

Open Science Grid

Linking Universities and Laboratories in National
Cyberinfrastructure



DOSAR Workshop
University of Texas, Arlington
Arlington, Texas
March 30, 2006

Paul Avery
University of Florida
avery@phys.ufl.edu

OSG Roots: “Trillium Consortium”

- Trillium = PPDG + GriPhyN + iVDGL
 - ◆ PPDG: \$12M (DOE) (1999 – 2006)
 - ◆ GriPhyN: \$12M (NSF) (2000 – 2005)
 - ◆ iVDGL: \$14M (NSF) (2001 – 2006)
 - ◆ Large science experiments (HEP/LHC, LIGO, SDSS)
- Total ~150 people with many overlaps between projects
 - ◆ Universities, labs, foreign partners
- Historically, a strong driver for funding agency collaboration
 - ◆ Inter-agency (NSF – DOE) + intra-agency (Directorate – Directorate)
- Coordination vital for meeting broad goals
 - ◆ CS research, developing/supporting Virtual Data Toolkit (VDT)
 - ◆ Multiple Grid deployments, using VDT-based middleware
 - ◆ Deployment of Grid3, a general purpose, national Grid
 - ◆ Unified entity when collaborating internationally

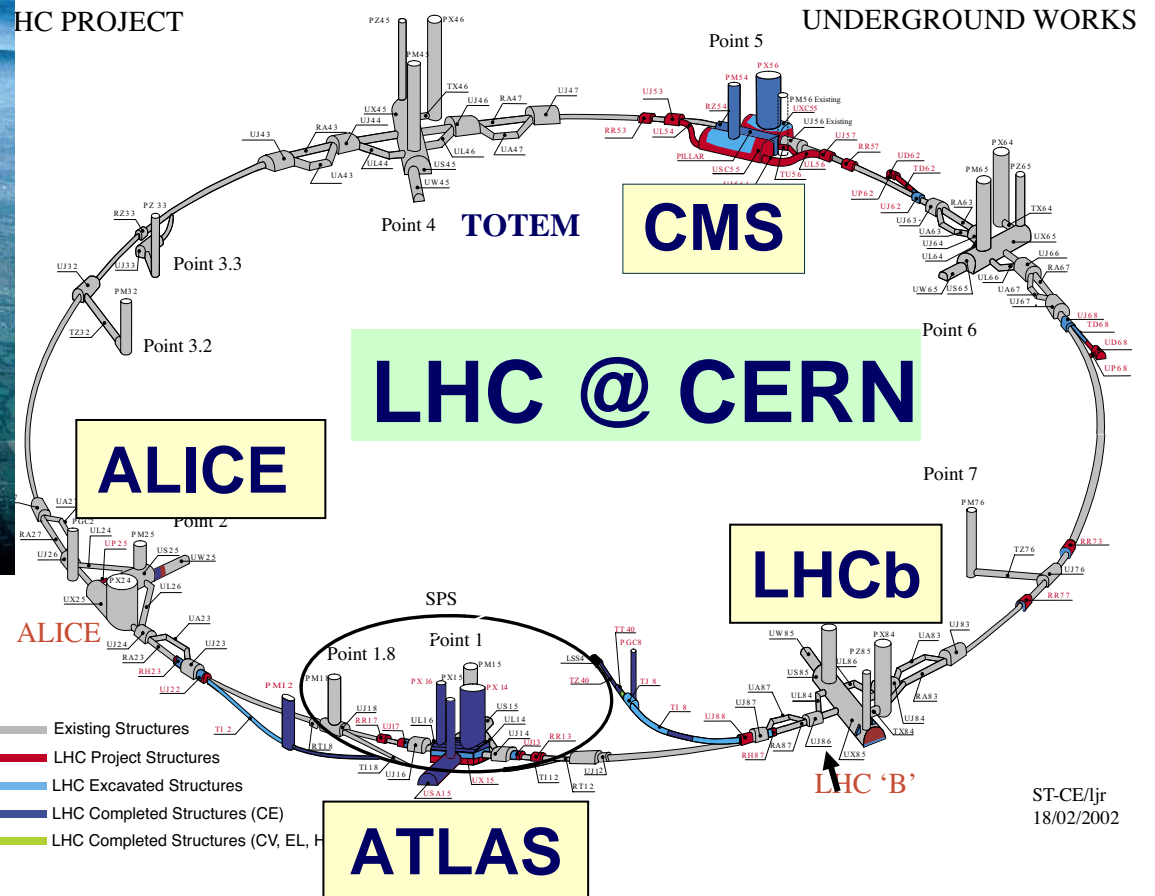




Scale of OSG Resources & Services Set by Large Hadron Collider (LHC) Expts.



★ 27 km Tunnel in Switzerland & France



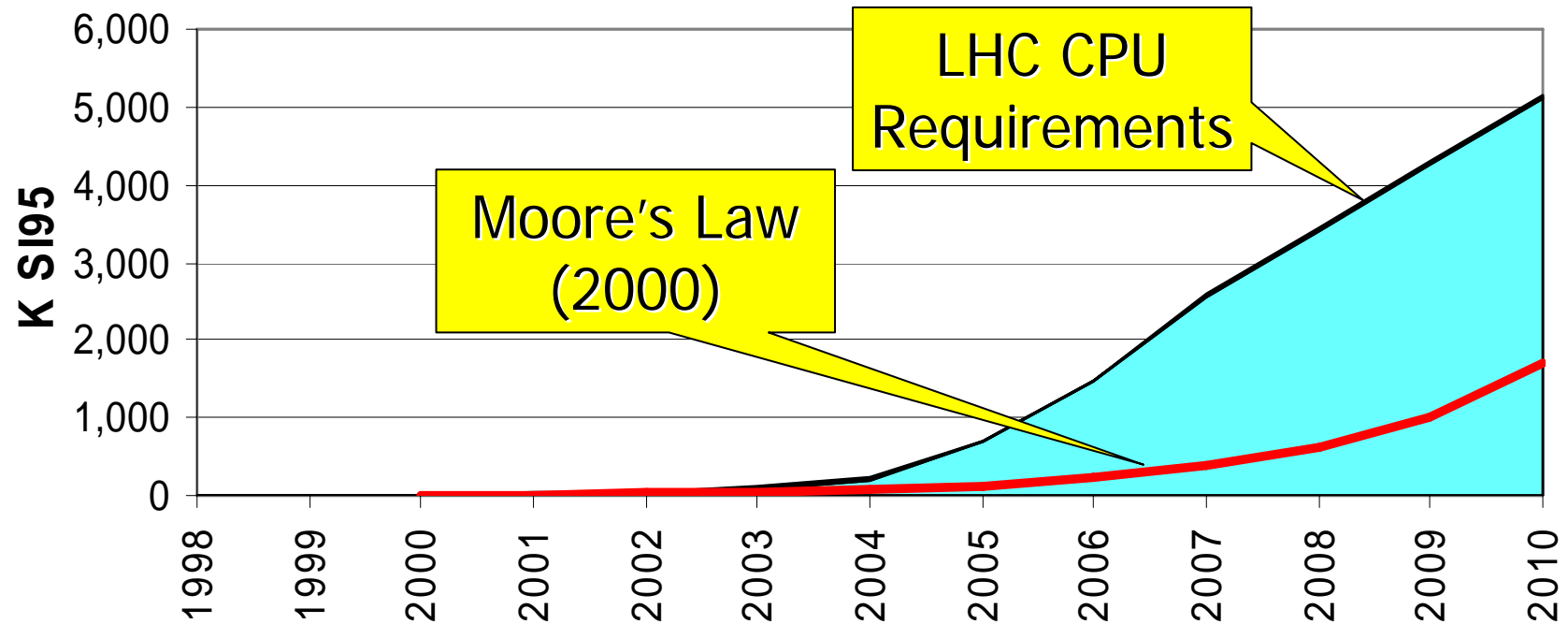
Search for

- Origin of Mass
- New fundamental forces
- Supersymmetry
- Other new particles
- 2007 – ?

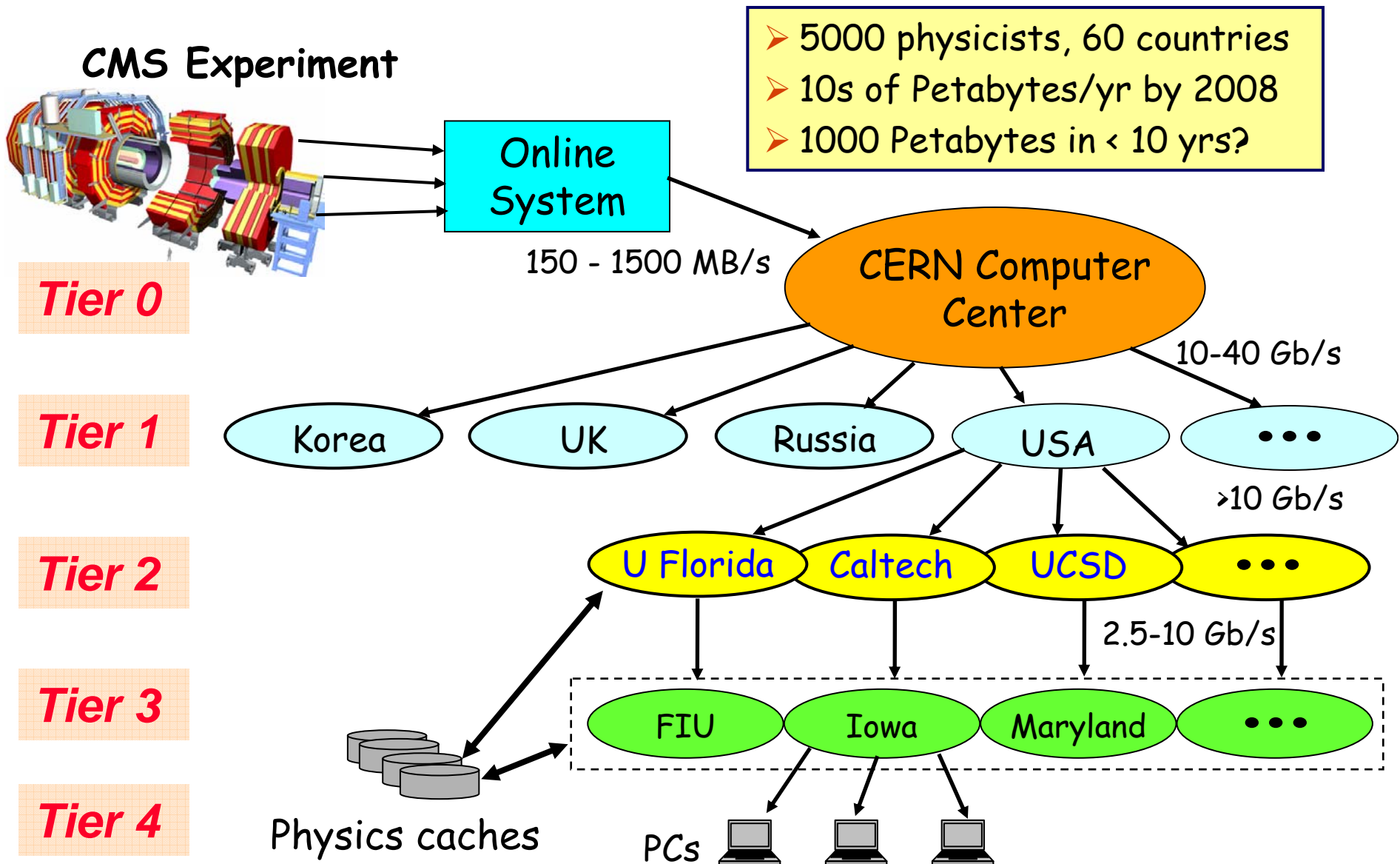
LHC: Beyond Moore's Law

Estimated CPU Capacity at CERN

1K SI95 = 10 Intel CPU (2 GHz)

