

# The International Virtual Data Grid Laboratory

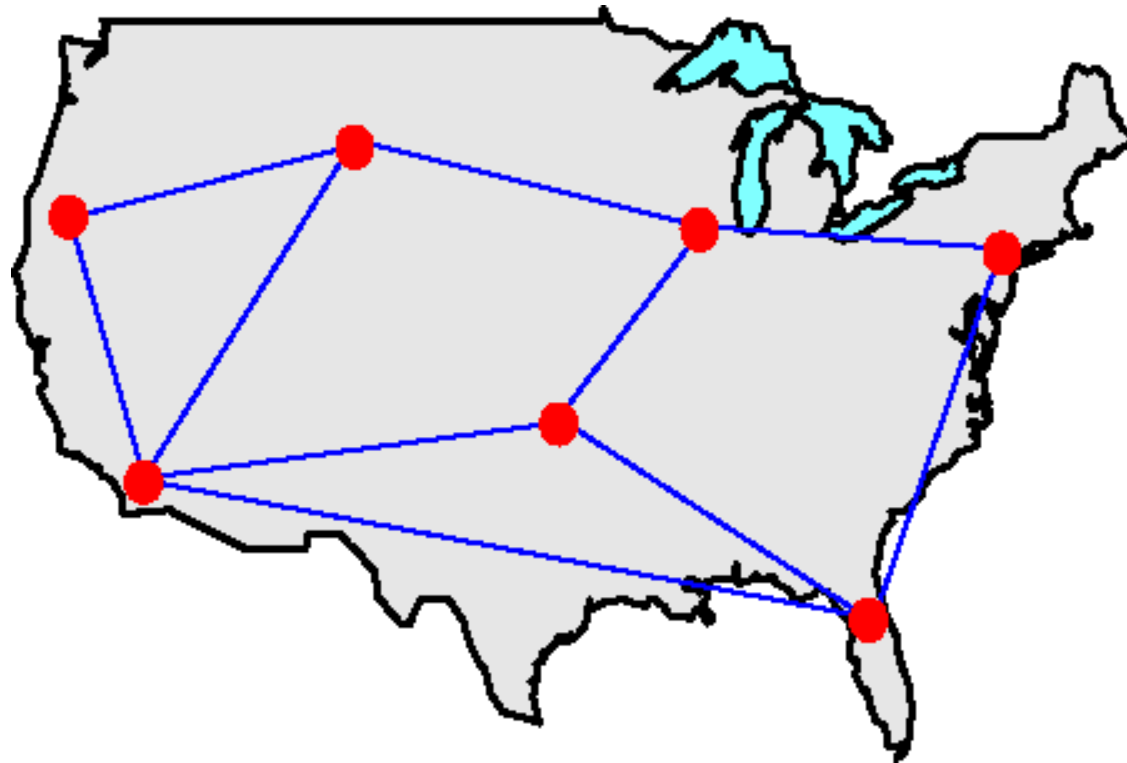


Paul Avery  
University of Florida  
<http://www.phys.ufl.edu/~avery/>  
avery@phys.ufl.edu

**AMPATH Meeting  
Valdivia, Chile  
April 12, 2002**

# The Grid Concept

- ➔ Grid: Geographically distributed computing resources configured for coordinated use
  - ◆ Fabric: Physical resources & networks provide raw capability
  - ◆ Middleware: Software ties it all together (tools, services, etc.)
- ➔ Goal: Transparent resource sharing





# Fundamental Idea: Resource Sharing

- ➔ Resources for complex problems are distributed
  - ◆ Advanced scientific instruments (accelerators, telescopes, ...)
  - ◆ Storage and computing
  - ◆ Groups of people
- ➔ Communities require access to common services
  - ◆ Scientific collaborations (physics, astronomy, biology, eng. ...)
  - ◆ Government agencies
  - ◆ Health care organizations, large corporations, ...
- ➔ Goal is to build “Virtual Organizations”
  - ◆ Make all community resources available to any VO member
  - ◆ Leverage strengths at different institutions
  - ◆ Add people & resources dynamically



# Some (Realistic) Grid Examples

## → High energy physics

- ◆ 3,000 physicists worldwide pool Petaflops of CPU resources to analyze Petabytes of data

## → Climate modeling

- ◆ Climate scientists visualize, annotate, & analyze Terabytes of simulation data

## → Biology

- ◆ A biochemist exploits 10,000 computers to screen 100,000 compounds in an hour

