – pleased to see Community led security session at this weeks CMS Tier-3 meeting – thanks Will.

Led by Mine Altunay

HTPC Focus this past year

HTPC presented past few days and later talks today..
From annual report:

High Throughput Parallel Computing (HTPC) as identified in the original proposal consists of submitting ensembles of small way parallel jobs to suitable resources. As such it is an emerging paradigm that combines the benefits of High Throughput Computing with small way parallel computing. One immediate benefit is that parallel HTPC jobs are far more portable than most parallel jobs since they do not depend on the nuances of parallel library software versions and machine specific hardware interconnects. For HTPC, parallel libraries are packaged and shipped simultaneously with job. This pattern allows for two additional benefits: First, there are no restrictions as to the method of parallelization, these can be MPI, OpenMP, Linda, or utilize other parallelization methods. Second, the libraries can be optimized for on-processor communication so that these jobs can run optimally on Multi-core hardware.

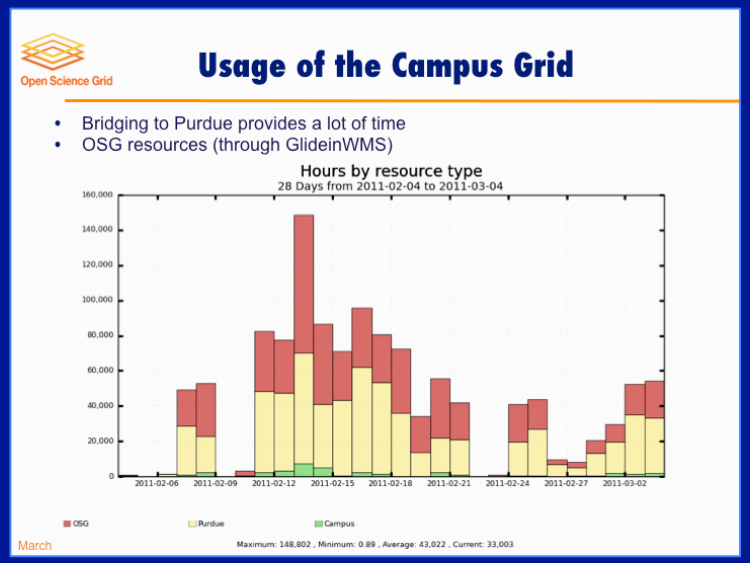

Federated Autonomous CyberInfrastructures:
Slide from several years ago
National
& Global Cyber-
Infrastructures
Campus Grids
Community Grids

