Globus and Grids

Jennifer M. Schopf
Argonne National Lab

Problem Solving in the 21st Century

- 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

Taking Sharing to the Next Level

- Sharing of communication
  - Telephones, mailing lists, collaboration tools
- Sharing of data and knowledge
  - Web, semantic web
- What about the rest of the infrastructure?
  - Services, computers, programs, sensors, …

Existing Technologies are Helpful, but Not Complete Solutions

- Peer-to-peer technologies
  - Limited scope and mechanisms
- Enterprise-level distributed computing
  - Limited cross-organizational support
- Databases
  - Vertically integrated solutions
- Web services
  - Not dynamic
- Semantic web
  - Limited focus

What's Missing is Support for ...

- Sharing & integration of resources, via
  - Discovery
  - Provisioning
  - Access (computation, data, …)
  - Security
  - Policy
  - Fault tolerance
  - Management
- In dynamic, scalable, multi-organizational settings

Building the Grid
(according to Ian Foster)

- Open source software
  - Globus Toolkit® , UK OGSA DAI, Condor, …
- Open standards
  - OGSA, other GGF, IETF, W3C standards, …
- Open communities
  - Global Grid Forum, Globus International, collaborative projects, …
- Open infrastructure
  - UK eScience, NSF Cyberinfrastructure, StarLight, AP-Grid, …

Globus and the Grid

- 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 Globus Project™

- A group of people with a common mission:

  "Make Grid computing an everyday reality"

- Housed at Argonne National Laboratory, Univ. of Chicago, and USC Information Sciences Institute

  – Led by Ian Foster (ANL, U-C), Carl Kesselman (ISI)

  – Includes researchers, software developers, software architects & designers, systems engineers, etc.

  – Collaborations (or at least acquaintances) with most Grid activities in the world

Globus Project Activities

- All activities contribute to our common mission
  - Research
  - Software Development (prototypes, reference implementations)
  - **Application consulting**
  - **Infrastructure consulting**

The Globus Project cont.

- Close collaboration with real Grid projects in both science and industry

- The Globus Toolkit®: Open source software base for building Grid *infrastructure* and *applications*

- Development and promotion of standard *Grid protocols and services* to enable interoperability and shared infrastructure

- Development and promotion of standard *Grid software APIs* to enable portability and code sharing

- Global Grid Forum: We co-founded GGF to foster Grid standardization and community
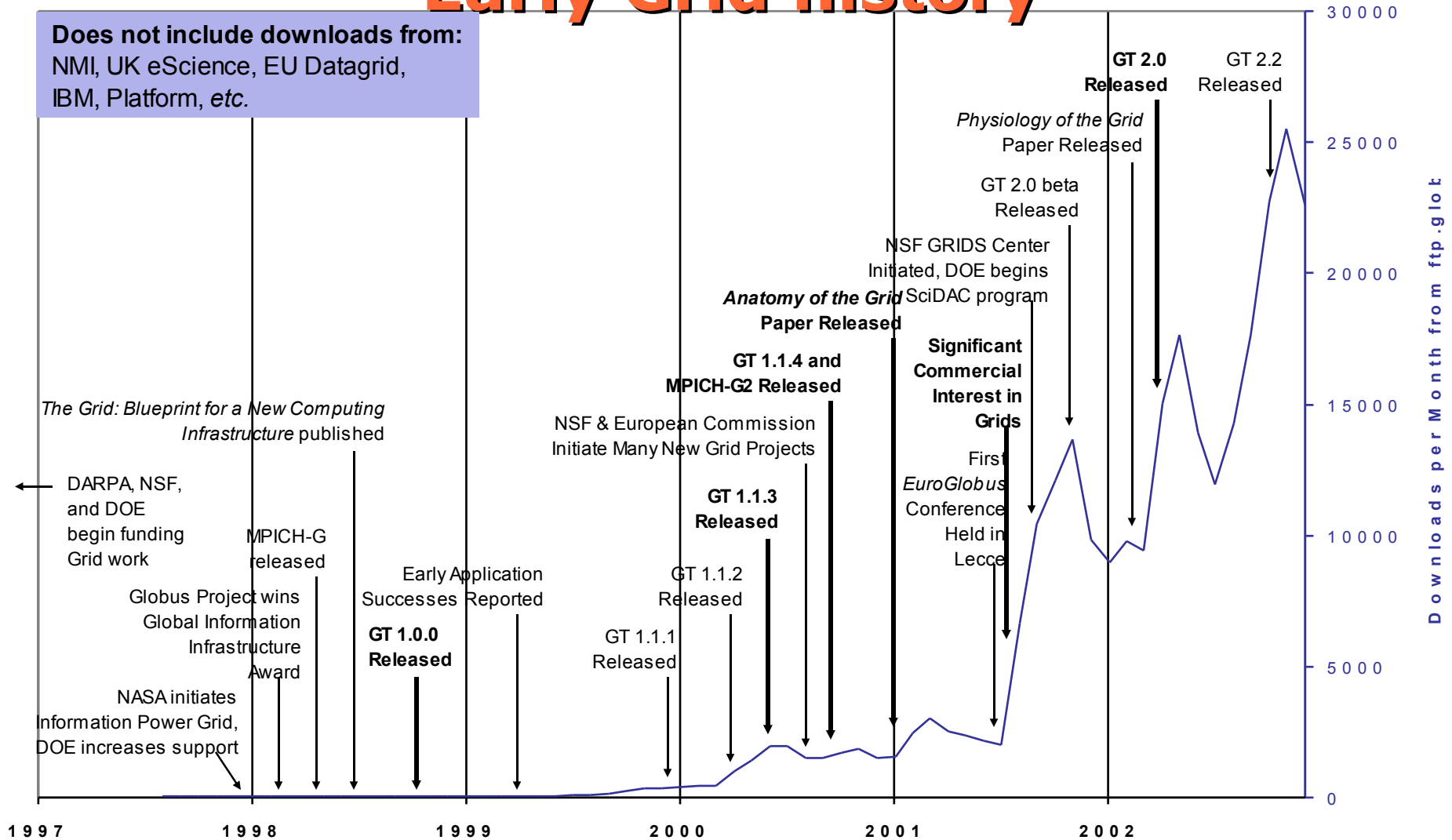
Globus Project Methodology