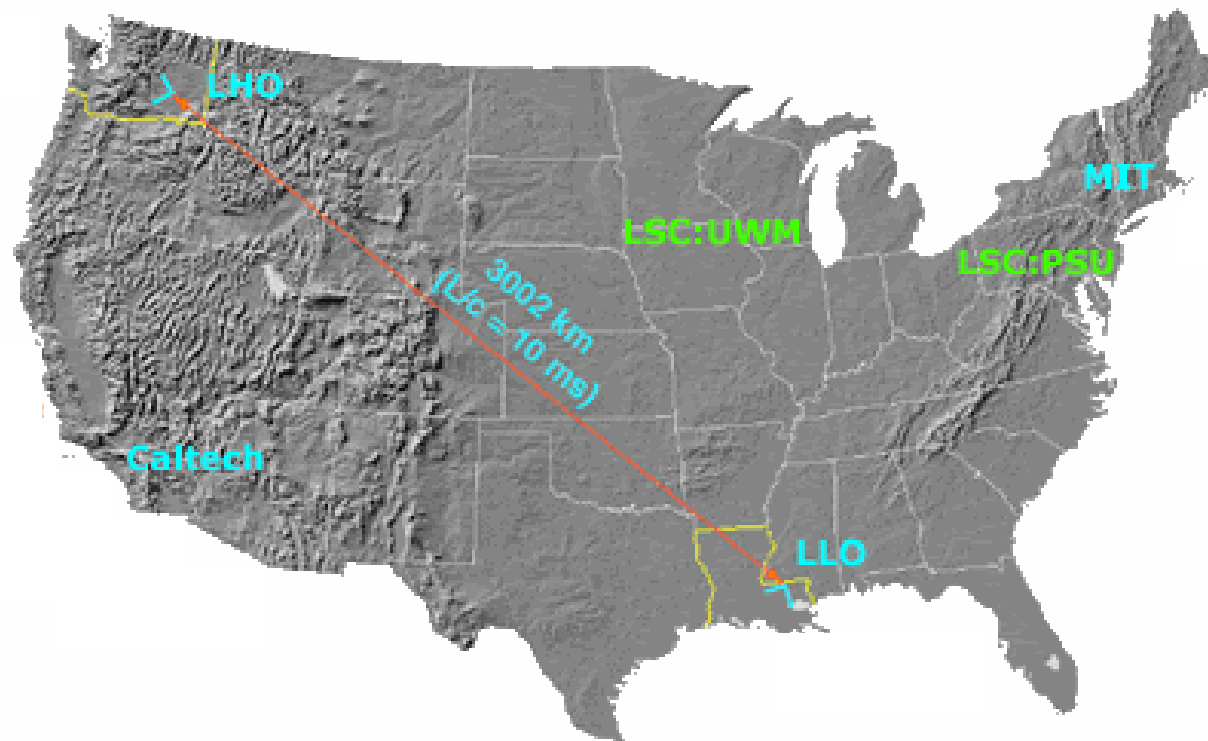


LHC Global Data Grid (2007+)



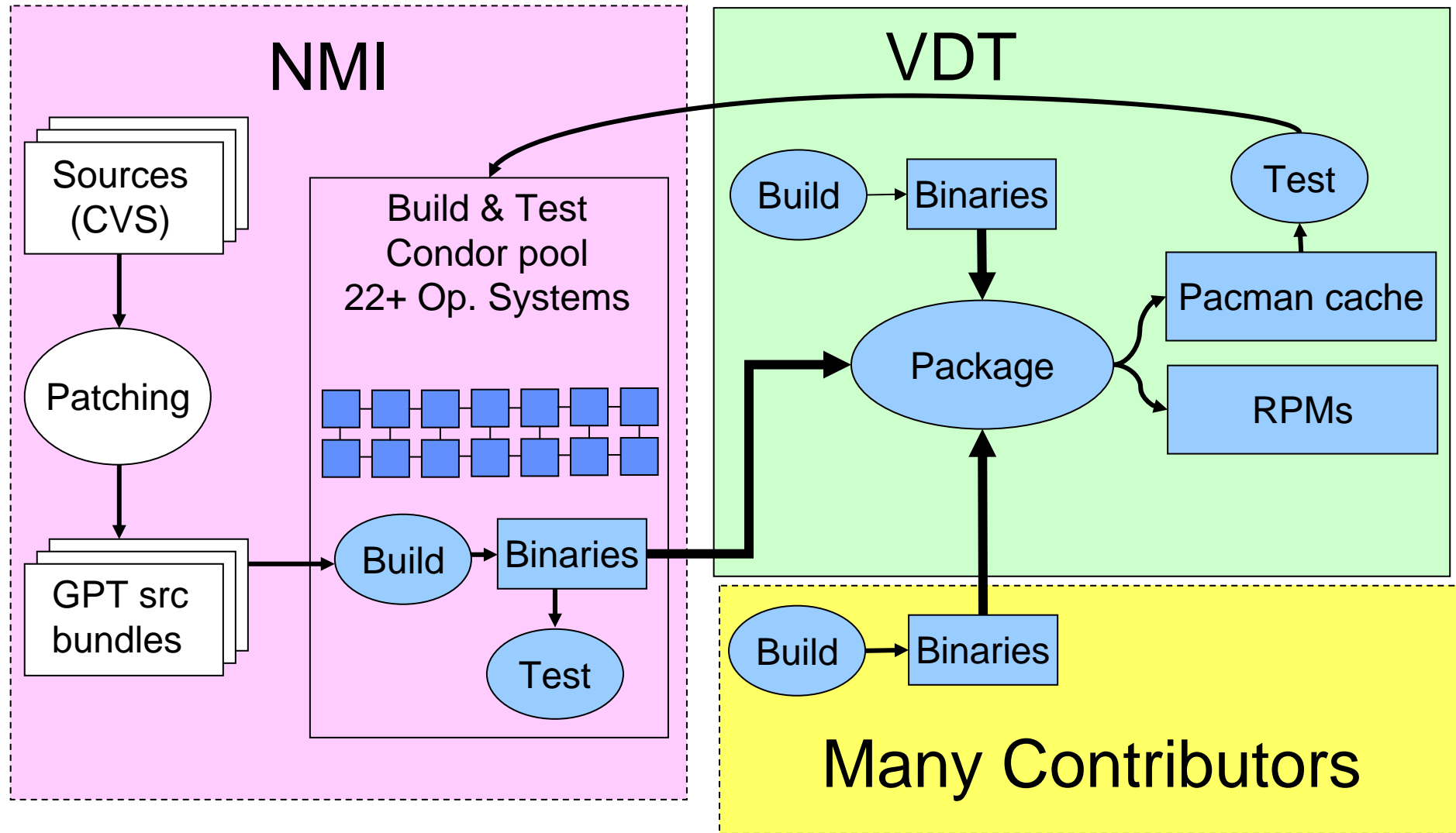
LIGO Grid

- LIGO Grid: 6 US sites + 3 EU sites (UK & Germany)



- * LHO, LLO: LIGO observatory sites
- * LSC: LIGO Scientific Collaboration

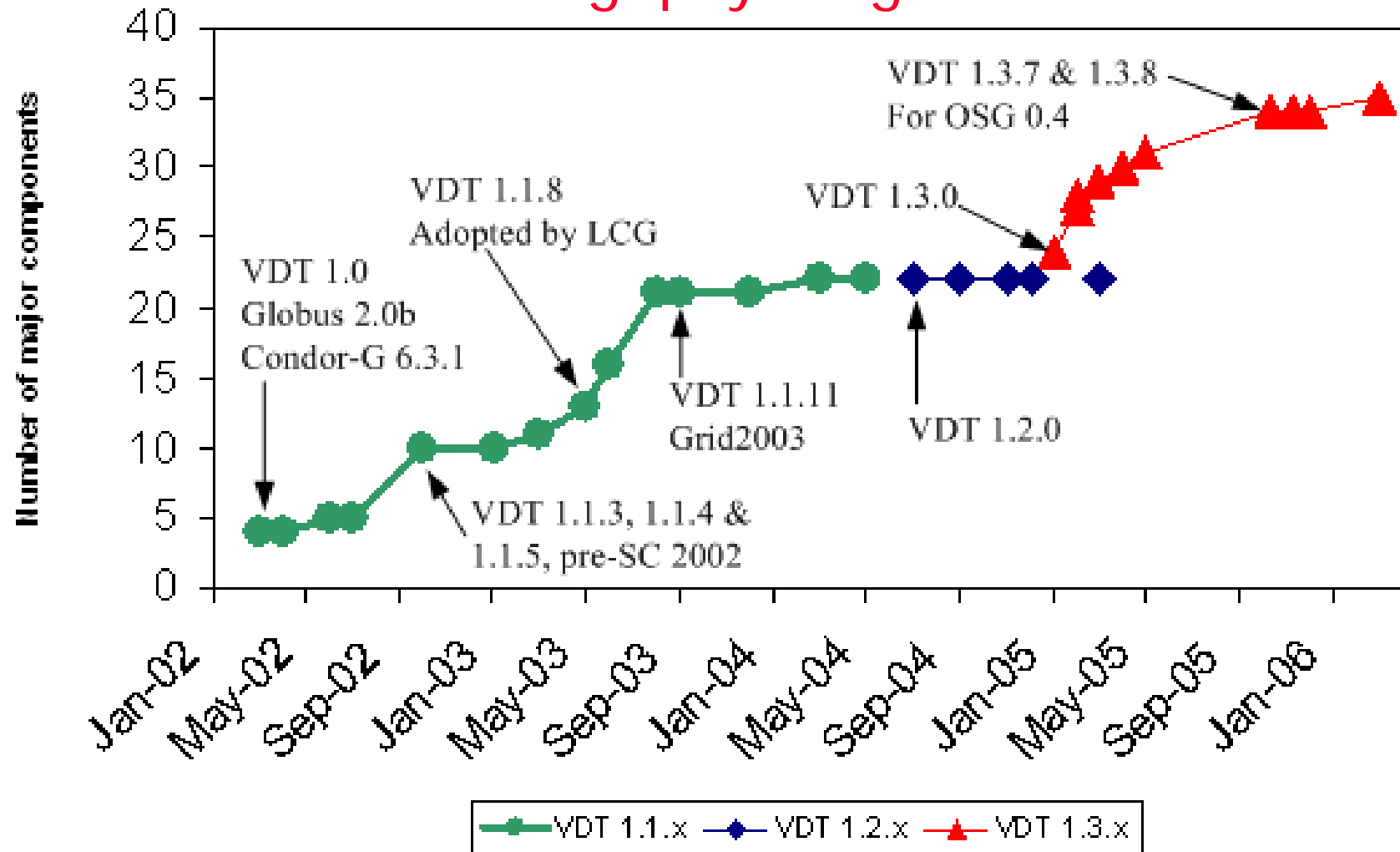
Common Middleware: Virtual Data Toolkit



VDT: Package, test, deploy, support, upgrade, troubleshoot

VDT Growth Over 4 Years (1.3.10 now)

www.griphyn.org/vdt/



Grid3: A National Grid Infrastructure

- October 2003 – July 2005
- 32 sites, 3,500 CPUs: Universities + 4 national labs
- Sites in US, Korea, Brazil, Taiwan
- Applications in HEP, LIGO, SDSS, Genomics, fMRI, CS



