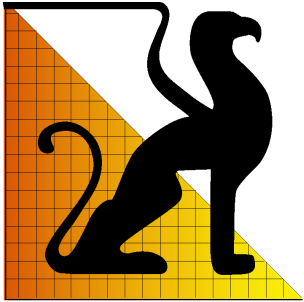
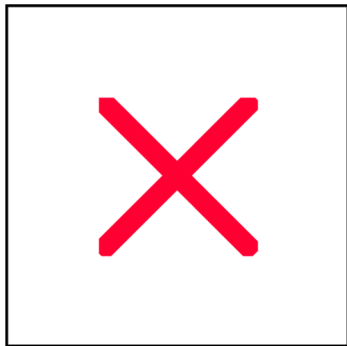


U.S. Grid Projects and Open Science Grid

GriPhyN



Data Intensive Science



UltraLight

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



Chinese American Networking
Symposium

Florida International University
December 1, 2004



CHEPREO

U.S. “Trillium” Grid Consortium

➤ Trillium = PPDG + GriPhyN + iVDGL

- ◆ Particle Physics Data Grid: \$12M (DOE) (1999 – 2004+)
- ◆ GriPhyN: \$12M (NSF) (2000 – 2005)
- ◆ iVDGL: \$14M (NSF) (2001 – 2006)

➤ Basic composition (~150 people)

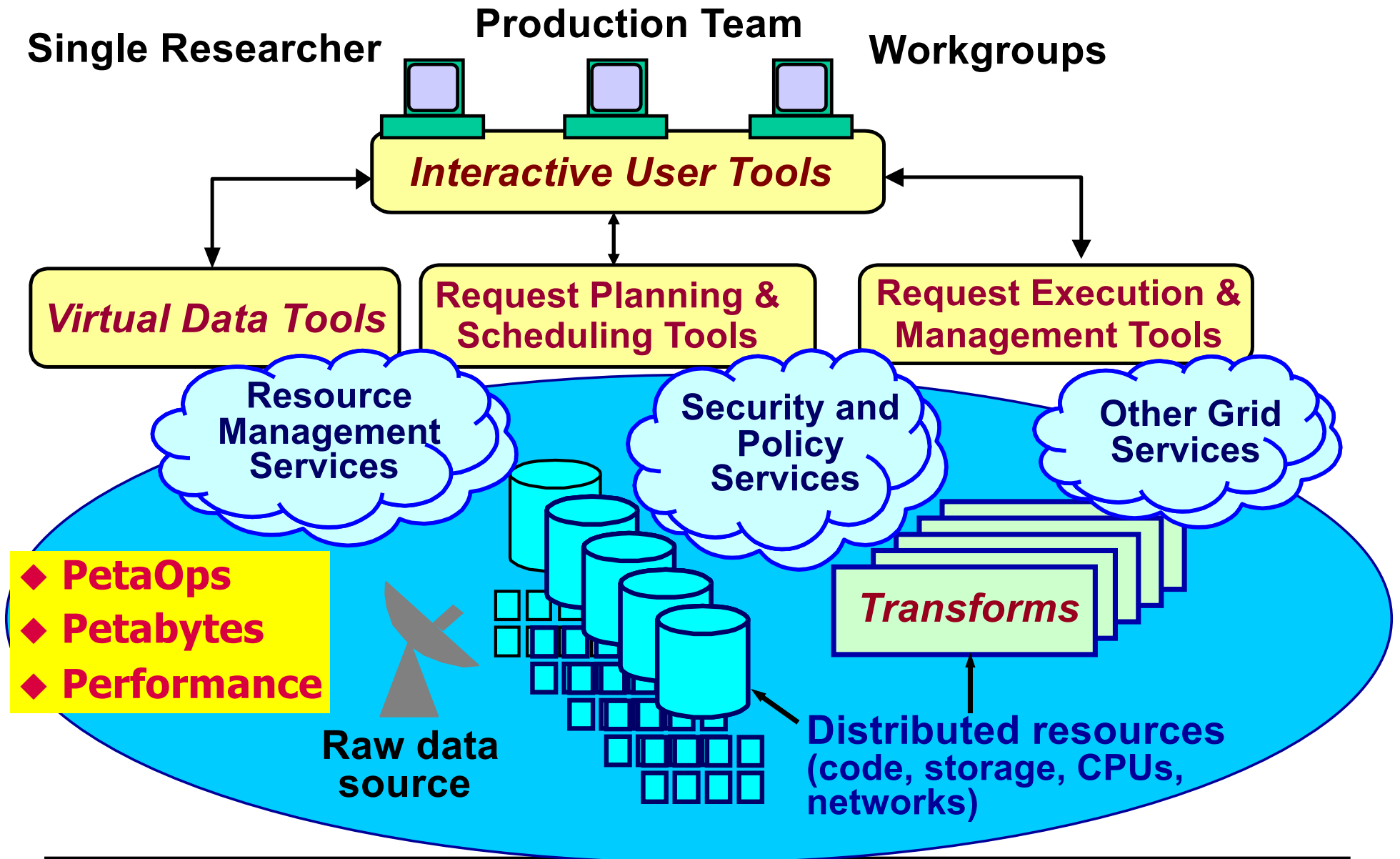
- ◆ PPDG: 4 universities, 6 labs
- ◆ GriPhyN: 12 universities, SDSC, 3 labs
- ◆ iVDGL: 18 universities, SDSC, 4 labs, foreign partners
- ◆ Expts: BaBar, D0, STAR, Jlab, CMS, ATLAS, LIGO, SDSS/NVO



➤ Complementarity of projects

- ◆ GriPhyN: CS research, Virtual Data Toolkit (VDT) development
- ◆ PPDG: “End to end” Grid services, monitoring, analysis
- ◆ iVDGL: Grid laboratory deployment using VDT
- ◆ Experiments provide frontier challenges
- ◆ Unified entity when collaborating internationally

Goal: Peta-scale Virtual-Data Grids for Global Science



Trillium Science Drivers

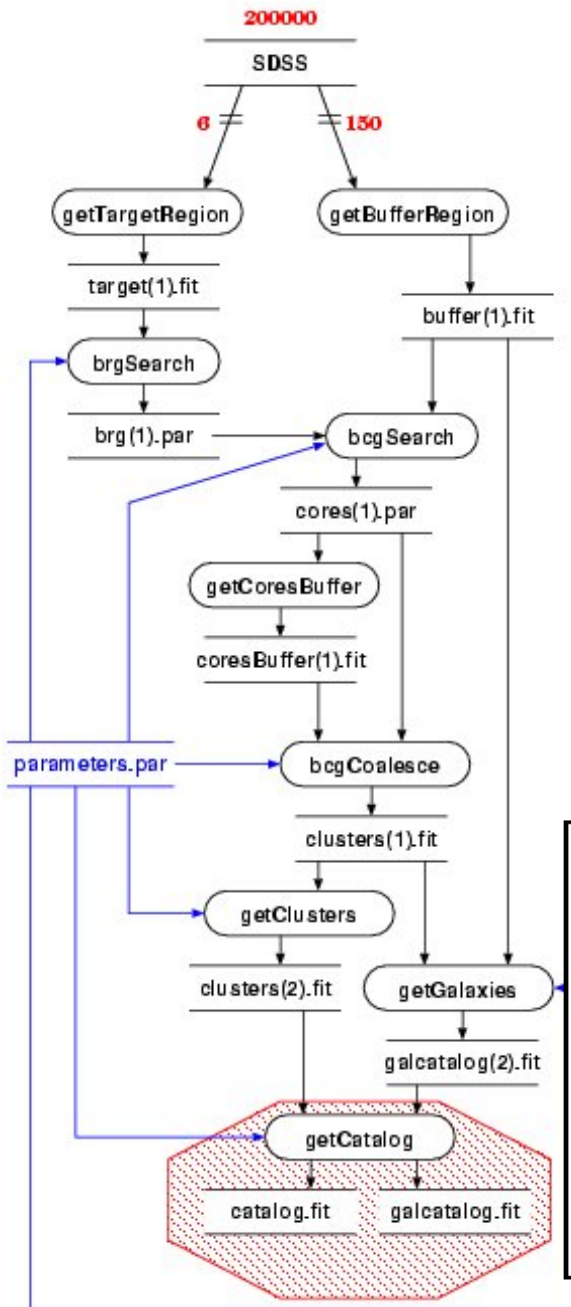
- | | | |
|--|----------------|------|
| ➤ Experiments at Large Hadron Collider
◆ 100s of Petabytes | 2007 - ? | 2009 |
| ➤ High Energy & Nuclear Physics expts
◆ ~1 Petabyte (1000 TB) | 1997 – present | 2007 |
| ➤ LIGO (gravity wave search)
◆ 100s of Terabytes | 2002 – present | 2005 |
| ➤ Sloan Digital Sky Survey
◆ 10s of Terabytes | 2001 – present | 2003 |
| | | 2001 |