- Identify theoretical applications or user communities.

- Establish collaborations with target users

- Identify key requirements of target users

- Identify common problems & requirements across many target users

- Develop architecture and designs for proposed technological solutions to common problems

- Implement usable versions of solutions

- Work with target users to integrate proposed solutions and evaluate results

- Propose standards to relevant communities

- Iterate…

Key Events in Early Grid history

Does not include downloads from: NMI, UK eScience, EU Datagrid, IBM, Platform, *etc.*

GT 2.0 Released

GT 2.2 Released

Physiology of the Grid Paper Released

GT 2.0 beta Released

NSF GRIDS Center Initiated, DOE begins SciDAC program

Anatomy of the Grid **Paper Released**

GT 1.1.4 and MPICH-G2 Released

Significant Commercial Interest in Grids

The Grid: Blueprint for a New Computing Infrastructure published

NSF & European Commission Initiate Many New Grid Projects

GT 1.1.3 Released

First *EuroGlobus* Conference Held in Lecce

DARPA, NSF, and DOE begin funding Grid work

MPICH-G released

Early Application Successes Reported

GT 1.1.2 Released

Globus Project wins Global Information Infrastructure Award

GT 1.0.0 Released

GT 1.1.1 Released

NASA initiates Information Power Grid, DOE increases support

Downloads per Month from ftp.globt

30000
25000
20000
15000
10000
5000
0

1997   1998   1999   2000   2001   2002

Who's using Globus?

Access Grid, AeroDB Trials, Astrophysics Simulation Collaboratory, ATLAS Data Challenge 1 Full Event Simulation and Reconstruction, Avaki, Aviation Safety Project, BioGridRunner , Biomedical Informatics Research Network (BIRN), Blood Flow Simulation Project, Butterfly.net, CCLRC e-Science Centre, Charles Schwab, CroatiaGrid (Grid for Scientific and Economic Development of Croatia), CrossGrid, DataSynapse, DOE Science Grid, Dutch Grid (ASCI), Earth System Grid, Entropia, Espresso Model Interface, European Union DataGrid, EZ-Grid. Fusion Grid, National Fusion Collaboratory (NFC), Genome Analysis and Databases Upload, Geodise , GlobeXplorer, GPDK, Grid Application Framework for Java (GAF4J), GRid Interoperability Project (GRIP), Grid-based Visualization Framework, GridFTP User Community, Gridlab , GridPort, GridSolve, GRIDSTART. HPCPortal and the UK e-Science Grid, IBM, ICENI - Imperial College e-Science Networked Infrastructure, Indiana-NCSA Science Portal, Java CoG Box, …