Grid3 Lessons Learned

- How to **operate** a Grid as a facility
 - ◆ Security, services, error recovery, procedures, docs, organization
 - ◆ Delegation of responsibilities (Project, VO, service, site, ...)
 - ◆ Crucial role of Grid Operations Center (GOC)
- How to **support** people \Leftrightarrow people relations
 - ◆ Face-face meetings, phone cons, 1-1 interactions, mail lists, etc.
- How to **test** and **validate** Grid tools and applications
 - ◆ Vital role of testbeds
- How to **scale** algorithms, software, process
 - ◆ Some successes, but “interesting” failure modes still occur
- How to **apply** distributed cyberinfrastructure
 - ◆ Successful production runs for several applications

Open Science Grid: July 20, 2005

- **VO based: Partnership of many organizations**
- **Production Grid: 50+ sites, 21,000 CPUs “present” (available but not at one time)**
- **Sites in US, Korea, Brazil, Taiwan**
- **Integration Grid: ~15 sites**



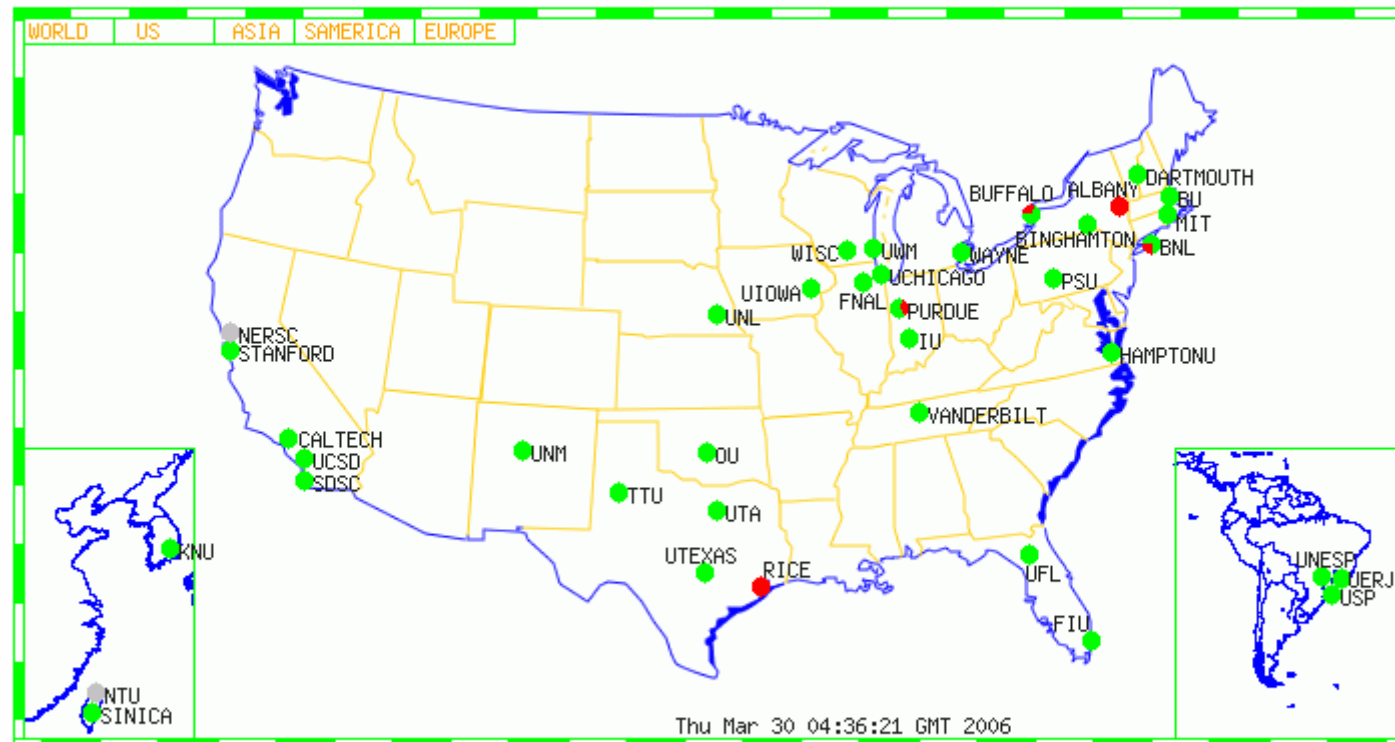
GridCat Status Monitoring



Open Science Grid

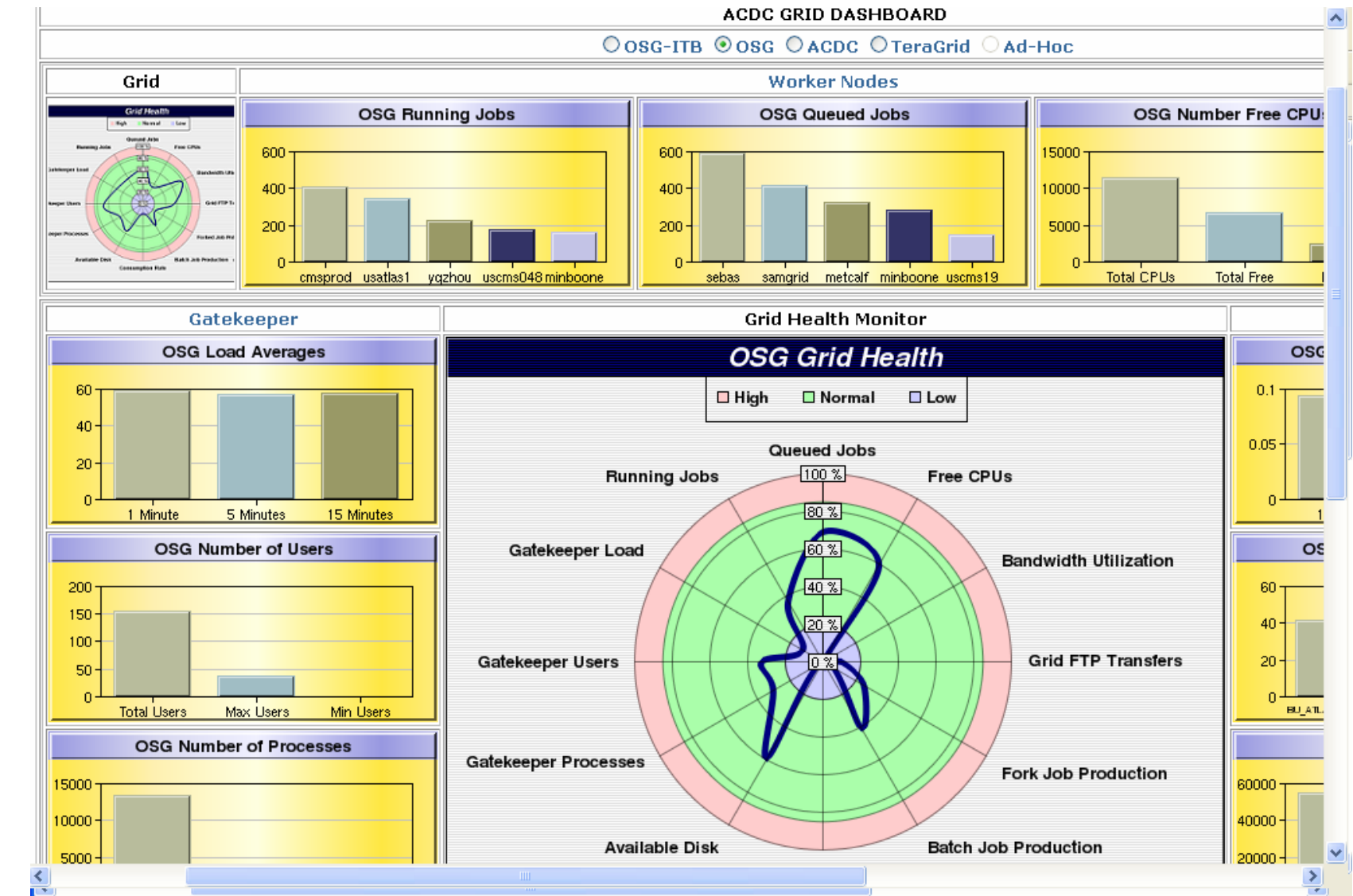
GridCat

Service: CS = Compute Service, SS = Storage Service, WS = Web Service GRAM



