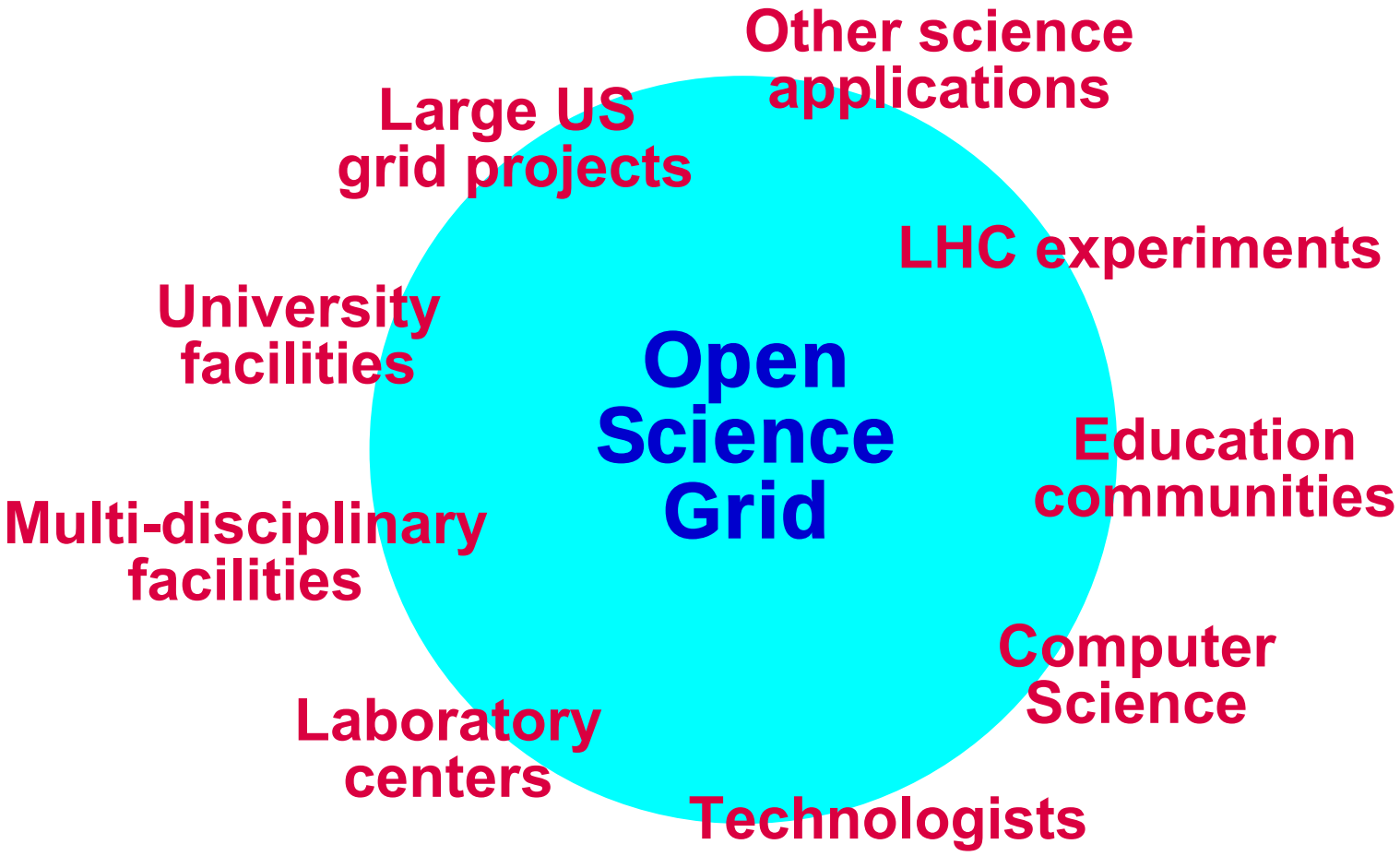
Open Science Grid

Linking Universities and Laboratories in National Cyberinfrastructure

ICFA Digital Divide Workshop
Cracow, Poland
Oct. 9, 2006

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

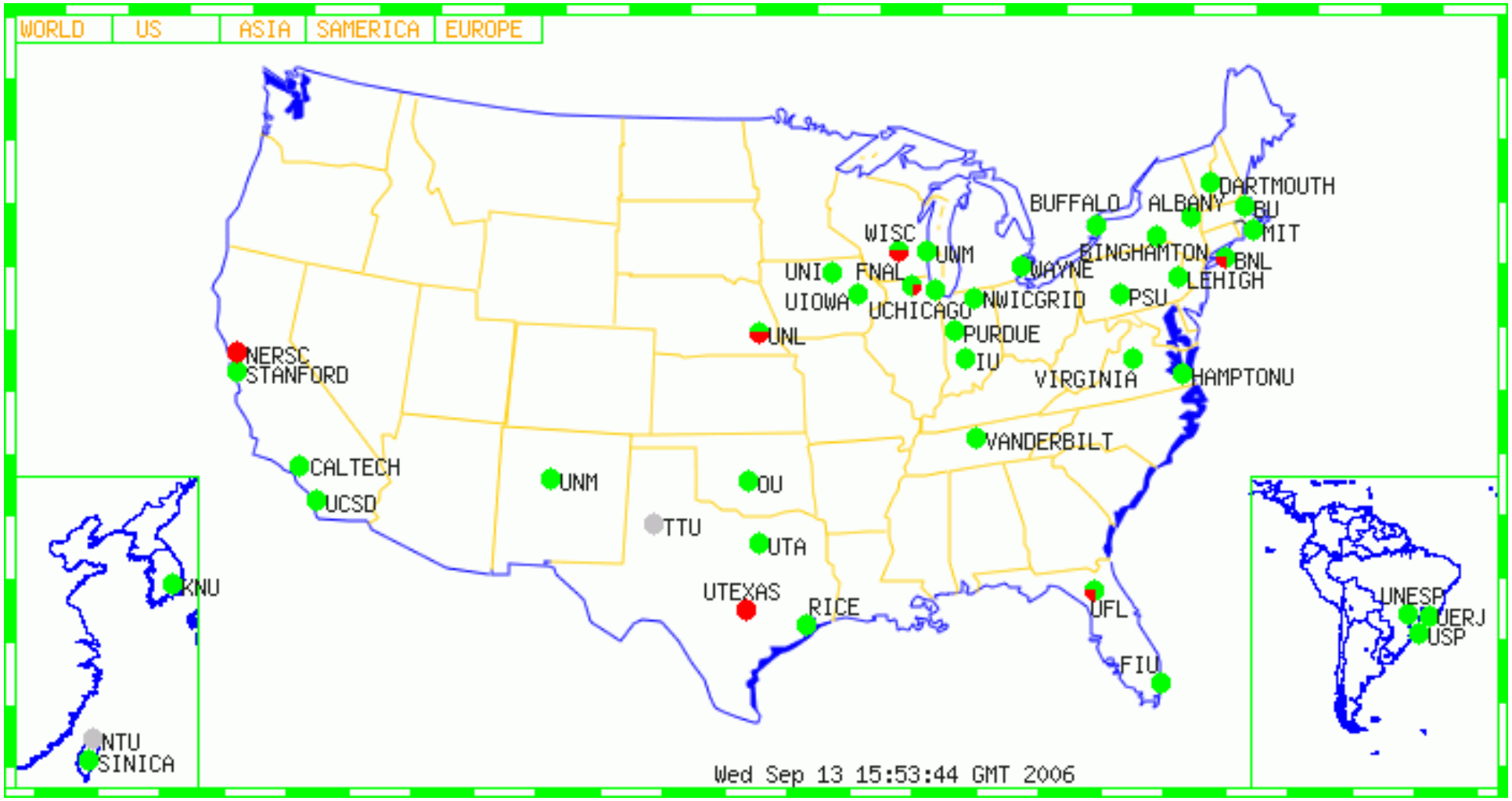
The Open Science Grid Consortium



Open Science Grid: July 20, 2005

- **Partnership of many organizations**
- **Production Grid: 60+ sites, 20,000+ CPUs "present"**
- **Sites in US, Korea, Brazil, Taiwan**

OSG Site Map (Sep. 2006)



General Comments About OSG

- OSG is a *Consortium* of projects, facilities, providers
  - Stakeholders represent important interests
- OSG *Project* was recently funded for $30M (2006 – 2011)
  - OSG Consortium manages the OSG Project
  - Value of constituent resources and operations far greater
- OSG was formed by *bottoms-up* activity, informed by history
  - Grid projects: GriPhyN, iVDGL, PPDG, UltraLight, CHEPREO, DISUN
  - Grid testbeds (2002 – 2004), Grid3 (2003 – 2005)
- OSG interfaces to *Virtual Organizations* (VO)
  - VOs responsible for support, authentication of members (scalability)
- OSG *does not* own resources
  - CPU, storage owned and managed by projects or sites
- OSG *integrates* technologies & middleware
  - Relies on S/W & technology creation by member projects or partners
  - Exploits NSF + DOE investments (NMI, Globus, Condor, ...)

OSG Participating Projects

Computer Science	Condor, Globus, SRM, SRB, dCache	
Physics	LIGO, Nuclear Physics, Tevatron, LHC	
Astrophysics	Sloan Digital Sky Survey, future astro projects	
Nanoscience	NanoHUB @ Purdue	
Bioinformatics	GADU @ Argonne Dartmouth Psychological & Brain Sciences	
Comp. Chemistry	ChemGrid	
University, laboratory & regional Grids	GRASE GLOW TACC MGRID UFGRID	Crimson Grid FermiGrid GROW SURA DOSAR

OSG Member Virtual Organizations

CDF	HEP experiment at FermiLab
CMS	HEP experiment at CERN
DES	Dark Energy Survey (Astro)
DOSAR	Regional grid in Southwest US
DZero	HEP experiment at FermiLab
FermiLab	HEP laboratory
fMRI	Functional MRI (Dartmouth)
GADU	Bioinformatics effort at Argonne
Geant4	Simulation project
GLOW	Campus grid (University of Wisconsin, Madison)
GRASE	Regional grid in Upstate NY
GridChem	Quantum chemistry grid
GridEx	Grid Exerciser
GROW	Campus grid (University of Iowa)

OSG Member Virtual Organizations (2)

I2U2	E/O effort (Interactions in Understanding the Universe)
iVDGL	Generic VO
LIGO	Gravitational wave experiment
Mariachi	Ultra-high energy cosmic ray experiment
MIS	OSG monitoring?
nanoHUB	Nanotechnology grid at Purdue
NWICG	Northwest Indiana regional grid
Ops	OSG Operations
OSG	Generic VO?
OSGEDU	OSG education/outreach
SDSS	Sloan Digital Sky Survey (Astro)
STAR	Nuclear physics experiment at Brookhaven
US-ATLAS	HEP experiment at CERN
LSU/CCT	Center for Computation and Technology

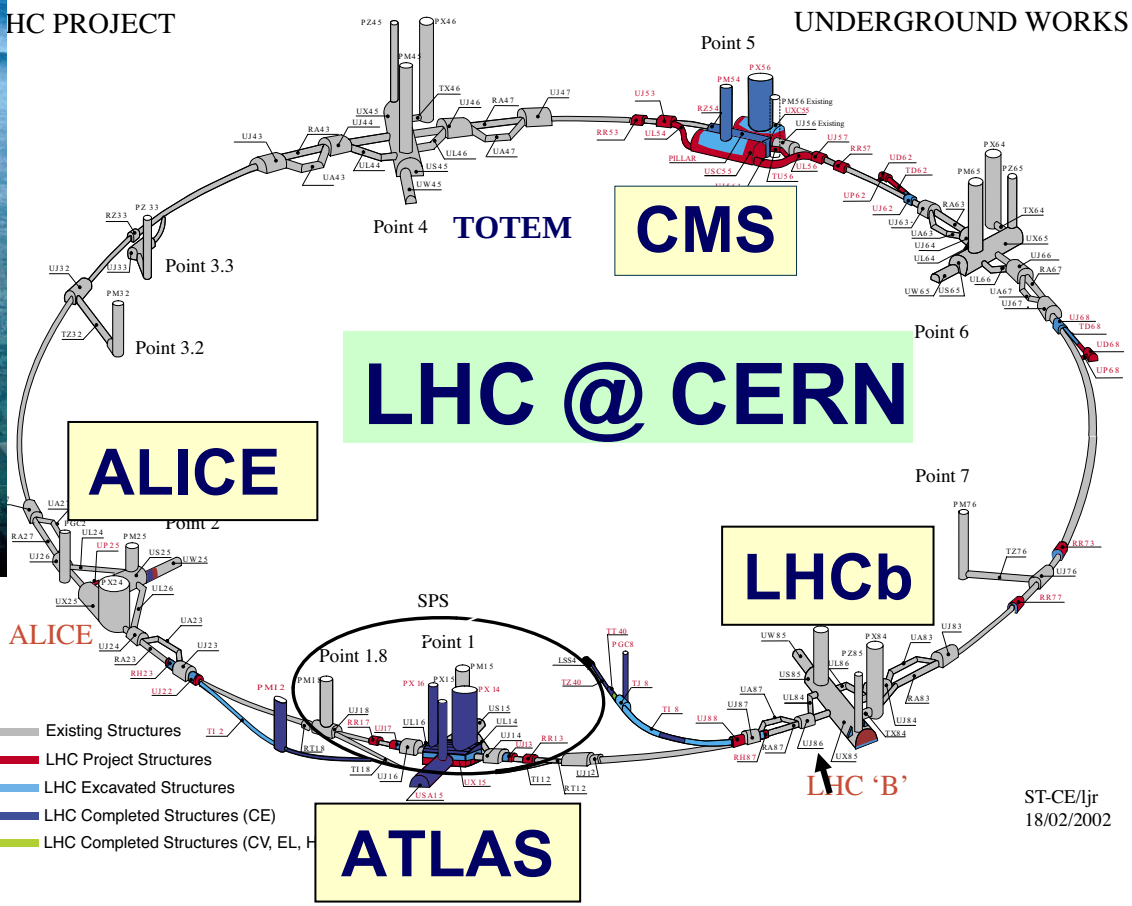
OSG Grid Partners

TeraGrid	• "DAC2005": run LHC apps on TeraGrid resources • TG Science Portals for other applications • Discussions on joint activities: Security, Accounting, Operations, Portals
EGEE	• EGEE/OSG part of Worldwide LHC Computing Grid • Joint Operations Workshops, defining mechanisms to exchange support tickets • Joint Security working group • US contributions to EGEE middleware gLITE
caBIG	• Cancer BioInformatics Grid; still being developed. More in late Fall 2006