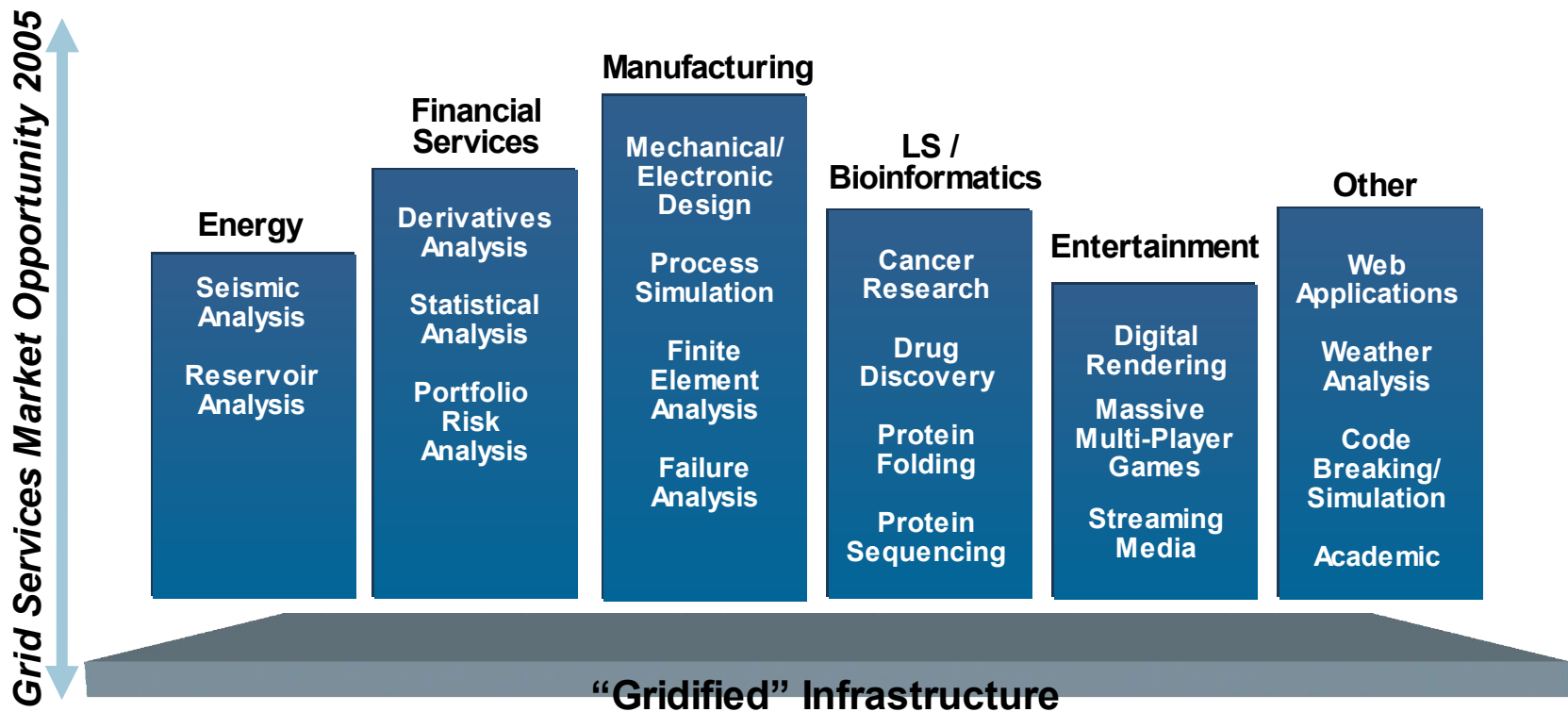**Globus Overview, Jennifer M. Schopf**

Who's using Globus, cont.

… JiPANG, KoreanGrid, Laser Gravitational Wave Observatory, Mgrid, Molecular Science Software Suite MS3, Mpich-G2, MyProxy CoG, NASA Information Power Grid (IPG) Infrastructure, NASA IPG Launch Pad, National Digital, Mammography Archive, NEESgrid, NERC DataGrid (NDG), Neuroscience: Assembling Visible Neurons for Simulations, Nimrod-g, NINF-G, Nordugrid, NPACI Grid, NSF Alliance Grid Infrastructure, NSF GrADS, NVO, Open Bioinformatics Grid, Oracle, Petascale Data Quest (PDQ), Platform Computing, ProActive, Purdue University, Reptor , Scientific Portal: Alliance Expedition, Southern California Earthquake Center (SCEC), Storage Resource Broker (SRB), Subsurface Science and Simulation for Environmental Cleanup, Symphony , TENT, The Texas Advanced Computing Center (TACC), TIGRE Testbed Portal, Top500 List, TRASC: A Globus Application Launcher, United Devices, CMS Grid, Virtual Observatory of China, XCAT, …. **AND MORE!**