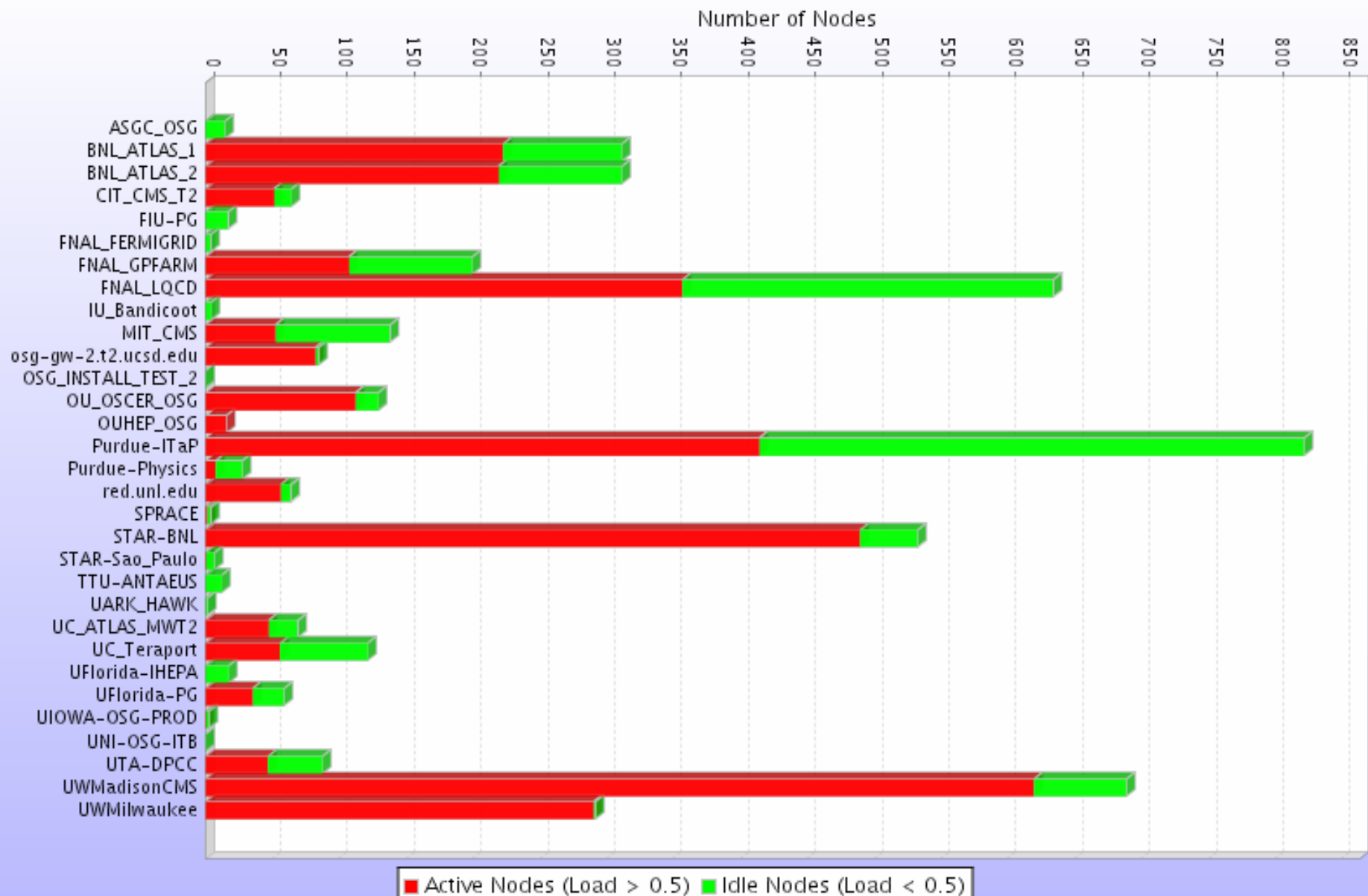
1 to 57 of 57 ◀▶ sort by: Service map: U.S.+Asia+S.America entries per page: 100 view: Summary

ACDC Grid Dashboard



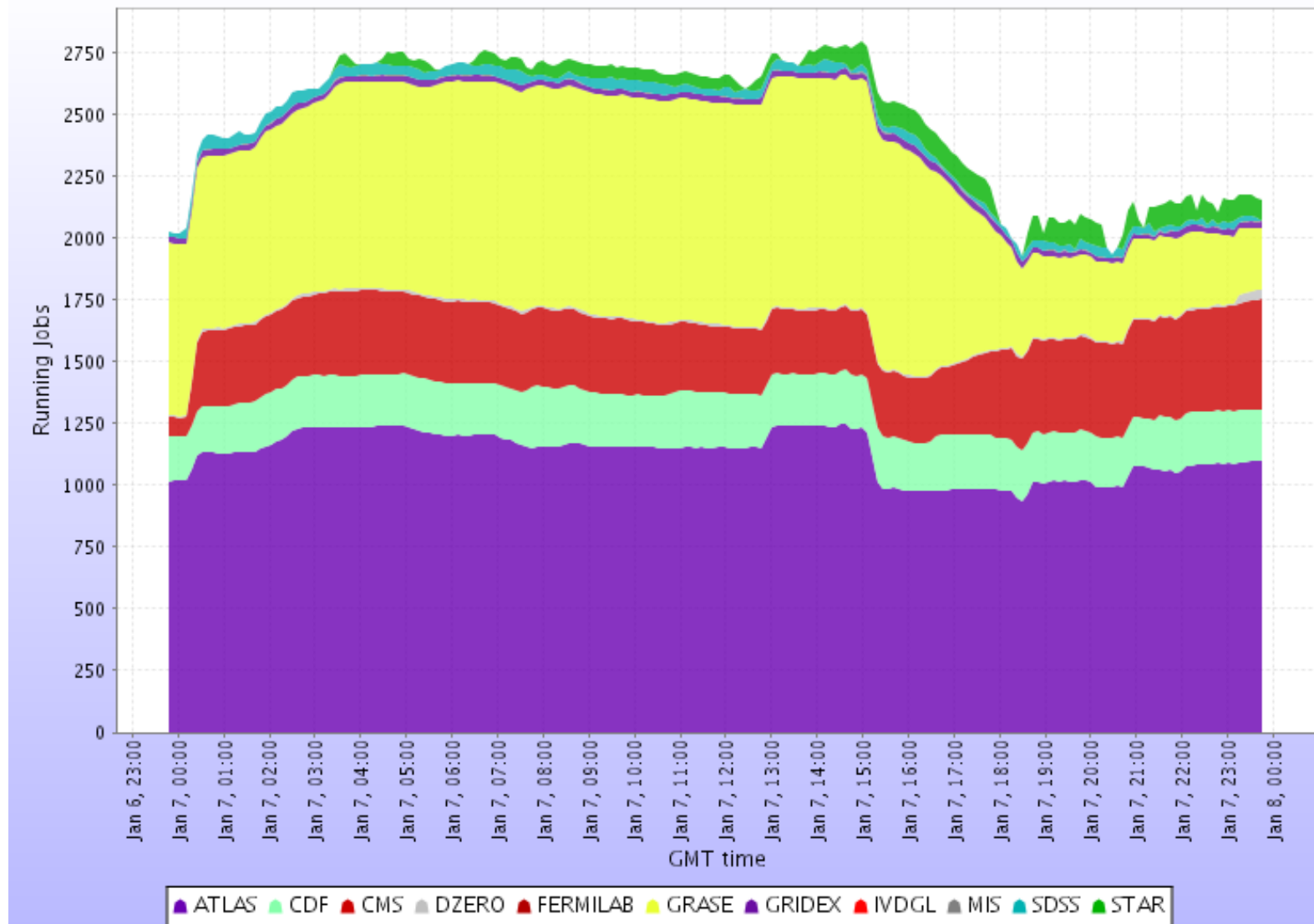
MonALISA Monitoring

Farms Usage



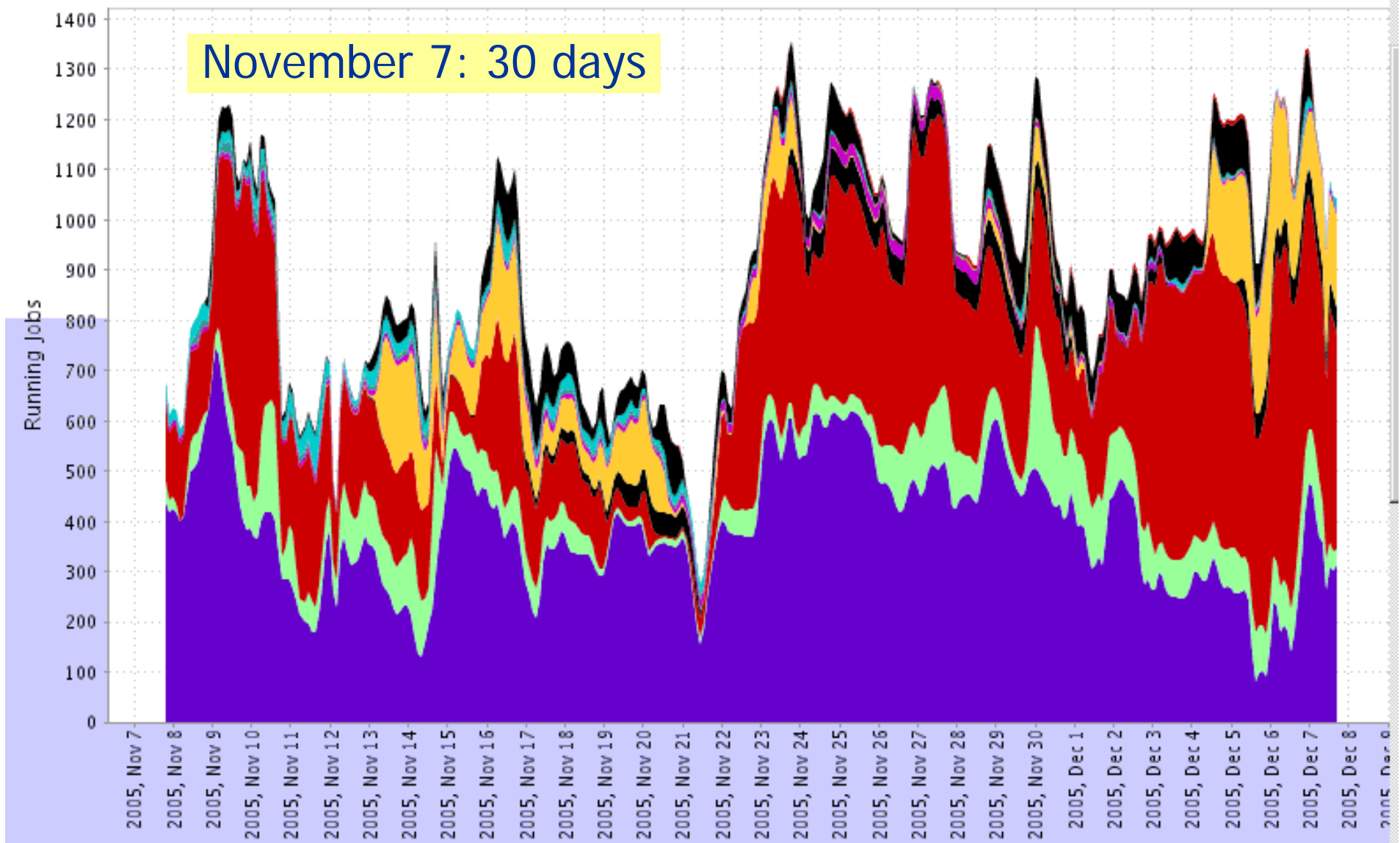
OSG Operations Snapshot (24 Hr)

Running Jobs



OSG Operations Snapshot (30 Day)

Total Jobs per VO



Registered VOs (incomplete)

VO Name	VO URL
CDF	http://www-cdf.fnal.gov
CMS	http://www.uscms.org
DES	https://www.darkenergysurvey.org/
DOSAR	http://www-hep.uta.edu/dosar/
DZero	http://www-d0.fnal.gov/
Fermilab	http://grid.fnal.gov/fermilab
fMRI	http://grid.dartmouth.edu
GADU	http://compbio.mcs.anl.gov/gaduvo/
geant4	www.cern.ch/geant4
GLOW	http://www.cs.wisc.edu/condor/glow/
GRASE	http://osg.ccr.buffalo.edu/grase
GridEx	http://www.cs.wisc.edu/condor/tools/exerciser/
GROW	http://www.uiowa.edu/~grow/
i2u2	https://fermigrd2.fnal.gov:8443/vomrs/i2u2/vomrs
iVDGL	http://www.ivdgl.org
LIGO	http://www.ligo.org
MIS	http://www.ivdgl.org/~ivdgl/mis-vo-privilege.html
nanoHUB	http://www.nanohub.org
SDSS	http://www.sdss.org
STAR	http://www.star.bnl.gov
USATLAS	http://www.usatlas.bnl.gov