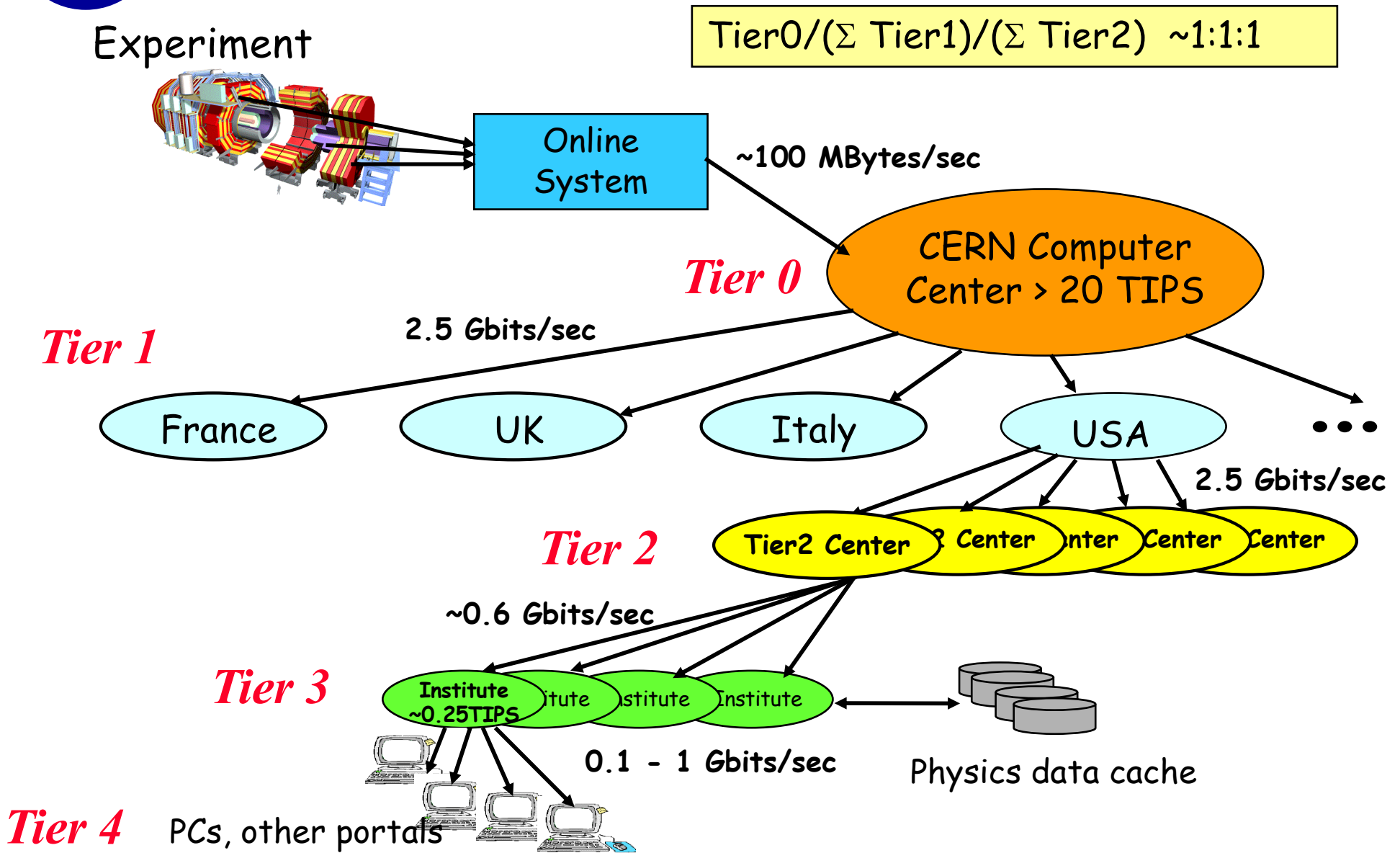
## → Engineering

- ◆ A multidisciplinary analysis in aerospace couples code and data in four companies to design a new airframe

