Enabling Grids for E-science in Europe

www.eu-egee.org

International Summer School on Grid Computing
Vico Equense, 13^{th} July 2005

Today's Wealth of Data: Are we ready for its challenges?

Malcolm Atkinson
Director
National e-Science Centre
www.nesc.ac.uk

EGEE is a project funded by the European Union under contract IST-2003-508833

Outline

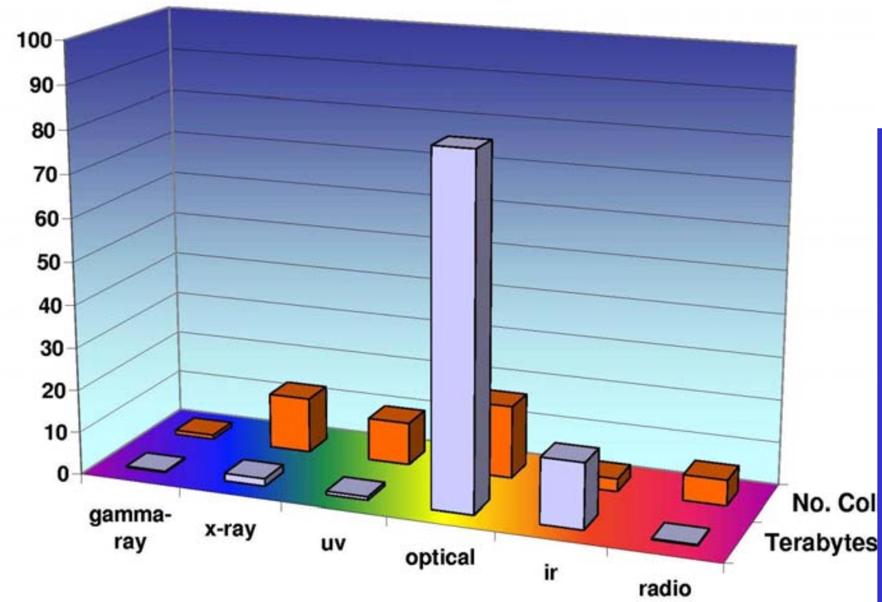
- **The Data Explosion**
  - **Central role of Data**
- **Examples & Discussion**
- **Growth of Data**
- **Challenges for Data Users**
- **Challenges for Data Creators**
- **Challenges for Distributed System Engineers**
- **What is e-Science?**
  - **Collaboration & Virtual Organisations**
  - **Structured Data at its Foundation**

Central Role of Data

- Accumulated from Instruments & Computation
  - Sky surveys, Ocean Surveys, Geo-surveys, …
    - ***The real record*** of experiments and phenomena
  - Crystallography, e-Chemistry, automated bio-labs, …
  - Dynamic observation – track a cosmic burst or a 1A Supernova
    - A thunder storm to diagnose tornado formation, …
  - Diagnostic engine data, Personal health data, …
  - Large physics experiments, …
- Shared Information
  - >400 biological DBs
    - FlyBase, PDB, Mouse Atlas, ZOOdb, biodiversity, …
- Results of a Stage of Processing
  - Derivatives become primary data for many users
  - Branching, Retries & iterative development, conserving results
- Definitive resource for scientific criticism