Creating & Registering a VO With OSG

- To form a Virtual Organization (VO) that participates in the Open Science Grid one needs the following:
 1. a **Charter statement** describing the purpose of the VO. This should be short yet concise enough to scope intended usage of OSG resources.
 2. at least one **VO participating Organization** that is a member or partner with the Open Science Grid Consortium.
 3. a **VO Membership Service** which meets the requirements of an OSG Release. This means being able to provide a full list of members' DNs to edg-mkgridmap. The currently recommended way to do this is to deploy the VDT VOMS from the OSG software package.
 4. a **support organization** (called a Support Center in OSG parlance) that will support the VO in OSG Operations. The Support Center should provide at least the following:
 - a written description of the registration process,
 - instructions for the members of the VO on how to complete the VO registration process,
 - instructions for the members of the VO on how to report problems and/or obtain help.
 5. completion of the **registration form** located [here](#) using [these instructions](#)



Vo Support Matrix

	ACCELERATOR	ASTRO	ATLAS	AUGER	CDF	CDMS	CMS	DES	DOSAR	DTEAM	DZERO	FERMILAB	FERMILAB-TEST	FMRI	GADU	GEANT4	GLOW	GRASE	GRIDEX	GROW	HYPERCP	I2U2	ILC	INDGL	KTEV	LIGO	LQCD	MINIBOONE	MINOS	MIPP	MIS	NANOHUB	NOVA	NUMI	OSG	PA TRIOT	SDSS	STAR	THEORY	USATLAS	USCMS	UNK
grid.rit.albany.edu																																										
rommel.cs.binghamton.edu																																										
gridgk01.racf.bnl.gov																																										
gridgk02.racf.bnl.gov																																										
agt.bu.edu																																										
acdc.ccr.buffalo.edu																																										
joplin.ccr.buffalo.edu																																										
u2-grid.ccr.buffalo.edu																																										
mama.ccr.buffalo.edu																																										
tier2b.cacr.caltech.edu																																										
osg01.grid.sinica.edu.tw																																										
cmsosgce.fnal.gov																																										
fermigrid1.fnal.gov																																										
quux.fnal.gov																																										
lqcd.fnal.gov																																										
cms-xen2.fnal.gov																																										
fnpgp-osg.fnal.gov																																										
cmisp4.fnal.gov																																										
tam01.fnal.gov																																										
hercules.hamptonu.edu																																										
bandicoot.uits.indiana.edu																																										
atlas.iu.edu																																										
grid3.avidd.iu.edu																																										
pdsfgrid2.nersc.gov																																										
boomer2.oscer.ou.edu																																										
ouhep0.nhn.ou.edu																																										
grid3.aset.psu.edu																																										
grid.physics.purdue.edu																																										
osg.rcac.purdue.edu																																										
t2cms02.sdsc.edu																																										
antaeus.hpcc.ttu.edu																																										
tp-osg.uchicago.edu																																										
tier2-osg.uchicago.edu																																										
ufloridaPG.phys.ufl.edu																																										
rtgrid1.its.uiowa.edu																																										
unlcompel1.unl.edu																																										
alliance.unl.edu																																										
red.unl.edu																																										
spgrid.if.usp.br																																										
atlas.dncc.uta.edu																																										

Green: DNs are mapped to this VO and compute element [clickable]
Yellow: No DNs are supported under this VO and compute element.
Black: No information

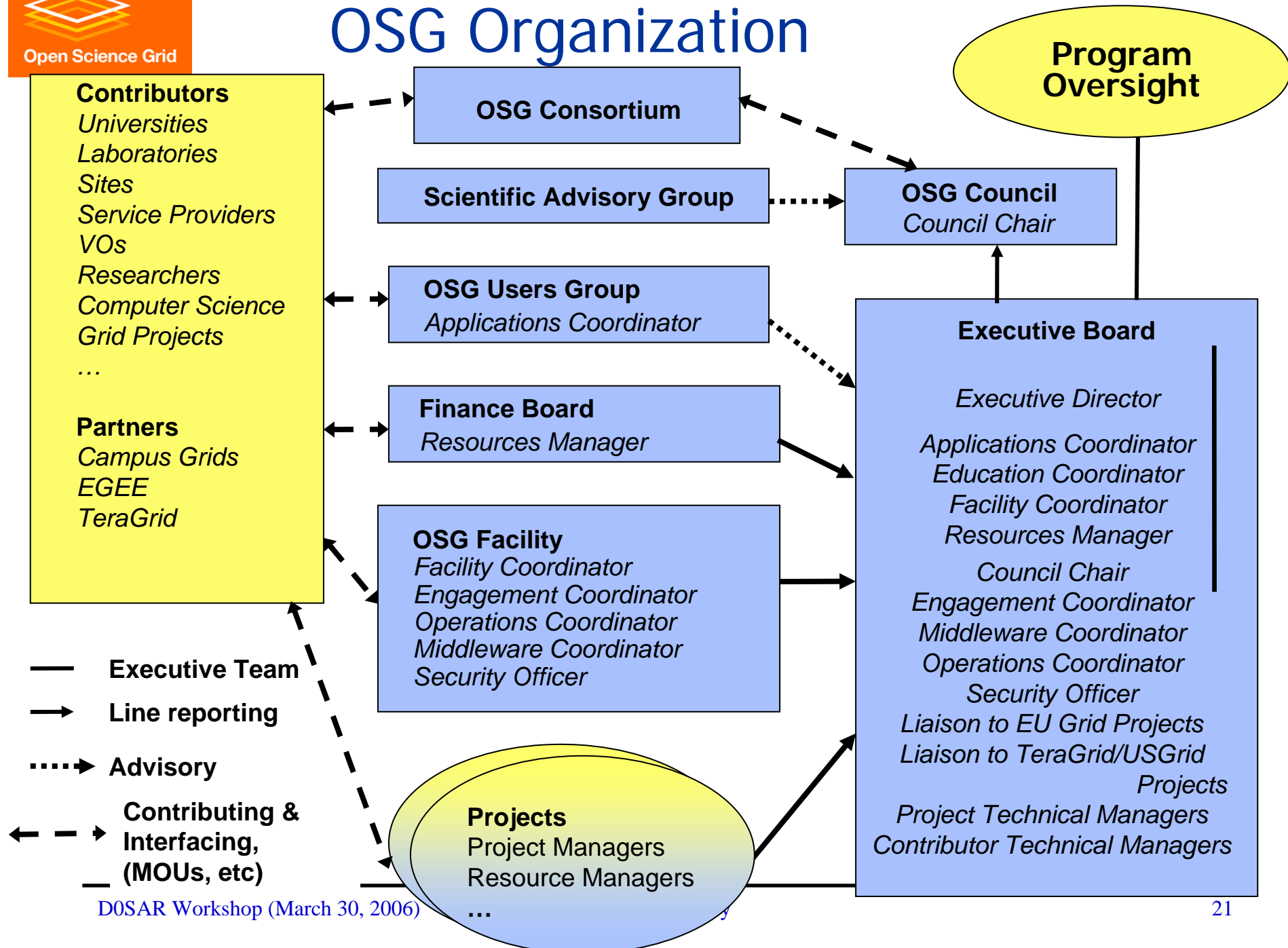
OSG Integration Testbed

- Test, validate new middleware & services
- Test, validate new applications
- Meets weekly (highly dynamic membership)