## → Many commercial examples

From Ian Foster

# Example: Global LHC Data Grid





# Grid Architecture and Globus Toolkit

- ➔ **Globus Project™ (Argonne + USC/ISI)**
  - ◆ O(40) researchers & developers
- ➔ **Globus Toolkit™ 2.0**
  - ◆ Reference implementation of core Grid protocols & services
  - ◆ Open source developer community
- ➔ **Globus Toolkit used by all Data Grid projects today**
  - ◆ US: GriPhyN, PPDG, TeraGrid, iVDGL
  - ◆ EU+Asia: EU-DataGrid + several EU, Asian national projects
- ➔ **Recent announcement of applying “web services” to Grids**
  - ◆ Keeps Grids in the commercial mainstream
  - ◆ Open source community: OGSA
  - ◆ Globus Toolkit 3.0 next year



# Background: Major Data Grid Projects

- ➔ Particle Physics Data Grid (US, DOE)
  - ◆ Data Grid applications for HENP expts.
- ➔ GriPhyN (US, NSF)
  - ◆ Petascale Virtual-Data Grids
- ➔ iVDGL (US, NSF)
  - ◆ Global Grid lab
- ➔ TeraGrid (US, NSF)
  - ◆ Dist. supercomp. resources (13 TFlops)
- ➔ European Data Grid (EU, EC)
  - ◆ Data Grid technologies, EU deployment
- ➔ CrossGrid (EU, EC)
  - ◆ Data Grid technologies
- ➔ DataTAG (EU, EC)
  - ◆ Transatlantic network, Grid applications
- ➔ Japanese Grid Project (APGrid?) (Japan)
  - ◆ Grid deployment throughout Japan

- ◆ Data intensive expts.
- ◆ Collaborations of application scientists & computer scientists
- ◆ Infrastructure devel. & deployment
- ◆ Globus based



# iVDGL: A Global Grid Laboratory

**“We propose to create, operate and evaluate, over a sustained period of time, an international research laboratory for data-intensive science.”**

*From NSF proposal, 2001*

## ➔ International Virtual-Data Grid Laboratory

- ◆ A global Grid laboratory (US, EU, Asia, South America, ...)
- ◆ A place to conduct Data Grid tests “at scale”
- ◆ A mechanism to create common Grid infrastructure
- ◆ A laboratory for other disciplines to perform Data Grid tests
- ◆ A focus of outreach efforts to small institutions

## ➔ U.S. part funded by NSF (2001-2006)

- ◆ \$13.7M (NSF) + \$2M (matching)





# iVDGL Components

- ➔ **Computing resources**
  - ◆ 2 Tier1 laboratory sites (funded elsewhere)
  - ◆ 7 Tier2 university sites      ⇒ software integration
  - ◆ 3 Tier3 university sites      ⇒ outreach effort
- ➔ **Networks**
  - ◆ USA (TeraGrid, Internet2, ESNET), Europe (Géant, ...)
  - ◆ Transatlantic (DataTAG), Transpacific, AMPATH?, ...
- ➔ **Grid Operations Center (GOC)**
  - ◆ Joint work with TeraGrid on GOC development
- ➔ **Computer Science support teams**
  - ◆ Support, test, upgrade GriPhyN Virtual Data Toolkit
- ➔ **Education and Outreach**
- ➔ **Coordination, management**



# iVDGL Components (cont.)

- ➔ High level of coordination with DataTAG
  - ◆ Transatlantic research network (2.5 Gb/s) connecting EU & US
- ➔ Current partners
  - ◆ TeraGrid, EU DataGrid, EU projects, Japan, Australia
- ➔ Experiments/labs requesting participation
  - ◆ ALICE, CMS-HI, D0, BaBar, BTeV, PDC (Sweden)