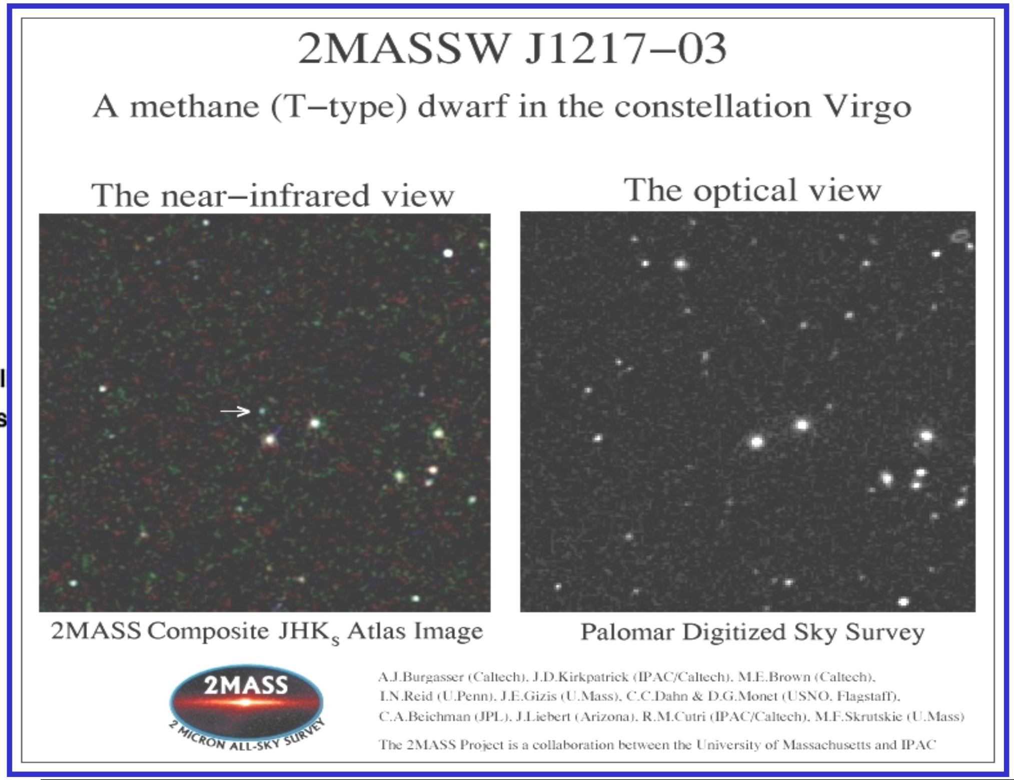
Composing Observations in Astronomy

Enabling Grids for E-science in Europe



No. & sizes of data sets as of mid-2002, grouped by wavelength

- 12 waveband coverage of large areas of the sky
- Total about 200 TB data
- Doubling every 12 months
- Largest catalogues near 1B objects



THE NEW YORK TIMES NATIONAL TUESDAY, JUNE 1, 1999

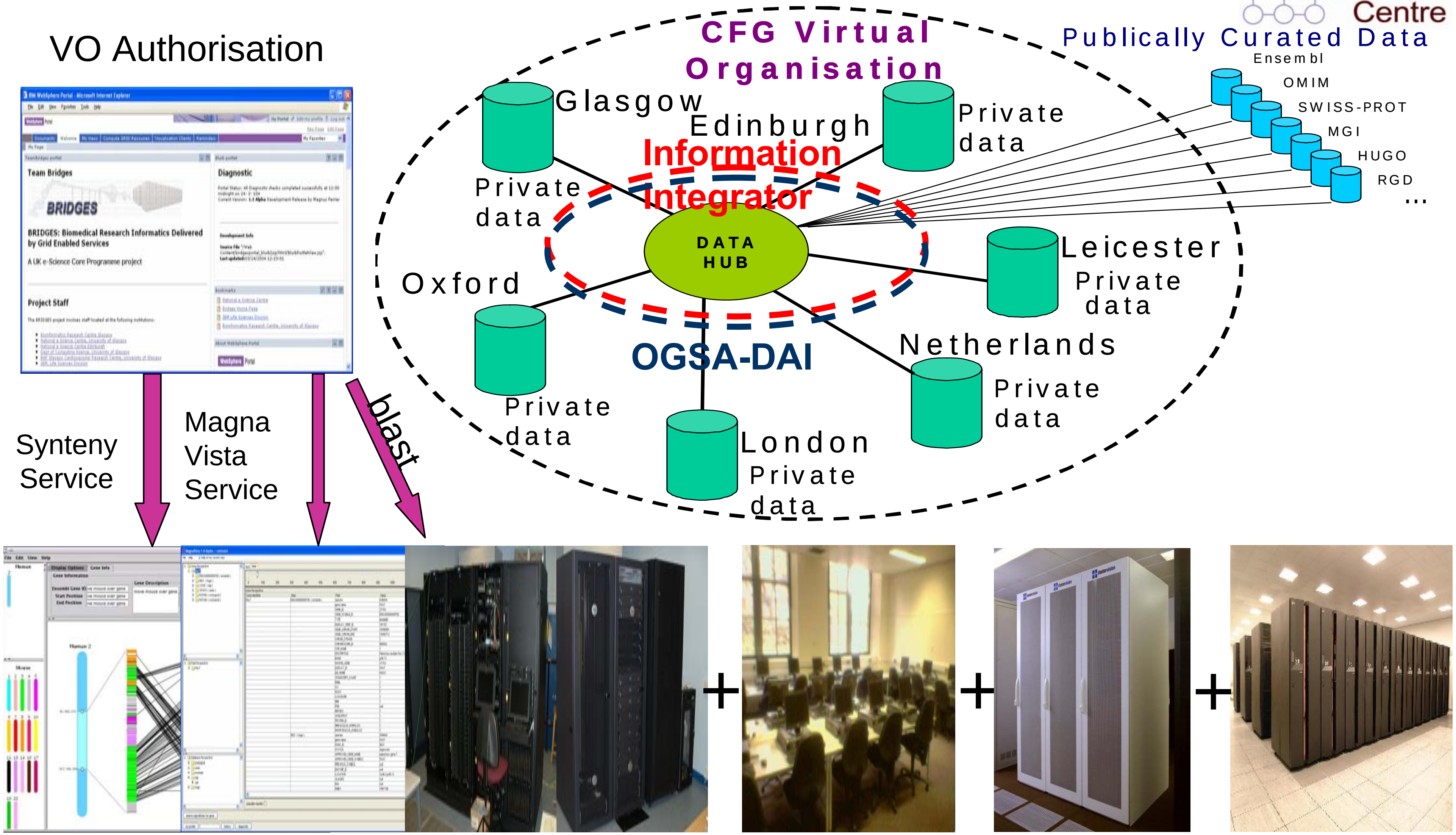
Astronomers Detect New Category of Elusive 'Brown Dwarfs'