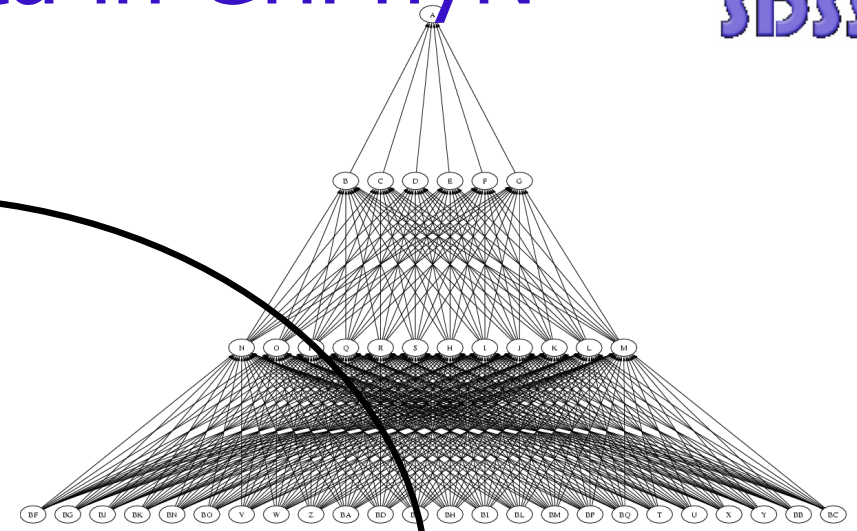
Future Grid resources

- **Massive CPU (PetaOps)**
- **Large distributed datasets (>100PB)**
- **Global communities (1000s)**

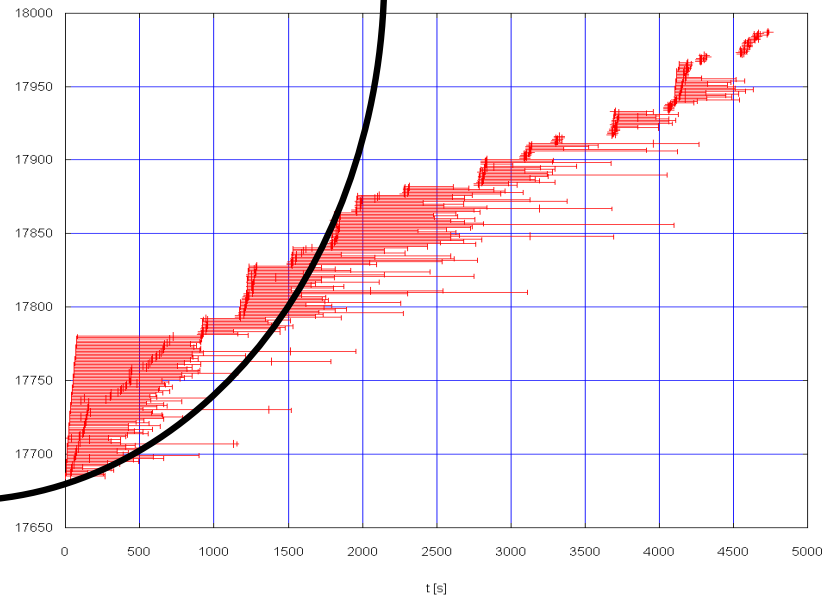
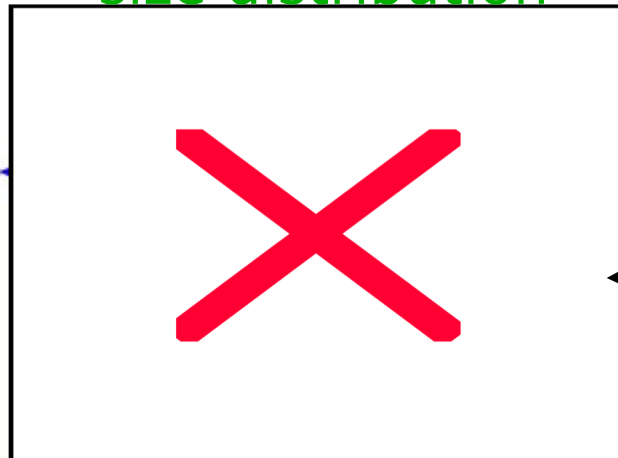
Sloan Digital Sky Survey (SDSS) Using Virtual Data in GriPhyN



Sloan Data



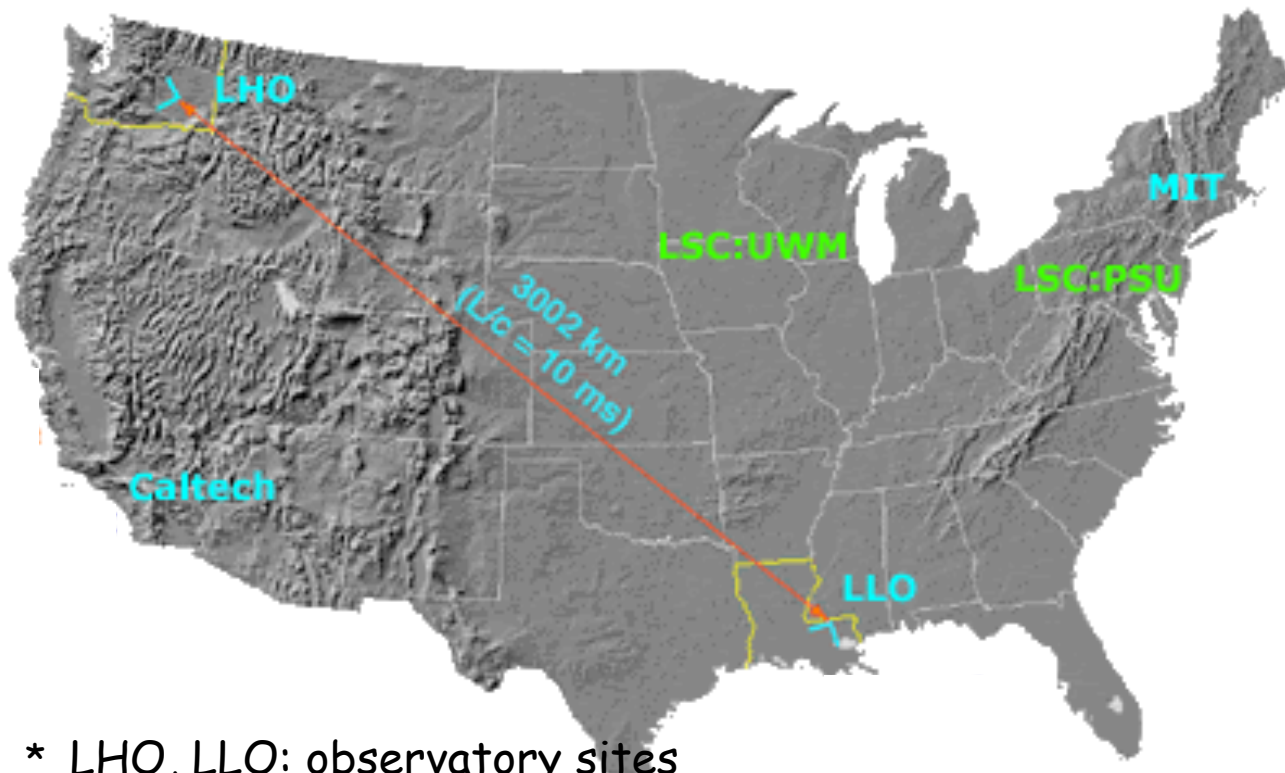
Galaxy cluster size distribution



The LIGO Scientific Collaboration (LSC) and the LIGO Grid

LIGO Grid: 6 US sites + 3 EU sites (Cardiff/UK, AEI/Germany)

