the globus project™
www.globus.org

Grid
Introduction and Overview

Ian Foster

Argonne National Lab
University of Chicago
Globus Project

www.mcs.anl.gov/~foster

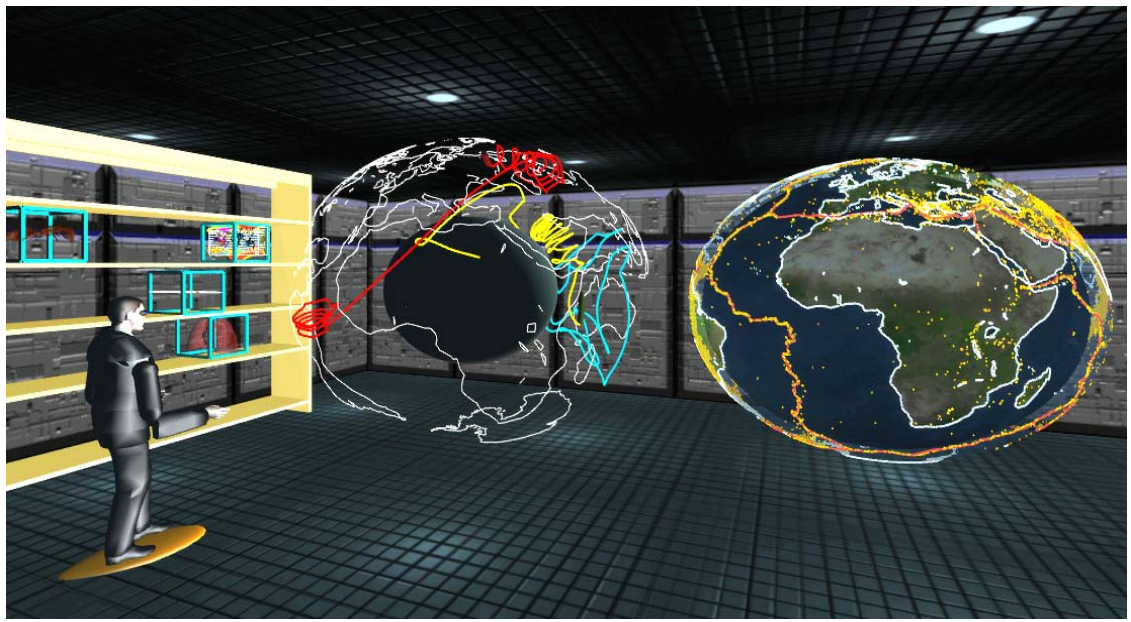


Image Credit: Electronic Visualization Lab, UIC

the globus project™
www.globus.org

Breaking News, August 14, 2003

BLACKOUT

BREAKING NEWS

Outage hits cities from NYC to Detroit

- New York City, Cleveland, Detroit, Toronto, Ottawa, other cities affected
- New York official: Outage result of electrical overload, not terrorism
- Niagara Mohawk power grid believed to be affected
- Ground stops ordered for airports in New York and Toronto because of outages for security screening

DEVELOPING STORY

- CNNRadio: Live coverage of power outage LIVE NOW

the globus project™
www.globus.org

Breaking News, August 14, 2013

BLACKOUT

Outage hits cities from NYC to Detroit

- New York City, Cleveland, Detroit, Toronto, Ottawa, other cities affected
- New York official: Outage computational overload, not terrorism
- Niagara Mohawk data grid believed to be affected
- Ground stops ordered for airports in New York and Toronto because of outages for security screening