By JOHN NOBLE WILFORD

CHICAGO, May 31 — Ambitious ... Apache Point, N.M. Dr. Michael Strauss and a graduate student, ... was a brown dwarf, but was not associated with a star companion. ... have been possible in hotter, younger objects. An estimate of their mass

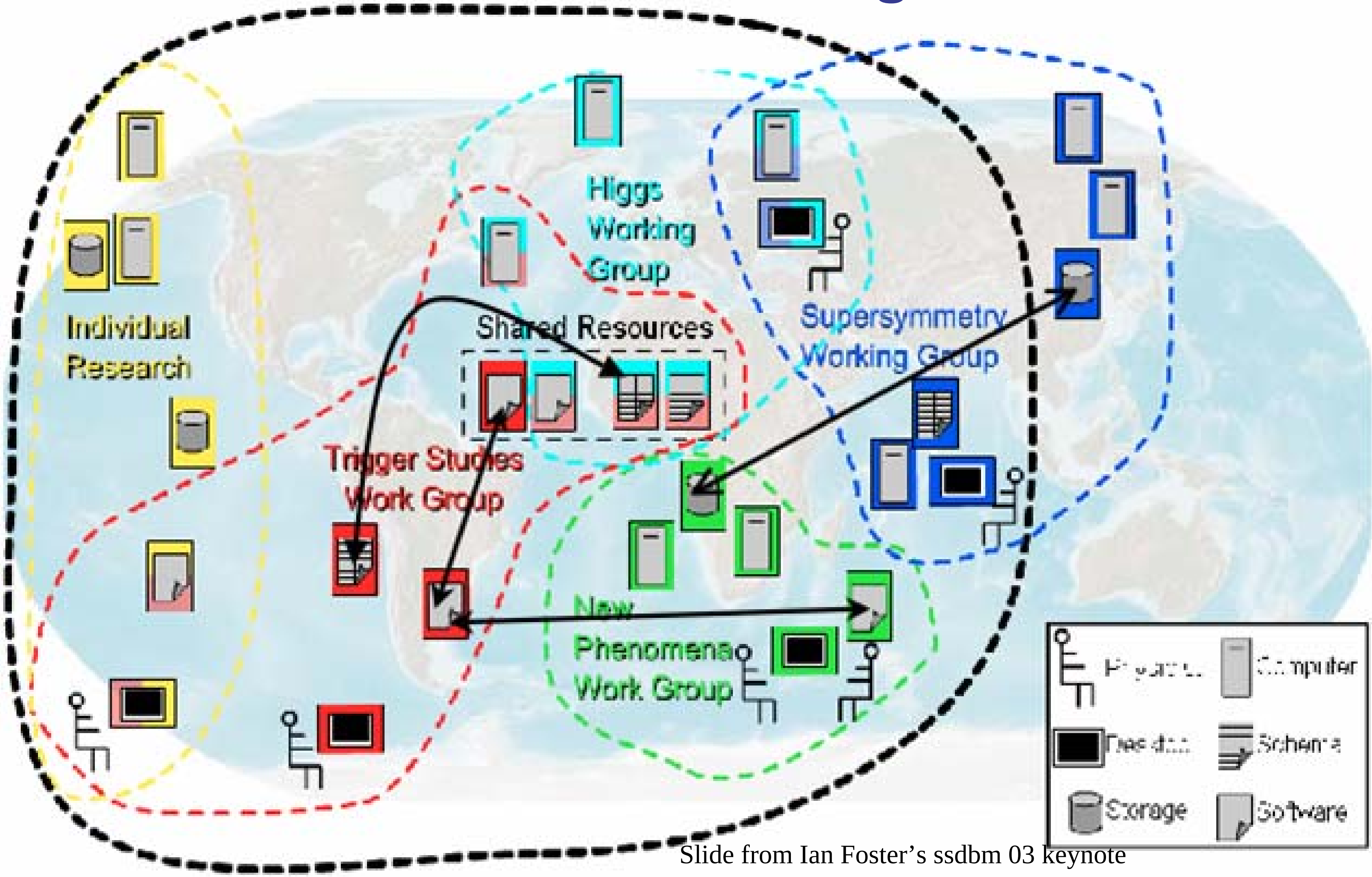
Data and images courtesy Alex Szalay, John Hopkins