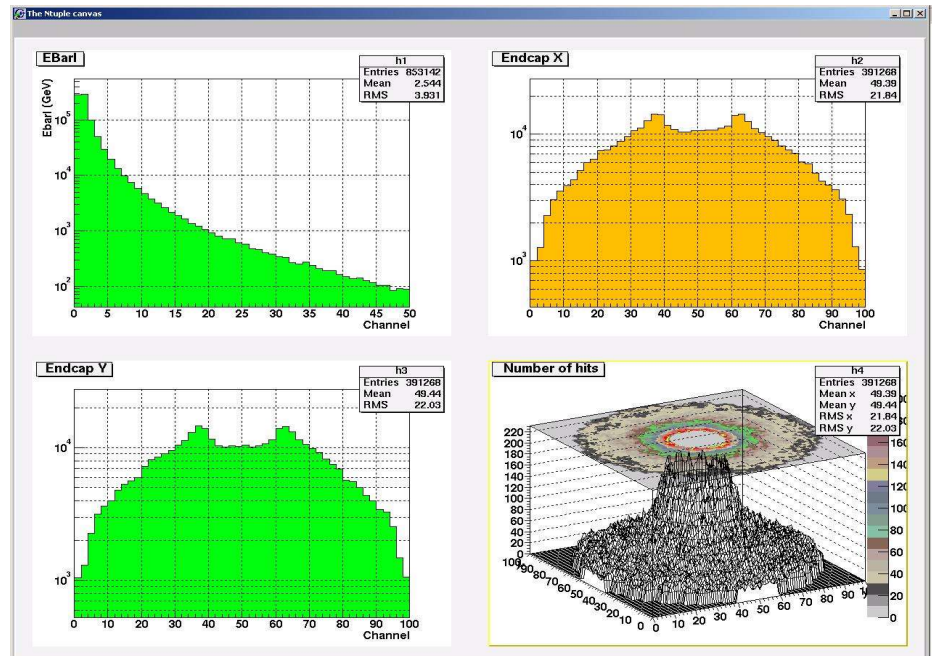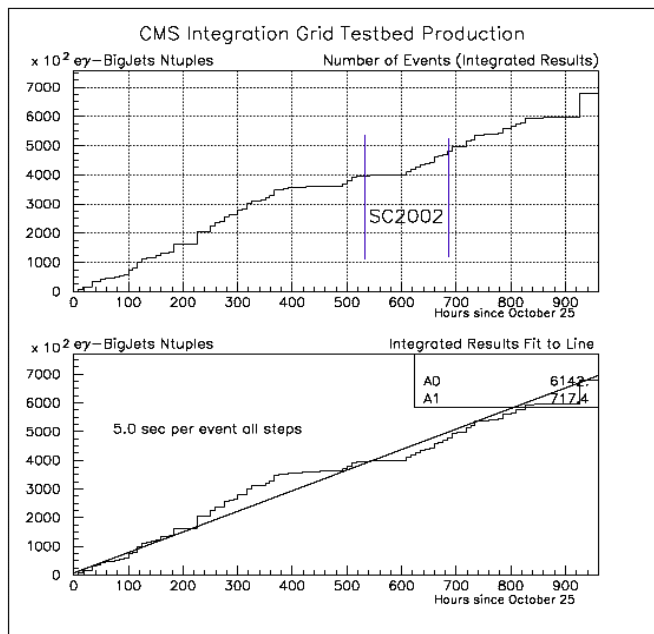# Industrial Perspective on Grids: A Wide Range of Applications

Unique by Industry with Common Characteristics

Grid Services Market Opportunity 2005 (vertical axis)

Energy
- Seismic Analysis
- Reservoir Analysis

Financial Services
- Derivatives Analysis
- Statistical Analysis
- Portfolio Risk Analysis

Manufacturing
- Mechanical/ Electronic Design
- Process Simulation
- Finite Element Analysis
- Failure Analysis

LS / Bioinformatics
- Cancer Research
- Drug Discovery
- Protein Folding
- Protein Sequencing

Entertainment
- Digital Rendering
- Massive Multi-Player Games
- Streaming Media

Other
- Web Applications
- Weather Analysis
- Code Breaking/ Simulation
- Academic