DEVELOPING STORY

- CNNRadio: Live coverage of data outage LIVE NOW

the globus project™ www.globus.org

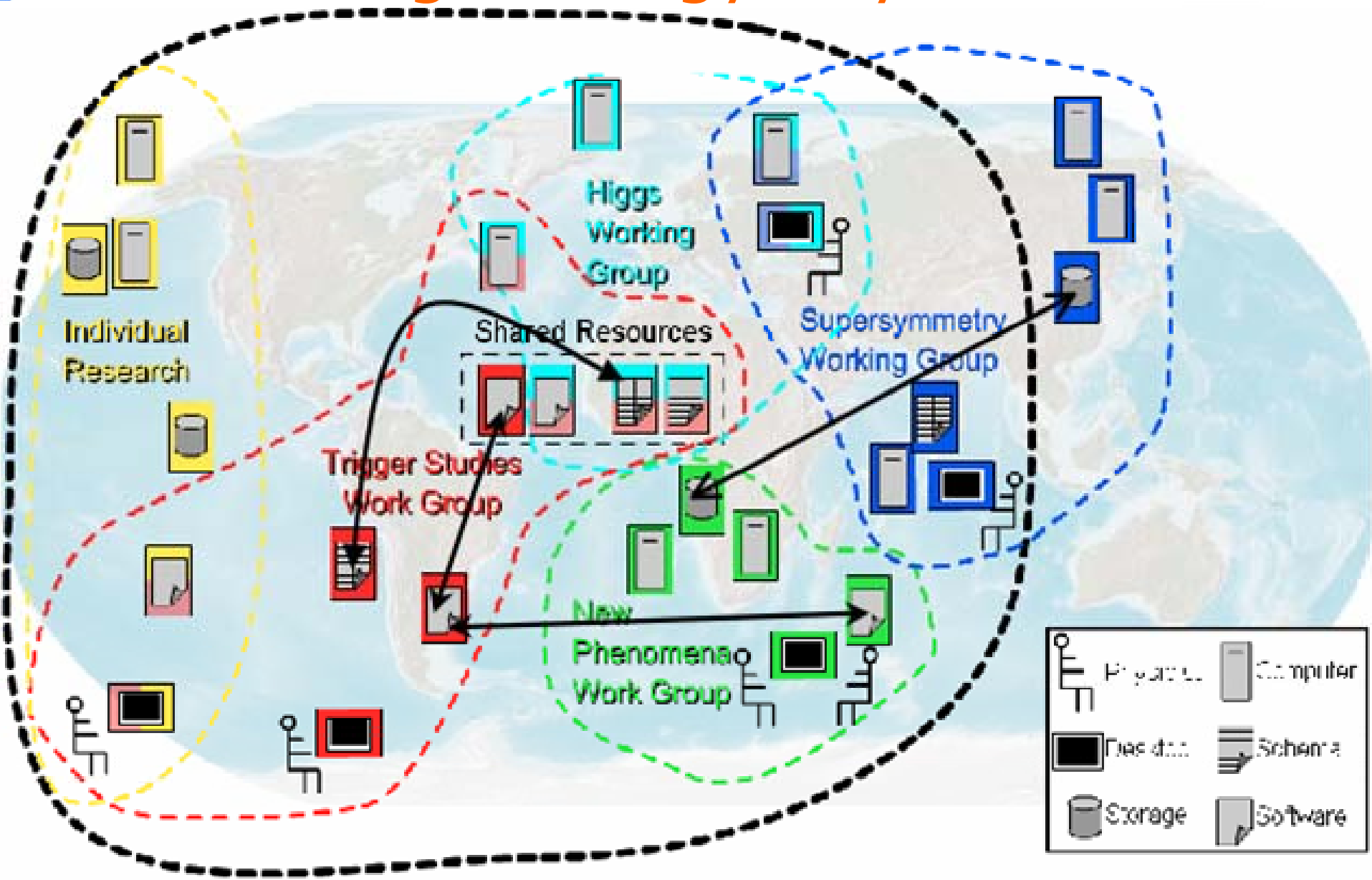
It's Easy to Forget How Different 2003 is From 1993

- Enormous quantities of data: Petabytes
  - For an increasing number of communities, gating step is not collection but analysis
- Ubiquitous Internet: 100+ million hosts
  - Collaboration & resource sharing the norm
- Ultra-high-speed networks: 10+ Gb/s
  - Global optical networks
- Huge quantities of computing: 100+ Top/s
  - Moore's law gives us all supercomputers

Consequence: The Emergence of Global Knowledge Communities

- Teams organized around common goals
  - Communities: "Virtual organizations"
- With diverse membership & capabilities
  - Heterogeneity is a strength not a weakness
- And geographic and political distribution
  - No location/organization possesses all required skills and resources
- Must adapt as a function of the situation
  - Adjust membership, reallocate responsibilities, renegotiate resources