Bridges Project



the globus project™
www.globus.org

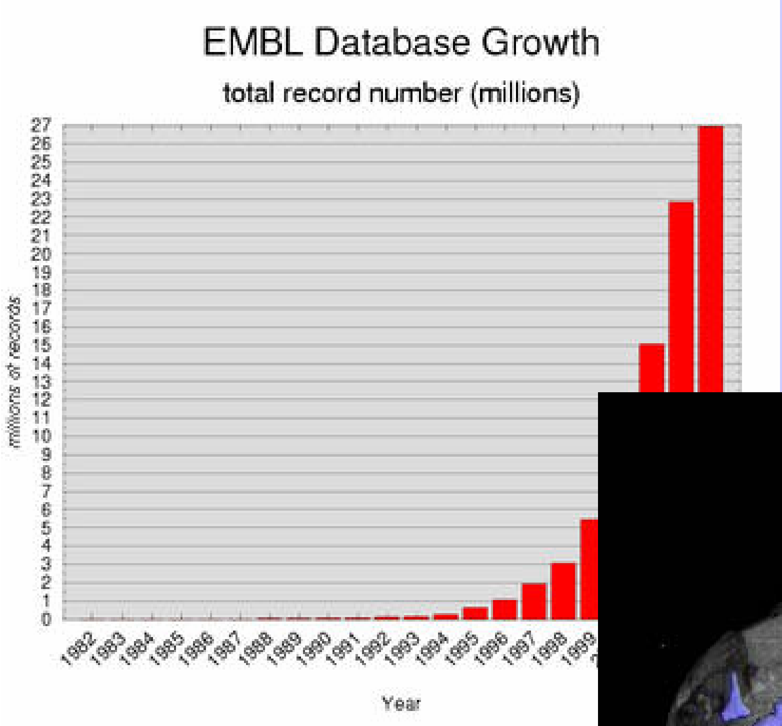
The Emergence of Global Knowledge Communities



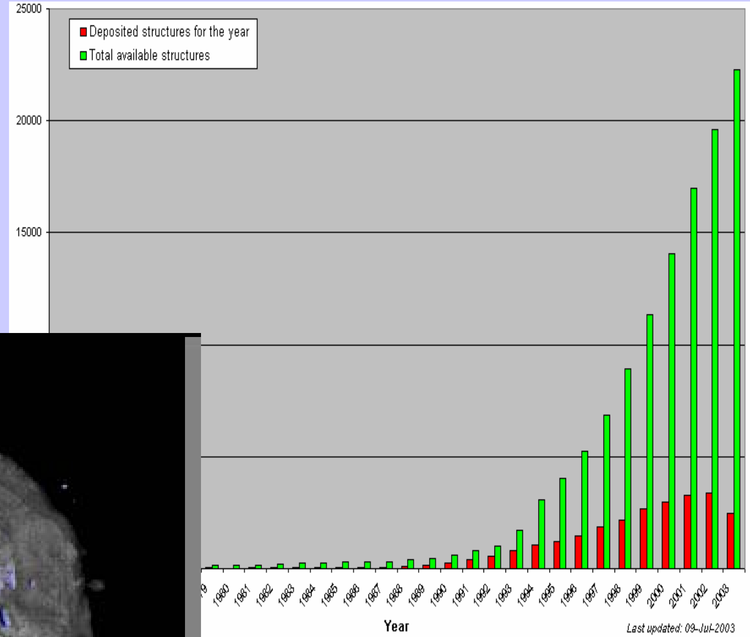
Slide from Ian Foster's ssdbm 03 keynote