"Gridified" Infrastructure

Sources: IDC, 2000 and Bear Stearns- Internet 3.0 - 5/01 Analysis by SAI

Example:
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
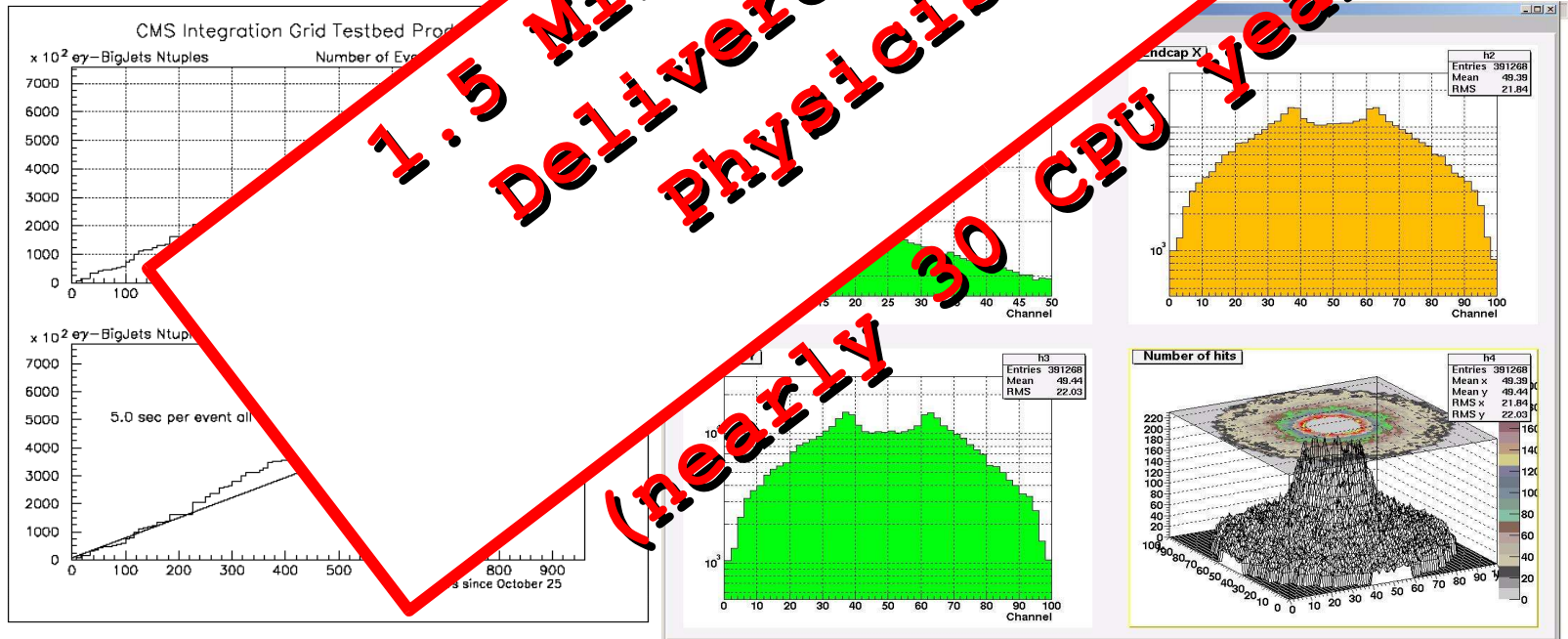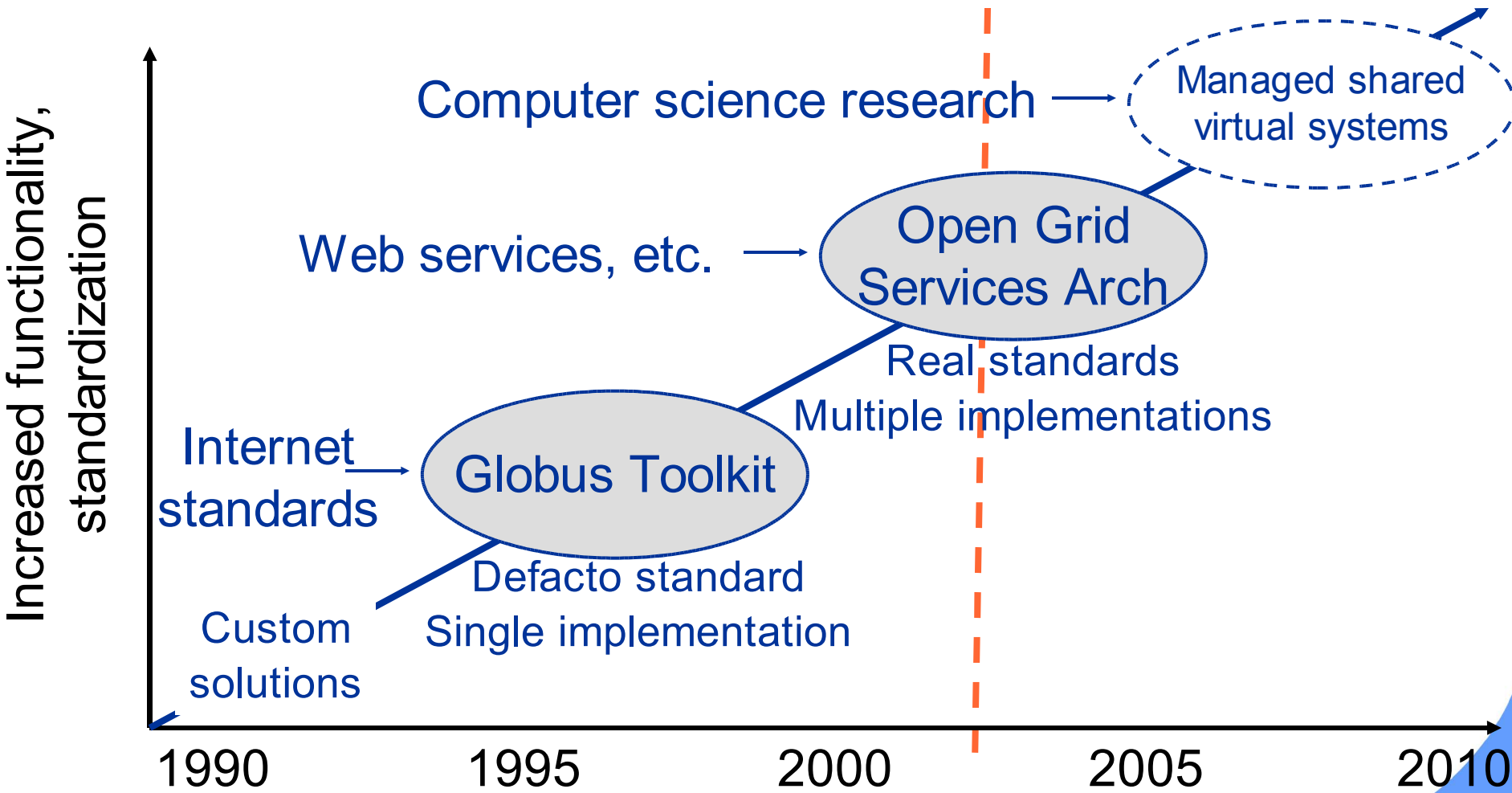 - Managed by a single person at the US-CMS Tier 1