For Example: High Energy Physics

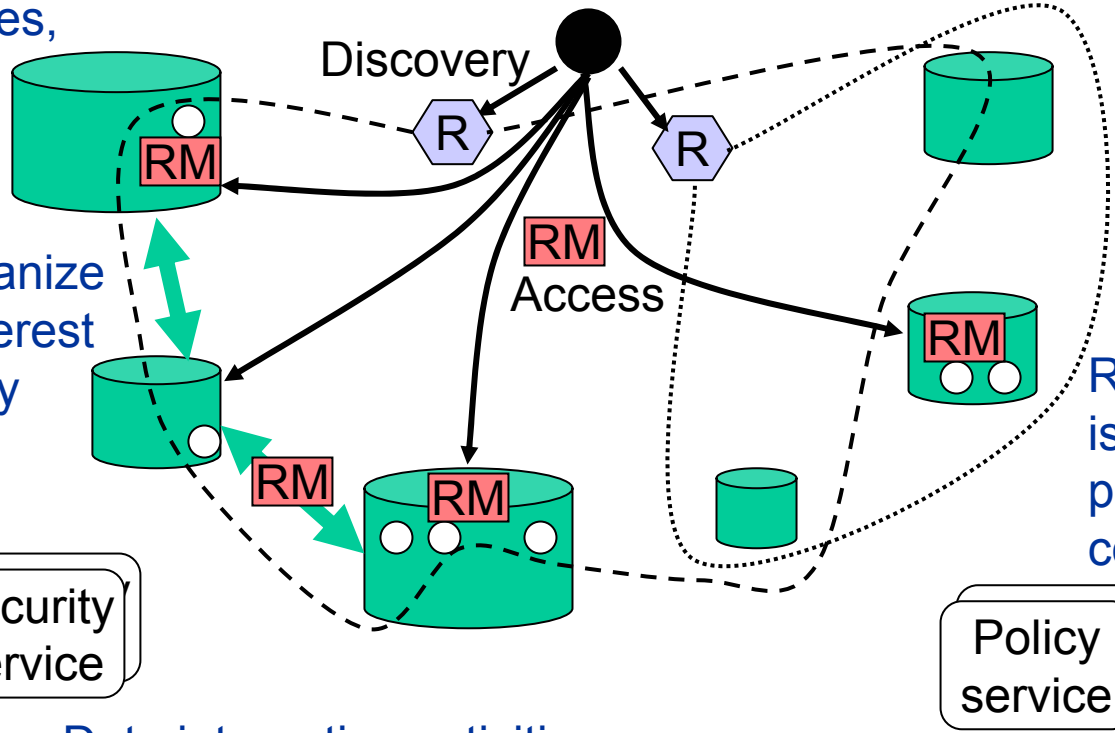


the globus project™ www.globus.org

Resource Integration as a Fundamental Challenge

Many sources of data, services, computation

Security & policy must underlie access & management decisions



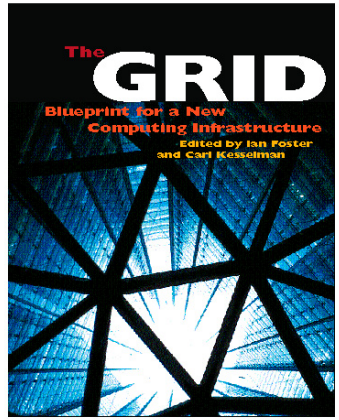
Registries organize services of interest to a community

Resource management is needed to ensure progress & arbitrate competing demands

Data integration activities may require access to, & exploration/analysis of, data at many locations

Exploration & analysis may involve complex, multi-step workflows

Grid Technologies Address Key Requirements



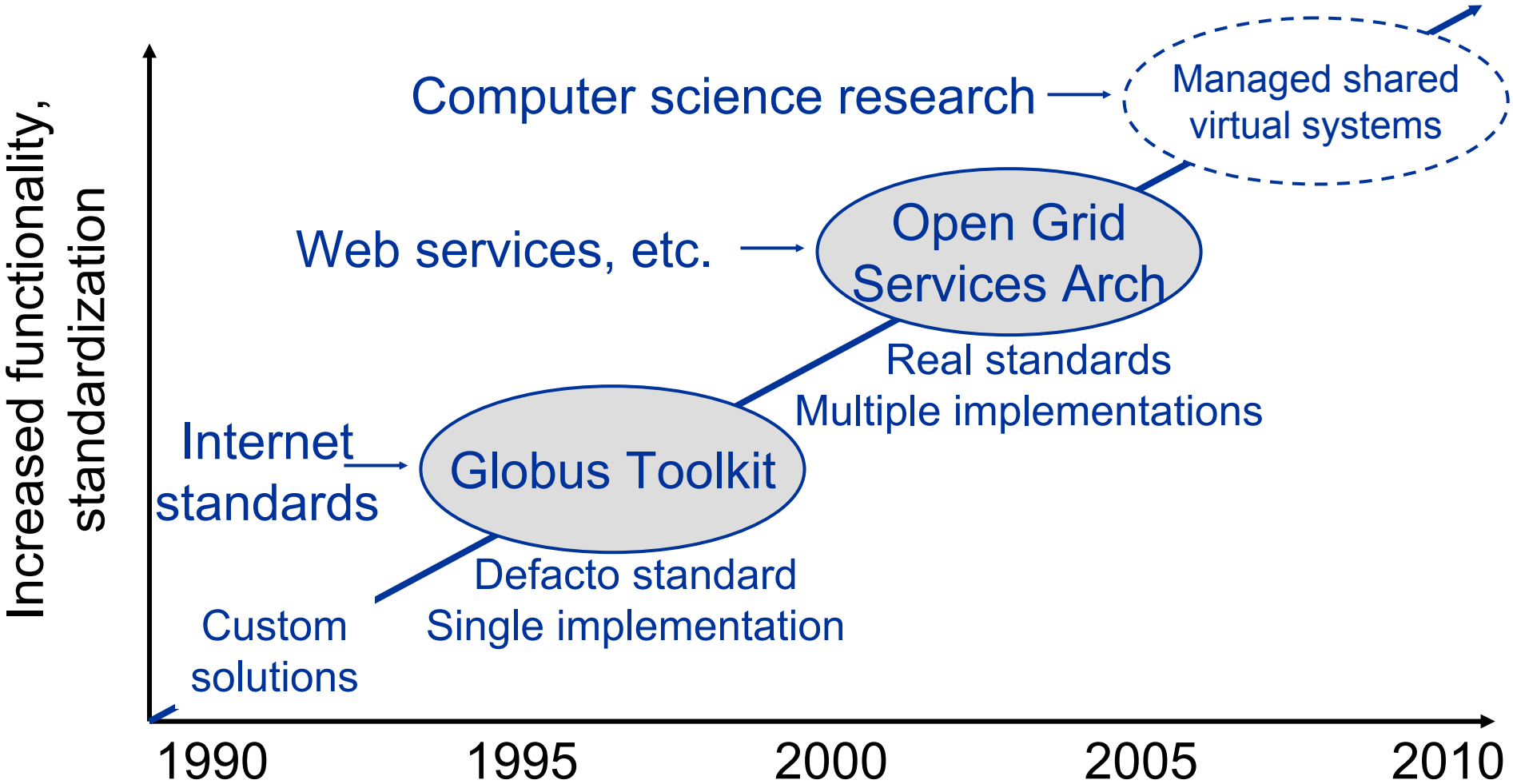
- Infrastructure ("middleware") for establishing, managing, and evolving multi-organizational federations
  - Dynamic, autonomous, domain independent
  - On-demand, ubiquitous access to computing, data, and services
- Mechanisms for creating and managing workflow within such federations
  - New capabilities constructed dynamically and transparently from distributed services
  - Service-oriented, virtualization