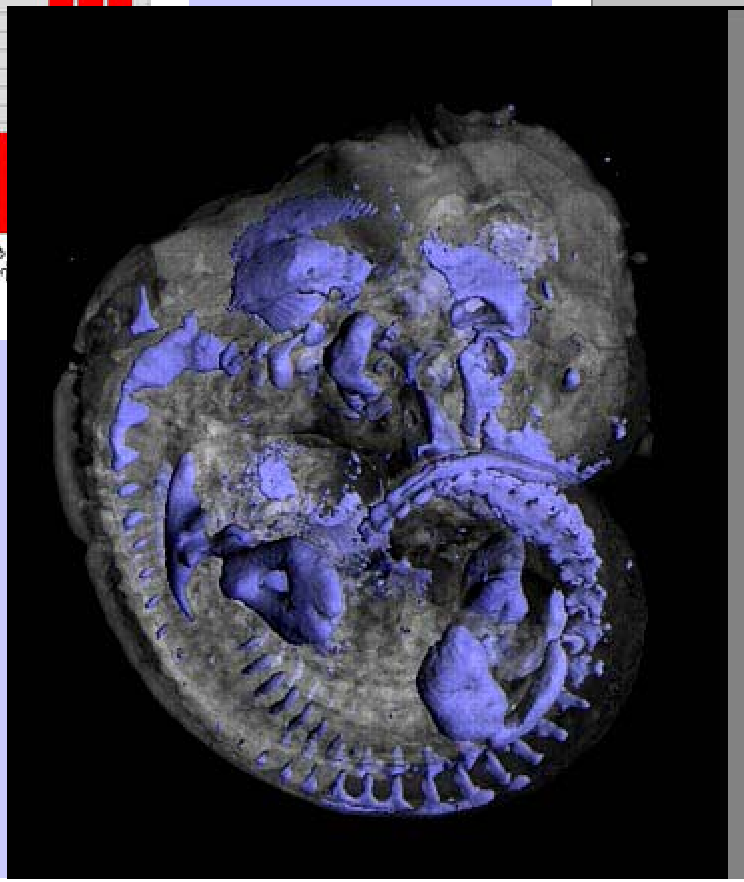
Database Growth

Bases 41,073,690,490



PDB Content Growth





MRC
Medical Research Council

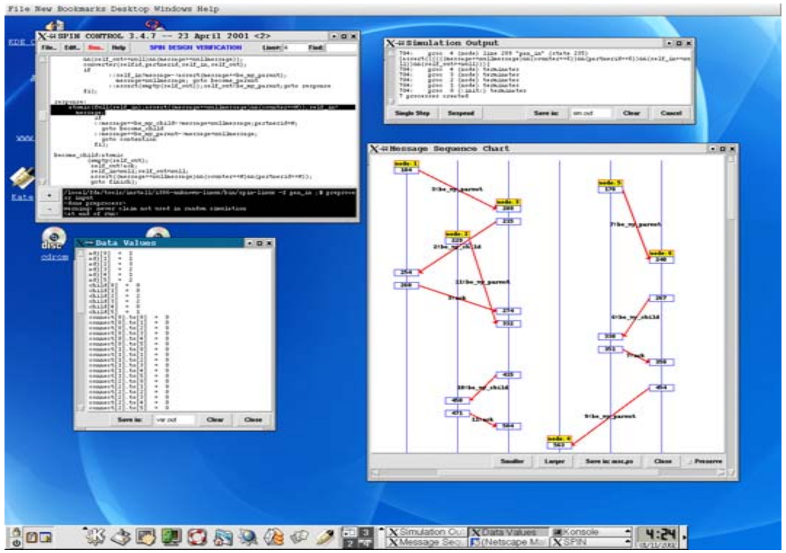
MRC HUMAN GENETICS UNIT
EDINBURGH

epcc

Biochemical Pathway Simulator

(Computing Science, Bioinformatics, Beatson Cancer Research Labs)

Walter Kolch

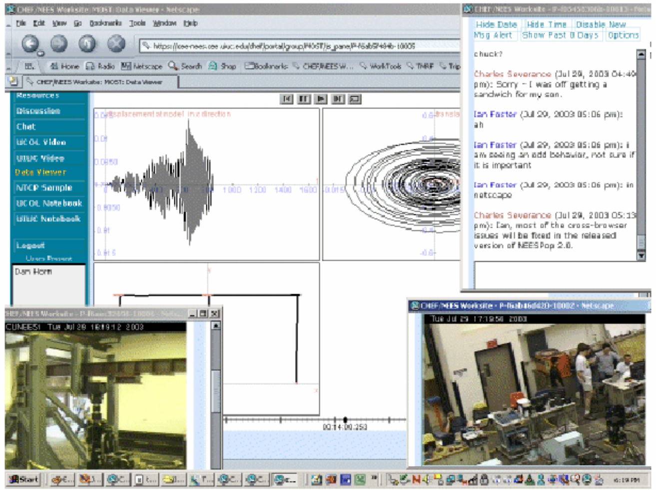


Closing the information loop – between lab and computational model.