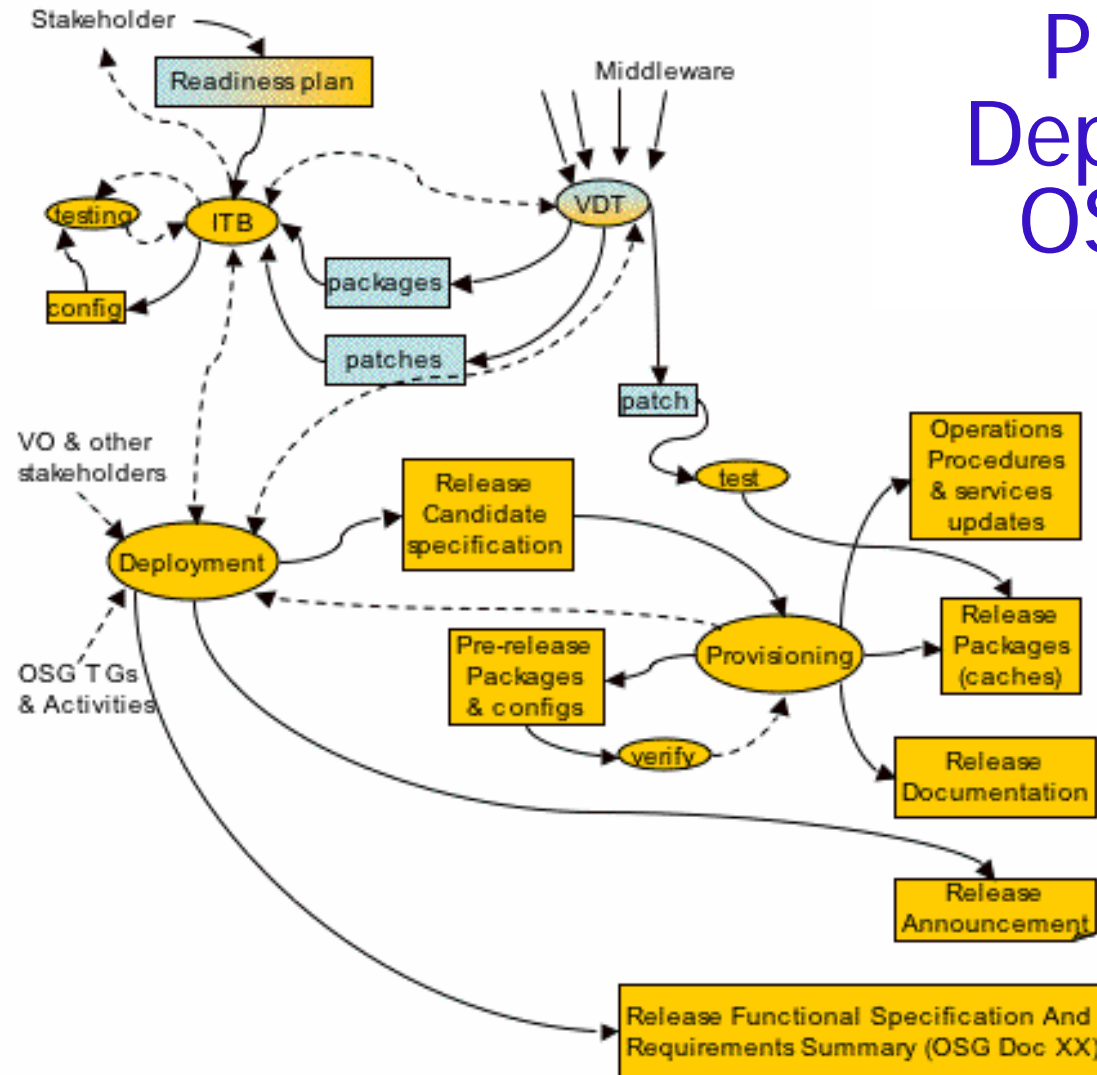




OSG Organization



Process for Deploying New OSG Service





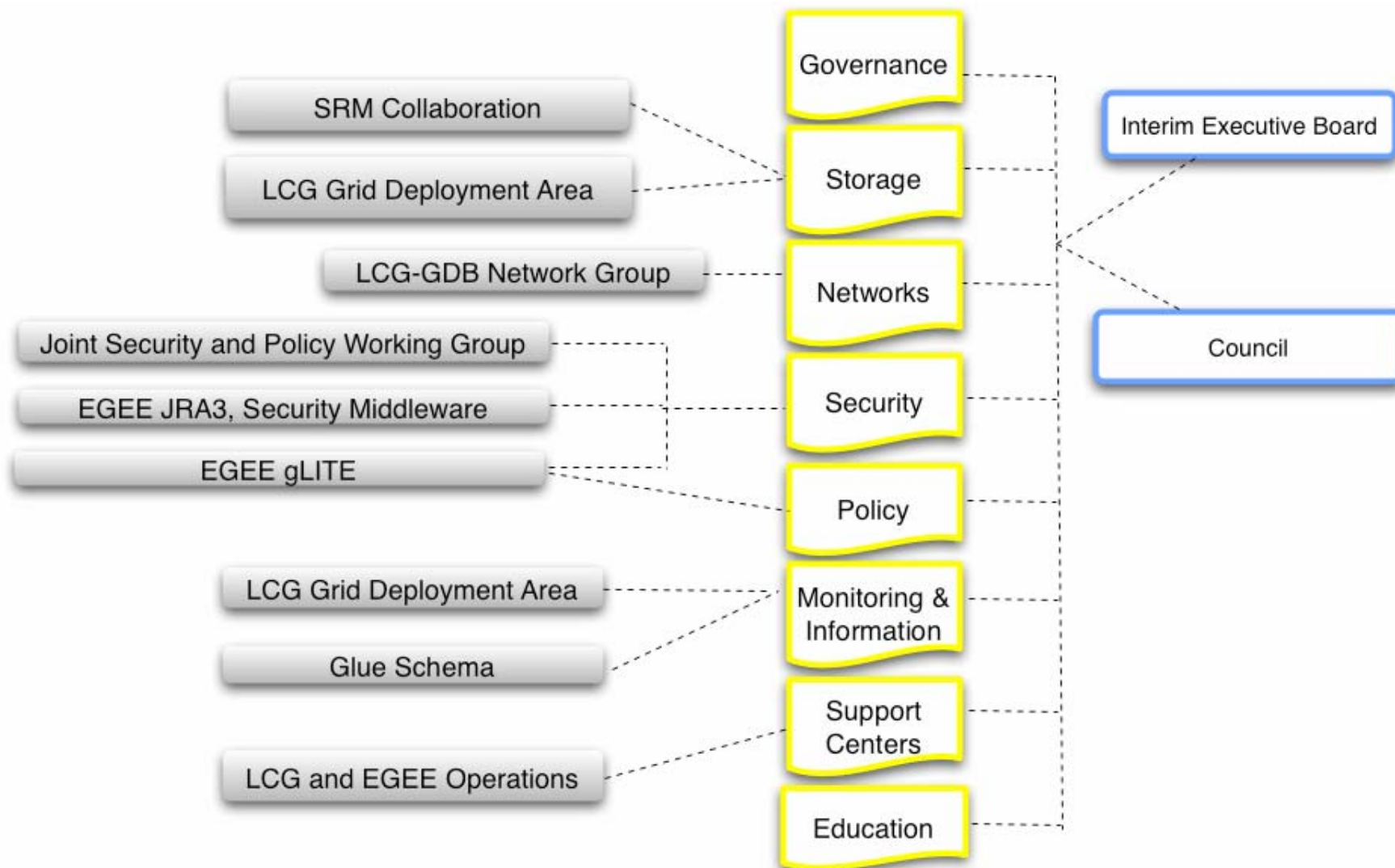
OSG Participating Disciplines

Computer Science	Condor, Globus, SRM, SRB	
Physics	LIGO, Nuclear Physics, Tevatron, LHC	Global Grids
Astrophysics	Sloan Digital Sky Survey	
Nanoscience	NanoHUB @ Purdue	
Bioinformatics	Argonne GADU project Dartmouth Psychological & Brain Sciences	BLAST, BLOCKS, gene sequences, etc Functional MRI
Comp. Chemistry	ChemGrid	
University campus Resources, portals, apps	➤CCR (U Buffalo) ➤GLOW (U Wisconsin) ➤TACC (Texas Advanced Computing Center) ➤MGRID (U Michigan) ➤UFGRID (U Florida) ➤Crimson Grid (Harvard) ➤FermiGrid (FermiLab Grid)	

OSG Grid Partners

TeraGrid	<ul style="list-style-type: none"> • “DAC2005”: run LHC apps on TeraGrid resources • TG Science Portals for other applications • Discussions on joint activities: Security, Accounting, Operations, Portals
EGEE	<ul style="list-style-type: none"> • Joint Operations Workshops, defining mechanisms to exchange support tickets • Joint Security working group • US middleware federation contributions to core-middleware gLITE
Worldwide LHC Computing Grid	<ul style="list-style-type: none"> • OSG contributes to LHC global data handling and analysis systems
Other partners	<ul style="list-style-type: none"> • SURA, GRASE, LONI, TACC • Representatives of VOs provide portals and interfaces to their user groups

Example of Partnership: WLCG and EGEE



OSG Activities

Blueprint	Defining principles and best practices for OSG
Deployment	Deployment of resources & services
Provisioning	Connected to deployment
Incidence response	Plans and procedures for responding to security incidents
Integration	Testing & validating & integrating new services and technologies
Data Resource Management (DRM)	Deployment of specific Storage Resource Management technology
Documentation	Organizing the documentation infrastructure
Accounting	Accounting and auditing use of OSG resources
Interoperability	Primarily interoperability between
Operations	Operating Grid-wide services

Networks



Evolving Science Requirements for Networks (DOE High Performance Network Workshop)

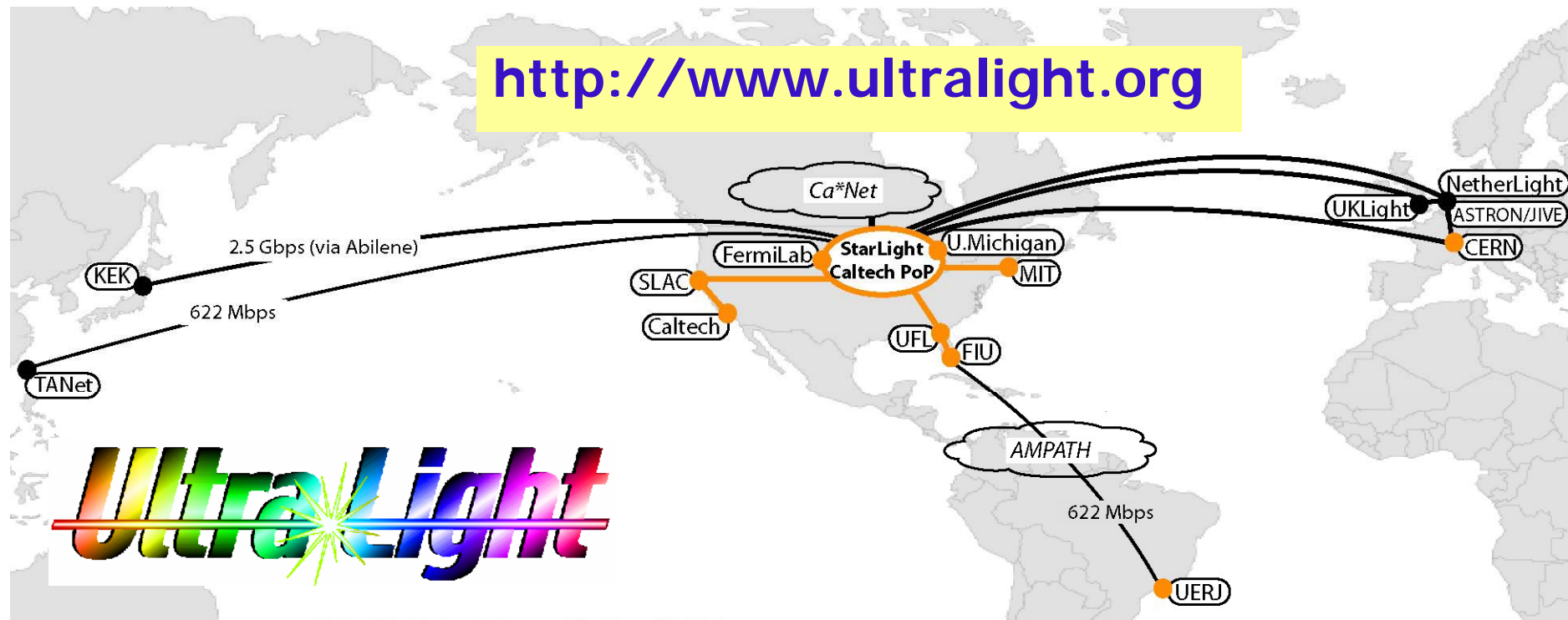
Science Areas	Today <i>End2End</i> Throughput	5 years End2End Throughput	5-10 Years End2End Throughput	Remarks
High Energy Physics	0.5 Gb/s	100 Gb/s	1000 Gb/s	High bulk throughput
Climate (Data & Computation)	0.5 Gb/s	160-200 Gb/s	N x 1000 Gb/s	High bulk throughput
SNS NanoScience	Not yet started	1 Gb/s	1000 Gb/s + QoS for Control Channel	Remote control and time critical throughput
Fusion Energy	0.066 Gb/s (500 MB/s burst)	0.2 Gb/s (500MB/20 sec. burst)	N x 1000 Gb/s	Time critical throughput
Astrophysics	0.013 Gb/s (1 TB/week)	N*N multicast	1000 Gb/s	Computational steering and collaborations
Genomics Data & Computation	0.091 Gb/s (1 TB/day)	100s of users	1000 Gb/s + QoS for Control Channel	High throughput and steering

See <http://www.doecollaboratory.org/meetings/hnpnw/>

UltraLight

Integrating Advanced Networking in Applications

<http://www.ultralight.org>



UltraLight

- UltraLight backbone (Native 10 GE)
- Connectivity to UltraLight's backbone (POS 10 Gbps)
- Partners sites
- Peer sites