The Grid World: Current Status

- Substantial number of Grid success stories
  - Major projects in science
  - Emerging infrastructure deployments
  - Growing number of commercial deployments
- Open source Globus Toolkit® a de facto standard for major protocols & services
  - Simple protocols & APIs for authentication, discovery, access, etc.: infrastructure
  - Large user and developer base
  - Multiple commercial support providers
- Global Grid Forum: community & standards
- Emerging Open Grid Services Architecture

the globus project™ www.globus.org

The Emergence of Open Grid Standards

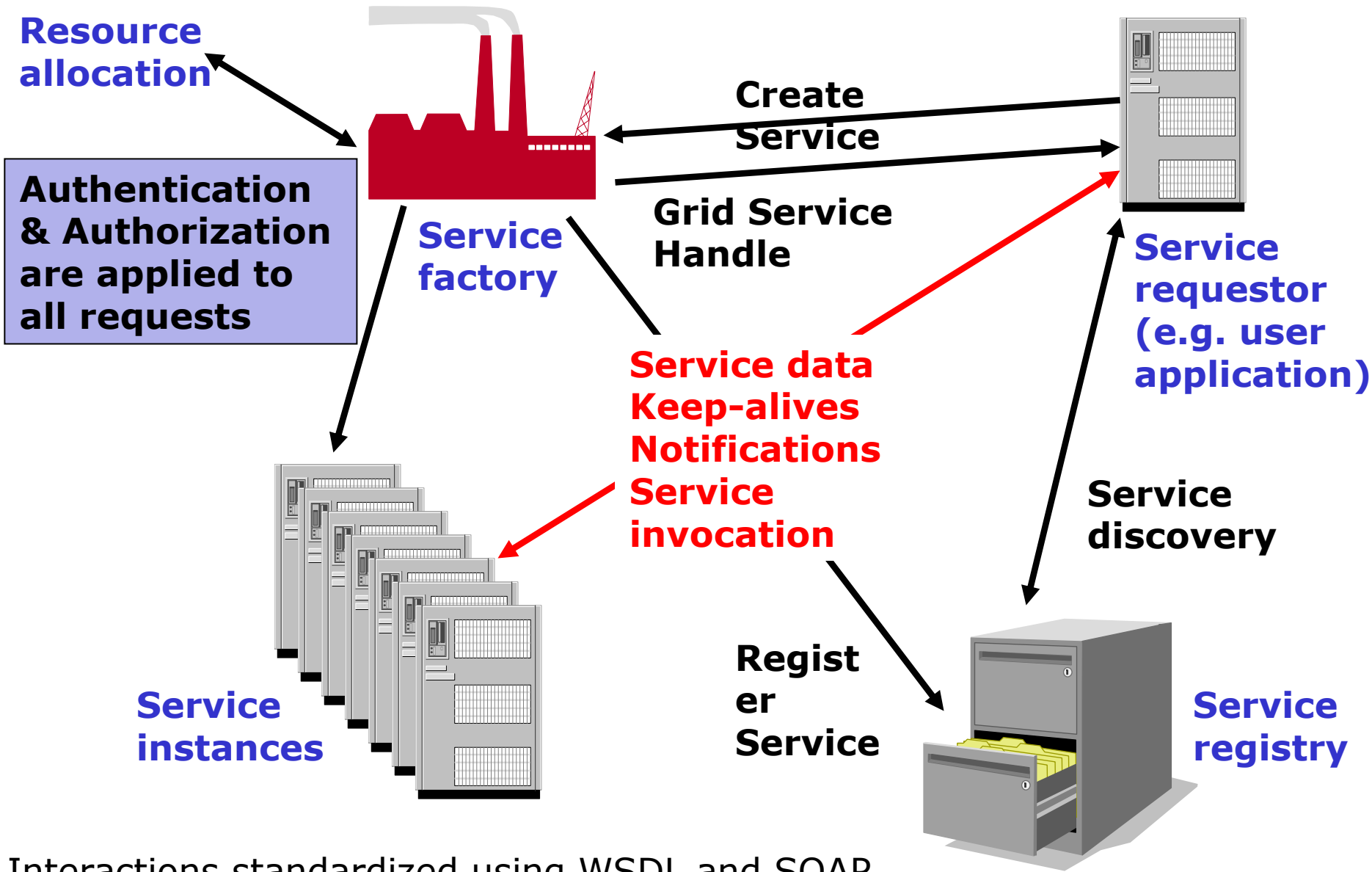


the globus project™ www.globus.org

Grid Evolution: OGSA (Open Grid Services Architecture)

- Goals
  - Refactor Globus protocol suite to enable common base and expose key capabilities
  - Service orientation to virtualize resources and unify resources/services/information
  - Embrace key Web services standards, leverage commercial efforts
- Result = standard interfaces & behaviors for distributed system mgmt: the Grid Service
  - Standardization within Global Grid Forum
  - GT3 open source implementation
- OGSA = Web services on steroids!