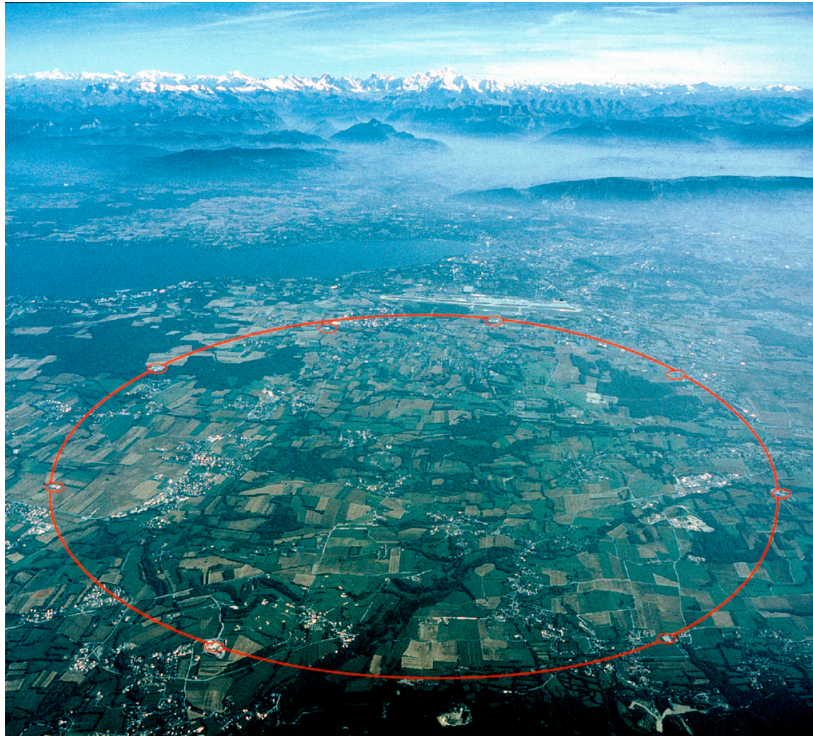
iVDGL has enabled LSC to establish a persistent production grid



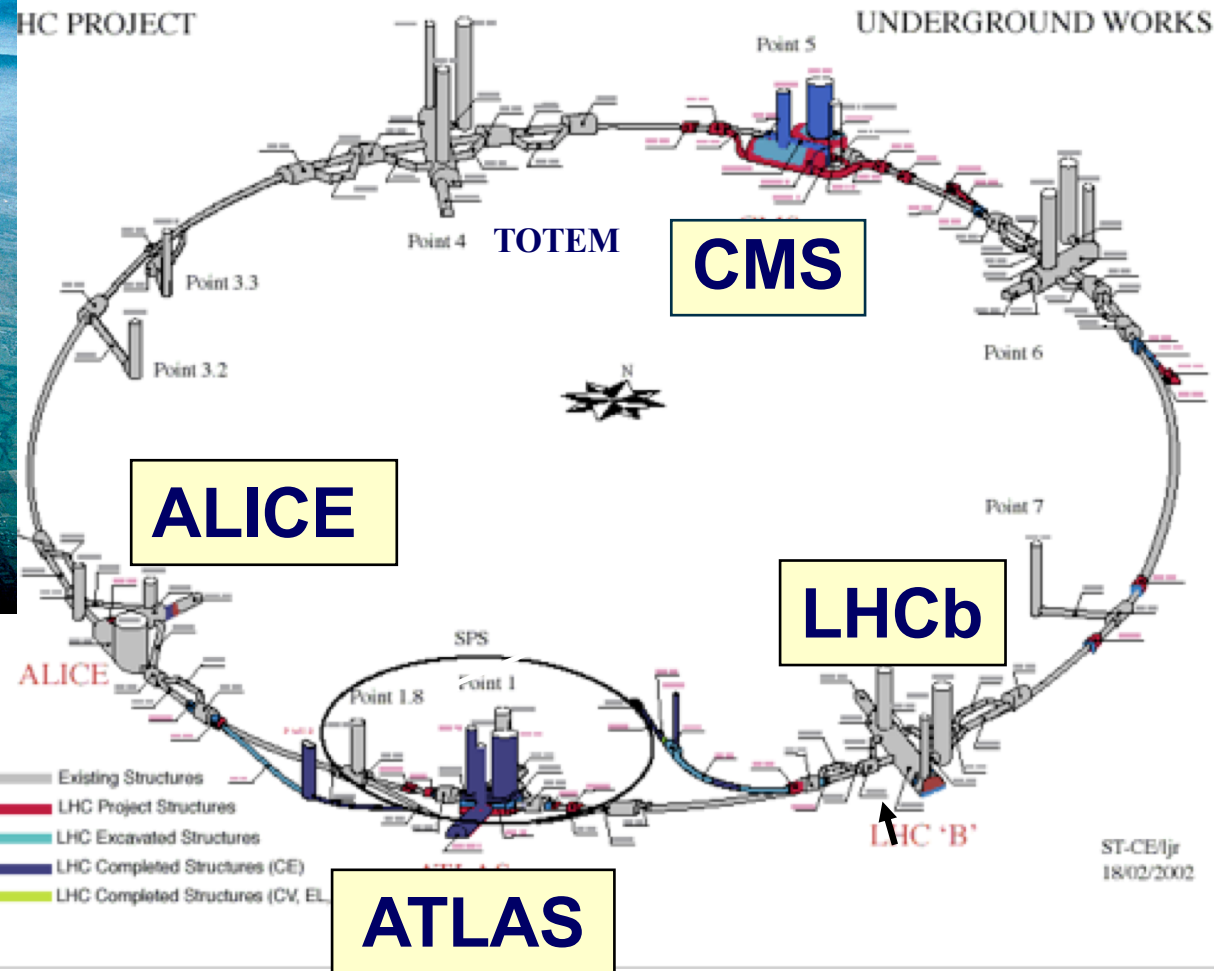
- * LHO, LLO: observatory sites
- * LSC - LIGO Scientific Collaboration - iVDGL supported