10 Gb/s+ network

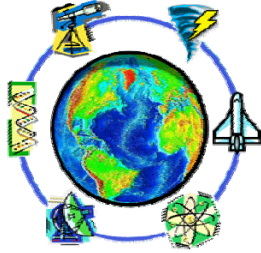
- Caltech, UF, FIU, UM, MIT
- SLAC, FNAL
- Int'l partners
- Level(3), Cisco, NLR

Training Outreach Communications

Grid Summer Schools

- June 2004: First US Grid Tutorial (South Padre Island, Tx)
 - ◆ 36 students, diverse origins and types
- July 2005: Second Grid Tutorial (South Padre Island, Tx)
 - ◆ 42 students, simpler physical setup (laptops)
- June 26-30: Third Grid Tutorial (South Padre Island, Tx)
- Reaching a wider audience
 - ◆ Lectures, exercises, video, on web
 - ◆ Students, postdocs, scientists
 - ◆ Coordination of training activities
 - ◆ More tutorials, 3-4/year
 - ◆ Agency specific tutorials





Grid Technology Cookbook

A guide to building and using grid resources

Current Timetable (2005 – 06)

▶ Acknowledgements

▶ Preface

▶ Introduction

▶ What Grids Can Do For You

▶ Grid Case Studies

▶ Technology For Grids

▶ Standards & Emerging Technologies

▶ Programming Concepts & Challenges

▶ Building Your Own Grid

▶ Installation Procedure Examples

▶ Typical Usage Examples

▶ Practical Tips

▶ Glossary

▶ Appendices

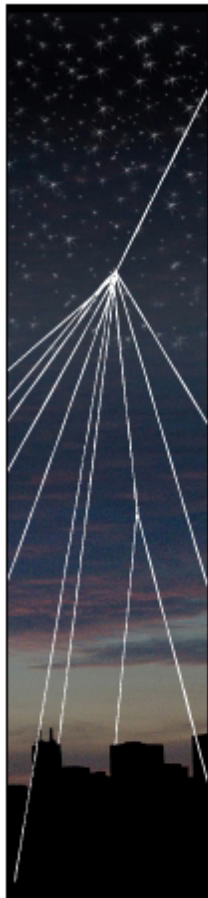
•Outline Development, Vetting	September-October
•Assemble Writing Teams	October-December
•Develop Web Structure	November-December
•Writing Process Underway	November-March
•Material Edited and Entered	December-April
•Review of First Draft	May
•Edits to First Draft Entered	Early June
•Review of Final Draft	Late June
•Release of Version 1	July 2006

QuarkNet/GriPhyN e-Lab Project

Cosmic Ray Collaboration

Join a national collaboration of high school students to study cosmic rays.

<http://quarknet.uchicago.edu/elab/cosmic/home.jsp>



Spending all your time in a shower?

When you're sleeping or sitting in class, cosmic rays shower the earth and everything on it.

What are cosmic rays?

Where do they come from?

Where do they hit?

Some cosmic rays have so much energy that scientists are not sure where they come from. A number of research projects are looking at this question.

Who are we?

We're a collaboration of high school students and teachers who use cosmic ray data to answer some of these questions. We work with scientists to provide cutting edge tools that use grid technology, graphs, and posters and collaborate with other students.

Who can join?

You! Think about steps you'd take to investigate cosmic rays. How have you started? What do you need to know? Can you collect and analyze data?

Username:

Password:

Login

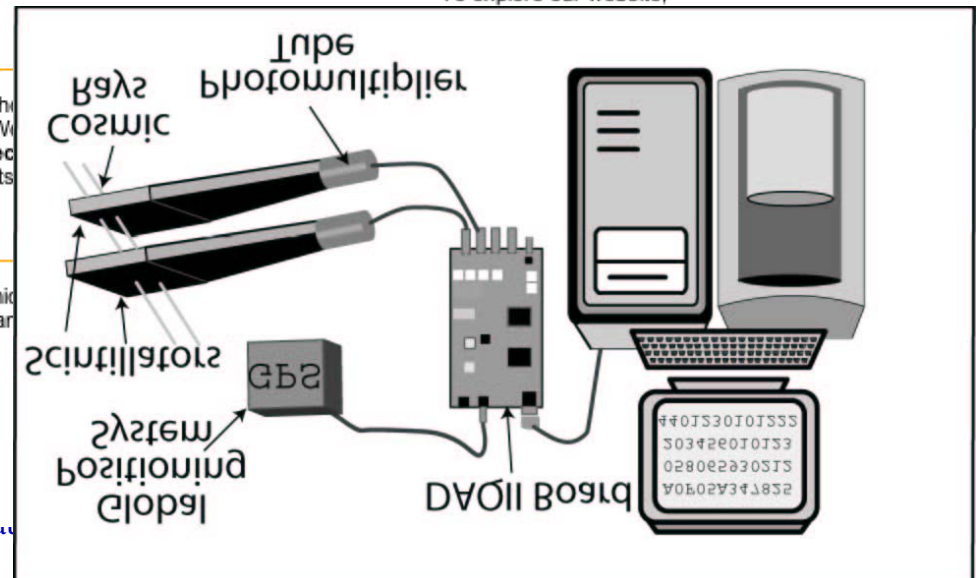
Need login info?

Ask your teacher.

Working on your own?

Contact quarknet@fnal.gov.

To explore our website,



CHEPREO: Center for High Energy Physics Research and Educational Outreach Florida International University



www.chepreo.org

- Physics Learning Center
- CMS Research
- Cyberinfrastructure
- WHREN network (S. America)

- Funded September 2003
- \$MPS, CISE, EHR, INT





- World Summit on Information Society
- HEP Standing Committee on Inter-regional Connectivity (SCIC)

- Global collaborations, Grids and addressing the Digital Divide
- Focus on poorly connected regions
- Brazil (2004), Korea (2005)

D0SAR Workshop (March 30, 2006)

Science Grid Communications

Broad set of activities

- (Katie Yurkewicz)
- News releases, PR, etc.
- Science Grid This Week
- OSG Monthly Newsletter

www.interactions.org/sgtw



Calendar/Meetings

March

7-8, Second CLEANER All-Hands Meeting, Arlington, Virginia

8-10, GridChem Workshop: Distributed Computational Chemistry (on the Grid), Austin, Texas

13-15, ISSSE 06: International Symposium on Secure Software Engineering, Washington, D.C.

26-28, PRAGMA 10: Pacific Rim Applications and Grid Middleware Assembly Tenth Workshop, Townsville, Queensland, Australia

[Full Calendar](#)

Feature Story

Simulating Supersymmetry



The ATLAS detector under construction.
Image © CERN

One of the discoveries eagerly anticipated by particle physicists working on the world's next particle collider is that of supersymmetry, a theoretical lost symmetry of nature. Supersymmetry, often called SUSY, predicts the existence of a superpartner particle for every known particle.

Why the big hunt for SUSY's "sparticles"? Recent experiments have suggested that most of the matter in our universe is not made of familiar atoms, but of some new sort of "dark matter." Discovering a hidden world of sparticles will shed light on the nature of this dark matter, connecting observations performed at earth-based accelerators with those performed by astrophysicists and cosmologists.

Physicist Sanjay Padhi, a Chancellor Fellow at the University of Wisconsin-Madison, searches for SUSY using the ATLAS detector at the Large Hadron Collider. Although the LHC and ATLAS won't start collecting experimental data until 2007, he and his colleagues are already hard at work generating the simulated data that is currently...

BBC Project Takes on Climate Change



Image Courtesy Climateprediction.net

It seems like new distributed computing projects are popping up every day, and it seems like the world is catching on to the possibilities made possible by harnessing the power of thousands of connected PCs. Case in point: Last month, the British Broadcasting Corp. teamed up with ClimatePrediction.net to launch a distributed computing project that is running, initially at least, concurrent with the channel's "Climate Chaos" season of programming.

Dubbed the "BBC Climate Change Experiment," this project, according to ClimatePrediction.net chief software architect Carl Christensen, is a little different than other similar projects. Unlike other distributed computing experiments where user computers will perform a "workunit" before moving on to another task, computers on the BBC experiment run the entire climate model—from start to finish. The experiment takes three months "on the fastest PCs out there today," he said, whereas tasks on other projects can be completed in a matter of hours.

There is a twofold reason for this

Image of the Week



Indian President A.P.J. Abdul Kalam using VRVS at CHEP06. (Click on image for larger version.)

Image Courtesy Phillippe Galvez

On February 17, Indian President A.P.J. Abdul Kalam visited the Computing in High Energy and Nuclear Physics (CHEP06) conference in Mumbai, India. The President's speech to the conference highlighted grid computing in India and around...



OSG Newsletter

Monthly newsletter

➤ (Katie Yurkewicz)

➤ 5 issues now

www.opensciencegrid.org/osgnews

DOSAR Workshop (March 30, 2006)



Open Science Grid News

FEBRUARY 2006

OSG HOME | SUBMIT NEWS | SUBSCRIBE | ARCHIVE | ABOUT OSG NEWS

Meetings and Events

Condor Week 2006
April 24-27, 2006

GGF17
May 9-12, 2006

HPDC 2006
June 19-23, 2006

[View Full Calendar](#)
[Add New Event](#)

Operations Report

The Operations team provides a bi-weekly report to the Council to keep them up to date and to bring issues to the table for discussion. Here are a few items from recent postings.



With CMS and ATLAS increasing their number of submitted jobs to a site, several scaling problems have arisen. These issues are being addressed by extensions to Condor and Condor-G, and with the Globus software which is run on the head node for the job-manager.

More than 45% of sites have upgraded to OSG 0.4.0, and 26% are reporting to the MonaLisa (ML) accounting. The daily usage reports are based on ML, so while it remains an optional component, if you want your site to be included in the accounting you will need to install and configure it. The operations team will be happy to help with this.

The education project MARIACHI and

From the Executive Director

Dear OSG Consortium and Friends,



I am very pleased to announce that Bill Kramer has been selected by the Council as its new chairperson. As head of the NERSC computing center, Bill brings a wealth of experience and understanding to our program, and we are already keeping him busy. As one of the new applications coordinators, Frank Würthwein continues to be part of the OSG's core team and we will continue to benefit from his contributions and insights. I look forward to working with each and every member of the Executive Board.

At the beginning of this month we submitted the OSG program of work as an unsolicited proposal to the NSF's Mathematical and Physical Sciences Directorate, and we are in the process of submitting the same proposal to the DOE SciDAC-2 program. The proposal focuses on three key areas: the OSG facility; education, outreach and training; and science-driven extensions.

The Consortium meeting saw the presentation and discussion of many aspects of the use and provisioning of the facility, including the contents and schedule of the next two OSG releases and VDT 1.3.10. The local organizers—Paul Avery, DeeDee Carver and Jorge Rodriguez—did a superb job. CMS is ramping up OSG activity once again and DZero is validating one site at a time to run SAMGrid-based reprocessing jobs. Mike Wilde is working with Soma Mukherjee and UTB on the logistics and schedule for this year's summer school; please contact him if you are interested in contributing.