Open Grid Services Infrastructure (OGSI)



Interactions standardized using WSDL and SOAP

OGSA Standardization & Implementation

- OGSI defines core interfaces and behaviors for manageable services
- Efforts are underway within GGF, OASIS, and other bodies to define standards for
  - Agreement negotiation
  - Common management model
  - Data access and integration
  - Security and policy
  - Etc., etc., etc.
- Supported by strong open source technology & major commercial vendors

Grid Infrastructure

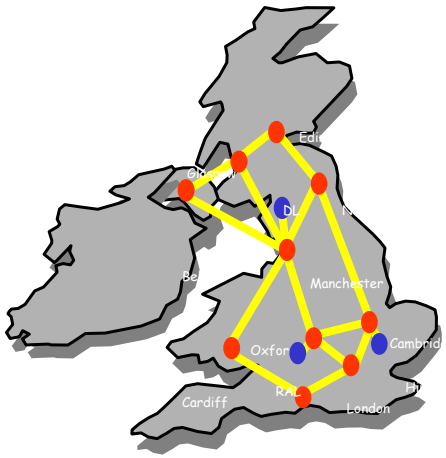
- Broadly deployed services in support of fundamental collaborative activities
  - Formation & operation of virtual organizations
  - Authentication, authorization, discovery, …
- Services, software, and policies enabling on-demand access to critical resources
  - Computers, databases, networks, storage, software services,…
- Operational support for 24x7 availability
- Integration with campus and commercial infrastructures

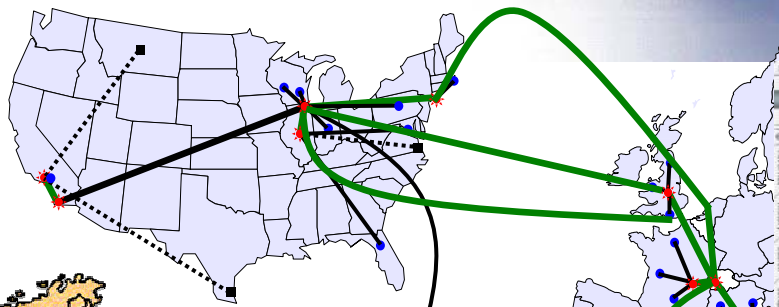
the globus project™
www.globus.org

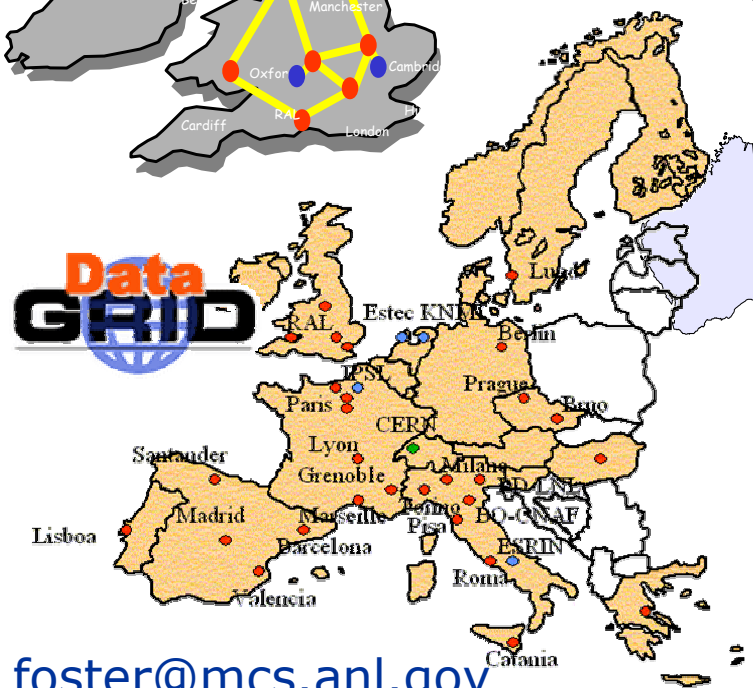
Open Infrastructure

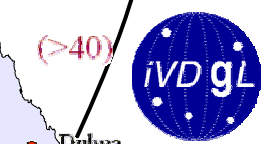
Building the National Virtual Collaboratory for Earthquake Engineering Research