Large Hadron Collider (LHC) @ CERN

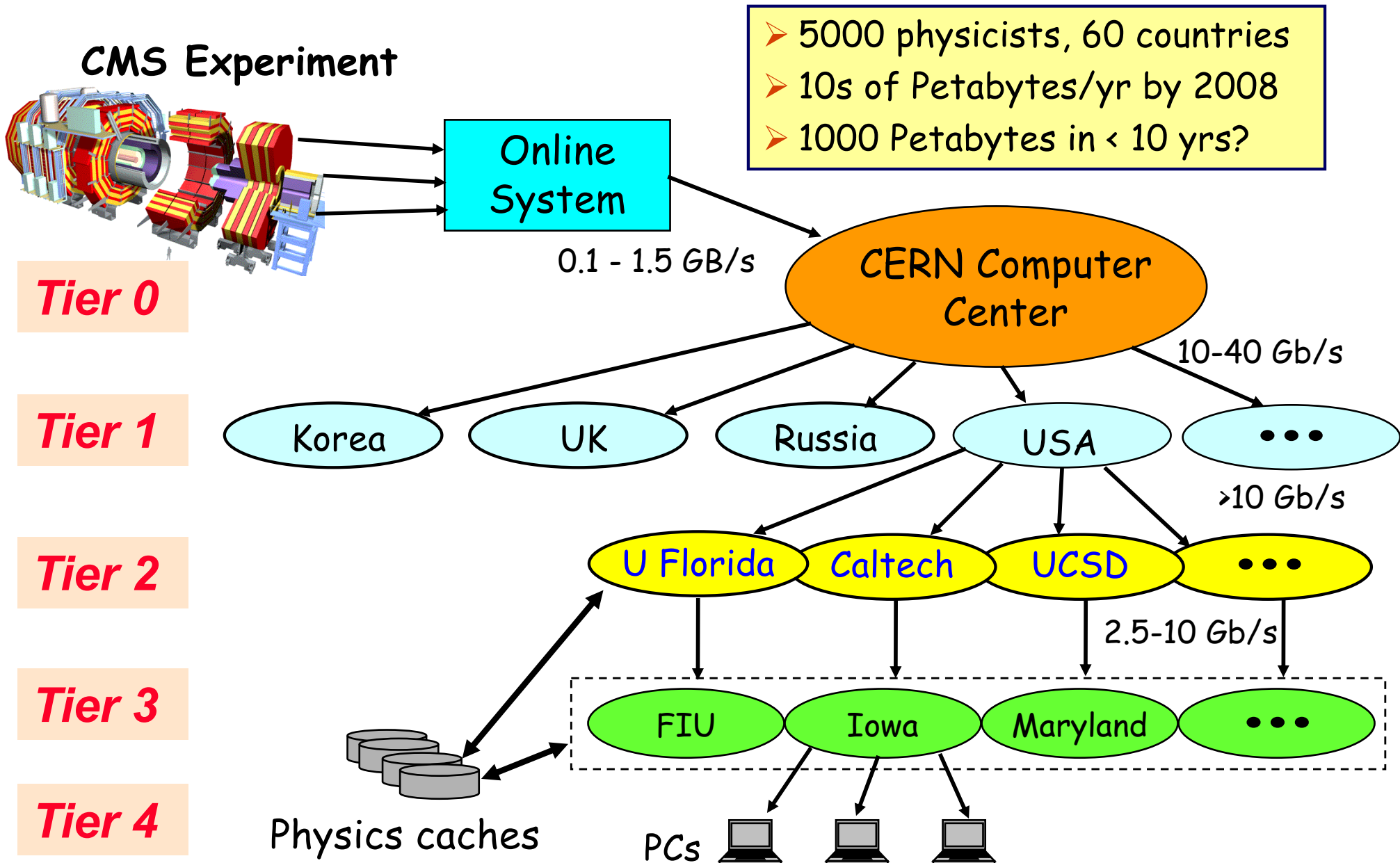


★ 27 km Tunnel in Switzerland & France



Search for Origin of Mass & Supersymmetry (2007 – ?)

LHC Global Data Grid

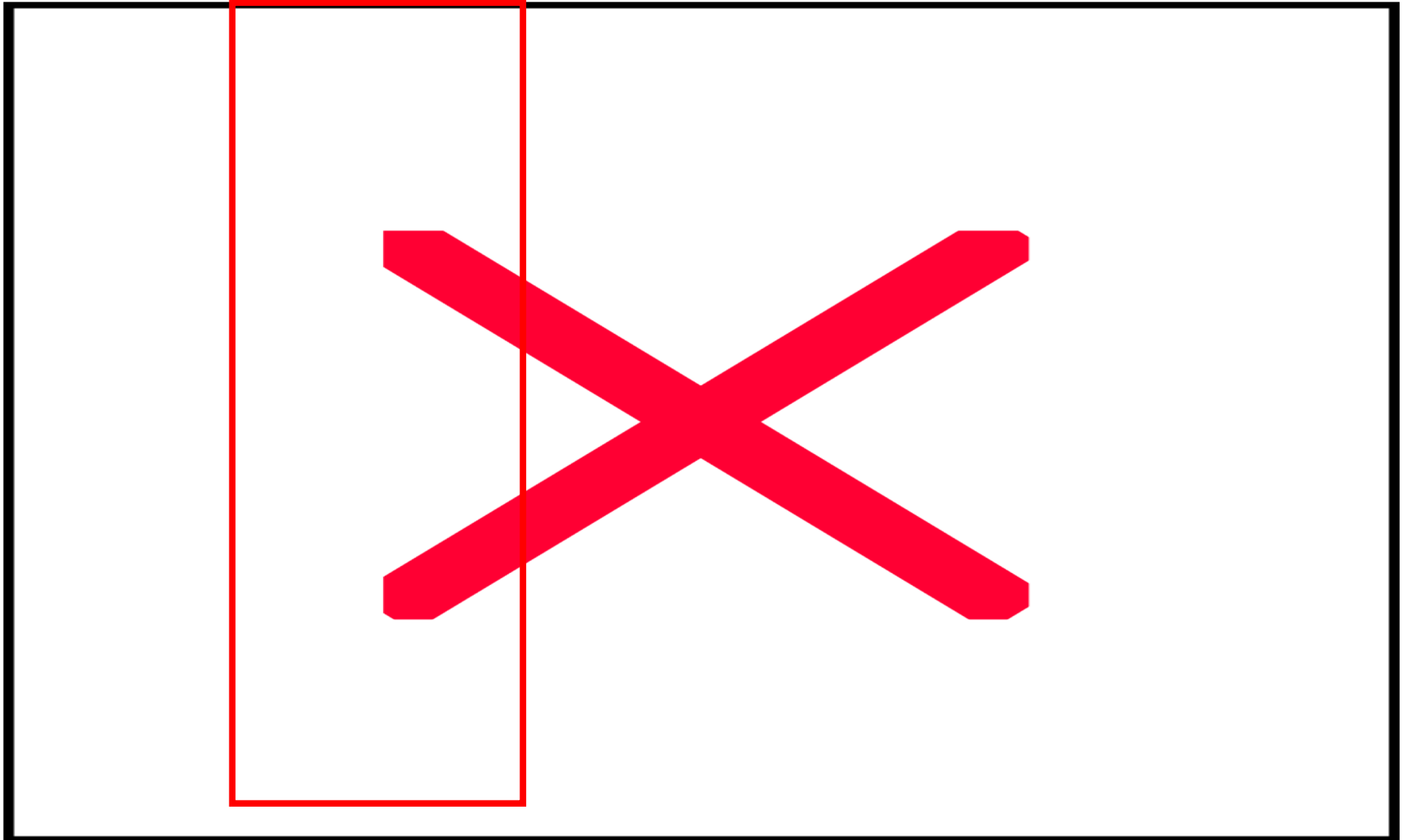


LCG: LHC Computing Grid

- **Global Grid infrastructure for LHC experiments**
 - ◆ Matched to decades long research program of LHC
- **Large scale resources**
 - ◆ Hundreds of resource sites throughout the world
 - ◆ Common resources, tools, middleware and environments
- **Operated and supported 24x7 globally**
 - ◆ A robust, stable, predictable, supportable infrastructure



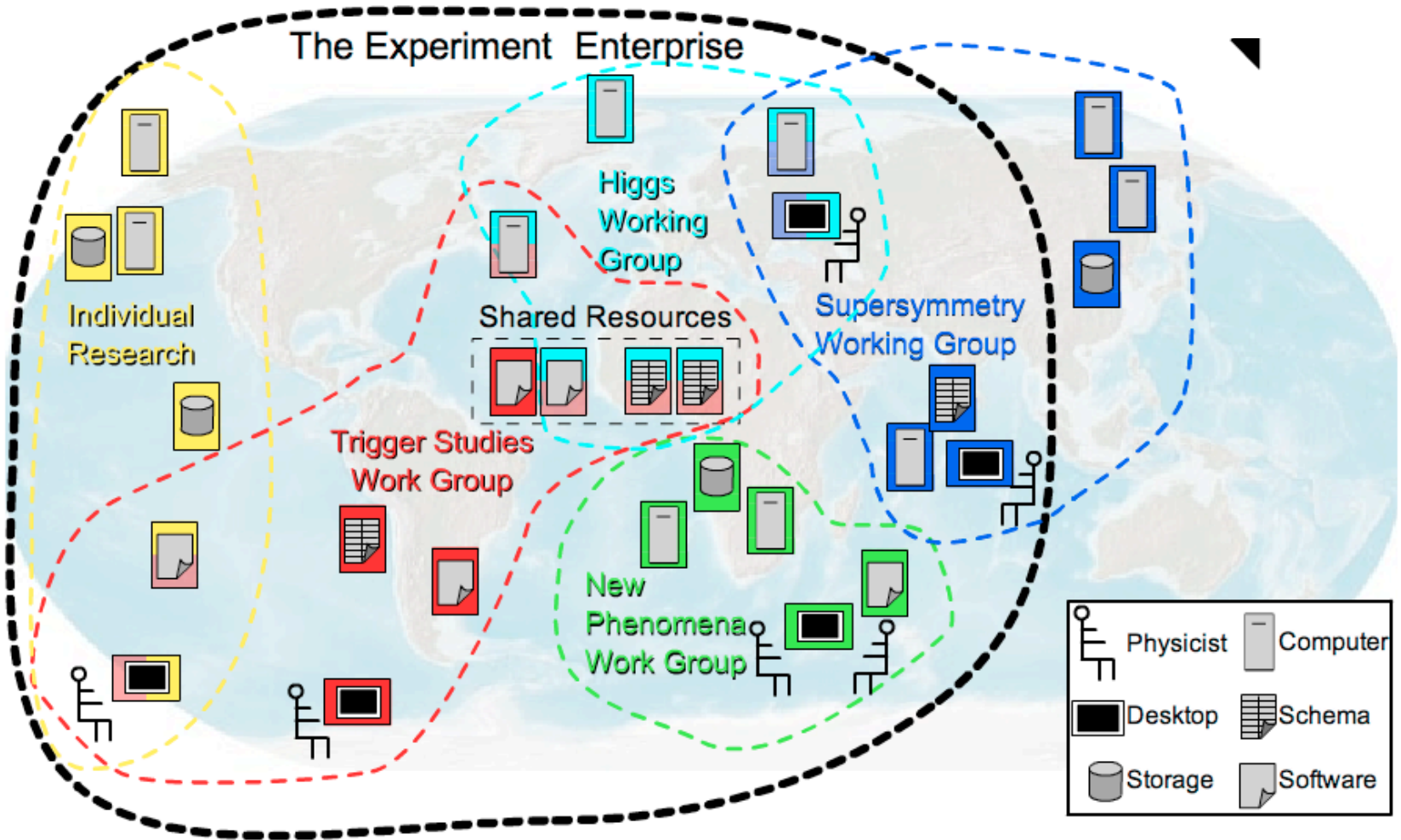
Network Bandwidth Needs (Gb/s)