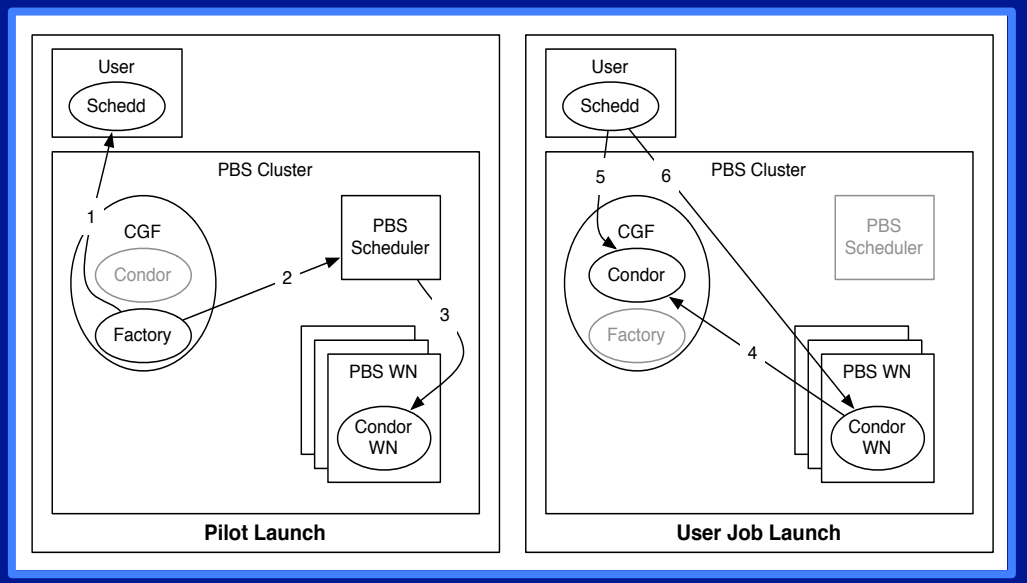
FY11 More Focused Effort on

Wisconsin, Nebraska, Purdue, NotreDame, Oklahoma, Langston

Led by Dan Fraser

Campus Infrastructures

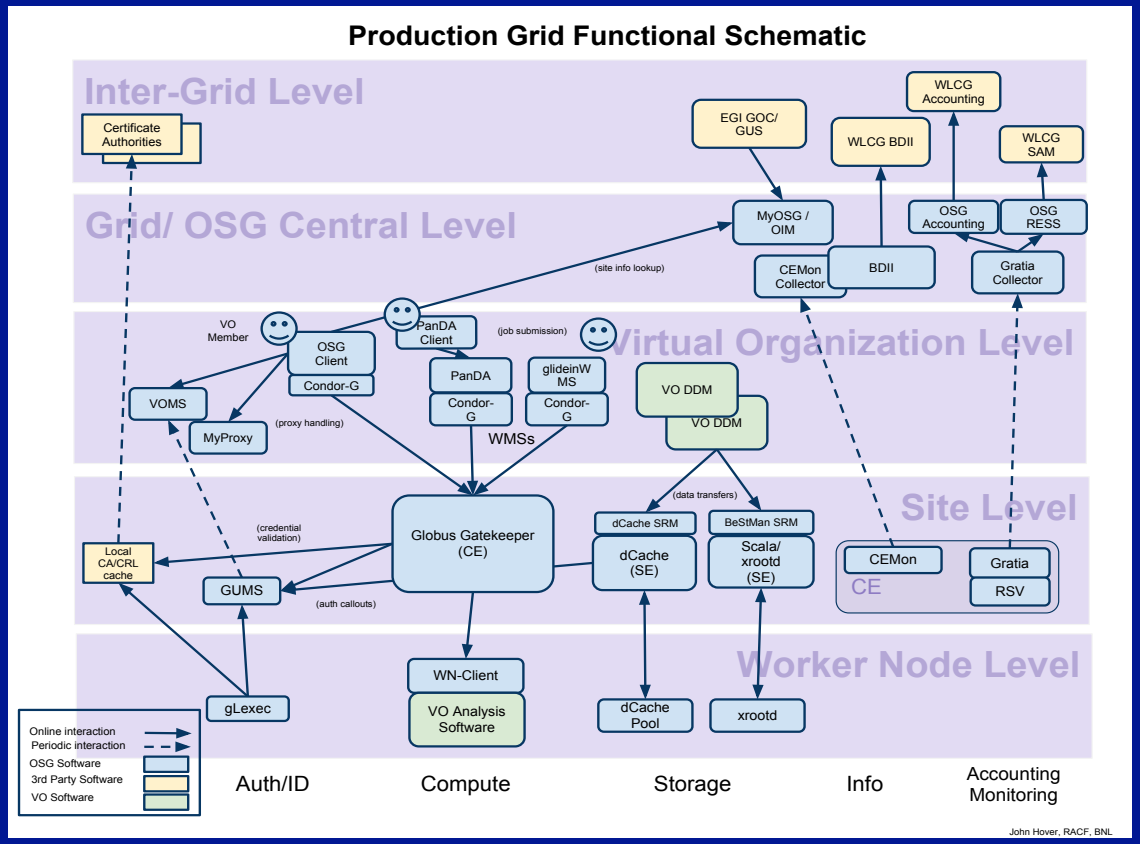




Extended Blueprint Reference

2 Production Architectures:

Production Grid
&
Campus Infrastructure



Much work from Brian Bockelman, John Hover, OSG Doc 314

Partners, Satellites, Supplements…

ANI & ESNET

Condor Sustainability & CorralWMS

EGI & EMI

ExTENCI & TeraGrid /XD

GLOBUS/CDIGS/ CEDPS

High Throughput Parallel Computing (HTPC)

Internet2, DYNES

WLCG

OSG and the Future:

Sustain and Extend

DOE HEP 1 Year extension, 4/11-3/12. for US LHC ++

Preparatory Workshops and Reports

Proposals

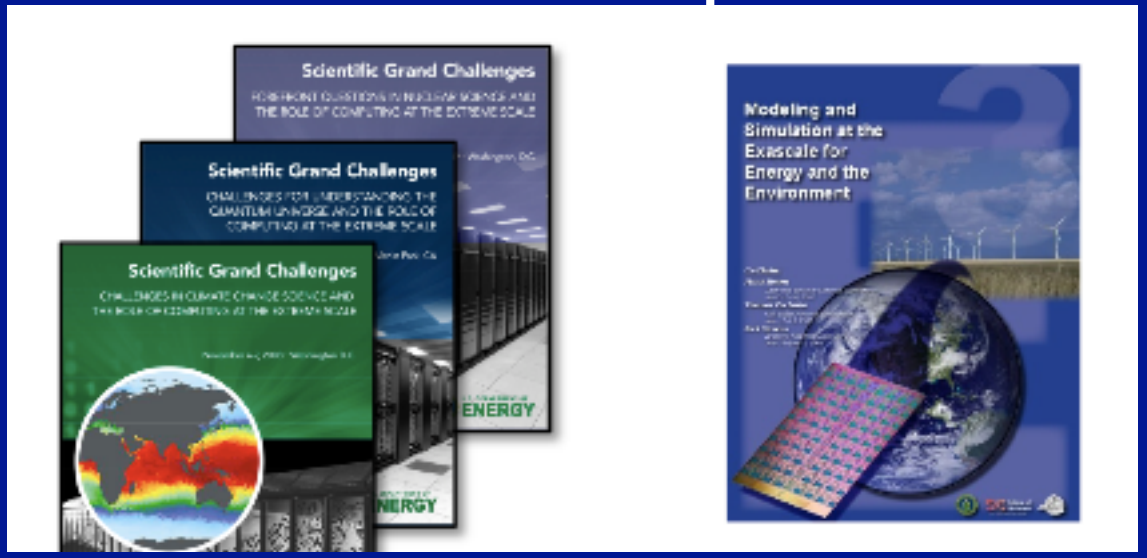
Participating in Preparatory Workshops..