NEESgrid

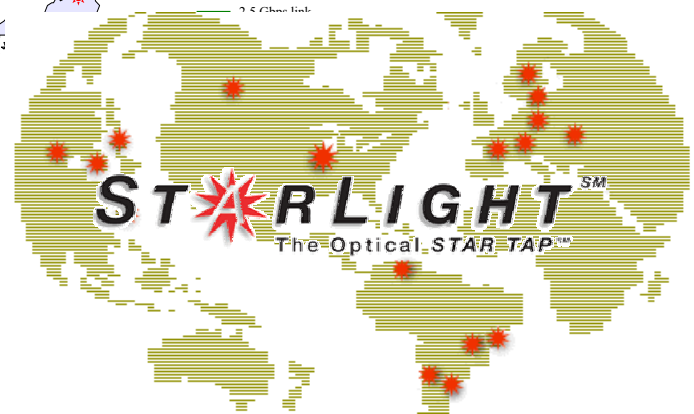








HEP sites
ESA sites

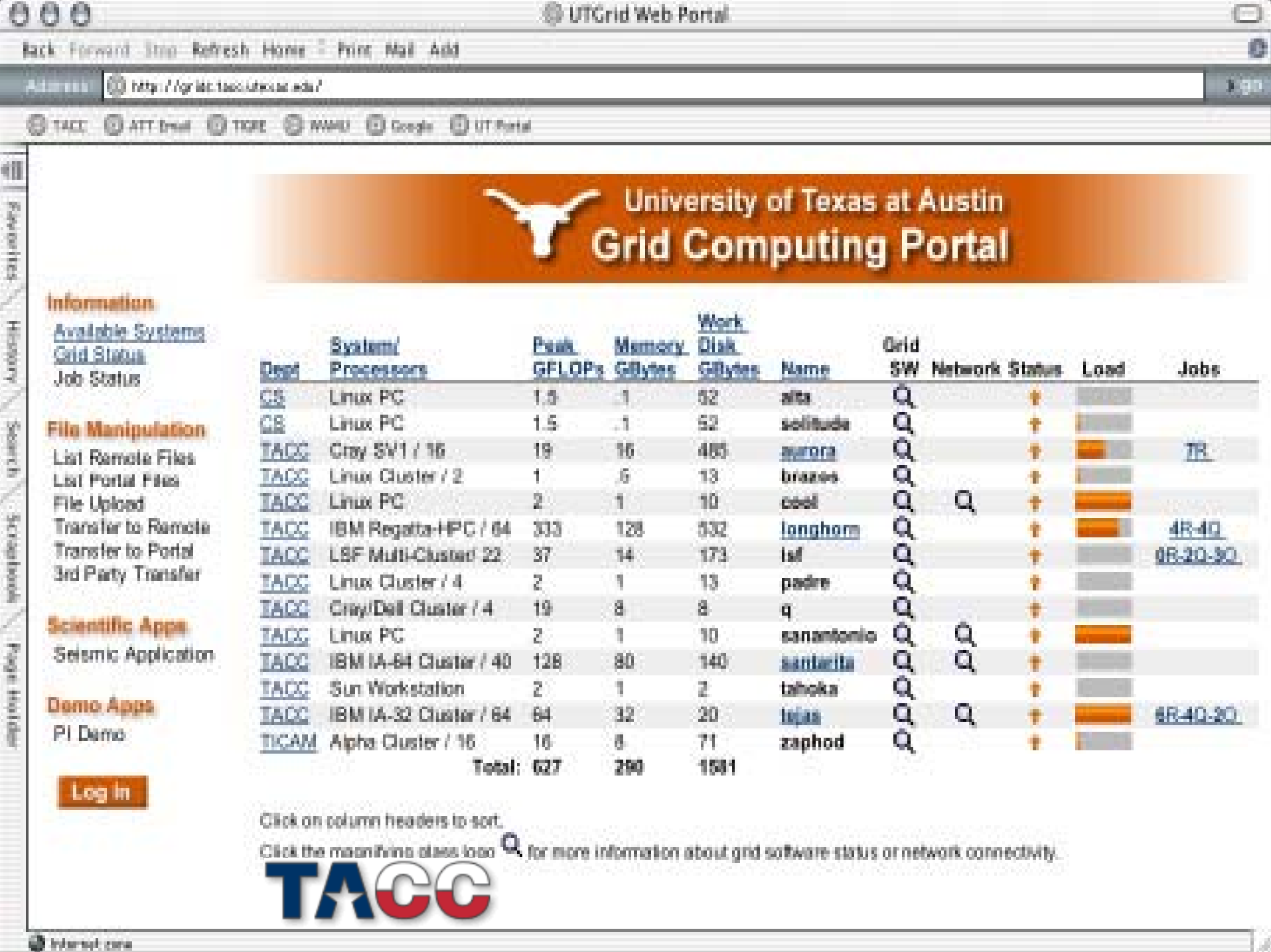


foster@mcs.anl.gov

the globus project™ www.globus.org

Where We Are: What We Can Do Today

- A core set of Grid capabilities are available and distributed in good quality form, e.g.
  - Globus Toolkit: security, discovery, access, data movement, etc.
  - Condor: scheduling, workflow management
  - Virtual Data Toolkit, NMI, EDG, etc.
- Deployed at moderate scales
  - WorldGrid, TeraGrid, NEESgrid, DOE SG, EDG, …
- Usable with some hand holding, e.g.
  - US-CMS event prod.: O(6) sites, 2 months
  - NEESgrid: earthquake engineering experiment


UTGrid Web Portal
Back Forward Stop Refresh Home Print Mail Add
Address http://gridlab.tacc.utexas.edu/
TACC ATT Email TIGRE WWU Google UT Portal
University of Texas at Austin
Grid Computing Portal
Information
Available Systems
Grid Status
Job Status
File Manipulation
List Remote Files
List Portal Files
File Upload
Transfer to Remote
Transfer to Portal
3rd Party Transfer
Scientific Apps
Seismic Application
Demo Apps
PI Demo
Log In
Dept
System/ Processors
Peak GFLOPs
Memory GBytes
Work Disk GBytes
Name
Grid SW
Network
Status
Load
Jobs
CS Linux PC 1.5 .1 52 alta
CS Linux PC 1.5 .1 52 solitude
TACC Cray SV1 / 16 19 16 485 aurora TR
TACC Linux Cluster / 2 1 .5 13 brazos
TACC Linux PC 2 1 10 cool
TACC IBM Regatta-HPC / 64 333 128 332 longhorn 4R-4Q
TACC LSF Multi-Clusted 22 37 14 173 isl 6R-2Q-3Q
TACC Linux Cluster / 4 2 1 13 padre
TACC Cray/Dell Cluster / 4 19 8 8 q
TACC Linux PC 2 1 10 sanantonio
TACC IBM IA-64 Cluster / 40 128 80 140 santarita
TACC Sun Workstation 2 1 2 tahoka
TACC IBM IA-32 Cluster / 64 64 32 20 lonestar 6R-4Q-3Q
TICAM Alpha Cluster / 16 16 8 71 zaphod
Total: 627 290 1581
Click on column headers to sort.
Click the magnifying glass icon for more information about grid software status or network connectivity.
TACC

the globus project™
www.globus.org

NEESgrid Earthquake Engineering Collaboratory

U.Nevada Reno

www.neesgrid.org

Remote Users
(Faculty, Students, Practitioners)