DTI Bioscience *Beacon Project*

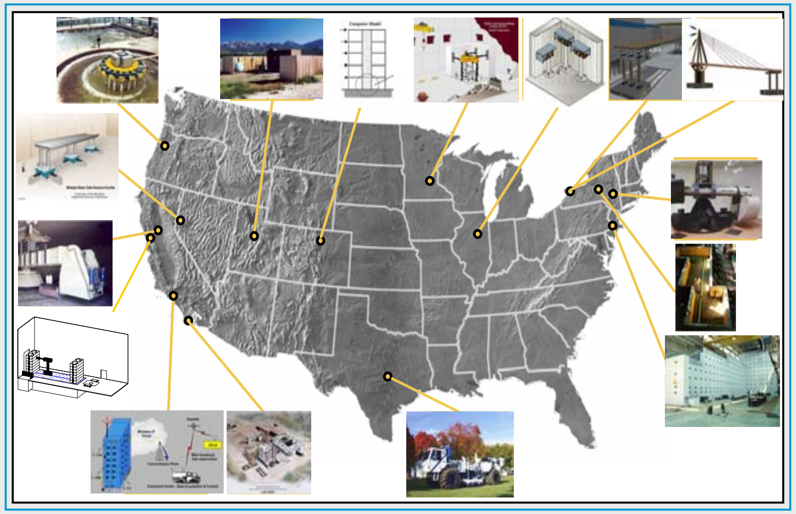
Now largest EU project in the Life Sciences – see http://www.cancerresearchuk.org/news/pressreleases/scottishscientists_22july04