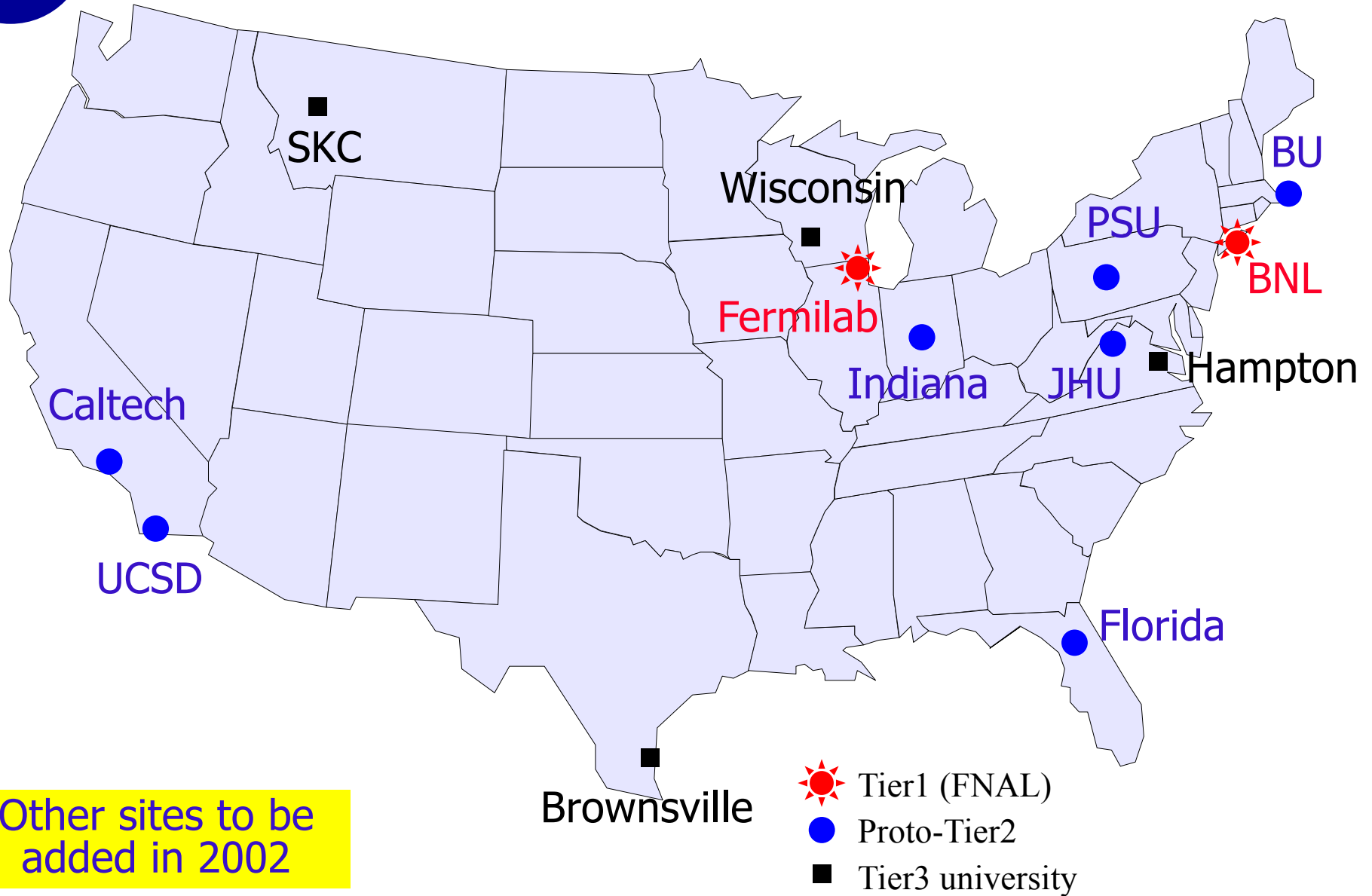


# Initial US iVDGL Participants

◆ U Florida	CMS	} <b>Tier2 / Software</b>
◆ Caltech	CMS, LIGO	
◆ UC San Diego	CMS, CS	
◆ Indiana U	ATLAS, GOC	
◆ Boston U	ATLAS	
◆ U Wisconsin, Milwaukee	LIGO	
◆ Penn State	LIGO	
◆ Johns Hopkins	SDSS, NVO	
◆ U Chicago/Argonne	CS	
◆ U Southern California	CS	
◆ U Wisconsin, Madison	CS	
◆ Salish Kootenai	Outreach, LIGO	} <b>Tier3 / Outreach</b>
◆ Hampton U	Outreach, ATLAS	
◆ U Texas, Brownsville	Outreach, LIGO	
◆ Fermilab	CMS, SDSS, NVO	} <b>Tier1 / Labs (funded elsewhere)</b>
◆ Brookhaven	ATLAS	
◆ Argonne Lab	ATLAS, CS	