Sincerely,
Ruth Pordes, OSG Executive Director

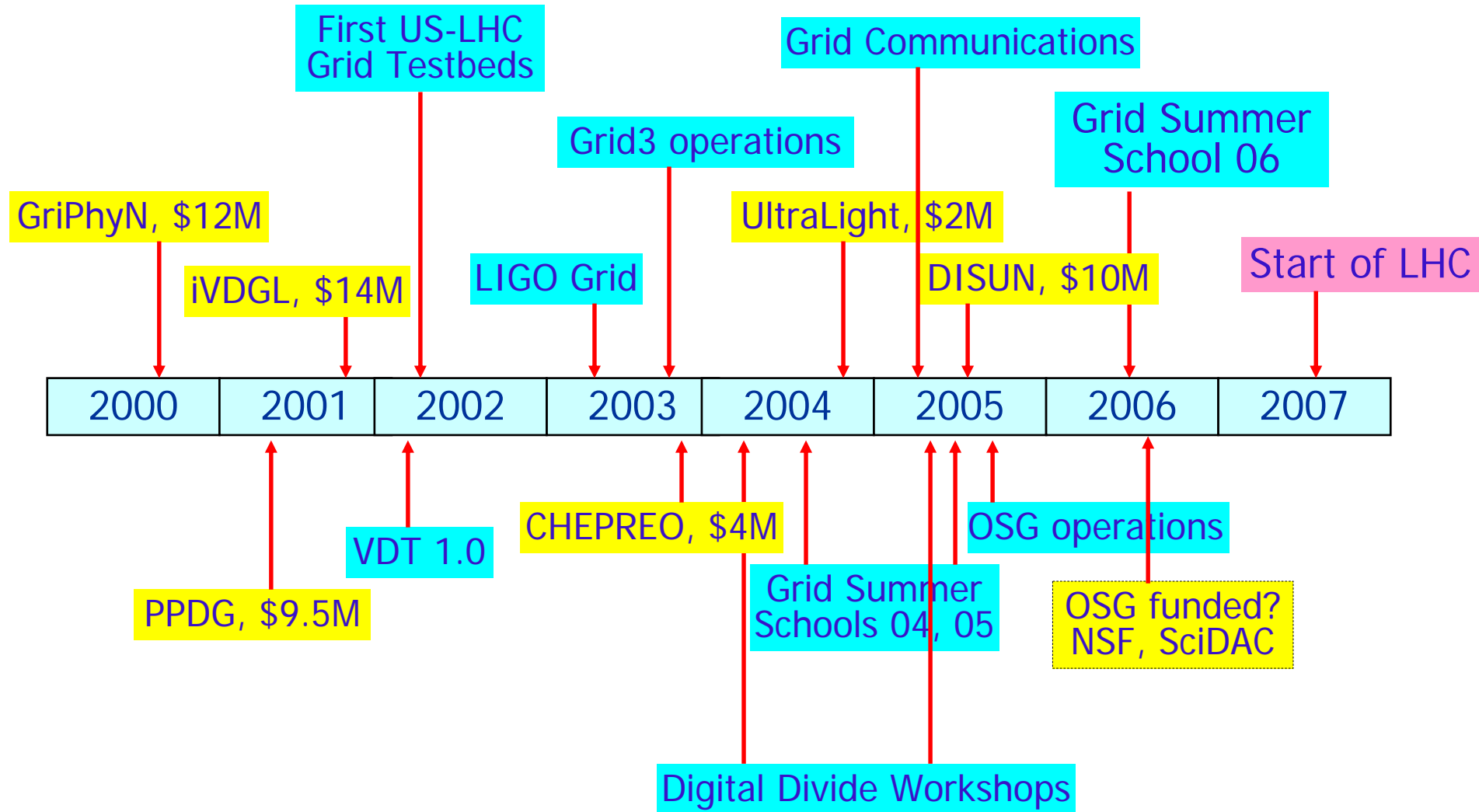
Applications - Reprocessing D0 Data

Category	Count
P17.07 Events	1411
P17.09 Events	1405
P17.09 Events Skimmed	587

Reprocessing status as of February 20.
(Click on image for larger version.)

D0's latest reprocessing of its Run IIa data used several OSG sites, which together processed more than 10 million events. D0 has used resources from collaborating institutions for several years for reprocessing, since their Fermilab resources are busy processing newly collected data and Monte Carlo simulations are always ongoing. The addition of OSG and LCG resources

Grid Timeline





OSG Consortium Meetings

- July 20, 2006: University of Wisconsin, Milwaukee
 - ◆ Kickoff meeting, ~100 attendees
 - ◆ Focus on getting off the ground with running jobs
- January 23, 2006: University of Florida (Gainesville)
 - ◆ ~110 people
 - ◆ Partnerships, organization, funding, operations, software infrastructure
- August 21-24, 2006: University of Washington (Seattle)
- January, 2007: TACC



Jan. 23-25 OSG Meeting

Monday January 23, 2006		
Plenary: Use of OSG Today		Session Chair: Paul Avery Union 282
9:00 am	Welcome & Status of OSG	Paul Avery, University of Florida
9:30am	Astrophysics	Kent Blackburn, Caltech
9:50am	Bioinformatics	Dina Sulakhe, ANL - Mark Green, University at Buffalo
10:10am	Accelerator Physics	Amber Boehnlein, Fermilab
10:30am	Operations Applications	Leigh Grundhoefer, Indiana University
10:50am	BREAK	Union Lobby
11:20 am	Portals and Campus/other Grid Communities	Sebastien Goasguen, Purdue University
11:40 am	Worldwide LHC Computing Grid	Les Robertson, CERN
12:00 pm	TeraGrid Interoperability	Dane Skow, UChicago/ANL
12:20	RENCI	Alan Blatecky, RENCi
12:40	LUNCH	Union Lobby
Plenary: OSG Partnerships and Regional Grids		Session Chair: Ruth Pordes Union 282
1:30 pm	GEON	Chaitan Baru, SDSC
1:50 pm	GridChem	John Connolly, Center for Computational Science
2:10 pm	University of Florida Grid	Charles Taylor
2:30 pm	SURA	Gary Crane, SURA
2:45 pm	New York State Cyberinstitute	Mark Green, University at Buffalo
3:00 pm	GLOW	Miron Livny, University of Wisconsin Madison
3:15 pm	HIPCAT/TIGRE	Jay Boisseau, TACC
3:30- 4:00 pm	BREAK	Physics Building
Plenary: Security		Session Chair: Bob Cowles Physics NPB 2205
4:30 pm	Security Report	Bob Cowles,, SLAC
5:10 pm	Introduction of OSG Security Officer	Don Petravick, Fermilab
5:20 pm	Proposed Security Center for Enabling Technology	Deb Agarwal, Brian Tierney, LBNL
5:50-10:30 pm	Security Program of Work	all

END



Grid Project References

➤ Open Science Grid

◆ www.opensciencegrid.org

➤ Grid3

◆ www.ivdgl.org/grid3

➤ Virtual Data Toolkit

◆ www.griphyn.org/vdt

➤ GriPhyN

◆ www.griphyn.org

➤ iVDGL

◆ www.ivdgl.org

➤ PPDG

◆ www.ppdg.net

➤ CHEPREO

◆ www.chepreo.org

➤ UltraLight

◆ www.ultralight.org

➤ Globus

◆ www.globus.org

➤ Condor

◆ www.cs.wisc.edu/condor

➤ WLCG

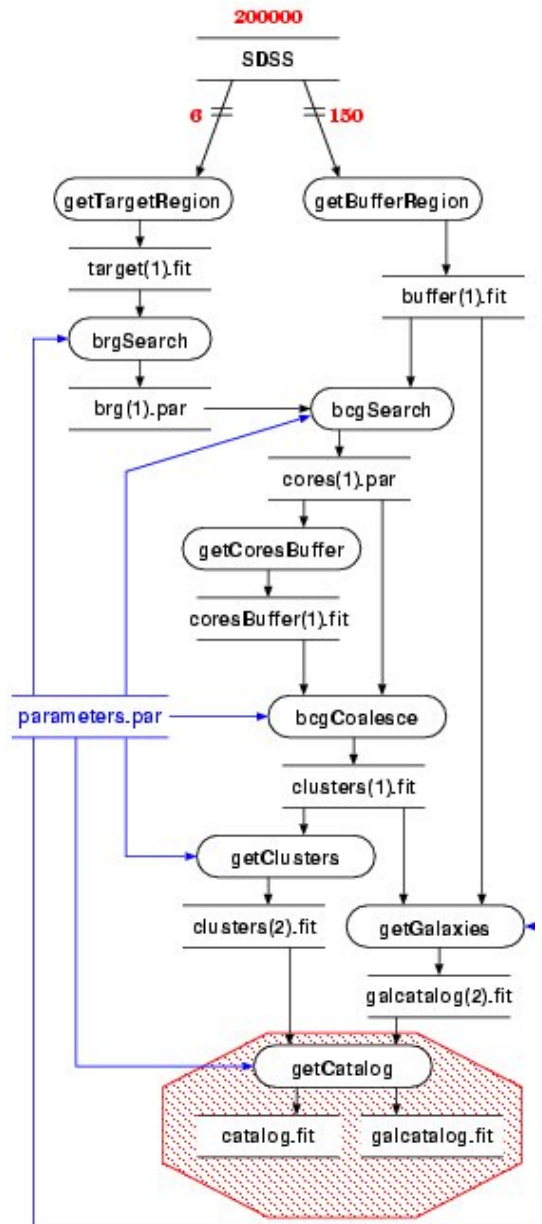
◆ www.cern.ch/lcg

➤ EGEE

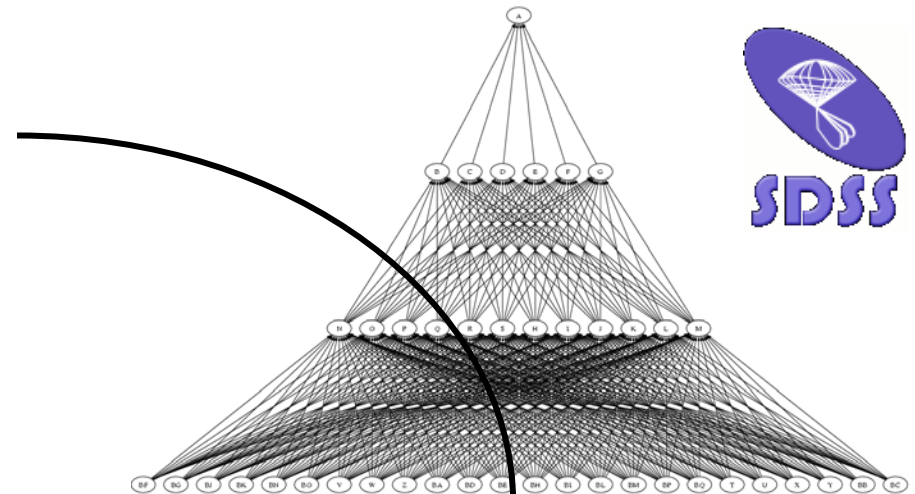
◆ www.eu-egee.org

Sloan Digital Sky Survey (SDSS)

Using Virtual Data in GriPhyN



Sloan Data



Galaxy cluster
size distribution