Community - WLCG data management jamboree, thinking towards some more targeted international co-working.

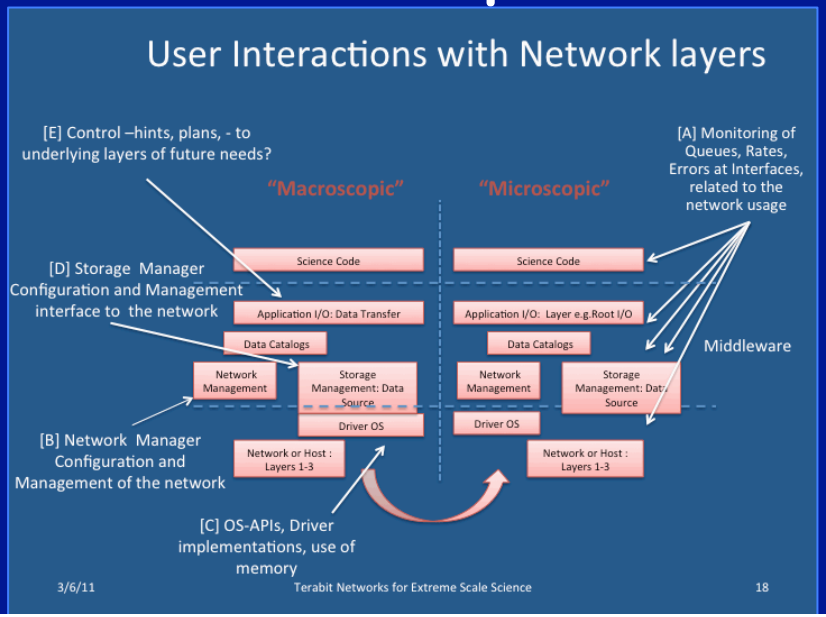
DOE - Exascale requirements, ESNET requirements gathering, Advanced Networks, KBASE, ..

NSF - Security and Software Infrastructure Institutes. Campus Bridging, Combustion, NEES data ..

Exascale reports



TeraBit Networks report being prepared



Report from the Workshops on Distributed Computing, Multidisciplinary Science, and the NSF's Scientific Software Innovation Institutes Program

October 2010

3 Outcome – Recommendation 1

The most significant outcome of the workshops was the vision (and key attributes) for:

> "***A US Software Infrastructure Institute that provides a national center of excellence for community based software architecture, design and production; expertise and services in support of software life cycle practices; marketing, documentation and networking services; and transformative workforce development activities.***"

The measure of success of such an institute should be the cost effectiveness (as measured in scholarly work) of our software infrastructure. This will be accomplished by a thriving and

The Institute will play a unique role by addressing organizational and life cycle elements not covered by infrastructure implementation and deployment projects that are driven by scientific objectives or technological trends.

4 Supporting Outcomes

To implement that vision attendees made the following more specific recommend

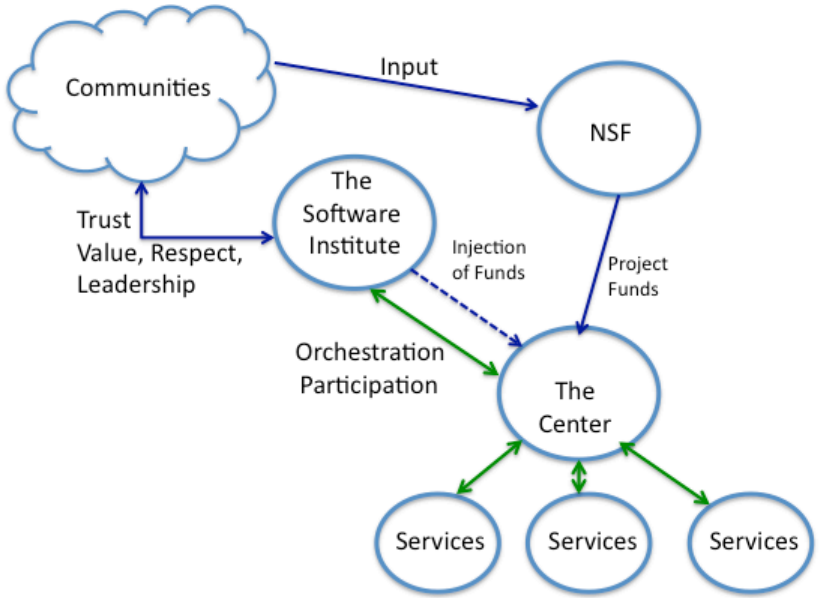
Recommendation 2. The Software Infrastructure Institute, together with its asso of excellence, should be **structured as an (Virtual) Organization** that provides a coordinated suite of high quality and dependable services that address the entire distributed software infrastructure.