OSG Sciences

Scale of OSG Resources & Services Set by LHC Experiments

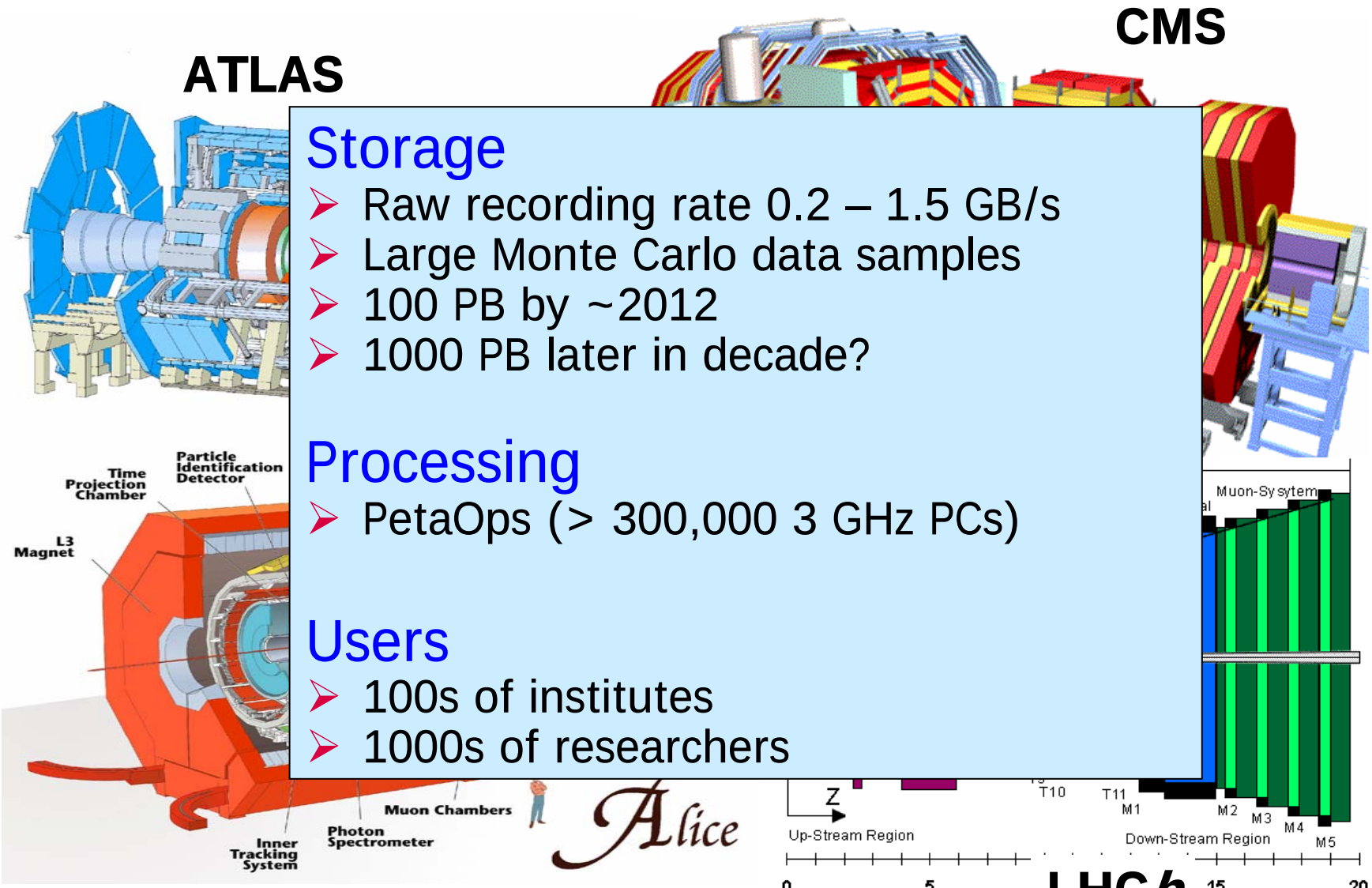
✸ **27 km Tunnel in Switzerland & France**



Search for

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

LHC Data and CPU Requirements



Storage

- Raw recording rate 0.2 – 1.5 GB/s
- Large Monte Carlo data samples
- 100 PB by ~2012
- 1000 PB later in decade?

Processing

- PetaOps (> 300,000 3 GHz PCs)

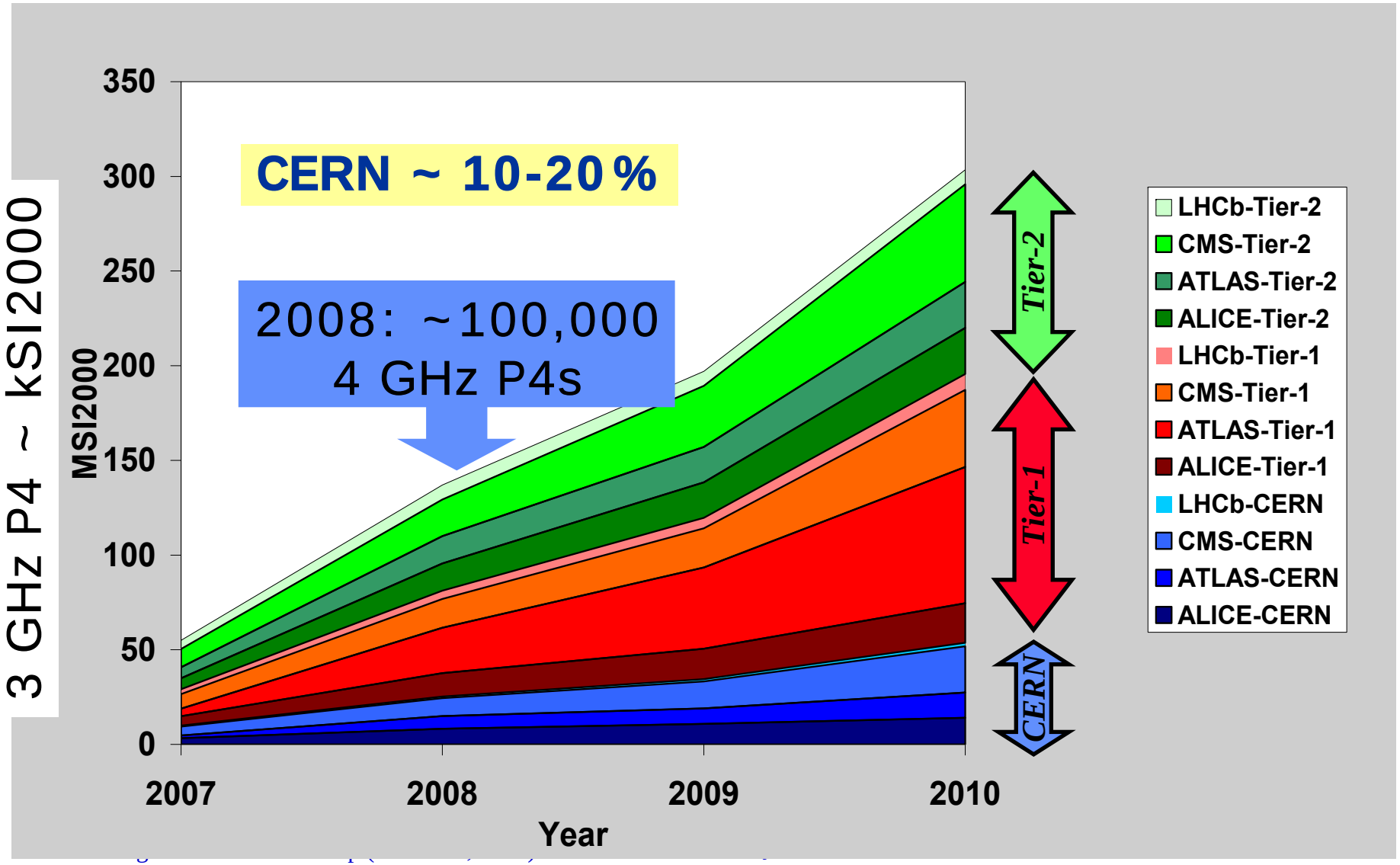
Users

- 100s of institutes
- 1000s of researchers

LHC Global Data Grid (2007+)

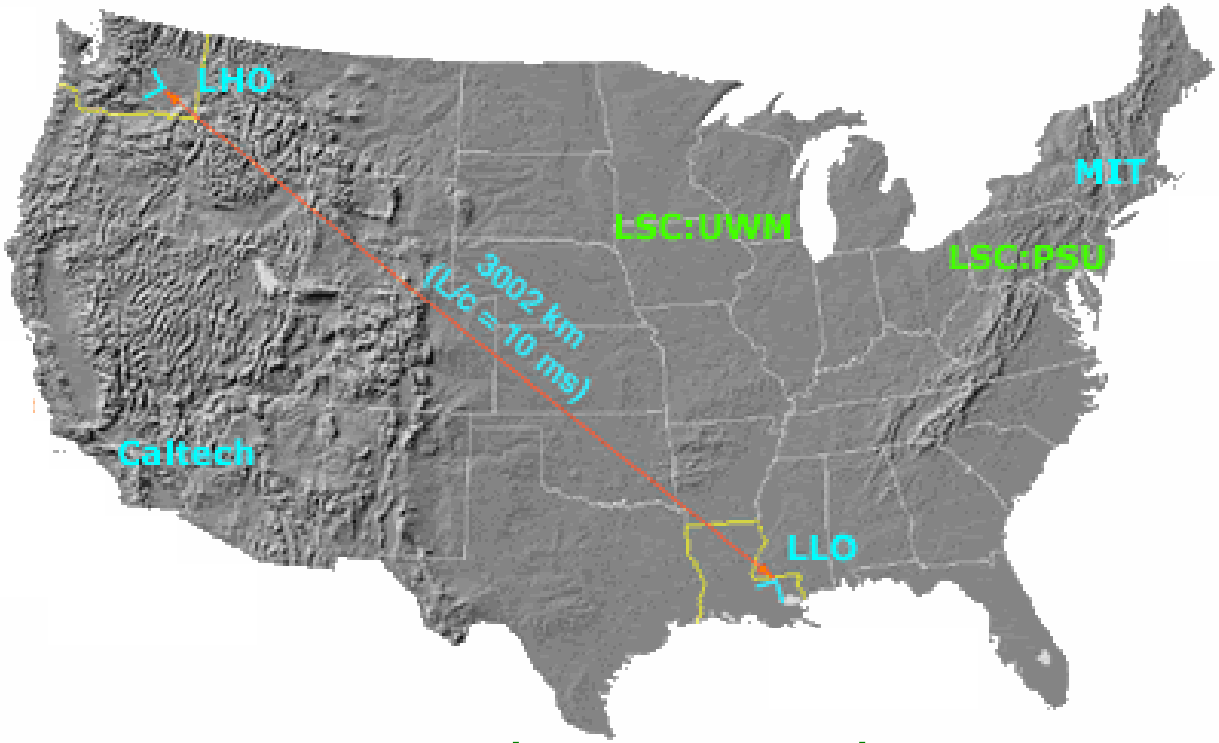
- 5000 physicists, 60 countries
- 10s of Petabytes/yr by 2008
- CERN / Outside = 10-20%

LHC Resource Requirements vs Time (U.S. portion has similar increase)



LIGO: Search for Gravity Waves

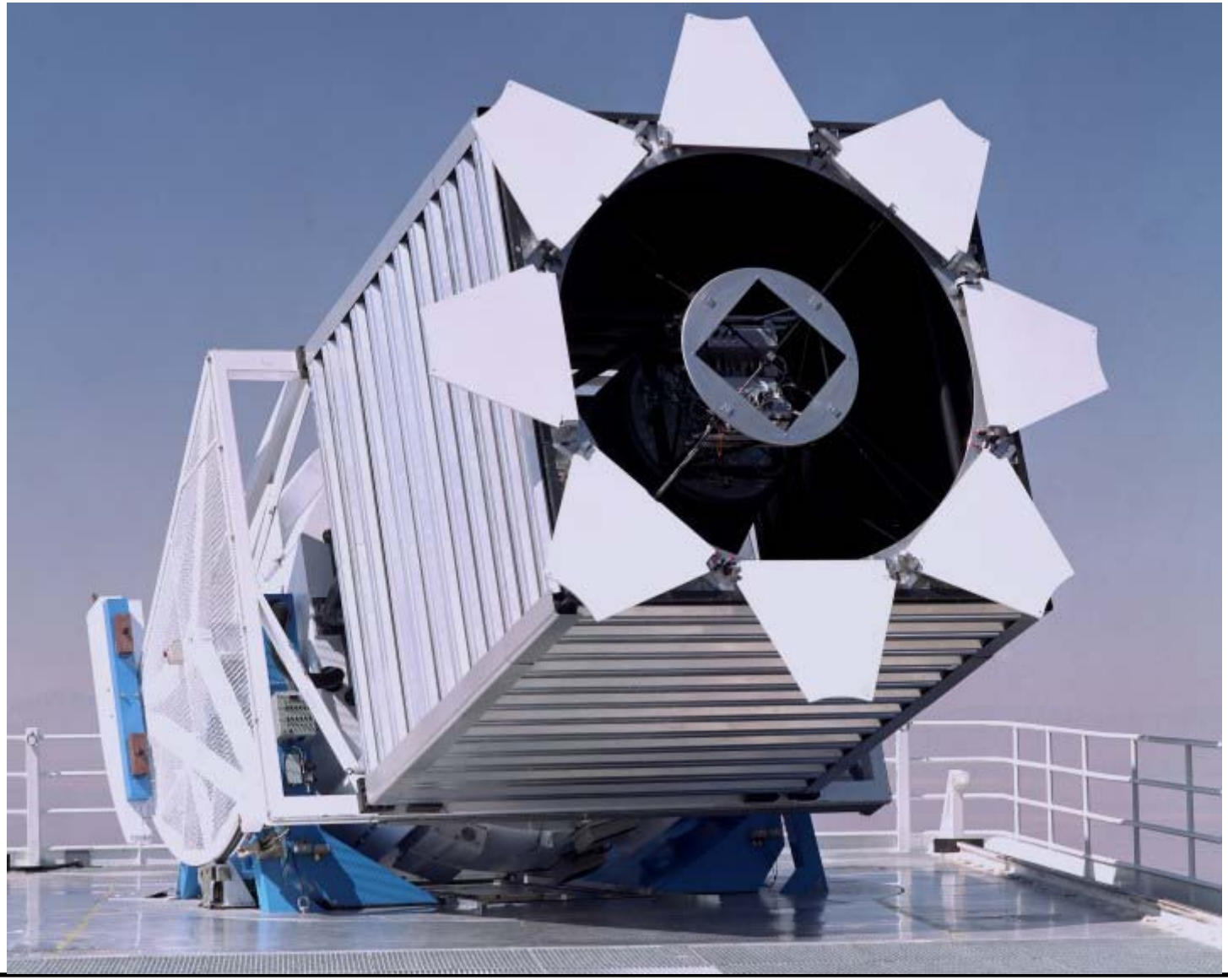
- LIGO Grid
  - 6 US sites
  - 3 EU sites (UK & Germany)



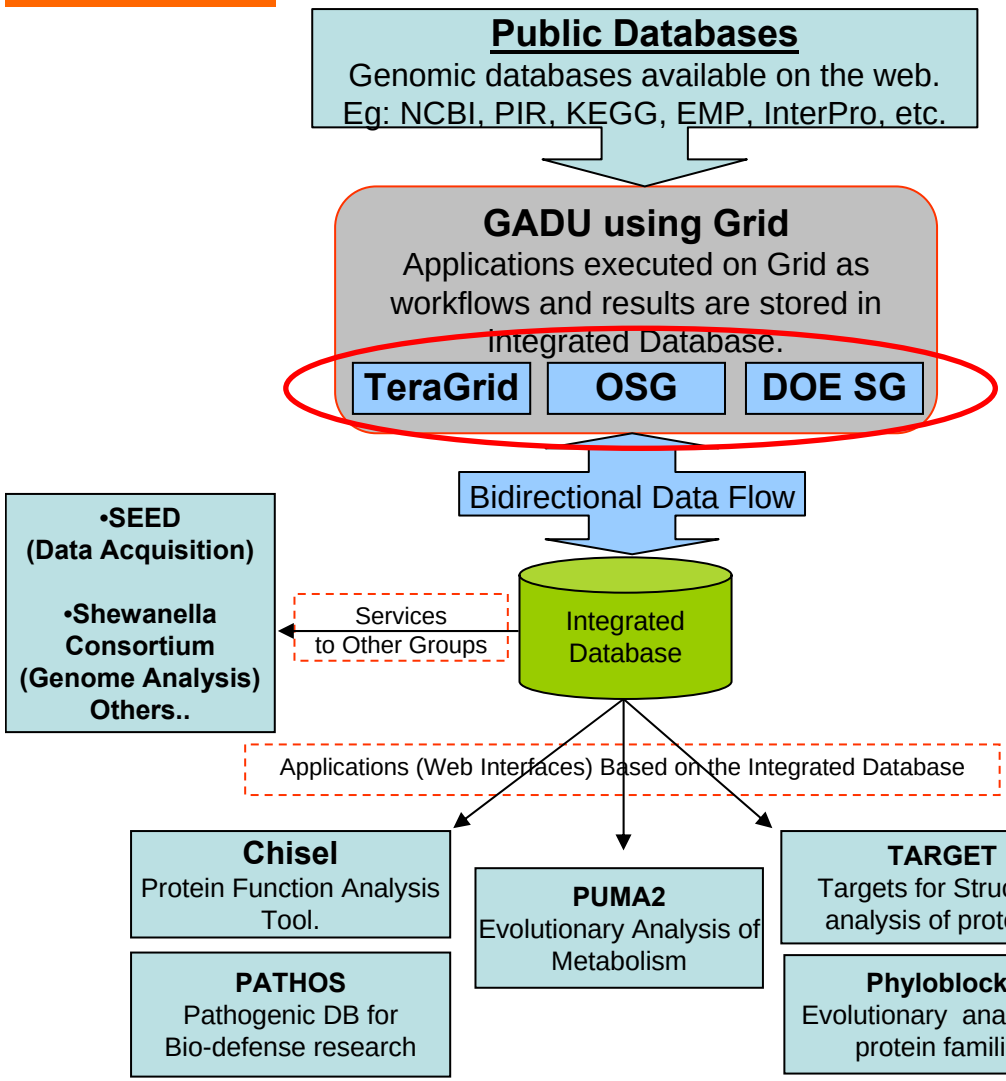
* LHO, LLO: LIGO observatory sites

* LSC: LIGO Scientific Collaboration

Sloan Digital Sky Survey: Mapping the Sky



Bioinformatics: GADU / GNARE



GADU Performs:

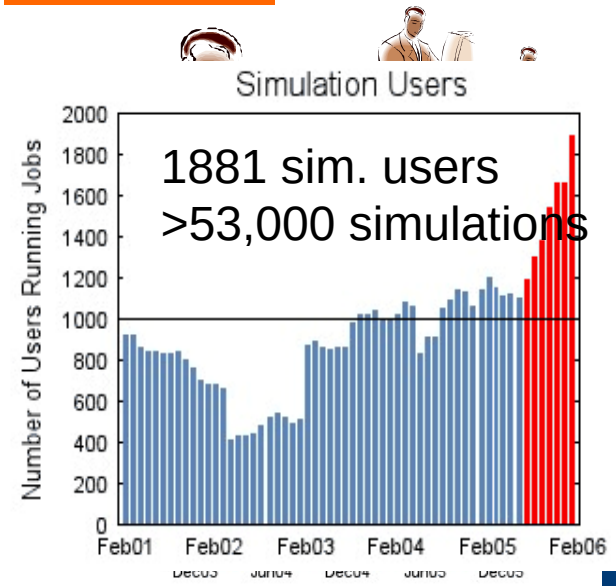
- ***Acquisition***: to acquire Genome Data from a variety of publicly available databases and store temporarily on the file system.
- ***Analysis***: to run different publicly available tools and in-house tools on the Grid using Acquired data & data from Integrated database.
- ***Storage***: Store the parsed data acquired from public databases and parsed results of the tools and workflows used during analysis.