CMS Event Simulation P........on

- Production Run on the Integrat...
  - Simulate 1.5 million full CM....
    studies: ~500 sec per ....
  - 2 months continuo.......sites
  - Managed by a ....s Tier 1

1.5 Million Events Delivered to CMS Physicists!

(nearly 30 CPU Years)

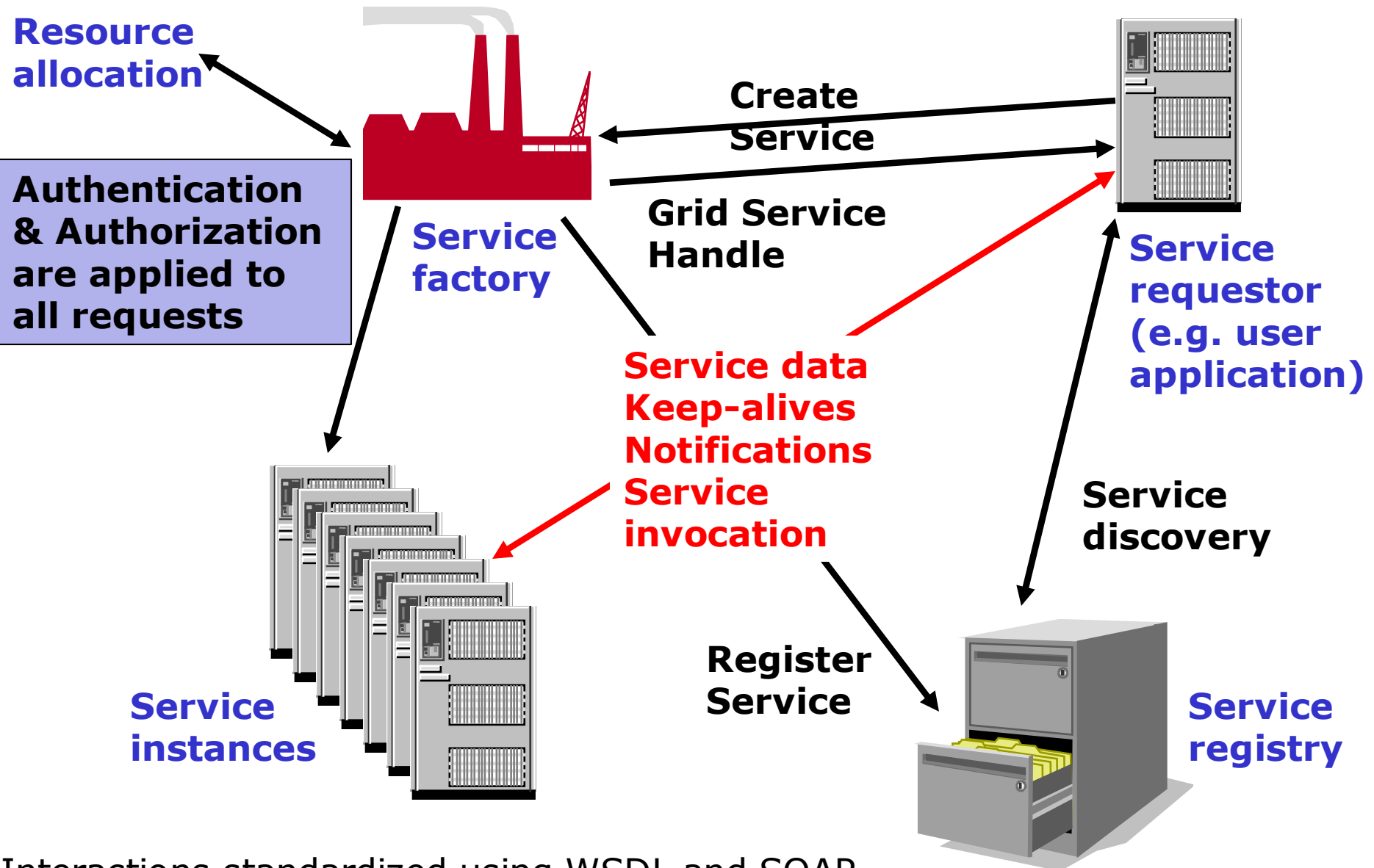# The Emergence of
Open Grid Standards

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)

Resource allocation

Authentication & Authorization are applied to all requests

Service factory

Create Service

Grid Service Handle

**Service data
Keep-alives
Notifications
Service invocation**

Service requestor (e.g. user application)

Service instances

Register Service

Service discovery

Service registry

Interactions standardized using WSDL and SOAP

OGSA Standardization & Implementation

- OGSI defines core interfaces and behaviors for manageable services

- Efforts are underway to define standards for
    - Agreement negotiation
    - Common management model
    - Data access and integration
    - Security and policy
    - Etc.