Instrumented Structures and Sites

Laboratory Equipment

High-Performance Network(s)

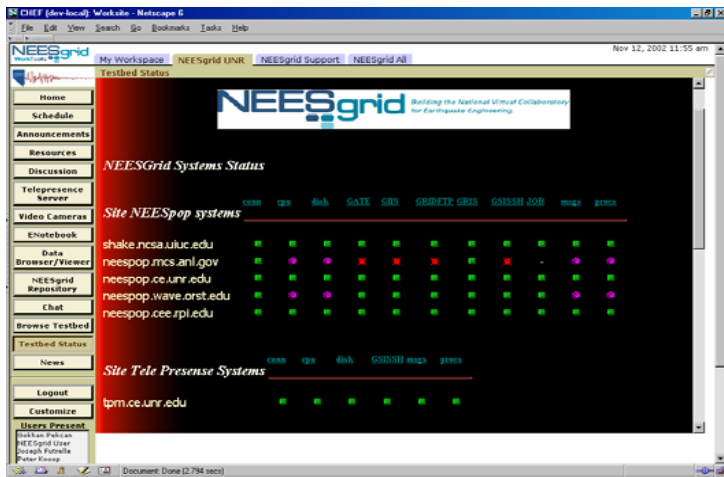
Field Equipment

Curated Data Repository

Global Connections
(fully developed FY 2005 – FY 2014)

Laboratory Equipment
(Faculty and Students)

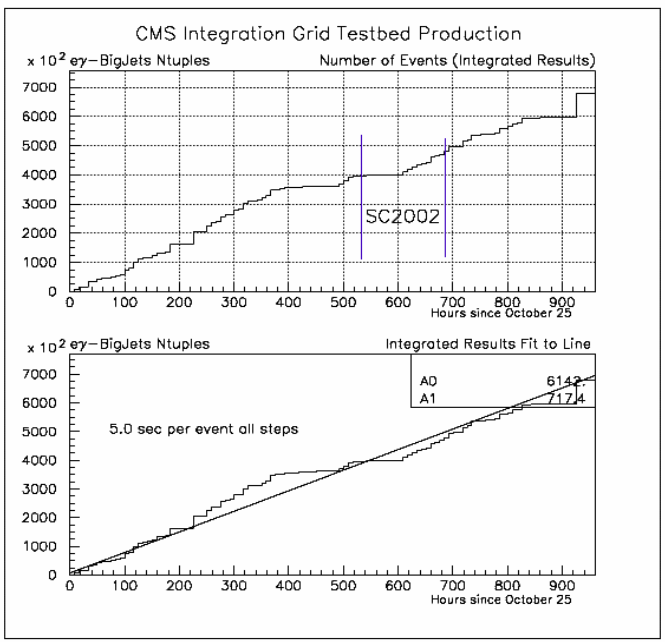
Remote Users:
(K-12 Faculty and Students)

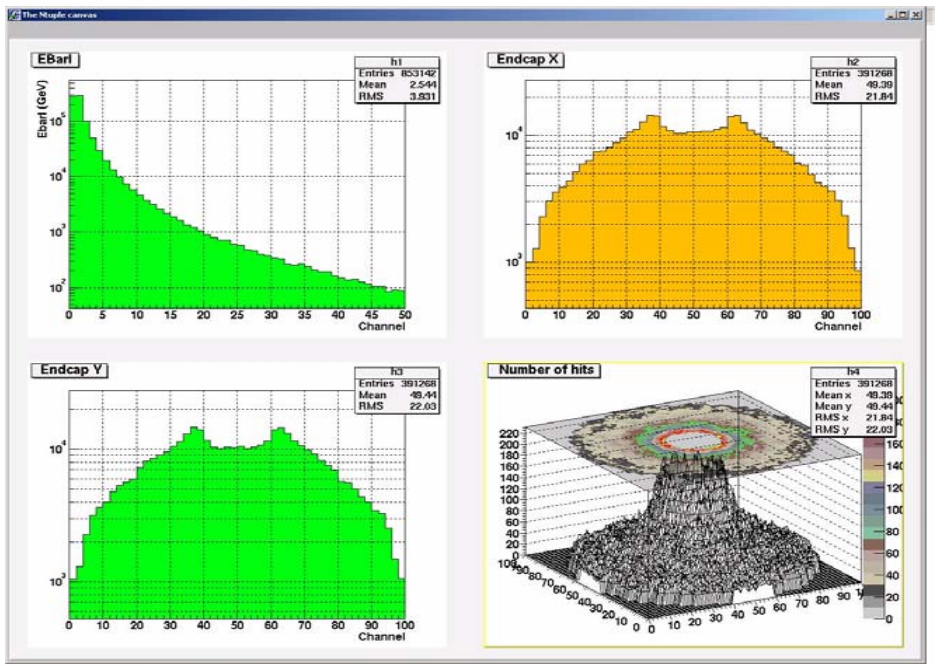


the globus project™ www.globus.org

CMS Event Simulation Production

- Production Run on the Integration Testbed
  - Simulate 1.5 million full CMS events for physics studies: ~500 sec per event on 850 MHz processor
  - 2 months continuous running across 5 testbed sites
  - Managed by a single person at the US-CMS Tier 1