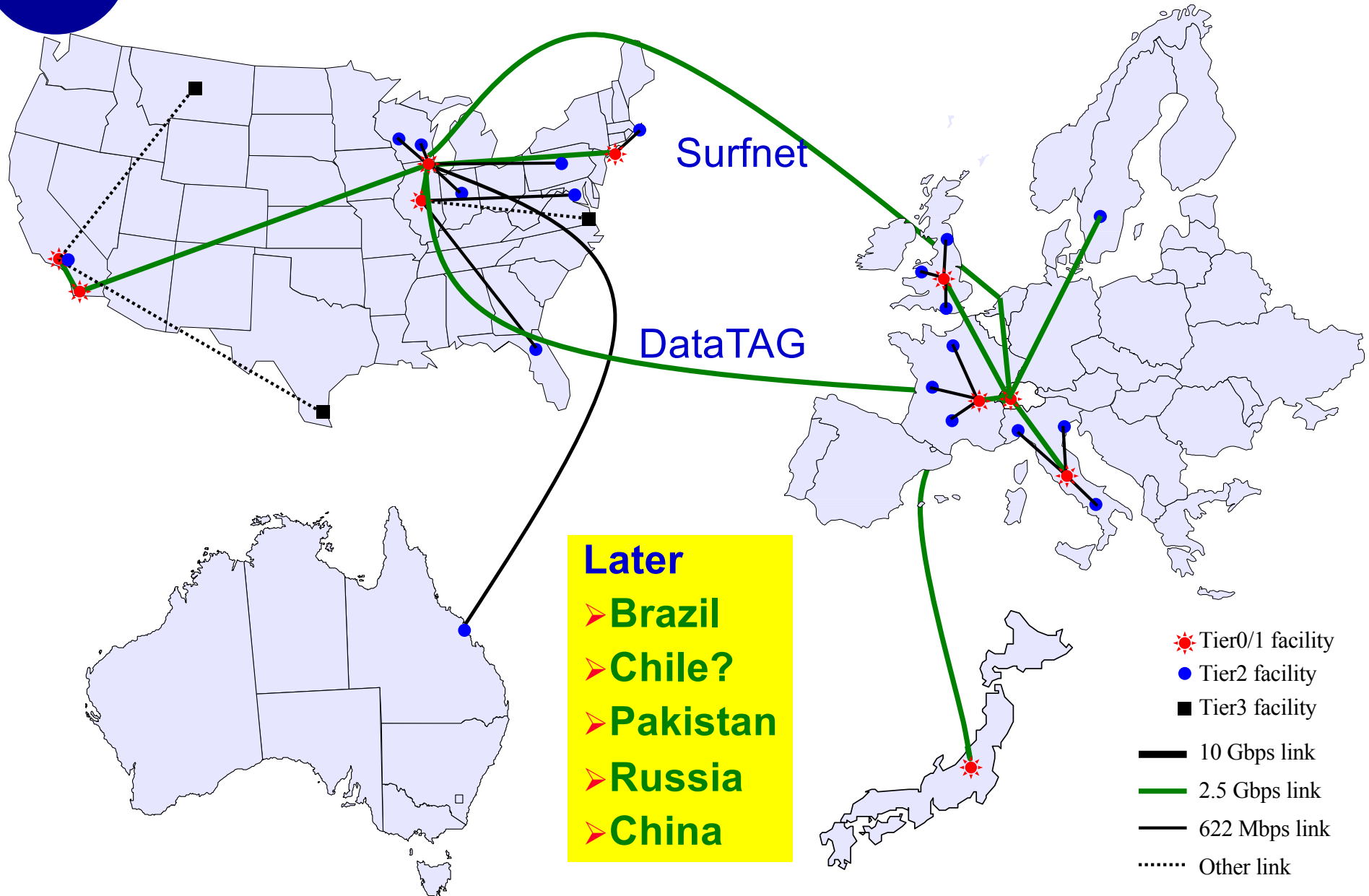


# Initial US-iVDGL Data Grid





# iVDGL Map (2002-2003)





# iVDGL Work Teams

## ➔ Facilities Team

- ◆ Hardware (Tier1, Tier2, Tier3)