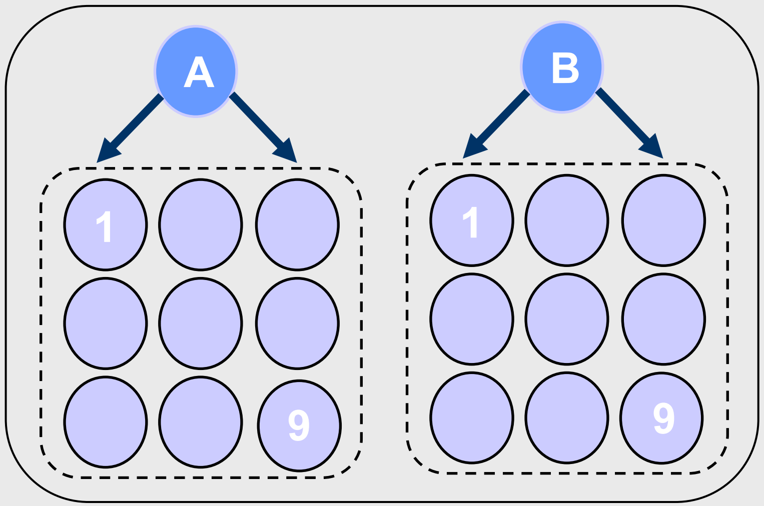
The Application-Infrastructure Gap



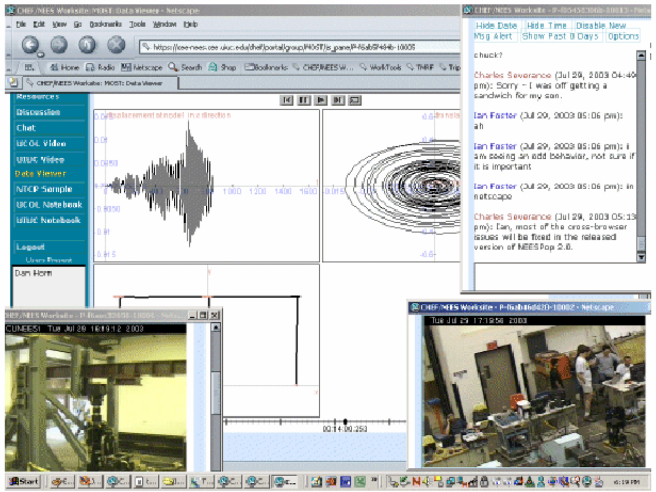
Dynamic and/or Distributed Applications

Shared Distributed Infrastructure



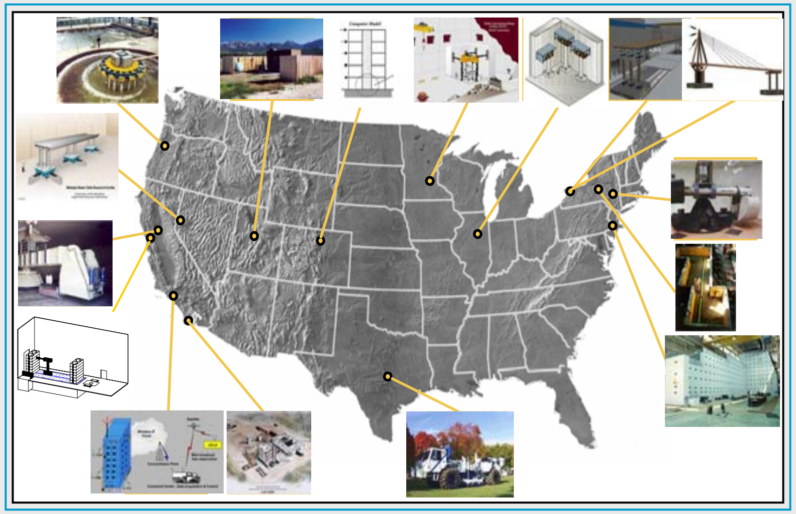


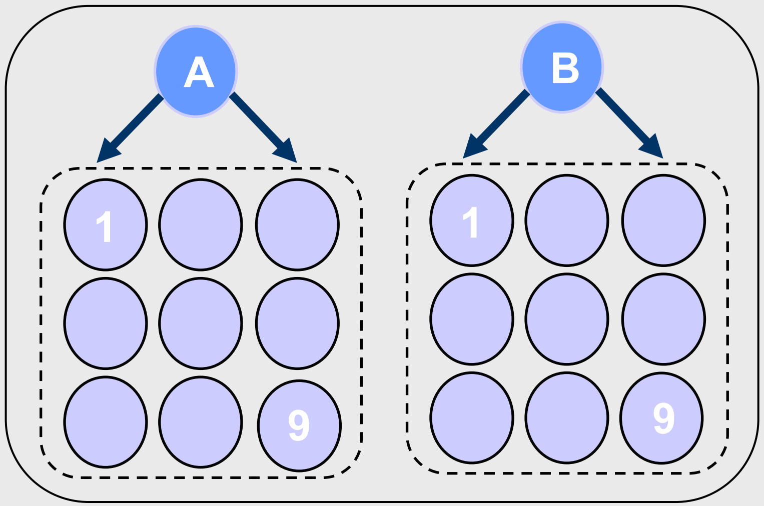
The Application-Infrastructure Gap



Dynamic and / or Distributed Applications

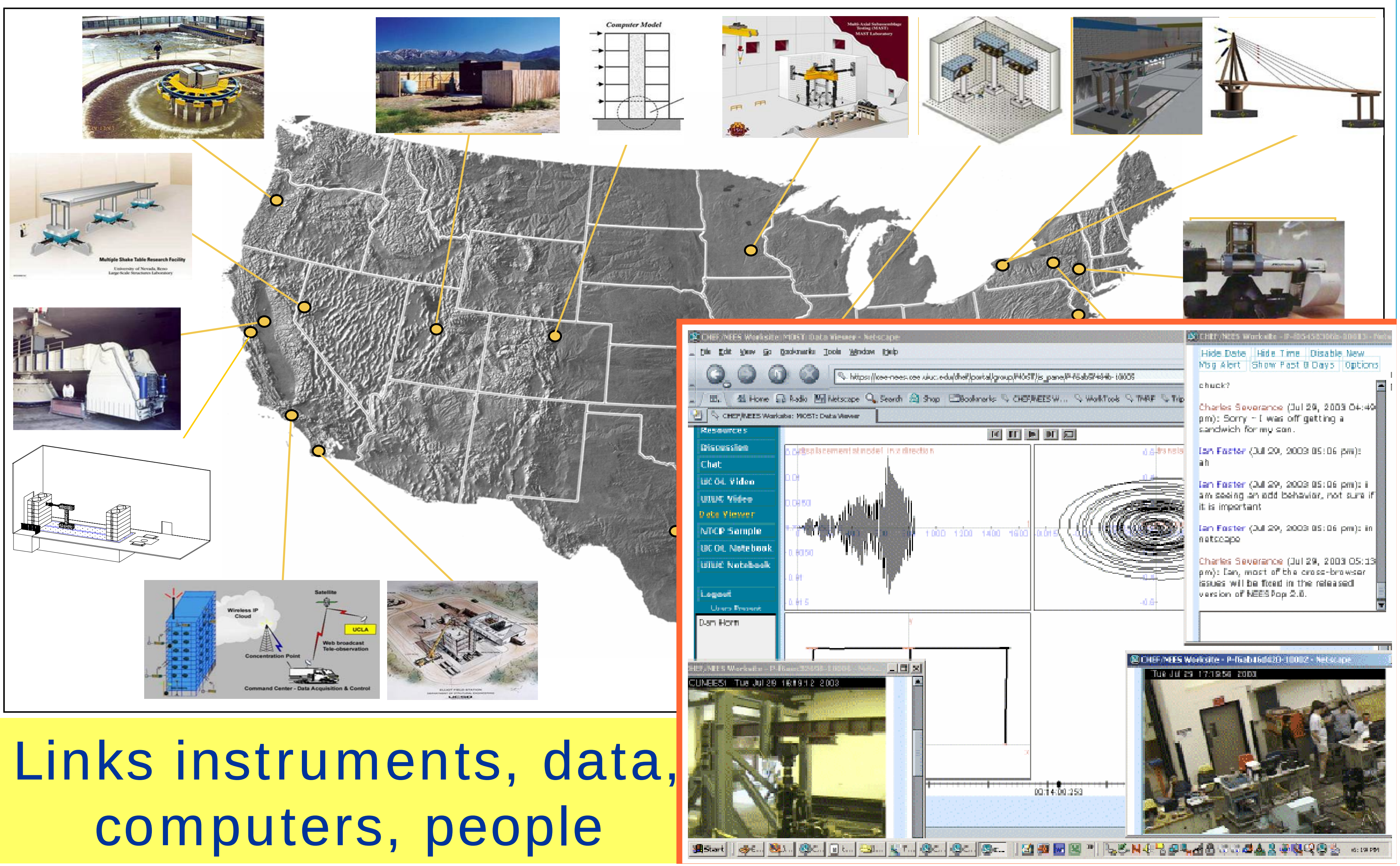
Shared Distributed Infrastructure





the globus toolkit® www.globustoolkit.org