the globus project™ www.globus.org

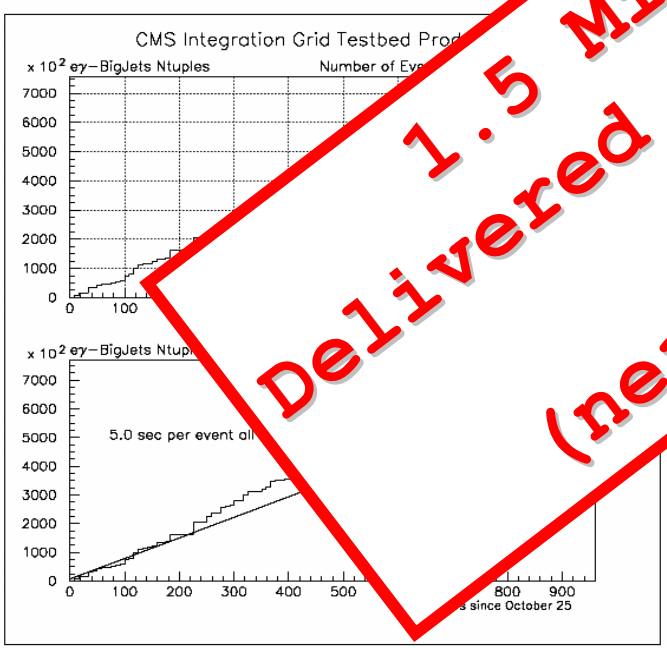
CMS Event Simulation Production

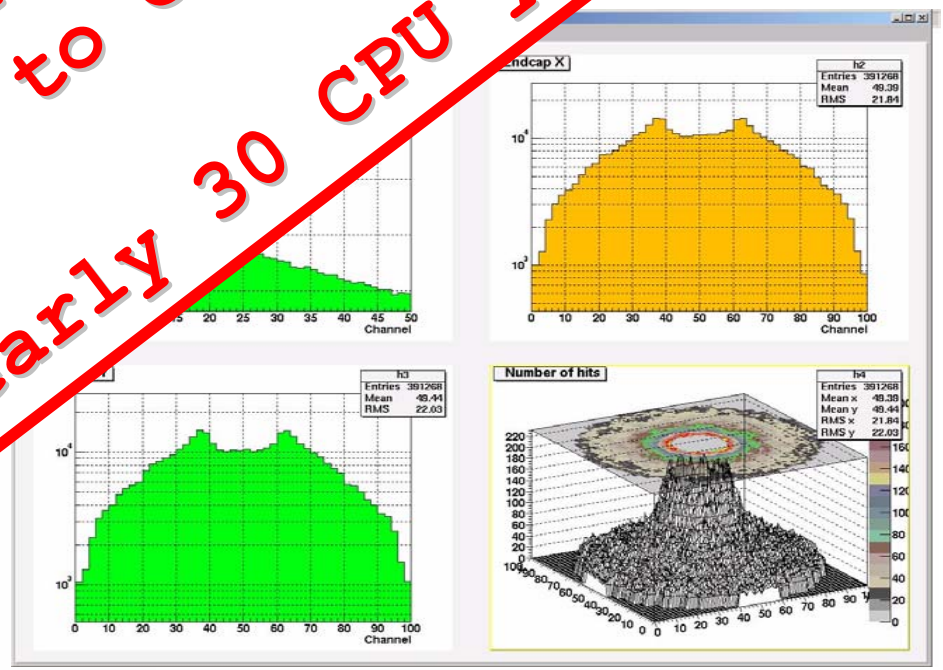
- Production Run on the Integrati
  - Simulate 1.5 million full CM
    studies: ~500 sec per e
  - 2 months continuou
  - Managed by a si

sites

Tier 1

1.5 Million Events
Delivered to CMS Physicists!
(nearly 30 CPU years)





the globus project™ www.globus.org

Where We Are: Key Areas of Concern

- Integration with site operational procedures
  - Many challenging issues
- Scalability in multiple dimensions
  - Number of sites, resources, users, tasks
- Higher-level services in multiple areas
  - Virtual data, policy, collaboration
- Integration with end-user science tools
  - Science desktops
- Coordination of international contributions
- Integration with commercial technologies

Summary: Grid Past, Present, Future

- Past
  - Origins and broad adoption in eScience, fueled by open source Globus Toolkit
- Present
  - Rapidly growing commercial adoption focused on intra-enterprise resource sharing
  - Increasingly large scale infrastructures
  - Open Grid Services Architecture (OGSA)
- Future
  - Key enabler of new applications & industries based on resource virtualization and distributed service integration

the globus project™
www.globus.org

For More Information

- GriPhyN, iVDGL, PPDG
  - www.griphyn.org, www.ivdgl.org, www.ppdg.net
- The Globus Project®
  - www.globus.org
- Global Grid Forum
  - www.ggf.org
- Background information
  - www.mcs.anl.gov/~foster

- **GlobusWORLD 2004**
  - **www.globusworld.org**
  - **Jan 20–23, San Fran**