Analysis by Globally Distributed Teams

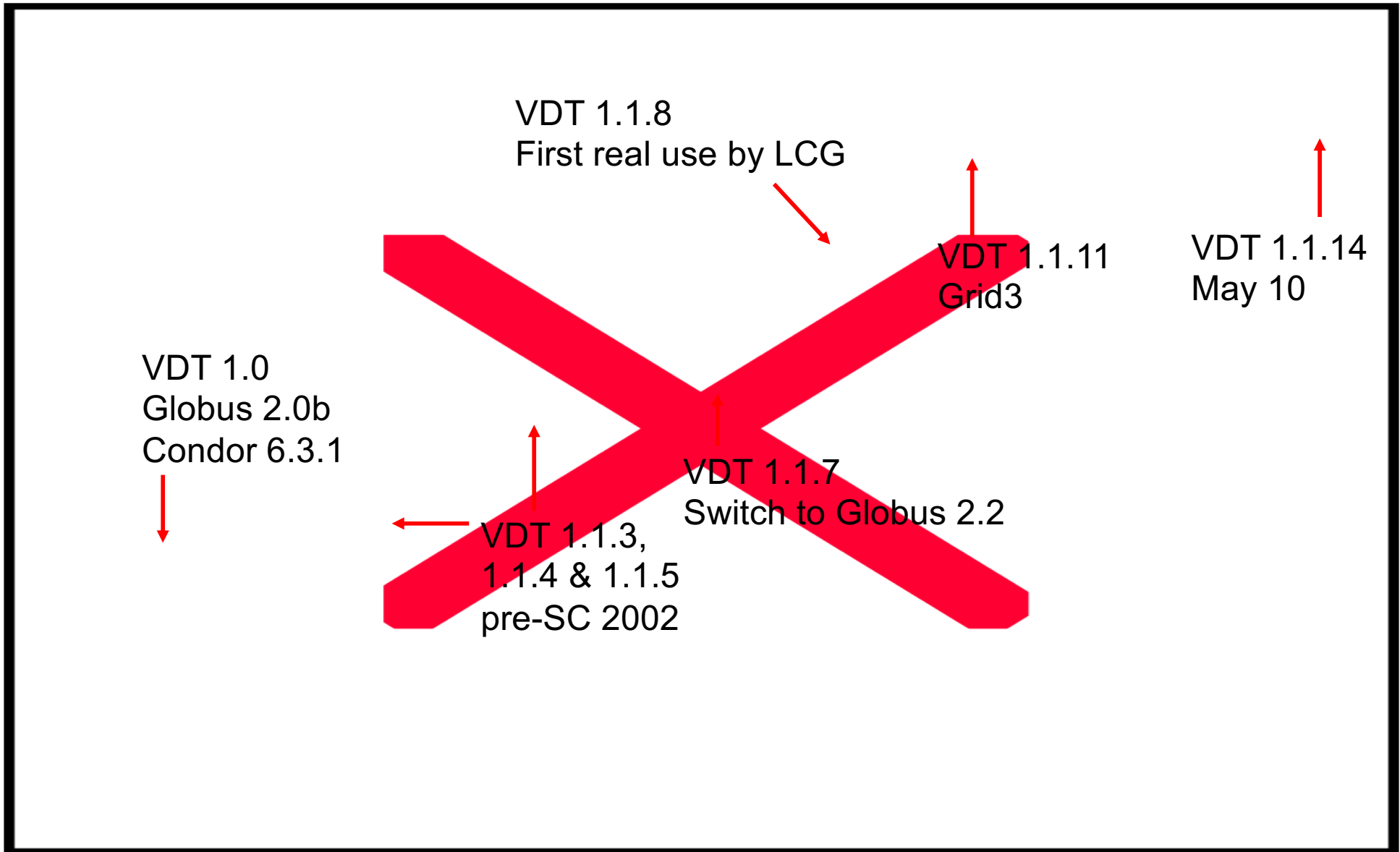
- **Non-hierarchical: Chaotic analyses + productions**
- **Superimpose significant random data flows**



Trillium Program of Work

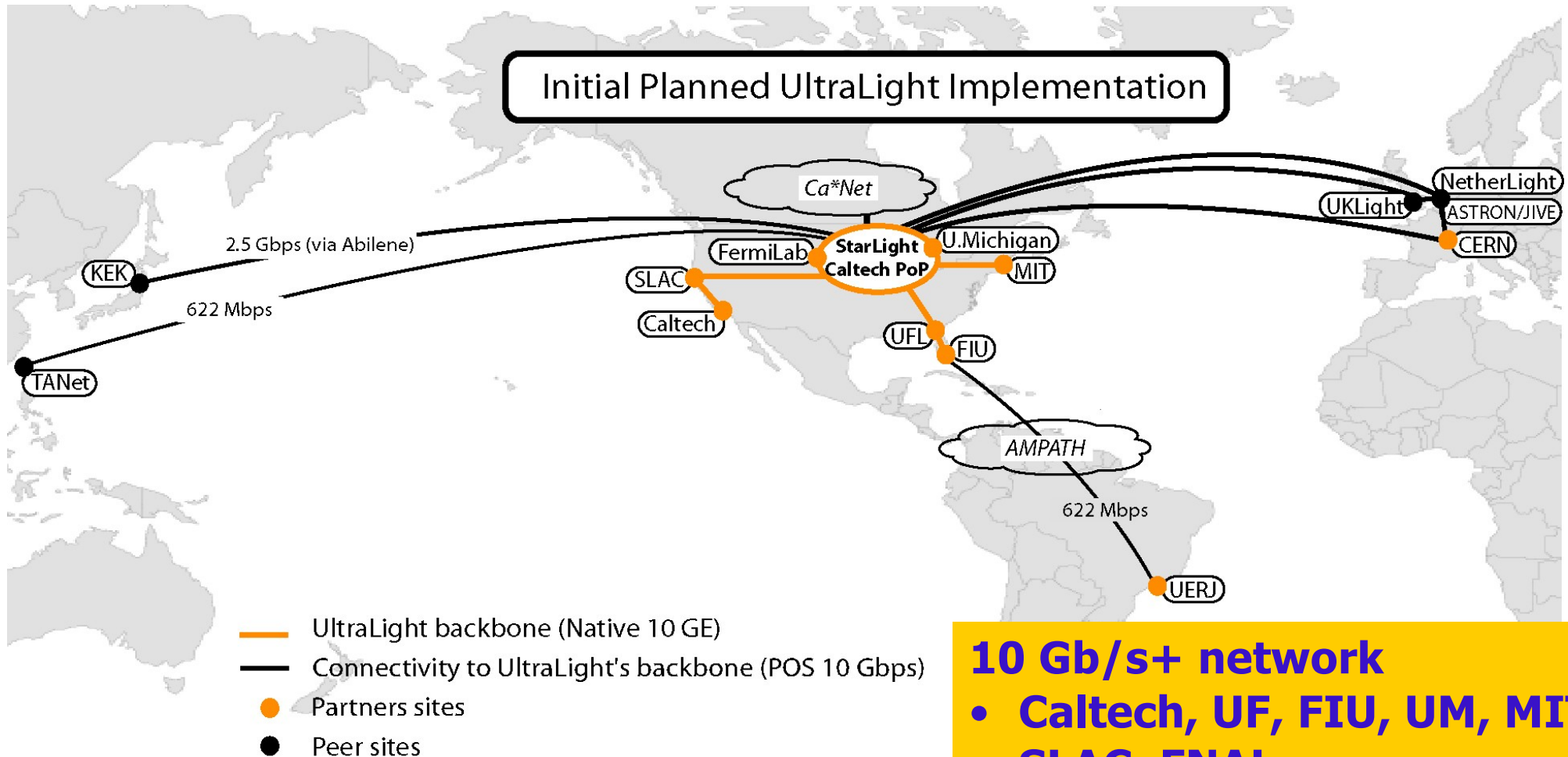
- Common experiments, leadership, participants
- CS research
 - ◆ Workflow, scheduling, virtual data
- Common Grid toolkits and packaging
 - ◆ Virtual Data Toolkit (VDT) + Pacman packaging
- Common Grid infrastructure: Grid3
 - ◆ National Grid for testing, development and production
- Advanced networking
 - ◆ Ultranet, UltraLight, etc.
- Integrated education and outreach effort
 - ◆ + collaboration with outside projects
- Unified entity in working with international projects
 - ◆ LCG, EGEE, Asia, South America