Integrated Database Includes:

- Parsed Sequence Data and Annotation Data from Public web sources.
- Results of different tools used for Analysis: Blast, Blocks, TMHMM, …

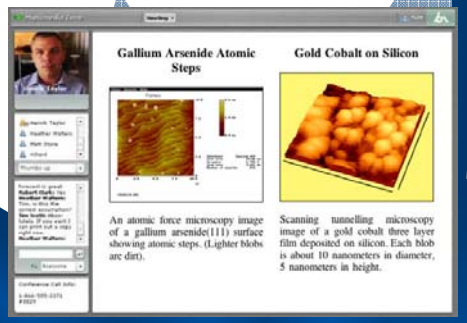
GNARE – Genome Analysis Research Environment

Nanoscience Simulations

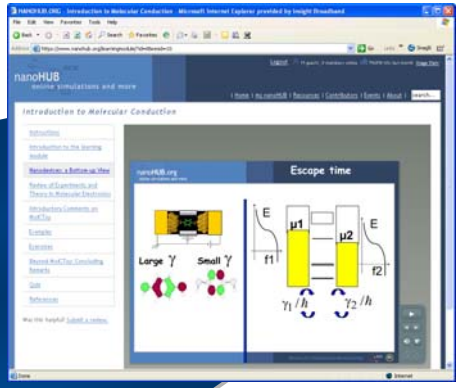


Real users and real usage
>10,100 users

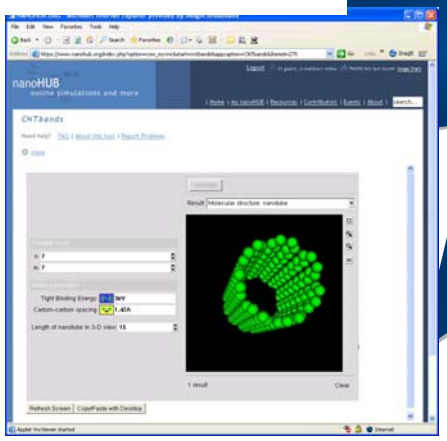
collaboration



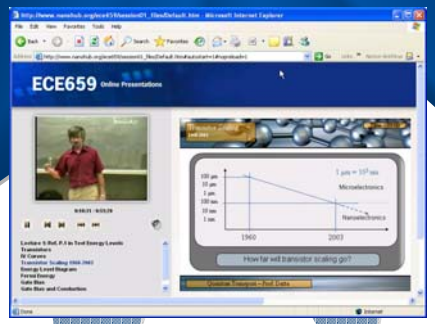
learning modules



nanoHUB.org



online simulation



courses, tutorials

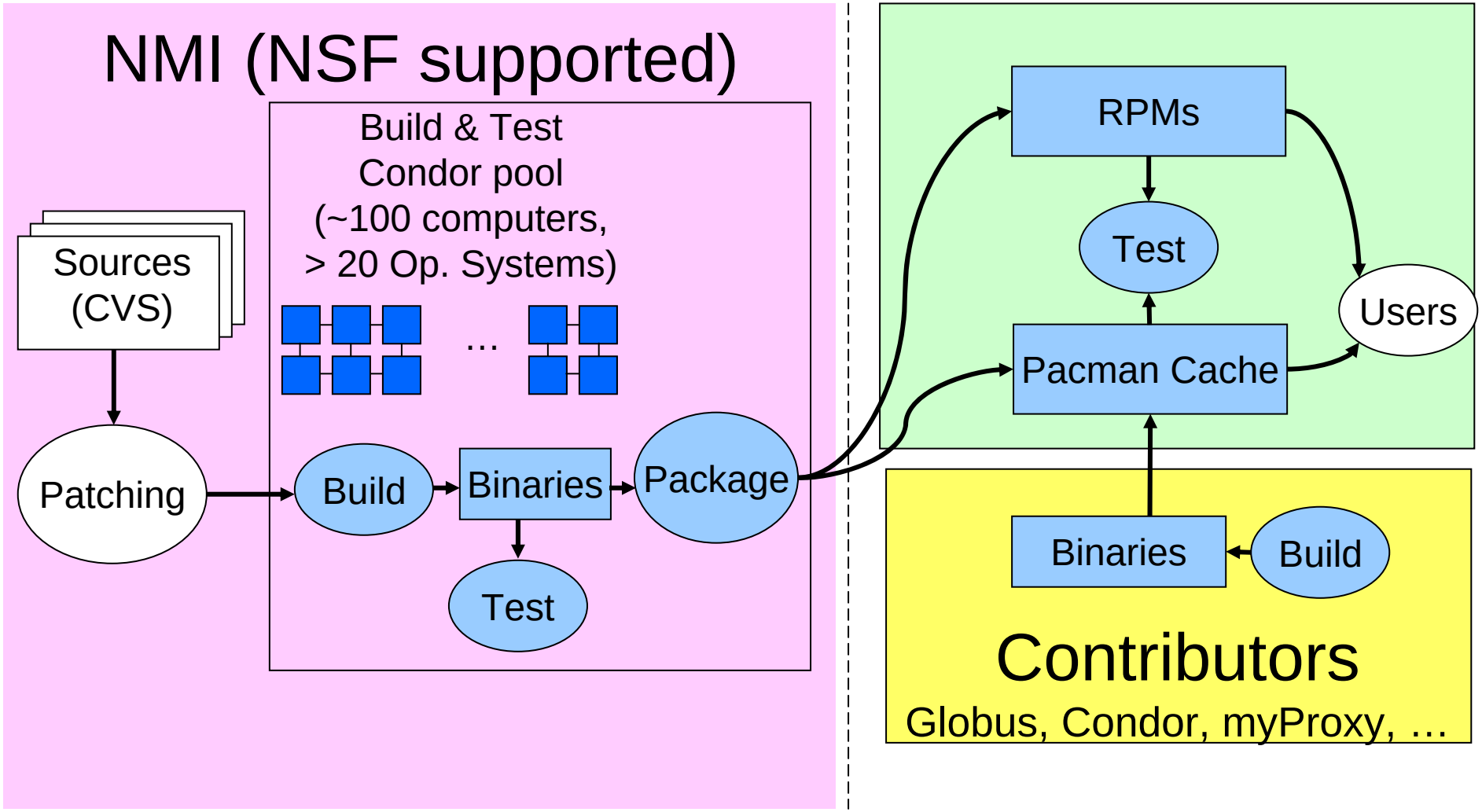
seminars

Strategy to Incorporate More Disciplines: OSG Engagement Effort

- Purpose: Bring non-physics applications to OSG
  - Led by RENCI (UNC + NC State + Duke)
- Specific targeted opportunities
  - Develop relationship
  - Direct assistance with technical details of connecting to OSG
- Feedback and new requirements for OSG infrastructure
  - (To facilitate inclusion of new communities)
  - More & better documentation
  - More automation
  - Bio Services and framework
- RENCI Bioportal for biology applications
  - Workflow nodes
  - Coordination with other OSG Bio activities

OSG and the Virtual Data Toolkit

Common Middleware: Virtual Data Toolkit



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

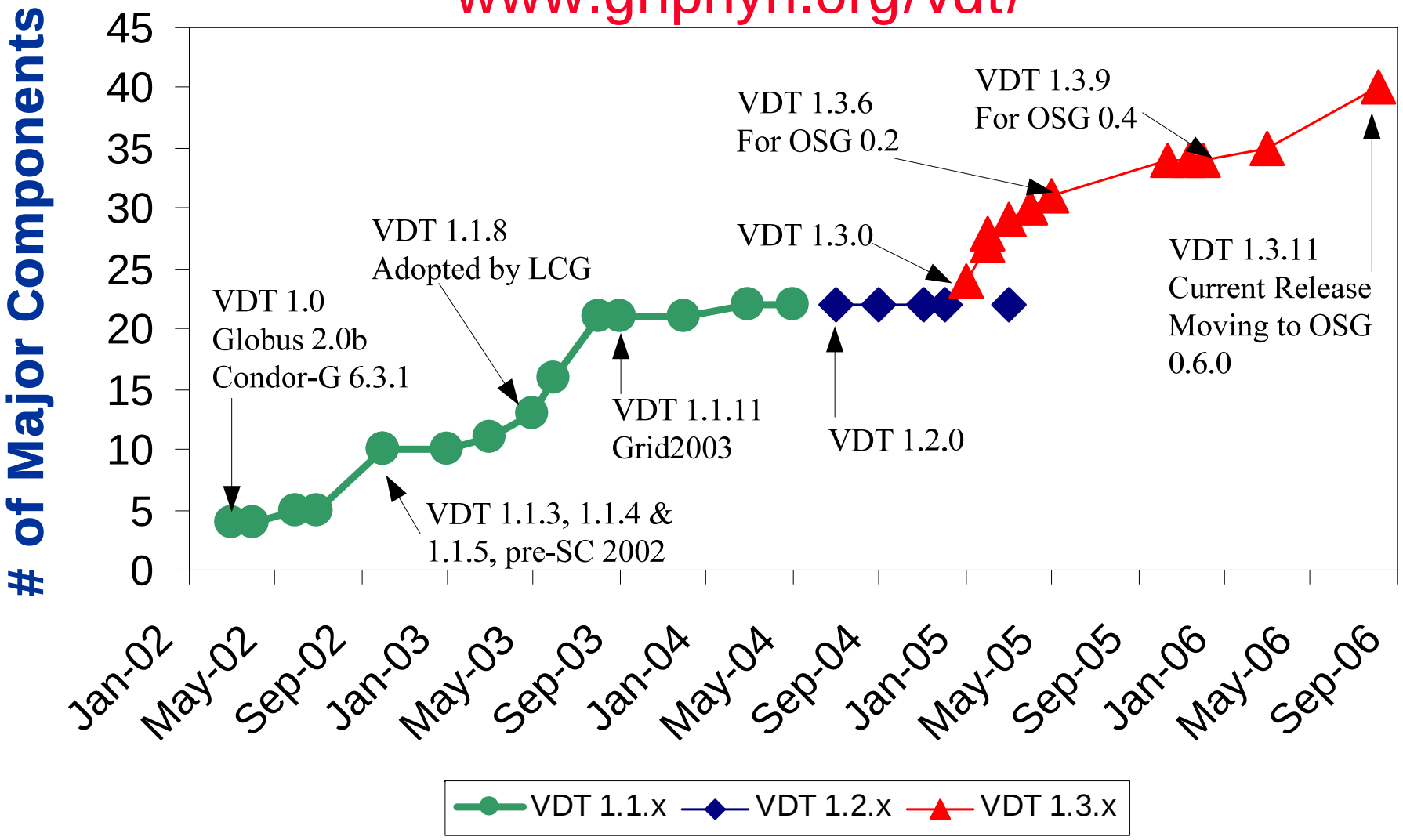
What the VDT Provides

- An integrated process for middleware integration
  - Figures out dependencies between software components
  - Works with providers for bug fixes
  - Provides automatic configuration
  - Packages it
  - Tests everything on multiple platforms

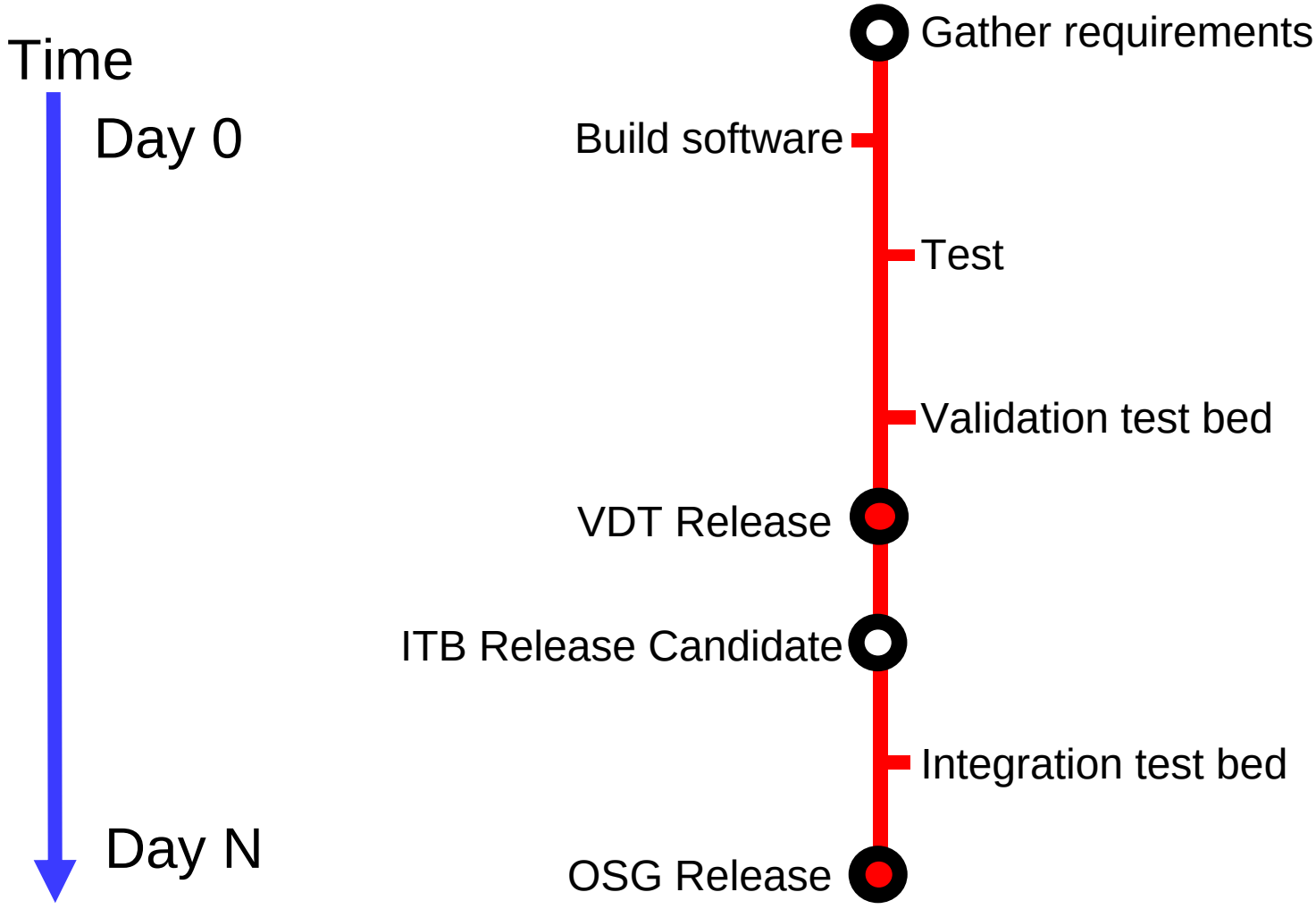
Far better than downloading individual components!