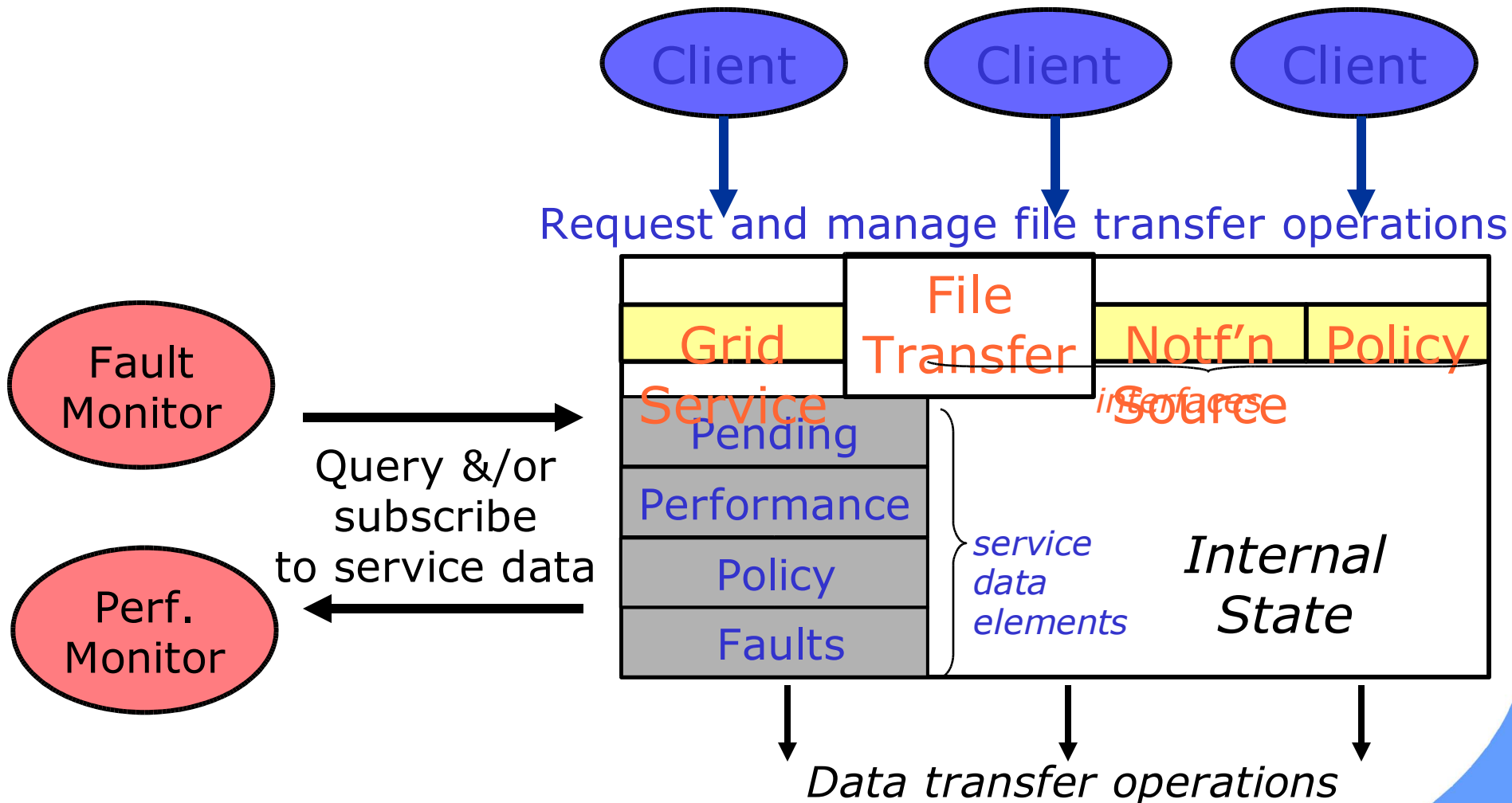- Supported by strong open source technology & major commercial vendors

Globus Toolkit v3 (GT3) Open Source OGSA Technology

- Implements and builds on OGSI interfaces
- Supports primary GT2 interfaces
  - Authentication, resource discovery, resource access, data movement, …
- Multiple platforms & hosting environments
  - J2EE, Java, C, .NET, Python
- New services
  - SLA negotiation, service registry, community authorization, data management, …
- Rapidly growing adoption and contributions

Example:
Reliable File Transfer Service

Client   Client   Client

Request and manage file transfer operations

File Transfer Source

Grid Service   Notf'n   Policy

interfaces

Fault Monitor

Query &/or subscribe to service data

Perf. Monitor

Pending

Performance

Policy

Faults

service data elements

Internal State

Data transfer operations

OGSA Future Directions

- OGSI leaves wide open many opportunities for new, higher-level management capabilities

  - Service Management capabilities

  - Service Level Agreements

  - "Autonomic computing"

- Major shift from applications using existing Grid services to applications being compositions of new Grid services

Next Steps for the Globus Toolkit

- Continue to serve as the vendor-neutral open platform of choice for Grid computing
  - A locus for commercial & noncommercial open source contributions
  - An enabler of commercial & noncommercial Grid computing solutions
- To this end, we will continue to
  - Work within the Global Grid Forum to define technical specifications and best practices
  - Work with industry and academia to implement and integrate Grid technologies

Globus Project
Goals for 2003-4

- Define and deliver key OGSA interfaces, e.g.
  - SLA management & provisioning services
  - Grid service security model & services
  - Data access and integration services
  - Monitoring and discovery services
- GT 3.O RELEASE June 2002!
  - Address transition & operations issues
- GT 3.2 release end of 2003, early 2004
  - New GridFTP server, Community access service, better index service

For More Information

- The Globus Project™
  - www.globus.org
- Technical articles
  - www.mcs.anl.gov/~foster
- Open Grid Services Arch.
  - www.globus.org/ogsa
- Global Grid Forum
  - www.gridforum.org

The **GRID**
Blueprint for a New Computing Infrastructure
Edited by Ian Foster and Carl Kesselman