VDT Growth Over 2.5 Years



UltraLight: 10 Gb/s Network

Funded by ITR2004

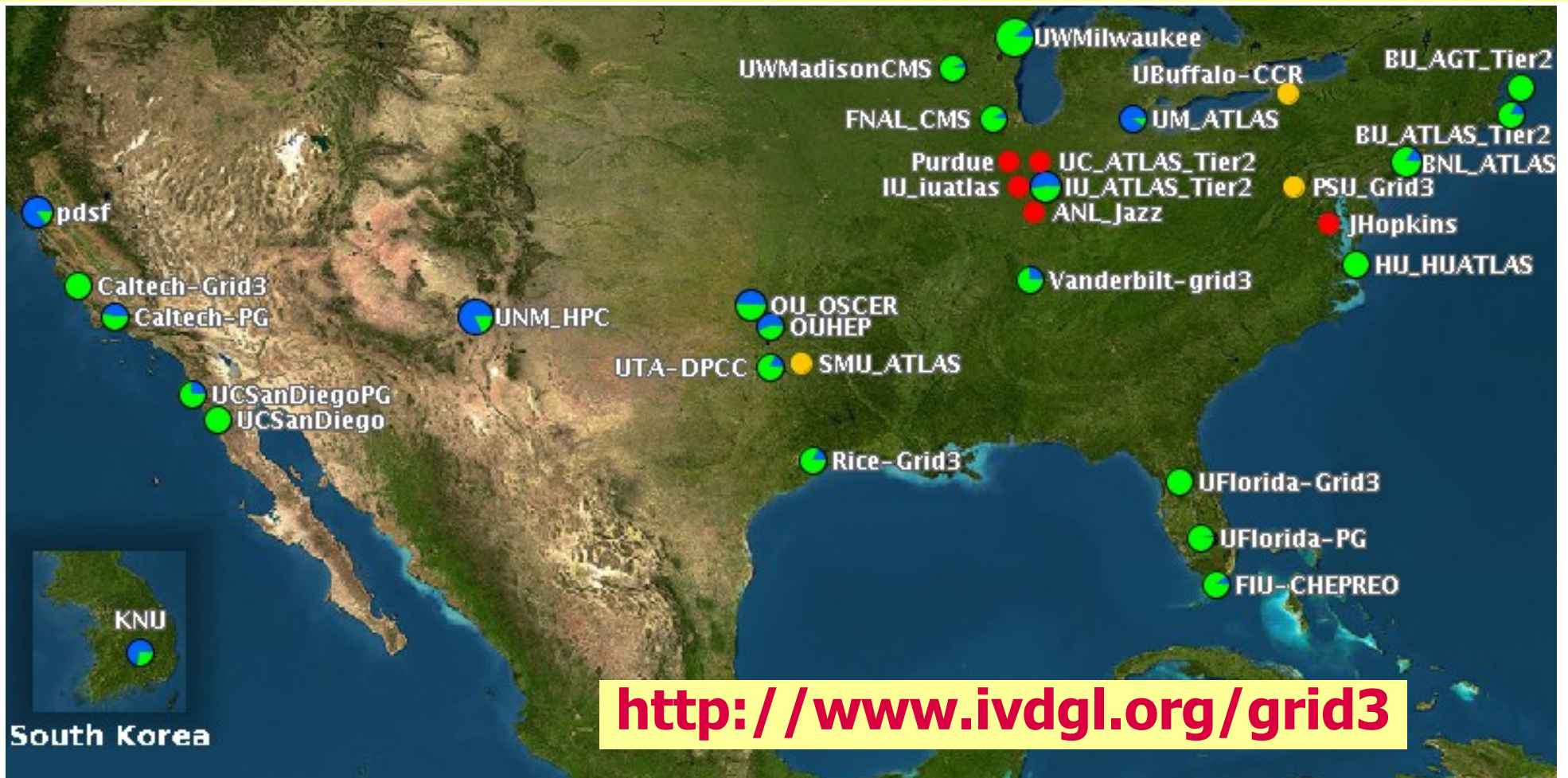


10 Gb/s+ network

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

Grid3: An Operational National Grid

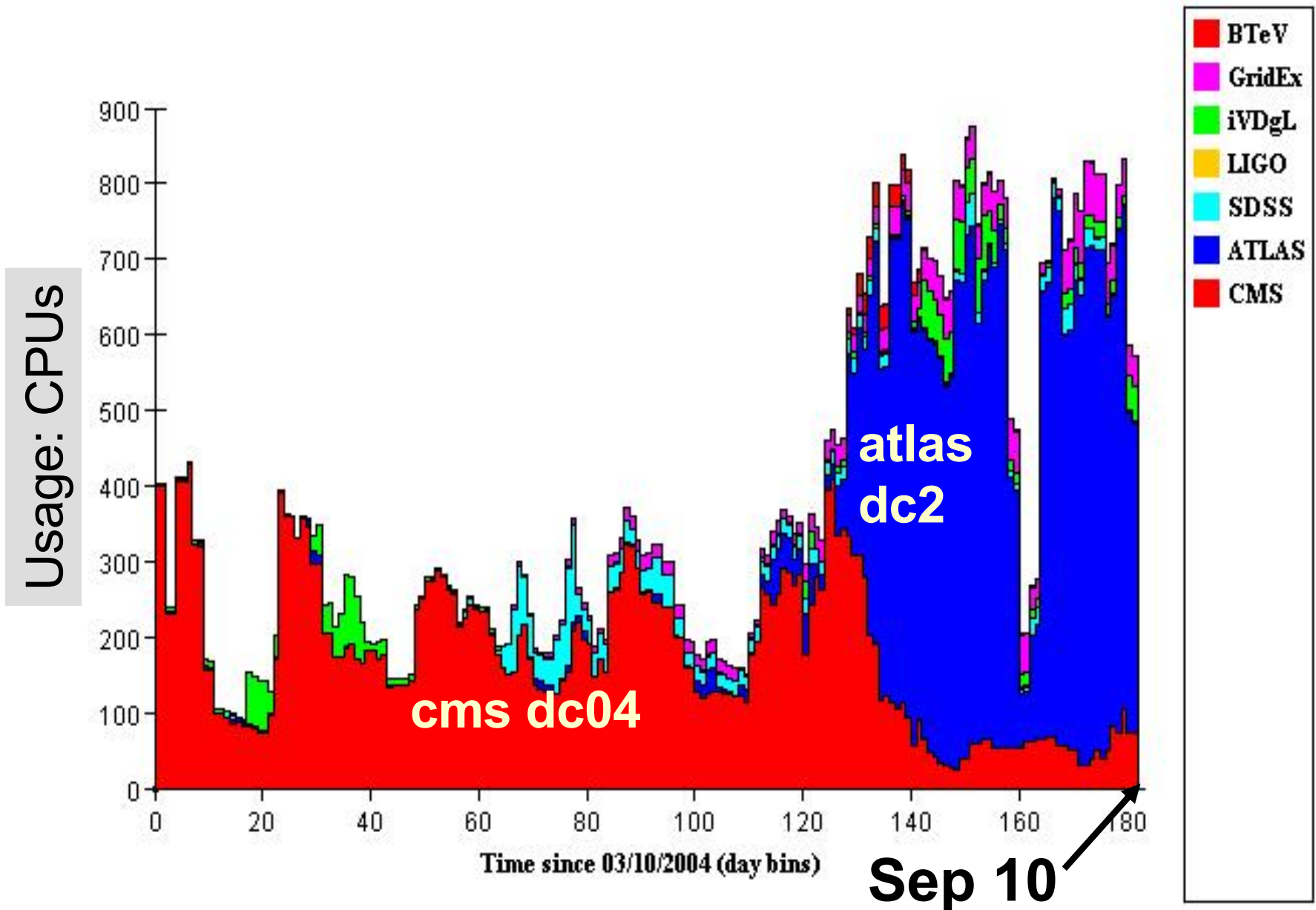
- 30 sites, 3500 CPUs: Universities + 4 national labs
- Part of LHC Grid
- Running since October 2003
- Applications in HEP, LIGO, SDSS, Genomics, CS



Grid2003 Applications

- **High energy physics**
 - ◆ US-ATLAS analysis (DIAL),
 - ◆ US-ATLAS GEANT3 simulation (GCE)
 - ◆ US-CMS GEANT4 simulation (MOP)
 - ◆ BTeV simulation
- **Gravity waves**
 - ◆ LIGO: blind search for continuous sources
- **Digital astronomy**
 - ◆ SDSS: cluster finding (maxBcg)
- **Bioinformatics**
 - ◆ Bio-molecular analysis (SnB)
 - ◆ Genome analysis (GADU/Gnare)
- **CS Demonstrators**
 - ◆ Job Exerciser, GridFTP, NetLogger-grid2003

Grid3 Shared Use Over 6 months



Open Science Grid

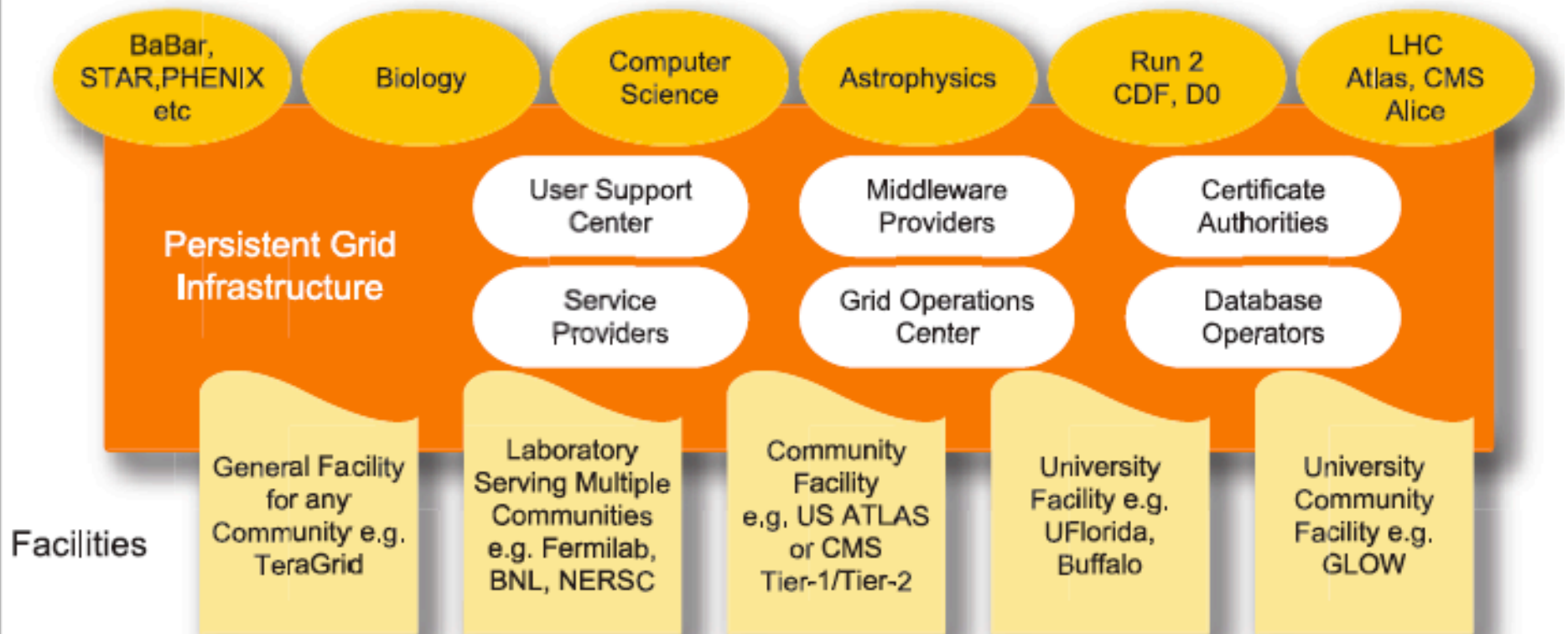
- **Build on Grid3 experience**
 - ◆ Persistent, production-quality Grid, national + international scope
- **Continue U.S. leading role in international science**
 - ◆ Grid infrastructure for large-scale collaborative scientific research
- **Create large computing infrastructure**
 - ◆ Combine resources at DOE labs and universities to effectively become a single national computing infrastructure for science
 - ◆ Grid3 → OSG-0 → OSG-1 → OSG-2 → ...
- **Maintain interoperability with LCG (LHC Grid)**
- **Provide opportunities for educators and students**
 - ◆ Participate in building and exploiting this grid infrastructure
 - ◆ Develop and train scientific and technical workforce

<http://www.opensciencegrid.org>



Open Science Grid

Applications



Education and Outreach

Grids and the Digital Divide

Rio de Janeiro, Feb. 16-20, 2004



NEWS:

Bulletin: ONE TWO
WELCOME BULLETIN
General Information
Registration
Travel Information
Hotel Registration
Participant List
How to Get UERJ/Hotel
Computer Accounts
Useful Phone Numbers
Program
Contact us:
Secretariat
Chairmen