VDT Growth Over 4 Years (1.3.11 now)

www.griphyn.org/vdt/



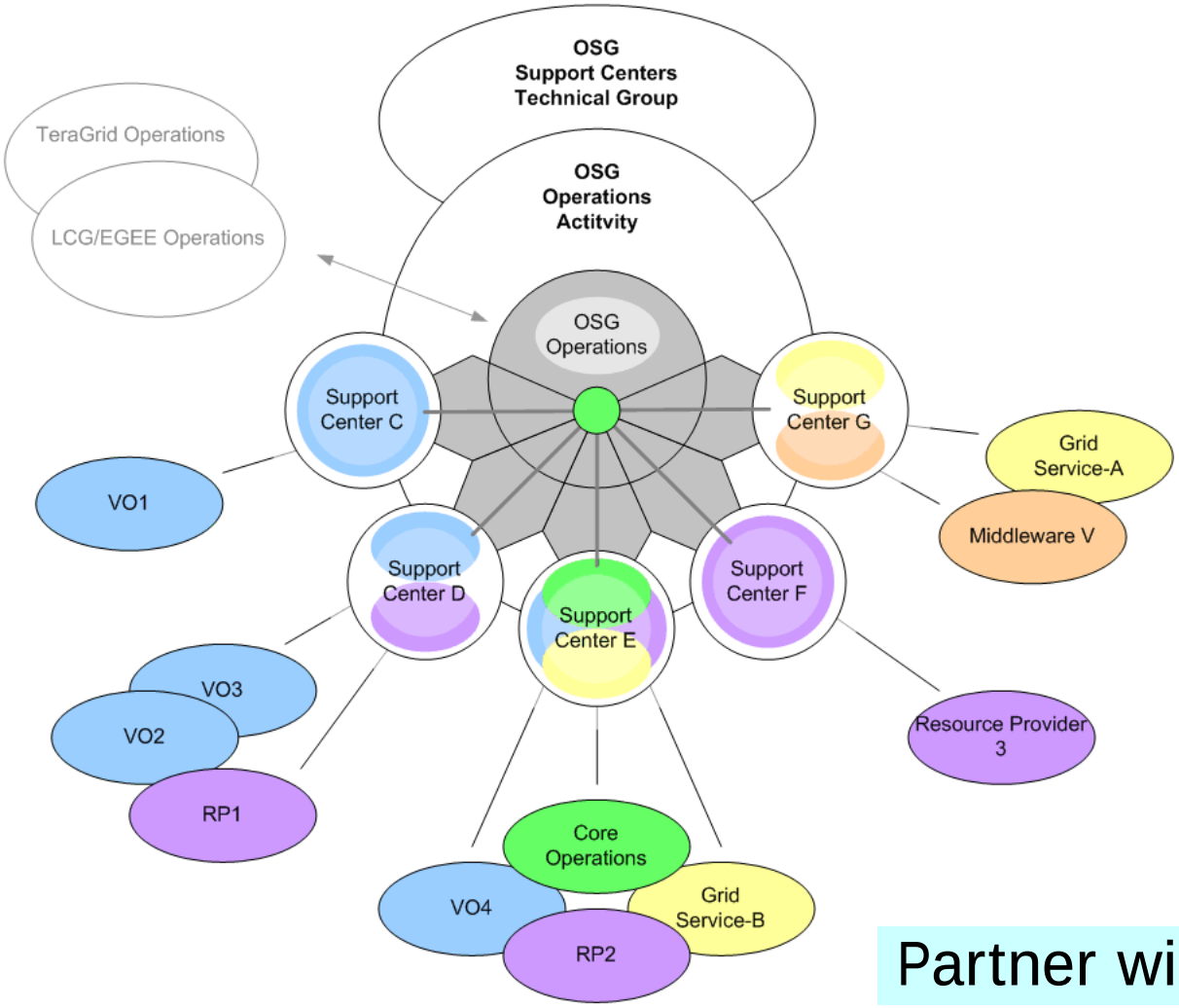
VDT Release Process (Subway Map)



From Alain Roy

OSG Operations And Usage

OSG Operations

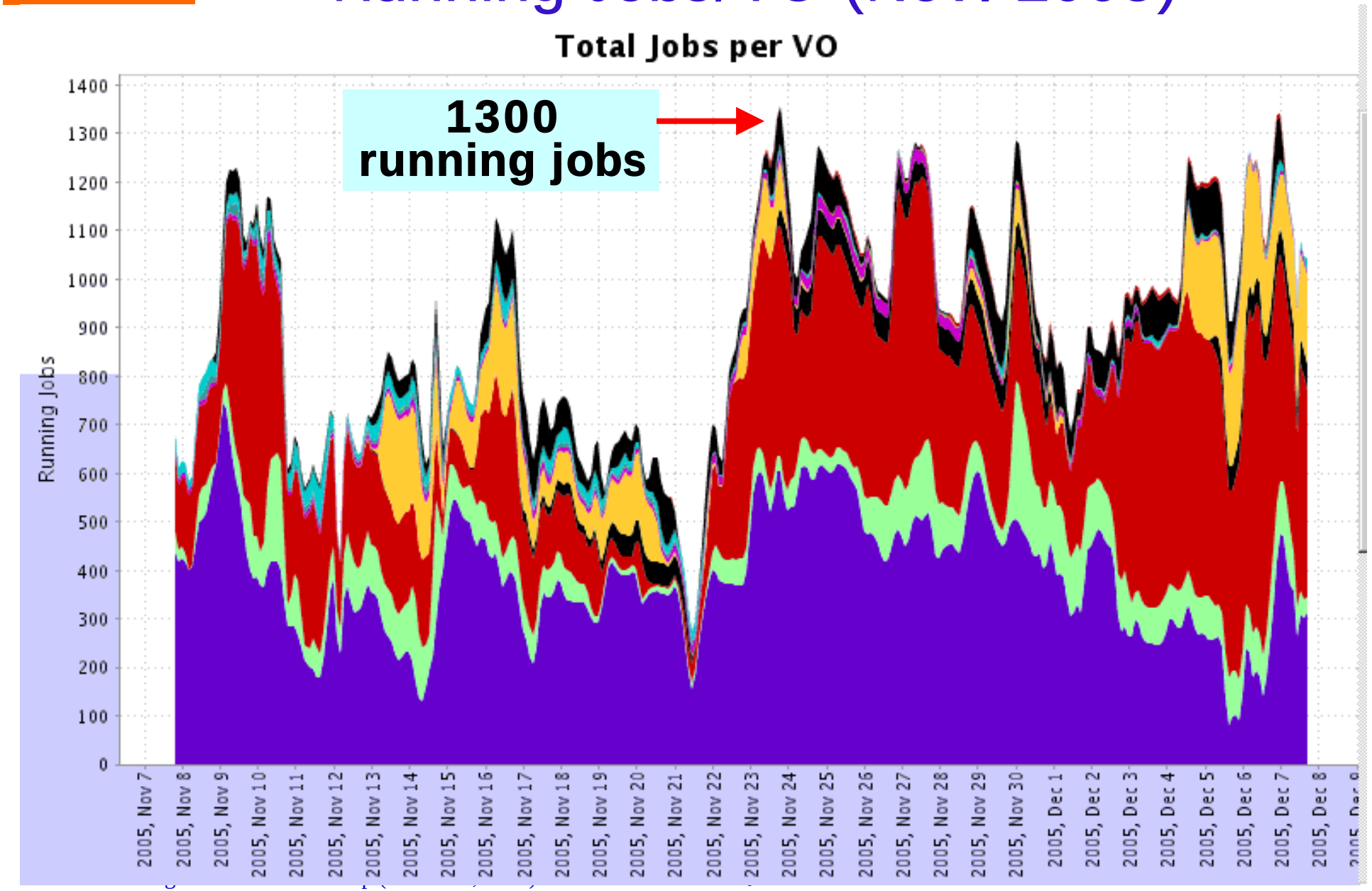


Distributed model

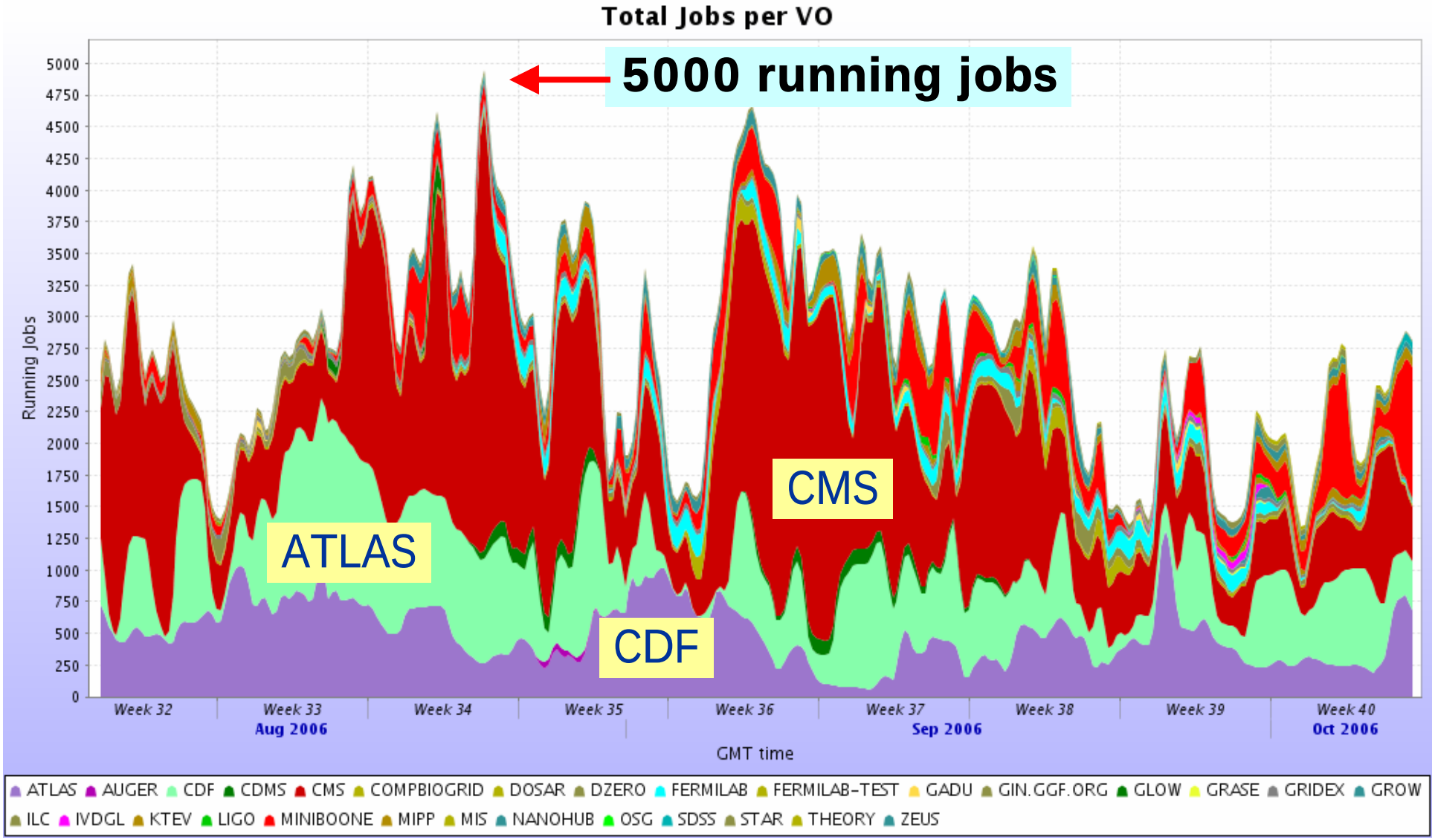
- Scalability!
- VOs, sites, providers
- Rigorous problem tracking & routing
- Security
- Provisioning
- Monitoring
- Reporting

Partner with EGEE operations

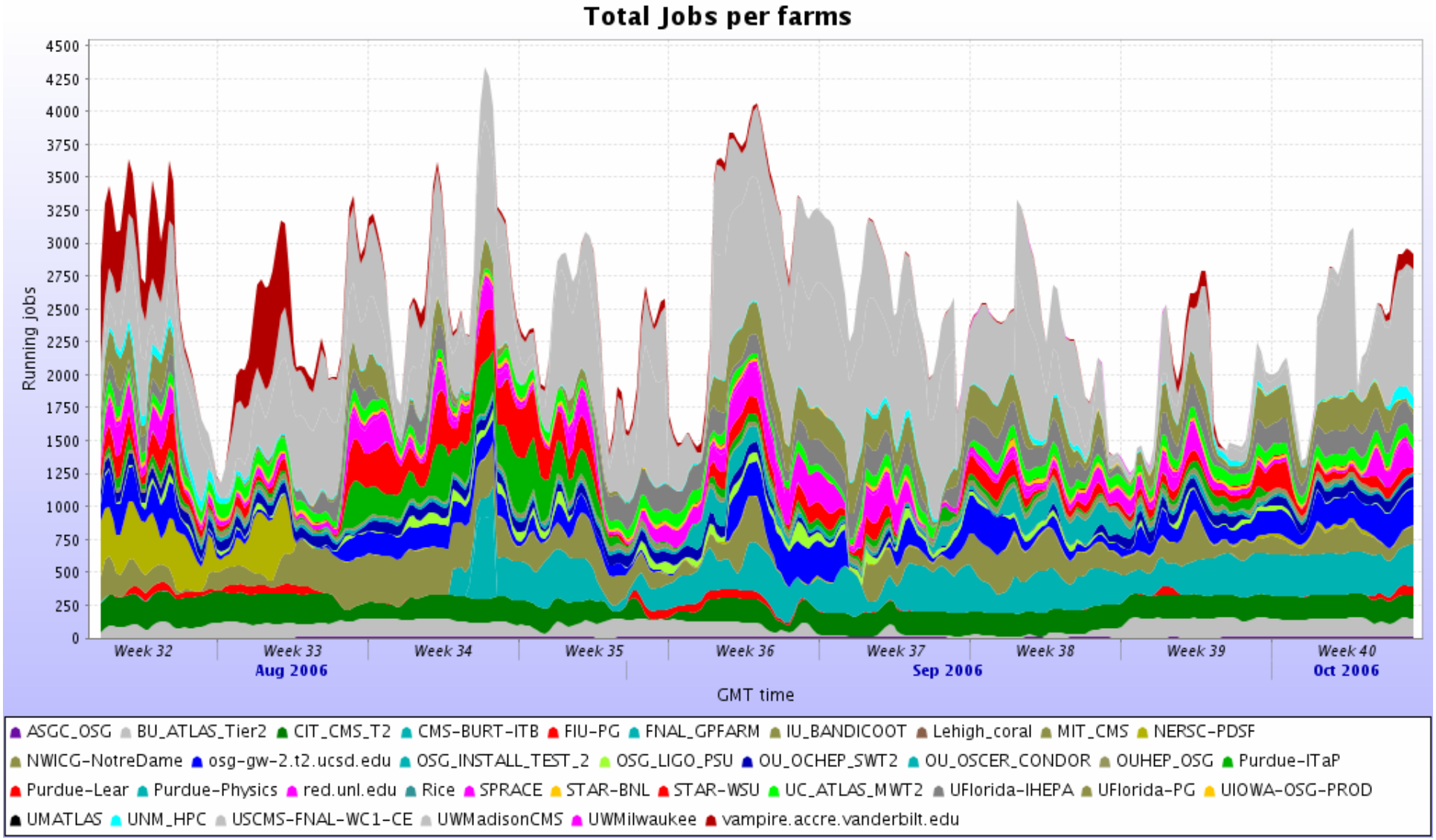
Running Jobs/VO (Nov. 2005)