Recommendation 3. Services that advance the quality, adoption, and the longev distributed software infrastructure should be provided by **teams of experts locat institutions with demonstrated leadership** in the areas of the service and stror to the sustainability of the service. While we envision a distributed multi-layer stru services, we argue that these services need to be managed by a well-organized a respected team that can provide and sustain leadership in the complex and rapid area of distributed software infrastructure.

Recommendation 4. Quality must be the guiding principal for the services provic with accountability that is based on **independent quantitative impact assessm** levels, allocation of effort and resources must be based on a professional and tra ranking of impact and cost.

Recommendation 5. The funding model of the Institute must allow **quick turn-a requests to fund short-term projects** that address critical deficiencies in the Sc Infrastructure that powers our science and research enterprise. The Institute mus quickly direct effort to a critical need. The Institute should adopt the metaphor of "supercomputer centers" in terms of assignment and allocation of the human resc to deliver the services to the DSI community.

Recommendation 6. The scope of the Institute must be to **offer software life cy and services** for distributed software infrastructure for a broad range of NSF programs, technology developers, and academic communities. The Institute must aim to improve the "accountability" in our software infrastructure enterprise by providing a recognized center of



Campus Bridging, CF21, CIF21, 1 of 6 ACCI Taskforces

http://pti.iu.edu/campusbridging/

The NSF should establish a national CI software roadmap. Through the Software Infrastructure for Sustained Innovation (SI^2) or other programs, the NSF should seek to systematically fund the creation and ongoing development and support of a suite of critical cyberinfrastructure software that identify and establish this roadmap, including CI software for authentication and access control; computing cluster management; data movement; data sharing; data, metadata, and provenance management; distributed computation / cycle scavenging; parallel computing libraries; network performance analysis / debugging; VO collaboration; and scientific visualization.

- The NSF should continue to invest in campus cyberinfrastructure through programs such as the Major Research Infrastructure (MRI) program, and do so in ways that achieve goals set in the Cyberinfrastructure Vision for 21st Century Discovery and a national CI software roadmap.

Strategic Recommendation to the NSF #3: The NSF should create a new program funding high-speed (currently 10 Gbps) connections from campuses to the nearest landing point for a national network backbone. The design of these connections must include support for dynamic network provisioning services and must be engineered to support rapid movement of large scientific data sets.

Strategic Recommendation to the NSF #4: The NSF should fund national facilities for at least short-term

Tactical Recommendation to the NSF #1:The NSF should fund the TeraGrid eXtreme Digital program, as currently called for in existing solicitations, and should continue to fund and invest in the Open Science Grid.

Tactical recommendation to the NSF #2: The NSF should commission a study of current reward structures and recommendations about the reward structure – particularly as regards promotion and tenure for faculty – that would better align reward structures as perceived by individual faculty members with the type of large, collaborative virtual organizations that the NSF asserts are required for successful approaches to pressing, large scale scientific problems and transformative research.

Tactical Recommendation to the NSF #3: The NSF should support joint efforts with organizations such as

Proposal to NSF 2011-2016

Submitted this past week.

..to transform the science and research computing landscape on our campuses through wide adoption of a new generation of DHTC technologies that support access to "any data, anytime, anywhere" to an expanded set of job and data services via a single identity, and enable the transformation of our core stakeholders computing capabilities from petascale to exascale .

Next Proposal?

DOE SciDAC-3 call from ASCR for Institutes.

- Tools and resources for lowering the barriers to effectively use state-of-the-art computational systems;
- Mechanisms for taking on computational grand challenges across different science application areas;
- Mechanisms for incorporating and demonstrating the value of basic research results from Applied Mathematics and Computer Science; and
- Plans for building up and engaging our nation's computational science research communities.

One of the primary metrics for the success of the SciDAC Institutes is the extent to which its deliverables are used by application scientists. An equally important metric is the extent to which Institute researchers actively collaborate and leverage their expertise in achieving that success.