## ➔ Core Software Team

- ◆ Grid middleware, toolkits

## ➔ Laboratory Operations Team

- ◆ Coordination, software support, performance monitoring

## ➔ Applications Team

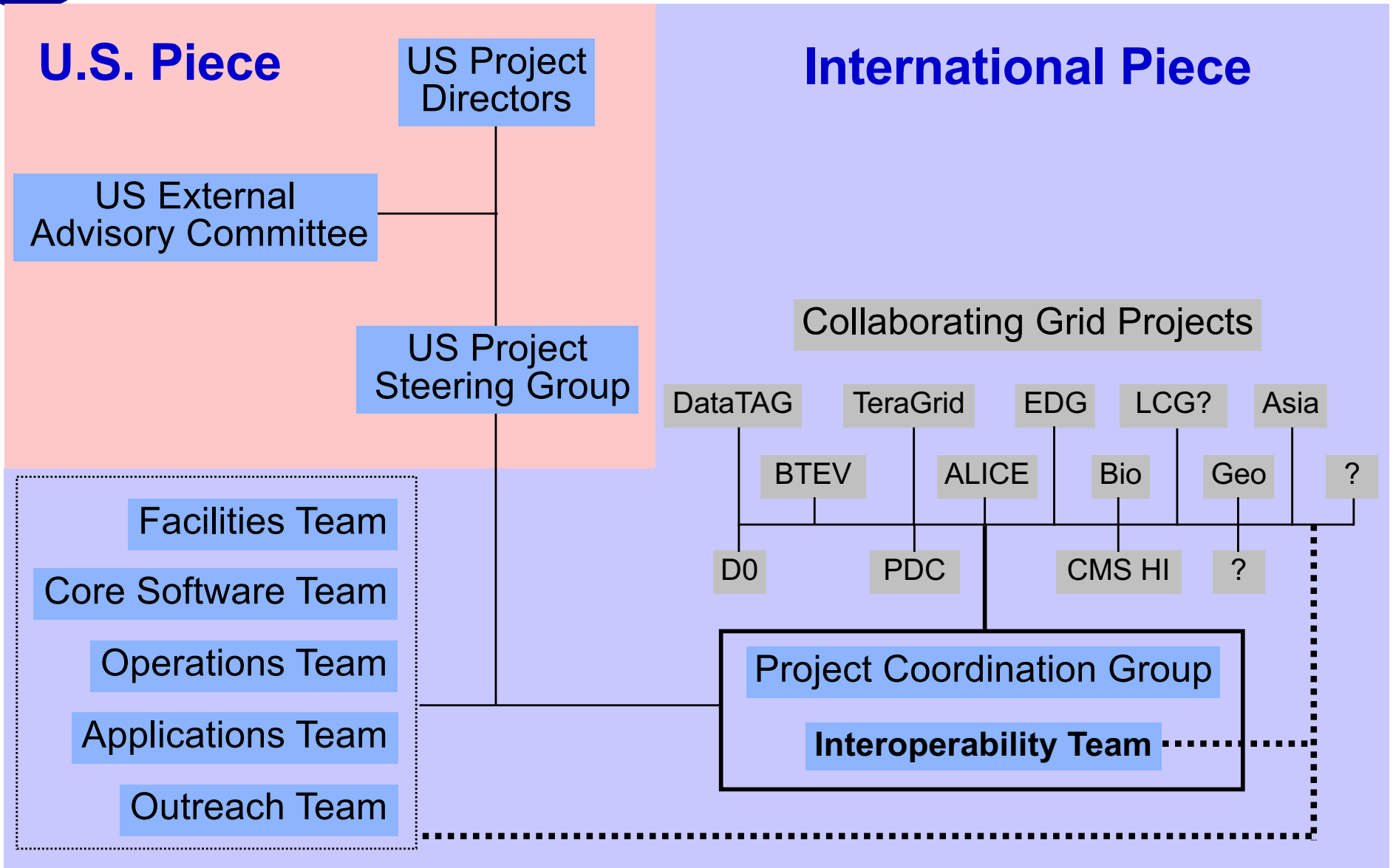
- ◆ High energy physics, gravity waves, virtual astronomy
- ◆ Nuclear physics, bioinformatics, ...

## ➔ Education and Outreach Team

- ◆ Primarily using Web tools
- ◆ Integrated with GriPhyN, connections to other projects
- ◆ Want to develop further international connections



# iVDGL Management and Coordination





# Grid References

- ➔ Grid Book
  - ◆ [www.mkp.com/grids](http://www.mkp.com/grids)
- ➔ Globus
  - ◆ [www.globus.org](http://www.globus.org)
- ➔ Global Grid Forum
  - ◆ [www.gridforum.org](http://www.gridforum.org)
- ➔ TeraGrid
  - ◆ [www.teragrid.org](http://www.teragrid.org)
- ➔ EU DataGrid
  - ◆ [www.eu-datagrid.org](http://www.eu-datagrid.org)
- ➔ PPDG
  - ◆ [www.ppdg.net](http://www.ppdg.net)
- ➔ GriPhyN
  - ◆ [www.griphyn.org](http://www.griphyn.org)
- ➔ iVDGL
  - ◆ [www.ivdgl.org](http://www.ivdgl.org)