Background

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

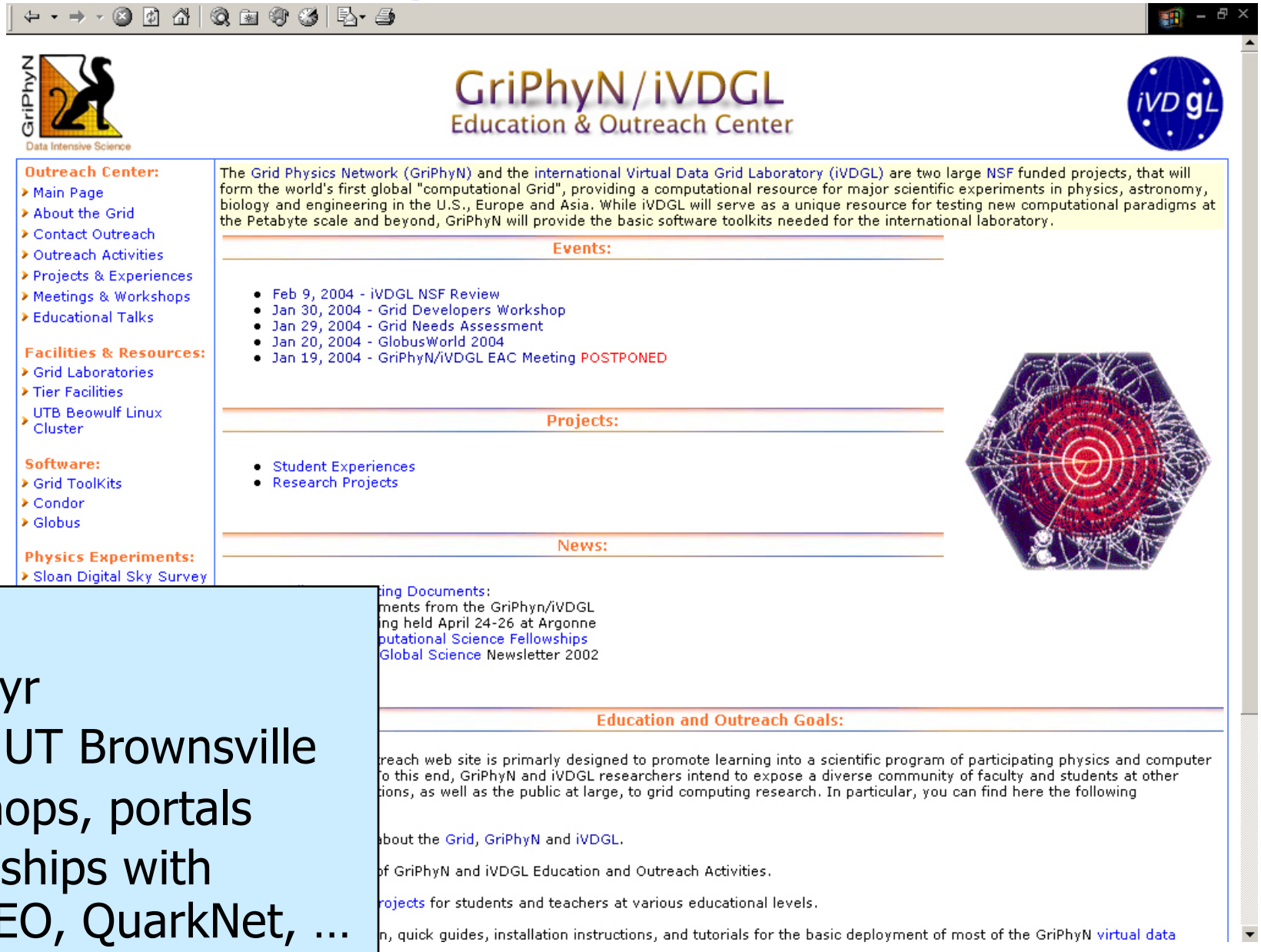
Themes

- Global collaborations, Grids and addressing the Digital Divide

Next meeting: May 2005 (Korea)

<http://www.uerj.br/lishep2004>

iVDGL, GriPhyN Education / Outreach



The screenshot shows a web browser window displaying the GriPhyN/iVDGL Education & Outreach Center website. The page features a navigation menu on the left, a main content area with sections for Events, Projects, and News, and a sidebar with a 'Basics' box. The website header includes the GriPhyN logo and the iVDGL logo. The main content area contains a paragraph about the Grid Physics Network (GriPhyN) and the international Virtual Data Grid Laboratory (iVDGL), followed by lists of events, projects, and news items. A large image of a network diagram is visible on the right side of the page.

Basics

- \$200K/yr
- Led by UT Brownsville
- Workshops, portals
- Partnerships with CHEPREO, QuarkNet, ...

Outreach Center:

- Main Page
- About the Grid
- Contact Outreach
- Outreach Activities
- Projects & Experiences
- Meetings & Workshops
- Educational Talks

Facilities & Resources:

- Grid Laboratories
- Tier Facilities
- UTB Beowulf Linux Cluster

Software:

- Grid ToolKits
- Condor
- Globus

Physics Experiments:

- Sloan Digital Sky Survey

Events:

- Feb 9, 2004 - iVDGL NSF Review
- Jan 30, 2004 - Grid Developers Workshop
- Jan 29, 2004 - Grid Needs Assessment
- Jan 20, 2004 - GlobusWorld 2004
- Jan 19, 2004 - GriPhyN/iVDGL EAC Meeting **POSTPONED**

Projects:

- Student Experiences
- Research Projects

News:

Education and Outreach Goals:

The outreach web site is primarily designed to promote learning into a scientific program of participating physics and computer science. To this end, GriPhyN and iVDGL researchers intend to expose a diverse community of faculty and students at other institutions, as well as the public at large, to grid computing research. In particular, you can find here the following information about the [Grid](#), [GriPhyN](#) and [iVDGL](#).

Information about the [Education and Outreach Activities](#) of GriPhyN and iVDGL Education and Outreach Activities.

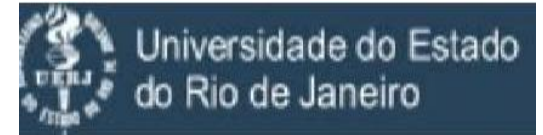
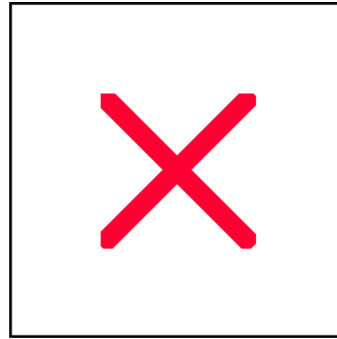
Information about [projects](#) for students and teachers at various educational levels.

Information about [tutorials](#), quick guides, installation instructions, and tutorials for the basic deployment of most of the GriPhyN [virtual data](#).

June 21-25 Grid Summer School

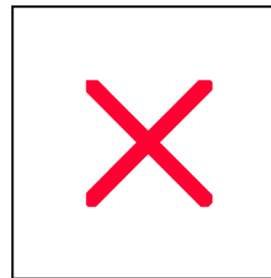
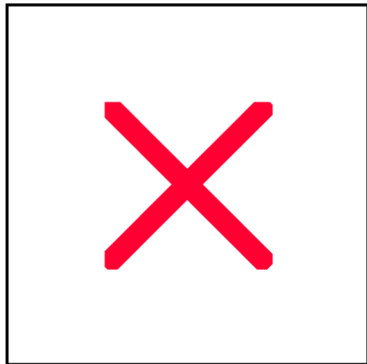
- **First of its kind in the U.S. (South Padre Island, Texas)**
 - ◆ 36 students, diverse origins and types (M, F, MSIs, etc)
- **Marks new direction for Trillium**
 - ◆ First attempt to systematically train people in Grid technologies
 - ◆ First attempt to gather relevant materials in one place
 - ◆ Today: Students in CS and Physics
 - ◆ Later: Students, postdocs, junior & senior scientists
- **Reaching a wider audience**
 - ◆ Put lectures, exercises, video, on the web
 - ◆ More tutorials, perhaps 3-4/year
 - ◆ Dedicated resources for remote tutorials
 - ◆ Create "Grid book", e.g. Georgia Tech
- **New funding opportunities**
 - ◆ NSF: new training & education programs





CHEPREO: Center for High Energy Physics Research and Educational Outreach

Florida International University



- Physics Learning Center
- CMS Research
- iVDGL Grid Activities
- AMPATH network (S. America)

- Funded September 2003
- \$4M initially (3 years)
- 4 NSF Directorates!

Grid Project References

➤ GriPhyN

◆ www.griphyn.org

➤ iVDGL

◆ www.ivdgl.org

➤ PPDG

◆ www.ppdg.net

➤ Grid3

◆ www.ivdgl.org/grid3

➤ Open Science Grid

◆ www.opensciencegrid.org

➤ CHEPREO

◆ www.chepreo.org

➤ UltraLight

◆ ultralight.cacr.caltech.edu

➤ Globus

◆ www.globus.org

➤ LCG

◆ www.cern.ch/lcg

➤ EU DataGrid

◆ www.eu-datagrid.org

➤ EGEE

◆ www.eu-egee.org

Trillium Grid Tools: Virtual Data Toolkit