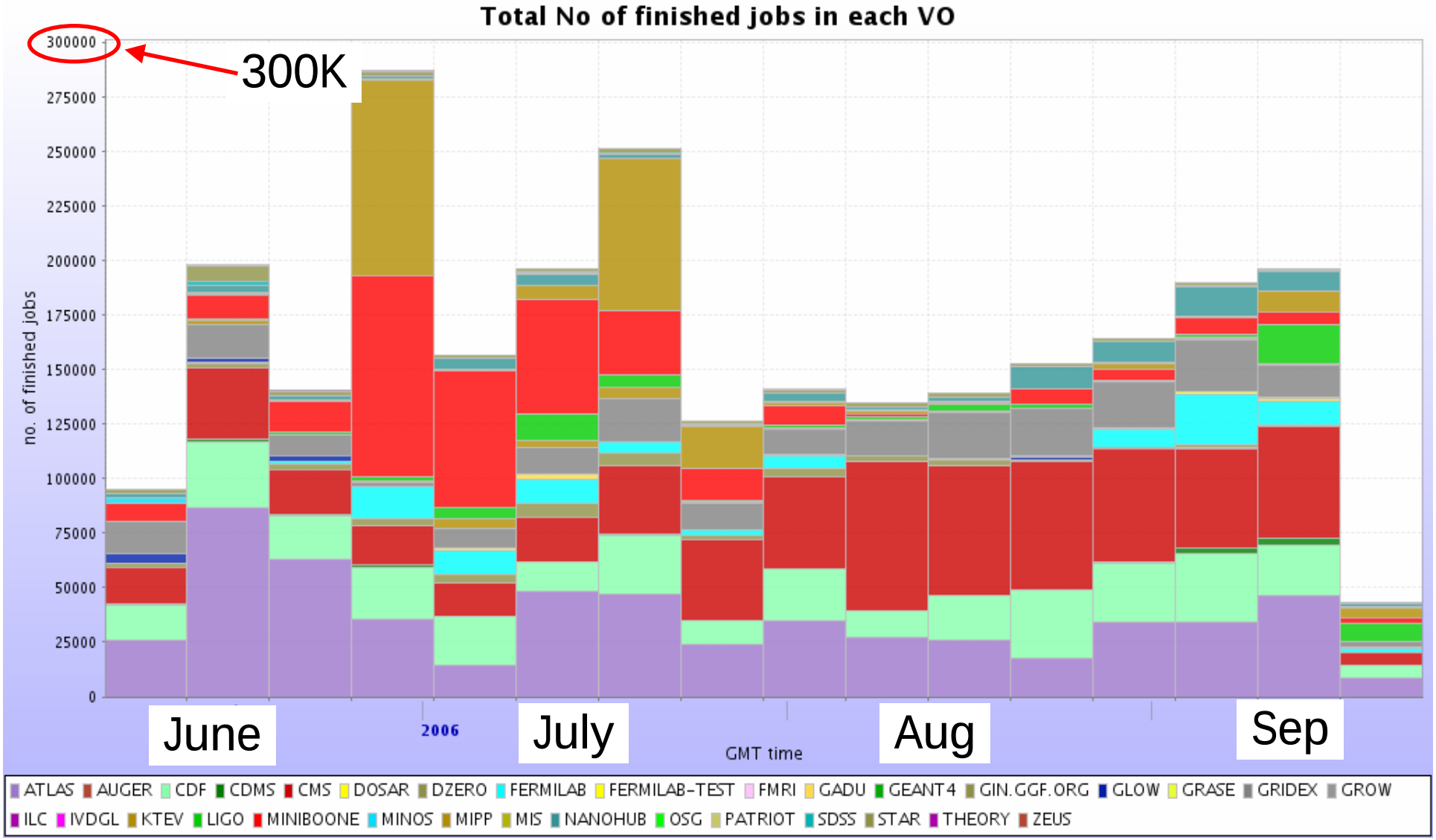
Running Jobs/VO (Aug. – Oct. 2006)



Running Jobs/Site (Aug. – Oct. 2006)



Completed Jobs/Week

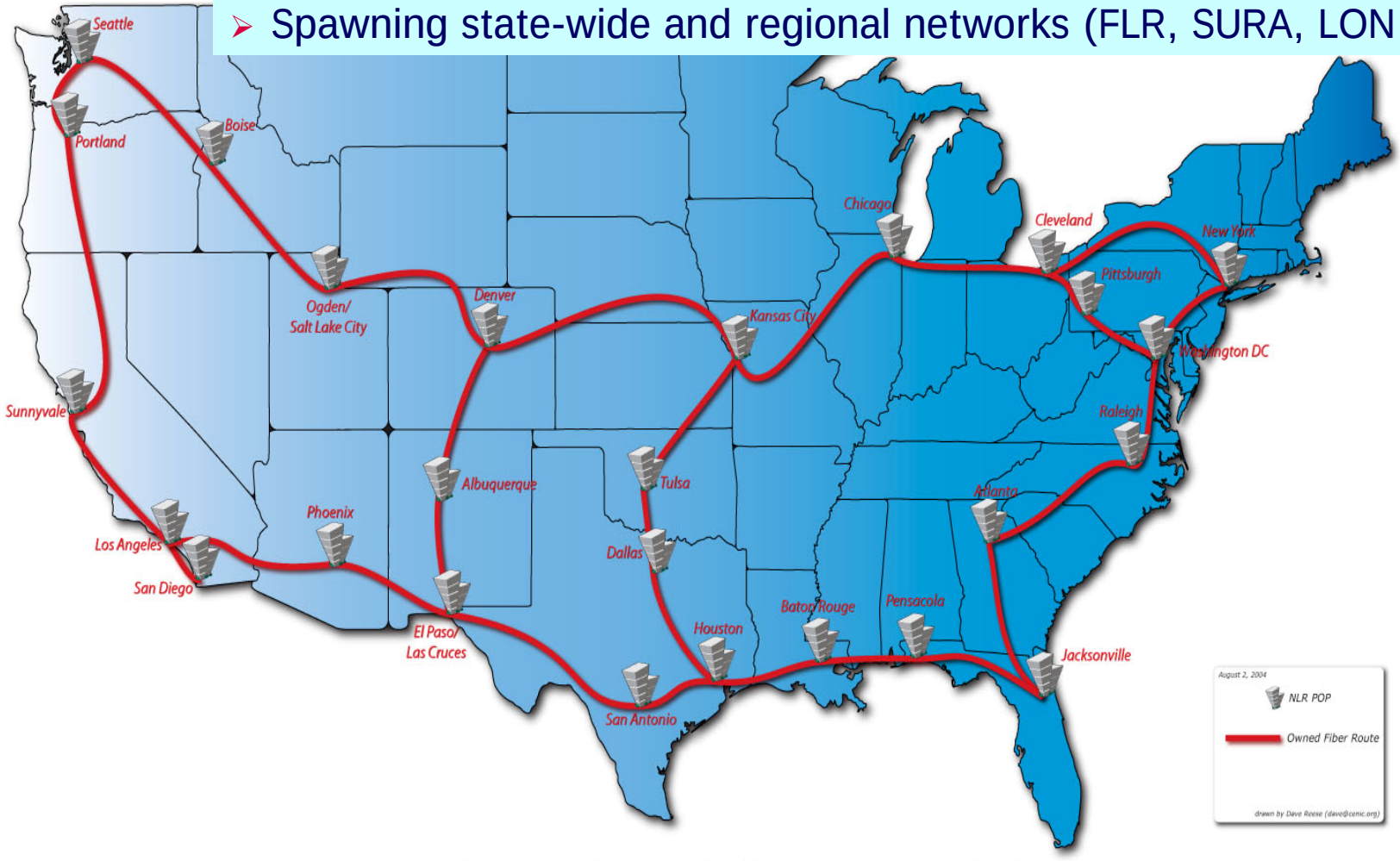


Collaborative Work with Optical Networks

Collaboration with Internet2 www.internet2.edu

Collaboration with National Lambda Rail www.nlr.net

- Optical, multi-wavelength community owned or leased "dark fiber" (10 GbE) networks for R&E
- Spawning state-wide and regional networks (FLR, SURA, LONI, ...)

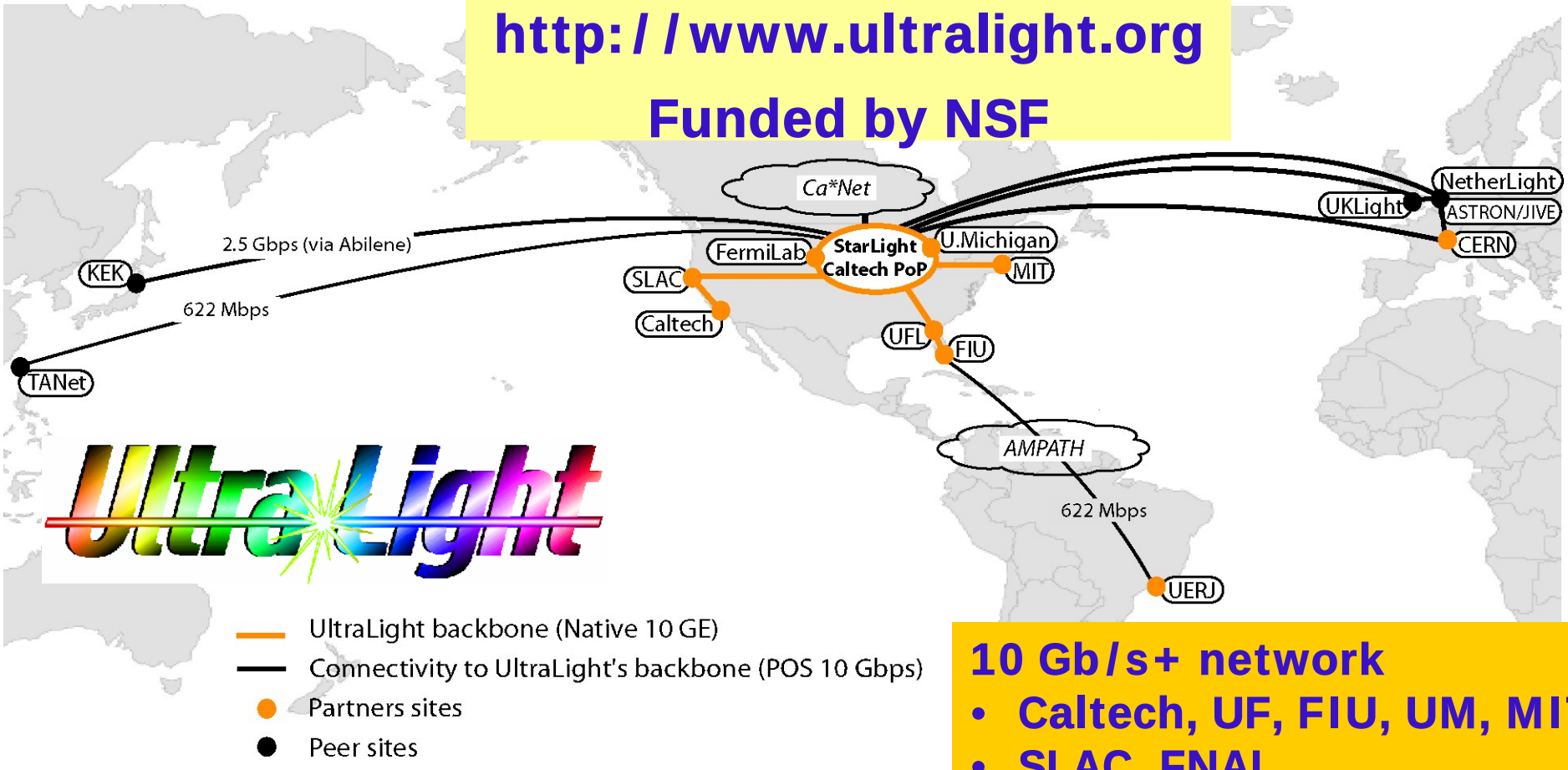


UltraLight

Integrating Advanced Networking in Applications

http://www.ultralight.org

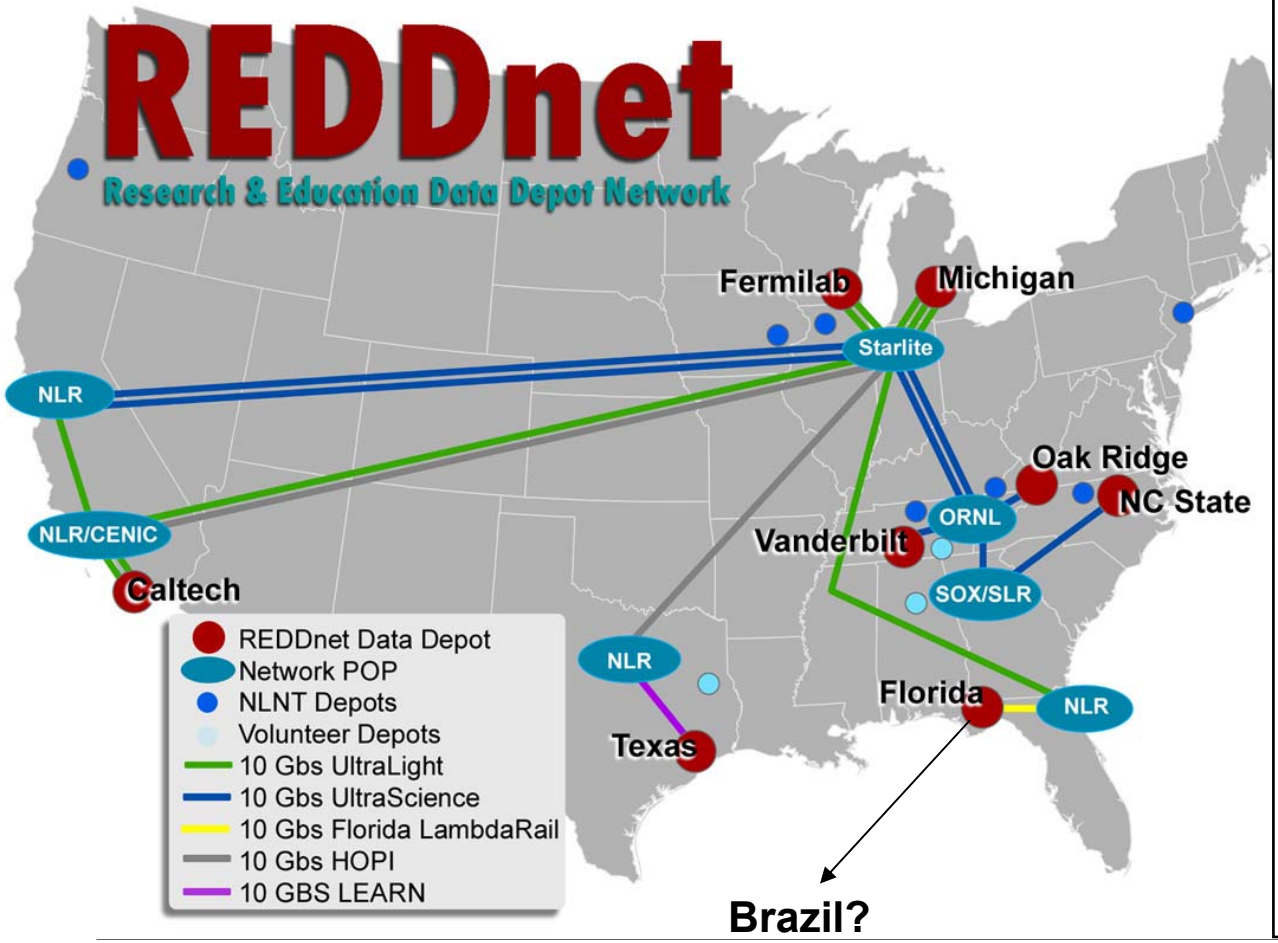
Funded by NSF



10 Gb/s+ network

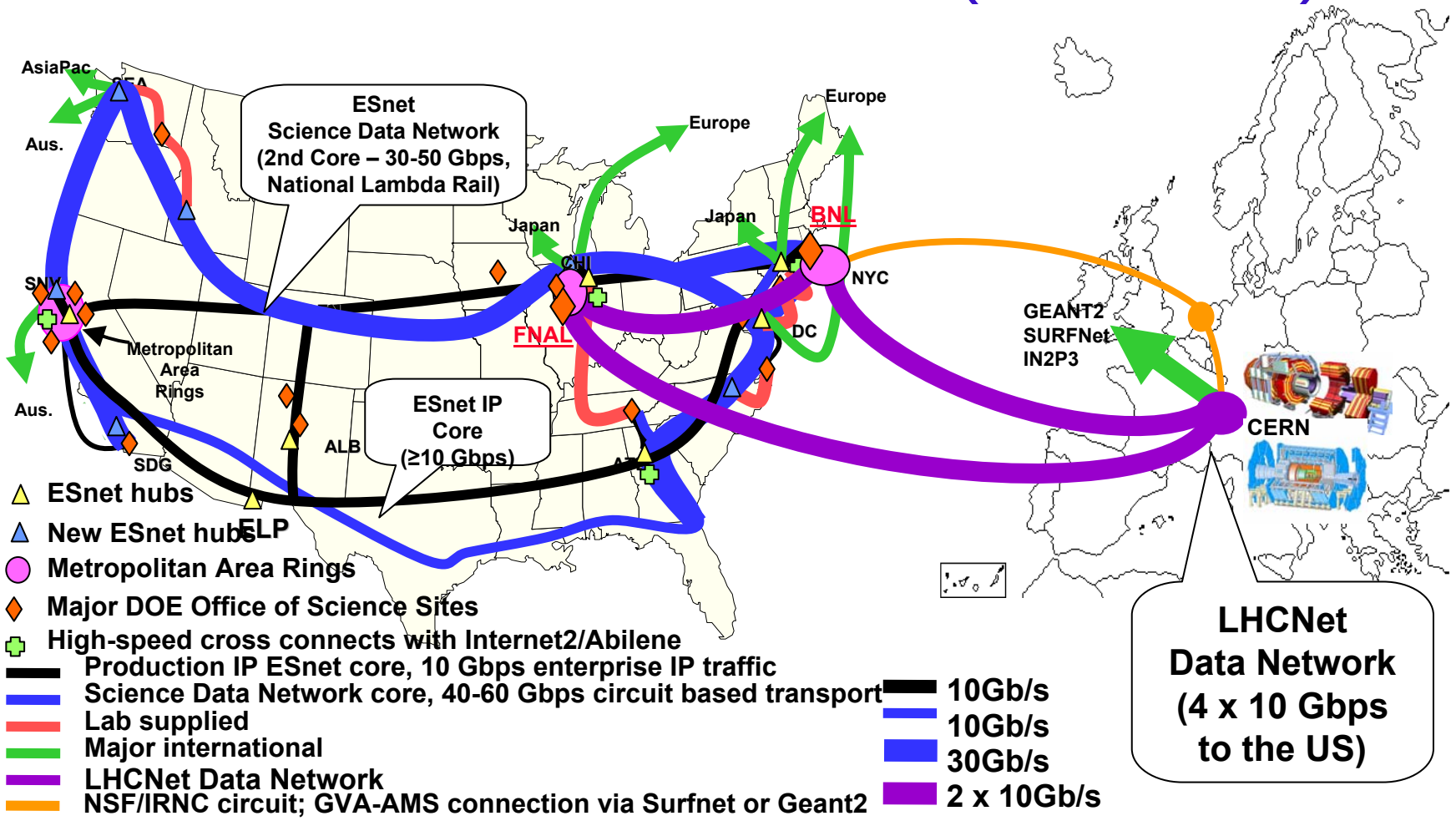
- **Caltech, UF, FIU, UM, MIT**
- **SLAC, FNAL**
- **Int'l partners**
- **Level(3), Cisco, NLR**

REDDnet: National Networked Storage



- NSF funded project
  - Vandebilt
- 8 initial sites
- Multiple disciplines
  - Satellite imagery
  - HEP
  - Terascale Supernova Initative
  - Structural Biology
  - Bioinformatics
- Storage
  - 500TB disk
  - 200TB tape

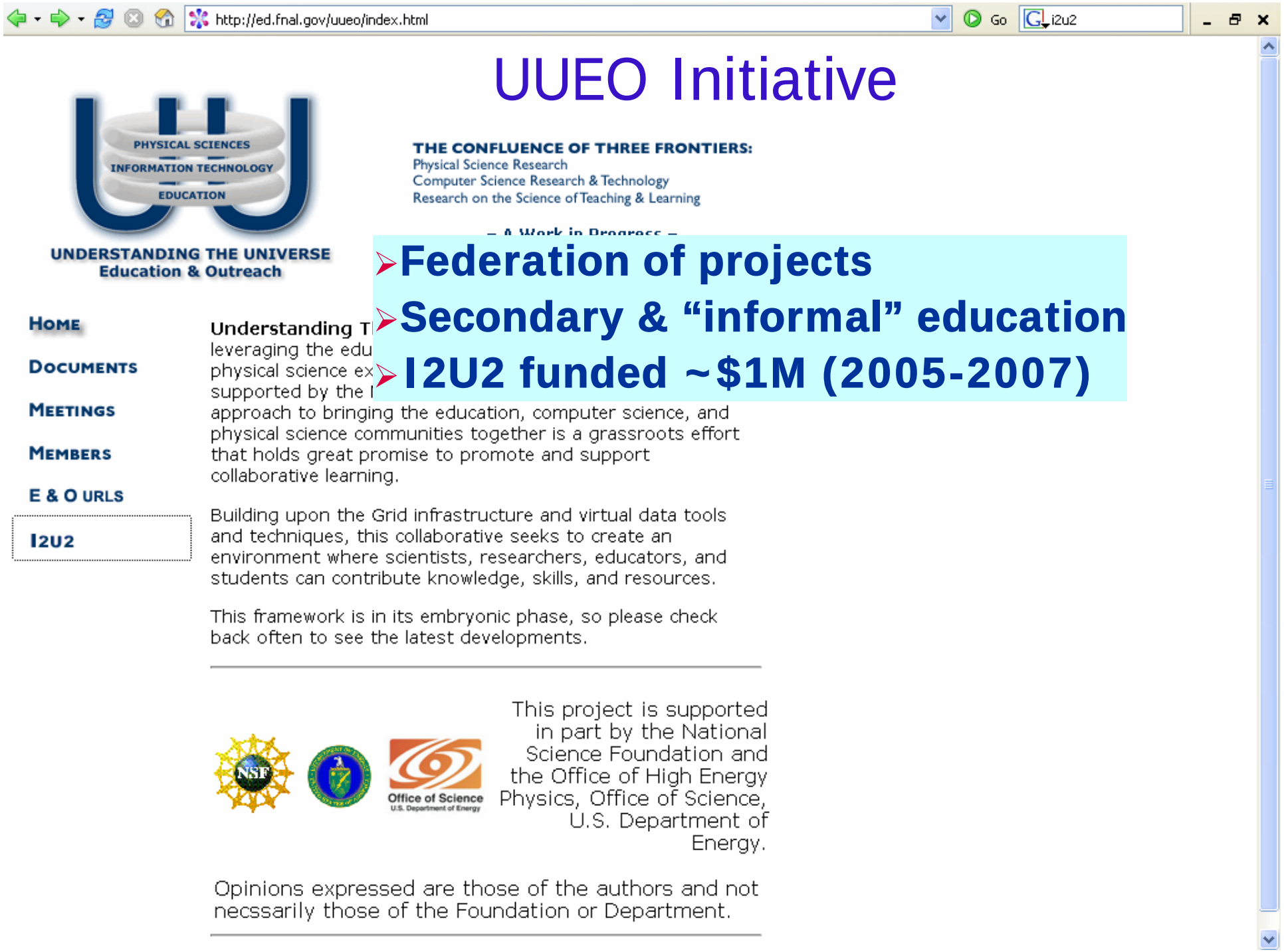
LHCNet: Transatlantic Link to CERN NSF/IRNC, DOE/ESnet (2006/2007)



OSG Training, Outreach, Communications

Grid Summer Schools

- Sponsored by iVDGL + UT Brownsville (2004, 2005, 2006)
  - 1 week @ South Padre Island, Texas
  - Lectures plus hands-on exercises to ~40 students
  - Students of differing backgrounds (physics + CS), minorities
- Aim to reach a wider audience
  - Experiment specific
  - Lectures, exercises, video, on web
  - Students, postdocs, scientists
  - More tutorials, 3-4/year
  - Agency specific tutorials



UUEO Initiative

- **Federation of projects**
- **Secondary & “informal” education**
- **I2U2 funded ~$1M (2005-2007)**

QuarkNet/GriPhyN e-Lab Project

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

Why cosmic rays?

Spending all your time in a shower?

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

Log in

Username:

Password:

Login

- **Analysis of high school cosmic ray data**
- **Now part of I2U2 program (www.i2u2.org)**

from. A number of reseach projects are looking at this question.

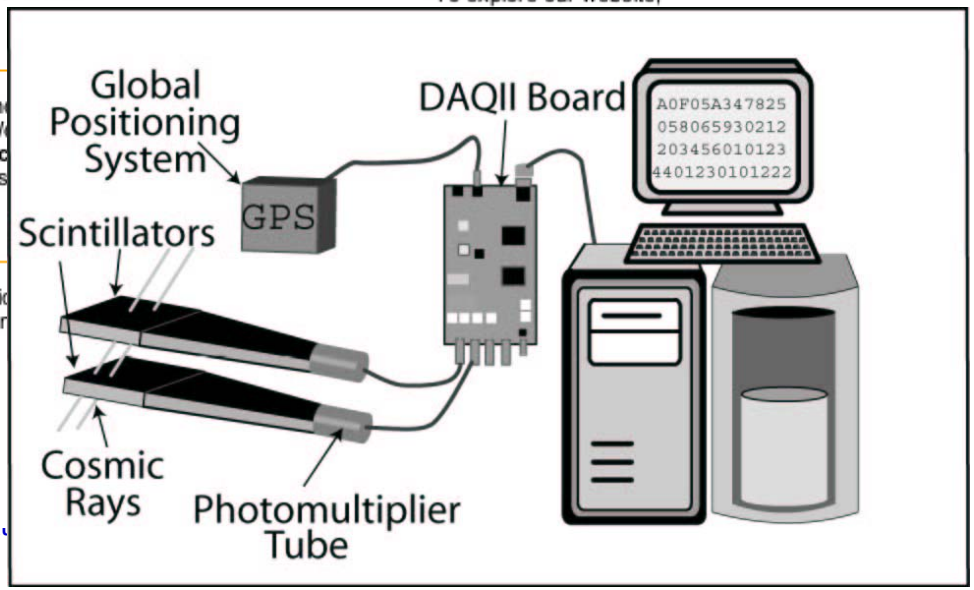
To explore our website,

Who are we?

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

Who can join?

You! Think about steps you'd take to investigate cosmi
started? What do you need to know? Can you collect ar



CHEPREO: Center for High Energy Physics Research and Educational Outreach

Florida International University

www.chepreo.org

Additional initiatives

- **CyberBridges**
- **Global CyberBridges**
- **Networking initiatives**
- **Etc.**

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

- 2003 – 2008 + more?
- $4M: MPS, CISE, EHR, OISE
- New faculty, postdocs

Science Grid Communications: Science Grid This Week

Science Grid This Week

- (Katie Yurkewicz)
- 1.5 years: >1000 subscribers
- Going international in Jan. 2007: "iSGTW"

www.interactions.org/sgtw

http://www.interactions.org/sgtw/

SCIENCE GRID THIS WEEK

MARCH 8, 2006 — ABOUT SGTW | SUBSCRIBE | ARCHIVE | CONTACT 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

Image of the Week

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

On Feburary 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

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

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

OSG Newsletter

Monthly newsletter
- (Katie Yurkewicz)
- 9 issues now

www.opensciencegrid.org/osgnews

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,

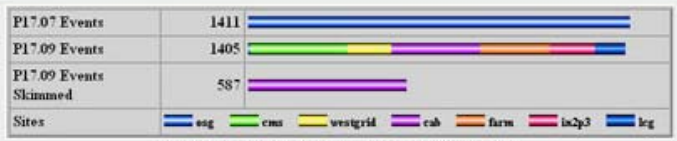
I am very pleased to announce that Bill Kramer has been selected by the Council as its new chairperson. As head of the NERSC computing center, Bill brings a wealth of experience and understanding to our program, and we are already keeping him busy. As one of the new applications coordinators, Frank Würthwein 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.

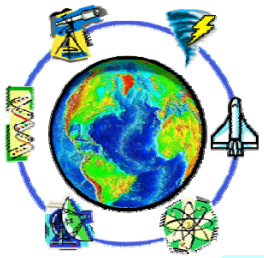
Sincerely,
Ruth Pordes, OSG Executive Director

Applications - Reprocessing D0 Data



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 Technology Cookbook

A guide to building and using grid resources

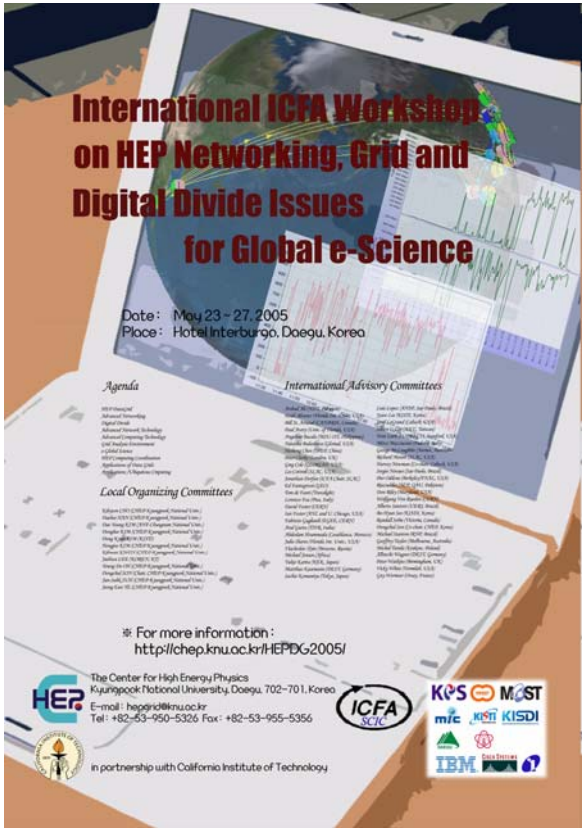
iVDGL + TATRC funded

(Mary Trauner, Mary Fran Yafchak)

- 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

Alleviating the Digital Divide



Background

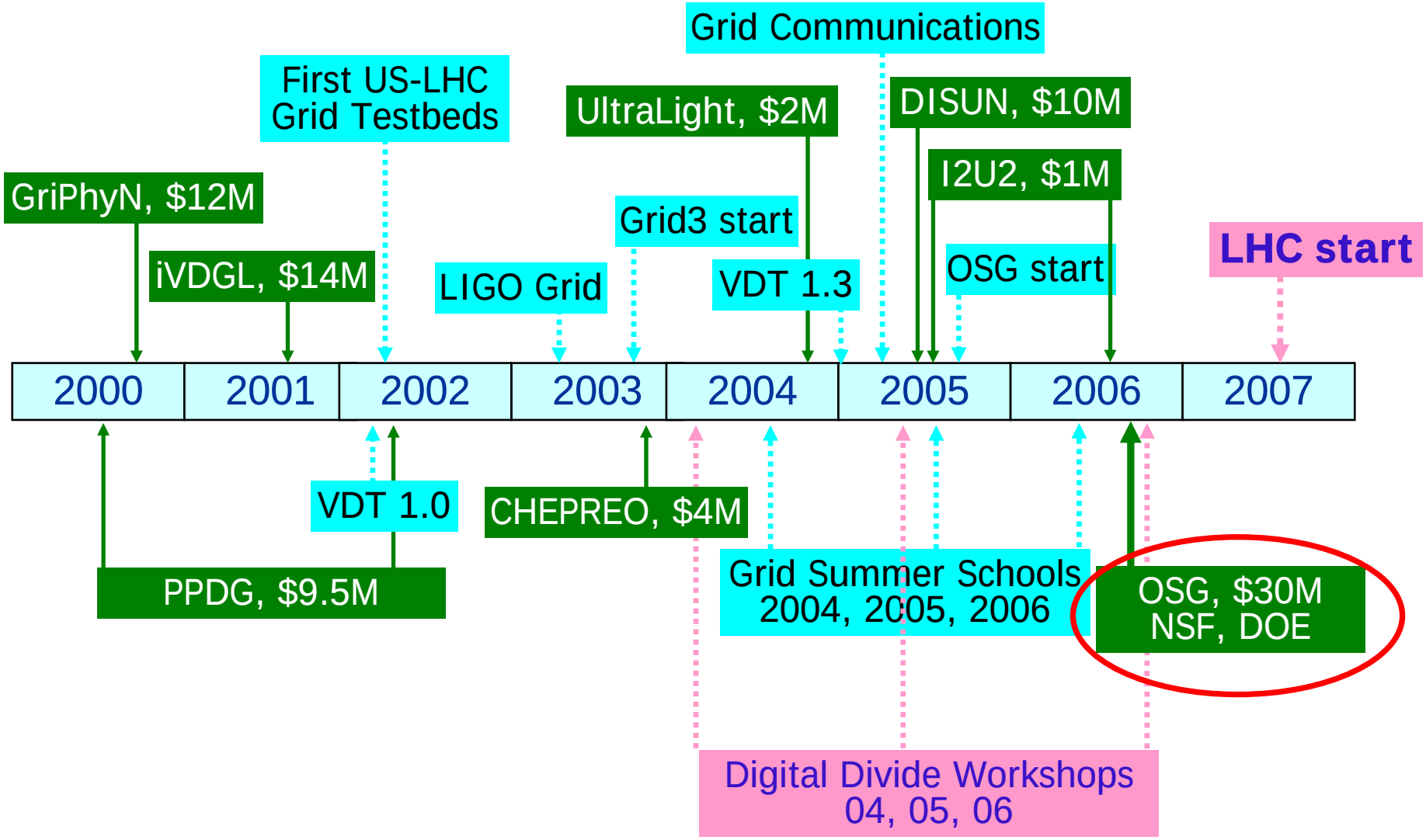
- ICFA/SCIC (Standing Committee on Inter-regional Connectivity)

Themes

- Global collaborations, Grids and addressing the Digital Divide
- Focus on poorly connected regions
- Brazil (2004), Korea (2005), Poland (2006)

OSG Today

OSG: Funding History & Milestones



OSG Project funding (Sep. 7):

- **$30M = $6.1M x 5 yrs**
- **~ 50% NSF: MPS, OISE, OCI**
- **~ 50% DOE: SciDAC II**

ciDAC | The SAMGrid Project

SciDAC
Scientific Discovery through Advanced Co

Office of Science

- Background
- Science Applications
  - Physics
  - Climate
  - Groundwater
  - Fusion Energy
  - Life Sciences
  - Materials & Chemistry
- SciDAC Institutes
- Enabling Technologies
  - Applied Mathematics
  - Computer Science
  - Visualization
- SciDAC Outreach
- Participating Orgs
- Grant Solicitations
  - FY2006
  - FY2005
  - FY2004
  - FY2001
- Collateral Materials
  - SciDAC Review magazine
  - '06 Progress Report (pdf)

Computational Science: Here, There, and Everywhere

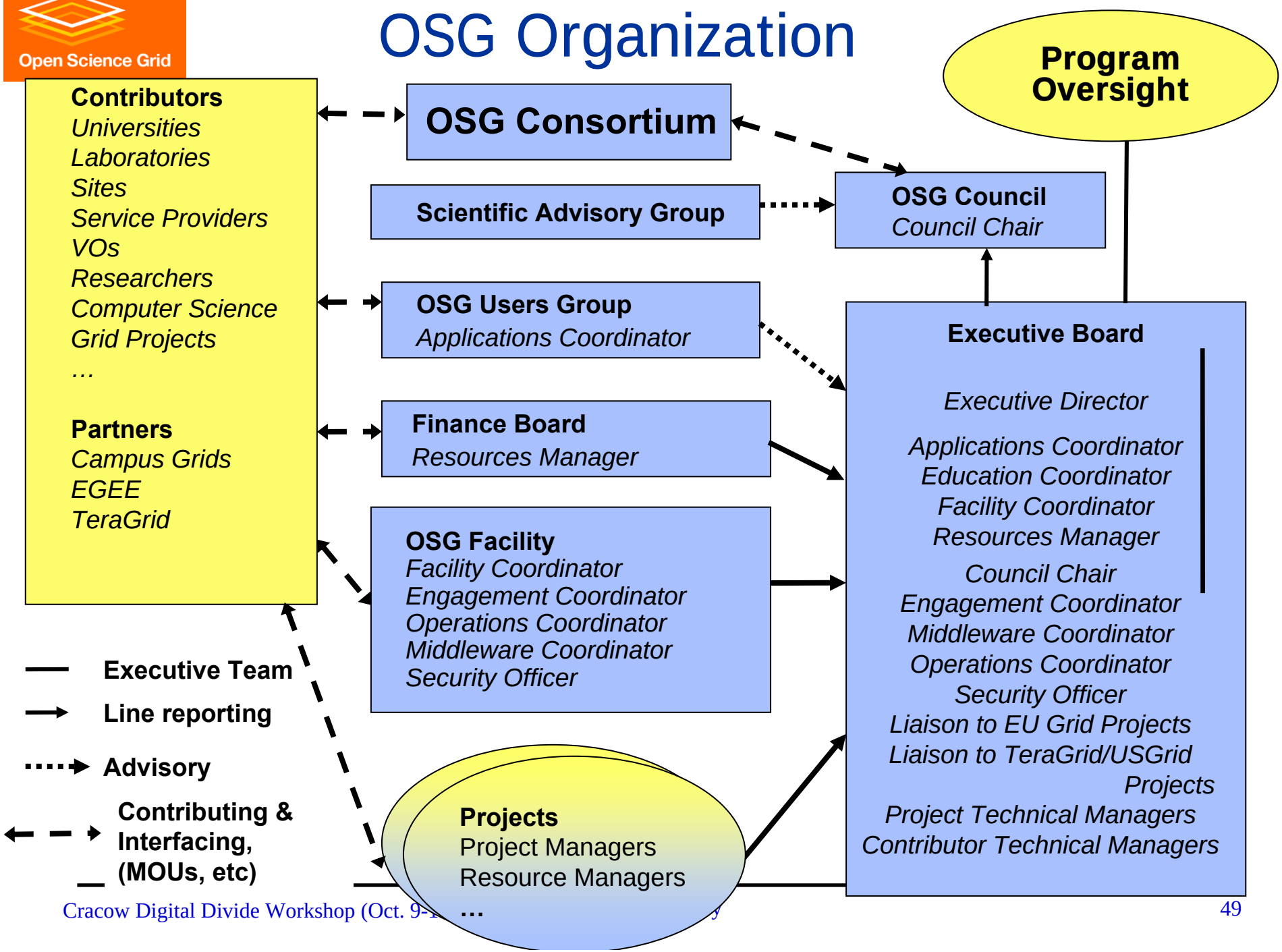
Stimulating new discoveries by providing scientists with effective and dependable access to a petascale distributed computational facility

www.scidac.gov/physics/petascale.html

This project brings together a unique ensemble of domain scientists, software developers and providers of computing resources who share a common goal: to stimulate new discoveries by providing scientists with effective and dependable access to the Open Science Grid (OSG), a national distributed computational facility. The massive amounts of data generated by the current and next generation of physics accelerators and detectors poses significant challenges to our computing and network infrastructure. The requirements in scale of resources, users, capacity and performance of the OSG facility are driven by the user communities, in particular the physics communities that are committed to the use of OSG. This project will maintain and operate a Petascale nationwide distributed facility that can grow to provide thousands of users at universities and DOE laboratories throughout the U.S. with effective access to massive computational and storage resources. Technical activities to engage train and include new researchers are integral parts of our program of work. The engagement activity will bring each new community to contribute to and benefit from the facility. New IT technologies are integrated and deployed in response to explicit needs and are evaluated in a real-life setting.

A reliable national infrastructure that can deal with the data management and analysis of petabytes of data from the next generation of physics accelerators and detectors is vital to maximizing the benefit of U.S. investments in these experiments. Without this effort, it would be difficult for the U.S to optimally exploit the Large Hadron Collider (LHC) competitively and to capitalize on the investment in the Laser Interferometer Gravitational-Wave Observatory (LIGO). This effort proposes a production environment for distributed data-intensive science provided through a consortium that consists of a unique ensemble of domain scientists, software developers and providers of computing resources using distributed computing tools and

OSG Organization



Project Execution Plan (PEP) - FTEs

	FTEs
Facility operations	5.0
Security and troubleshooting	4.5
Software release and support	6.5
Engagement	2.0
Education, outreach & training	2.0
Facility management	1.0
Extensions in capability and scale.	9.0
Staff	3.0
Total FTEs	**33**

OSG Project Effort Distribution: Year 1

Developing procedures and structures for coherent project

- Each institution must sign Statement of Work. Taking place now.
- Each individual submits open monthly written reports. Fall 2006
- Finance Board reviews the accounts and deliverables. FB exists
- Executive Board reviews plans and achievements. EB exists
- Activities covered by the Project Plan and WBS. PEP & WBS exist
- Effort distribution reviewed & potentially modified each year.

Boston	0.5
BNL	3.0
CalTech	2.5
Columbia	0.5
Cornell	0.0
FermiLab	7.0
Indiana U	3.0
LBNL	1.8
RENCI	1.5
SLAC	0.5
UCSD	2.0
U of Chicago	3.5
U of Florida	0.5
U of Iowa	1.0
Wisconsin	6.5
Totals	**33.8**

OSG PEP - High Level Milestones

2006Q3	Release OSG software stack version 0.6.0
2006Q3	Project baseline review
2006Q4	Sign off on OSG Security Plan.
2006Q4	Meet operational metrics for 2006.
2007Q1	Accounting reports available for users and resource owners.
2007Q2	Production use of OSG by one additional science community.
2007Q2	OSG-TeraGrid: software releases based on same NMI software base.
2007Q2	Release OSG software version 0.8.0: Complete extensions for LHC data taking.
2007Q2	Support for ATLAS and CMS data taking.
2007Q3	1 year Project Review.
2007Q4	Meet 2007 deliverables as defined by science stakeholders.
2007Q4	Meet operational metrics for 2007.
2007Q4	Release OSG software version 1.0
2008Q2	Production use of OSG by 2 additional science communities.
2008Q3	OSG-TeraGrid: production service interoperation.
2008Q3	2nd year Project Review.
2008Q4	Meet 2008 deliverables as defined by science stakeholders.
2008Q4	Meet operational metrics for 2008.
2009Q2	Support for all STAR analysis (10,000 jobs/day).
2010Q1	Support for data taking with order of magnitude increase in LIGO sensitivity.

Security, Safety, Risk Management

- Assess, monitor & respond to security issues ⇒ Security Officer
- Each site responsible for local security and incident reporting
- OSG security plan modeled on NIST process

Area of Risk	Description	Probability of Occurrence	Impact of Occurrence	Strategy for Mitigation
Schedule	Missed milestones from internal or external activities	High	Medium	Monthly review of deliverables and milestones; Technical project leaders are members of OSG Executive Board. Well-developed integration and system test program. Consistent project management; an experienced, professional staff;
Operational	Security Incident	High	Medium	Security policies and procedures operational and exercised. Ongoing attention to NIST security management, operations and technology for OSG assets. Include security issues in all activities and meetings.
Operational	Faster growth and expansion than expected.	Low	Medium	Maintain flexibility and agility to reprioritize and move effort to facility and operations tasks.
Organizational	Withdrawal of significant stakeholder	Low	High	Monthly reporting and analysis of use of OSG and satisfaction of the stakeholders. Energetic communication and proactive follow up by Executive Team with users.
Organizational	Changes in staff	Moderate	Moderate	Team approach and cross-training to have technical depth in all activities. Build and test procedures included in project plan to validate access to all codes, and provide documentation of distributions.
Technical	Technical solution not available	Moderate	Moderate	Ongoing tracking of requirements and delivery. Early testing of needed capabilities on testbeds.
Technical	Support for software dropped by developer.	High	Medium	OSG software and release teams have some technical knowledge about all components that are built and tested. Availability of the source code for an OSG distribution. Include external project technical leads in the Executive Board
Technical	Technology or infrastructure fails in its capabilities.	High	Moderate	Frequent integration and validation of technology use at scale and with realistic use cases. Definition and review of system performance metrics.

Scaling of LHC & LIGO in 2008-2009

- Data distribution
  - Routinely >1 GB/Sec at ~10-20 sites
- Workflows
  - >10,000 batch jobs per client
- Jobs/Day
  - >20,000 per VO with >99% success rate
- Accessible Storage
  - >10PB
- Facility Availability/Uptime
  - >99.x% with no single points of failure

Continued focus on OSG Core Competencies

- Integration
  - Software, Systems, Virtual Organizations
- Operations
  - Common support & grid services
- Inter-Operation with other grids
  - TeraGrid, EGEE, caBIG, ...
  - Bridging administrative & technical boundaries

- With validation, verification and diagnosis at each step
- With integrated security operations and management

Core Operations and Common Support

Join OSG: 1-2-3

1. VO registers with Operations Center, users register with VO
2. Sites register with Operations Center
3. VOs and sites provide Support Center Contact and join Ops groups

The OSG VO

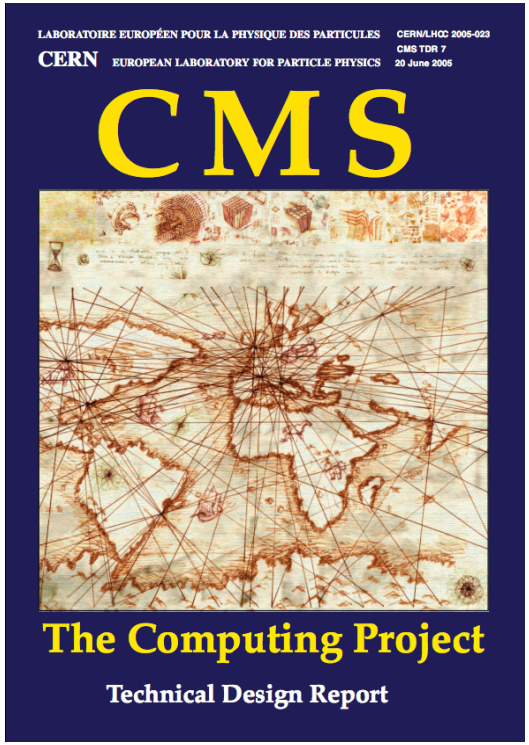
- Individuals & small groups
- Managed by OSG
- Good learning environment

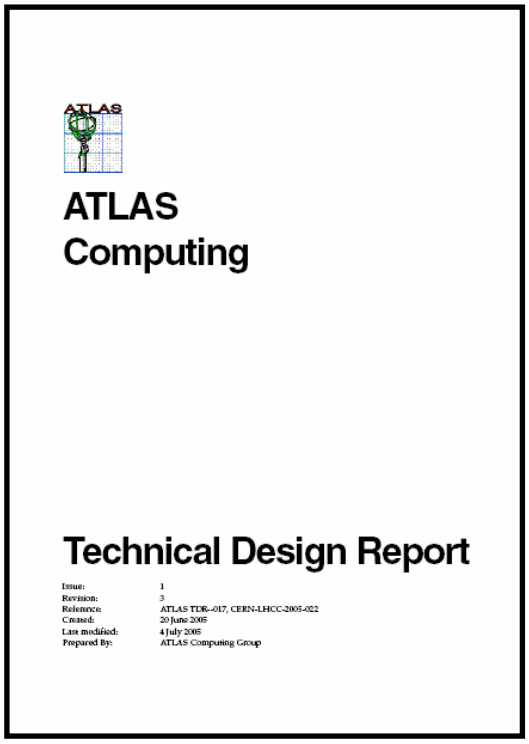
END

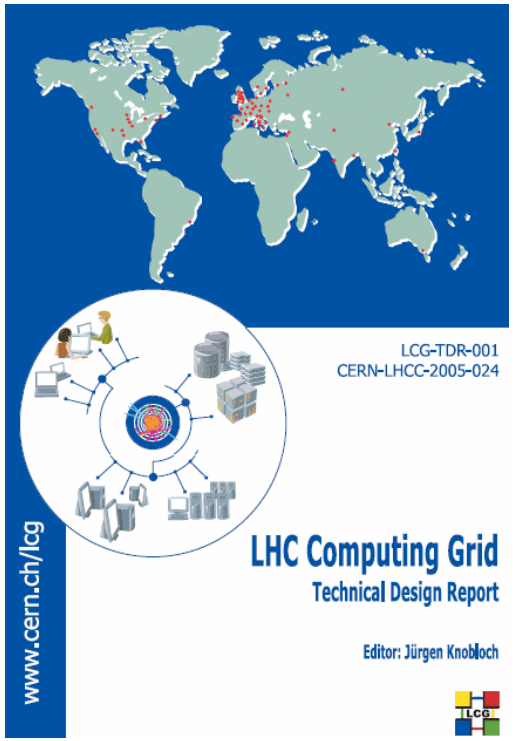
Future Astronomy OSG Projects

- Fermilab Experimental Astrophysics Group (EAG) has 4 projects planned for Open Science Grid
  - Fitting SDSS Quasar Spectra by genetic algorithm
  - Simulation effort for Dark Energy Survey (DES)
  - Search for Near Earth Asteroids (NEOs) in the SDSS Imaging data
  - The Co-addition of the SDSS Southern Stripe (COADD)

Integral Role of Computing at LHC: TDRs







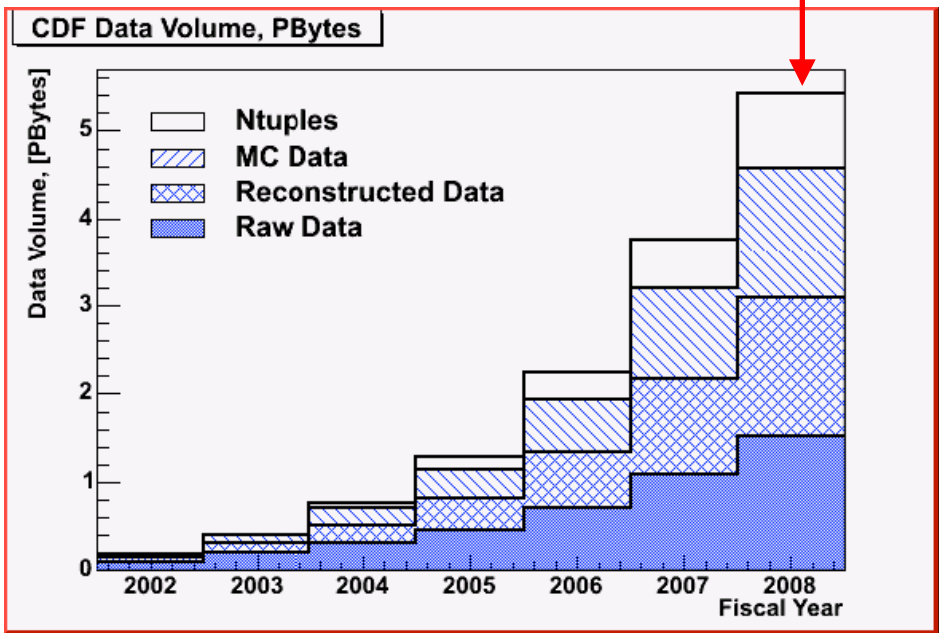
100s of pages apiece

- CPU
- Storage
- International optical networks

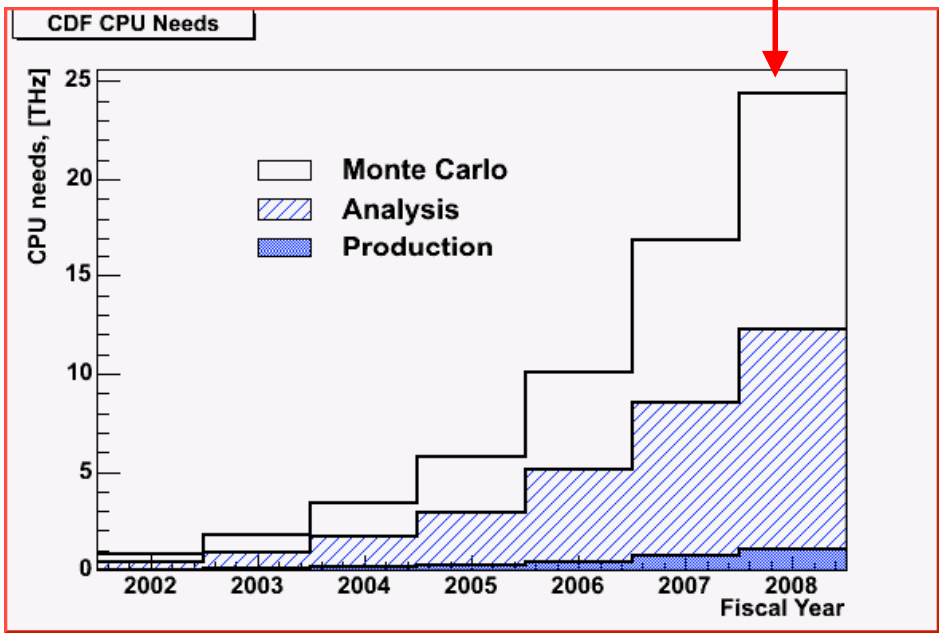
CPU, Storage Projections of Current HEP Expts.

- **Primarily driven by increasing datataking rates**
- **Similar increases in other disciplines**

2008 Data volume: 5.7 PB

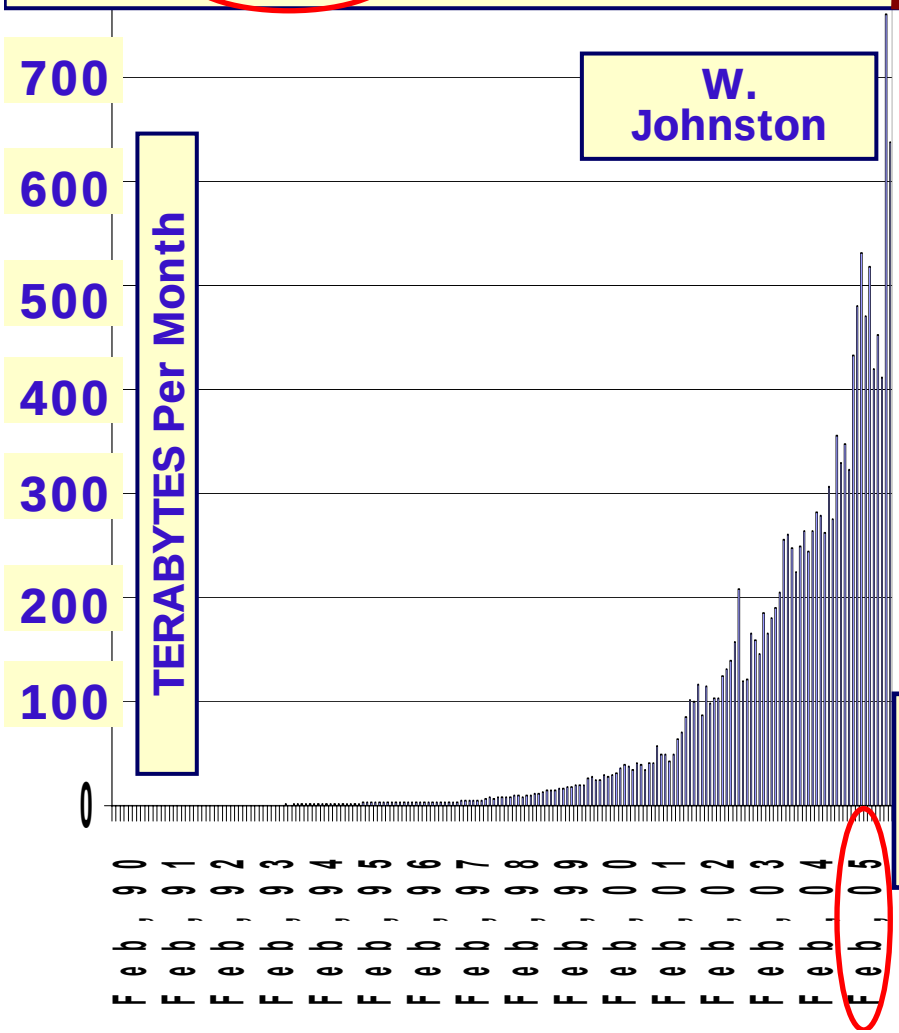


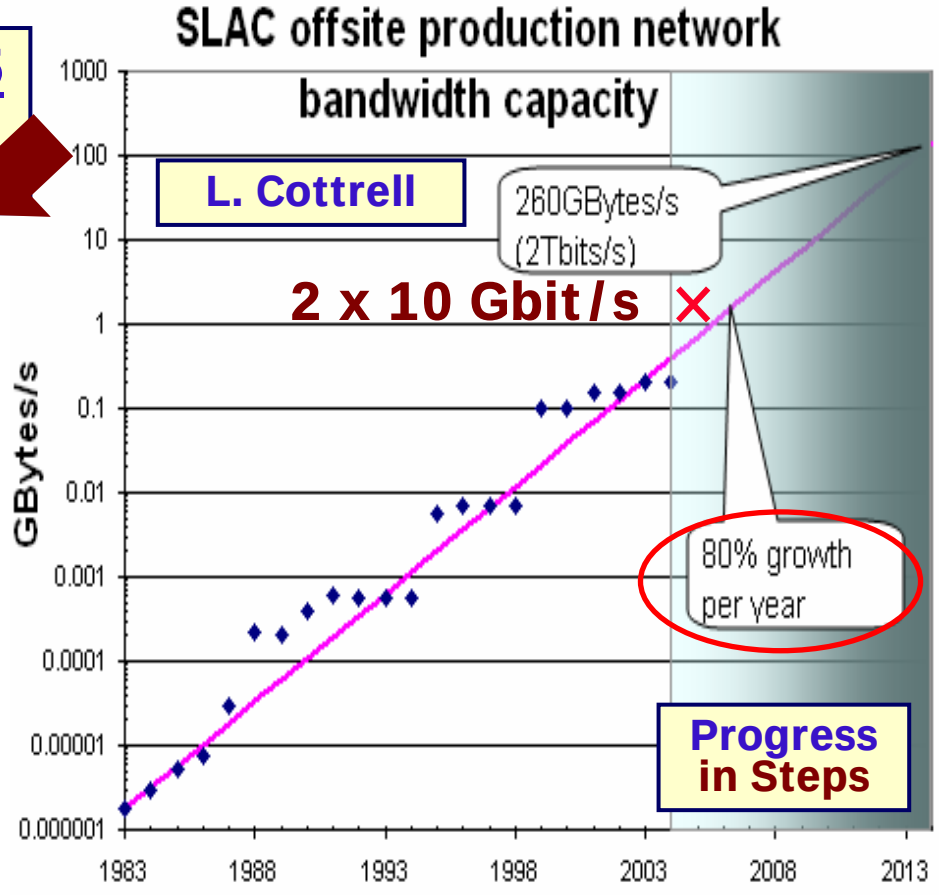
2008: ~8,000 3 GHz CPUs



Long Term Trends in Network Traffic Volumes: 300-1000X/10Yrs

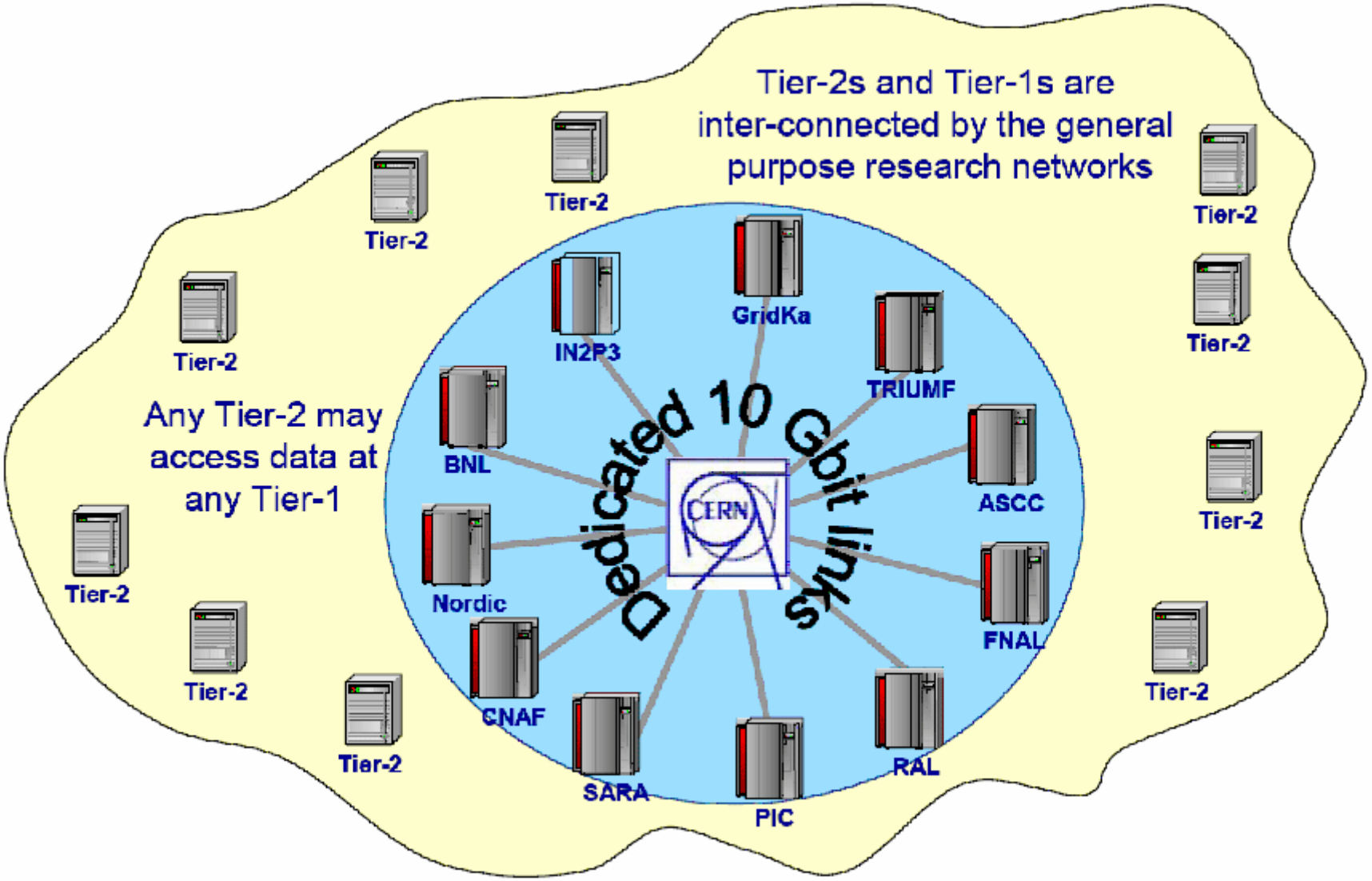
ESnet Accepted Traffic 1990 – 2005
Exponential Growth:
Avg. +82%/yr for the Last 15 Years





- 2005 SLAC Traffic ~400 Mbps
- Growth in steps (ESNet Limit): ~ 10X/4 years

LHC Tier0–Tier1–Tier2 Network Connections



OSG Roots: "Trillium Consortium"

- Trillium Grid projects: PPDG + GriPhyN + iVDGL
  - PPDG: $12M (DOE) (1999 – 2006)
  - GriPhyN: $12M (NSF) (2000 – 2005)
  - iVDGL: $14M (NSF) (2001 – 2006)
  - Large science experiments: ATLAS, CMS, LIGO, SDSS
  - Supplements + new projects : UltraLight, CHEPREO, DISUN ($17M)
- Total ~150 people with many overlaps between projects
  - Universities, labs, SDSC, 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
  - Unified entity when collaborating internationally

GLORIAD: 10 Gbps Global Optical Ring (Complete by March 2007)

China, Russia, Korea, Japan, US, Netherlands

US: NSF IRNC Program

GLORIAD Circuits Today

- **10 Gbps Hong Kong-Daejon-Seattle**
- **10 Gbps Seattle-Chicago-NYC**
- **622 Mbps Moscow-AMS-NYC**
- **2.5 Gbps Moscow-AMS**
- **155 Mbps Beijing-Khabarovsk-Moscow**
- **2.5 Gbps Beijing-Hong Kong**
- **1 GbE NYC-Chicago (CANARIE)**