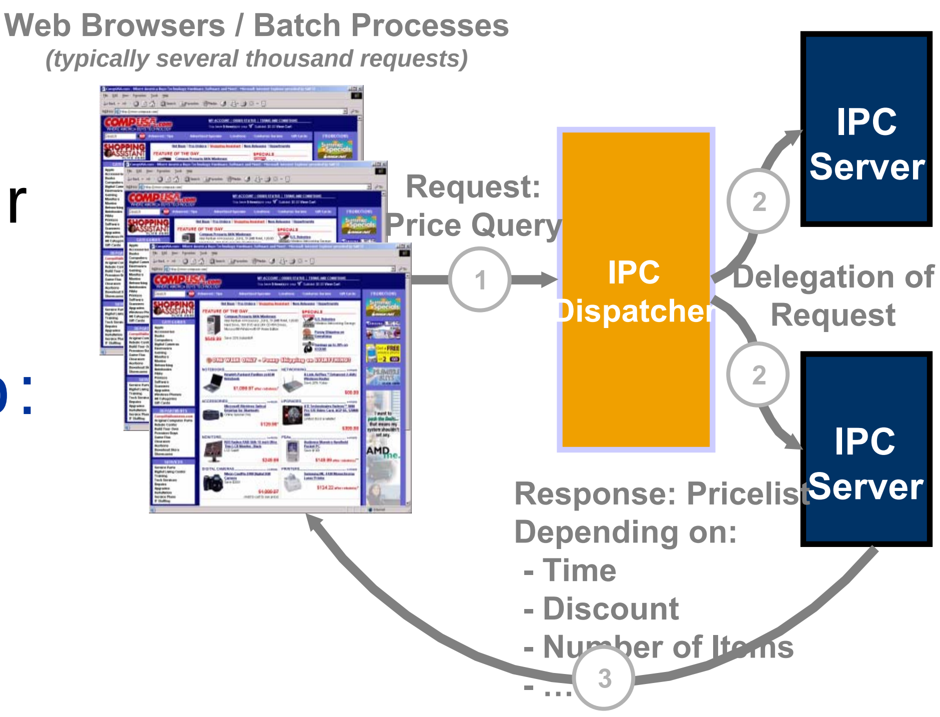
A Typical eScience Use of Globus: Network for Earthquake Eng. Simulation



Links instruments, data, computers, people

An eBusiness Use of Globus: SAP Demonstration @ GlobusWorld

- 3 Globus-enabled applns:
  - CRM: Internet Pricing Configurator (IPC)
  - CRM: Workforce Management (WFM)
  - SCM: Advanced Planner & Optimizer (APO)
- Applications modified to:
  - Adjust to varying demand & resources
  - Use Globus to discover & provision resources

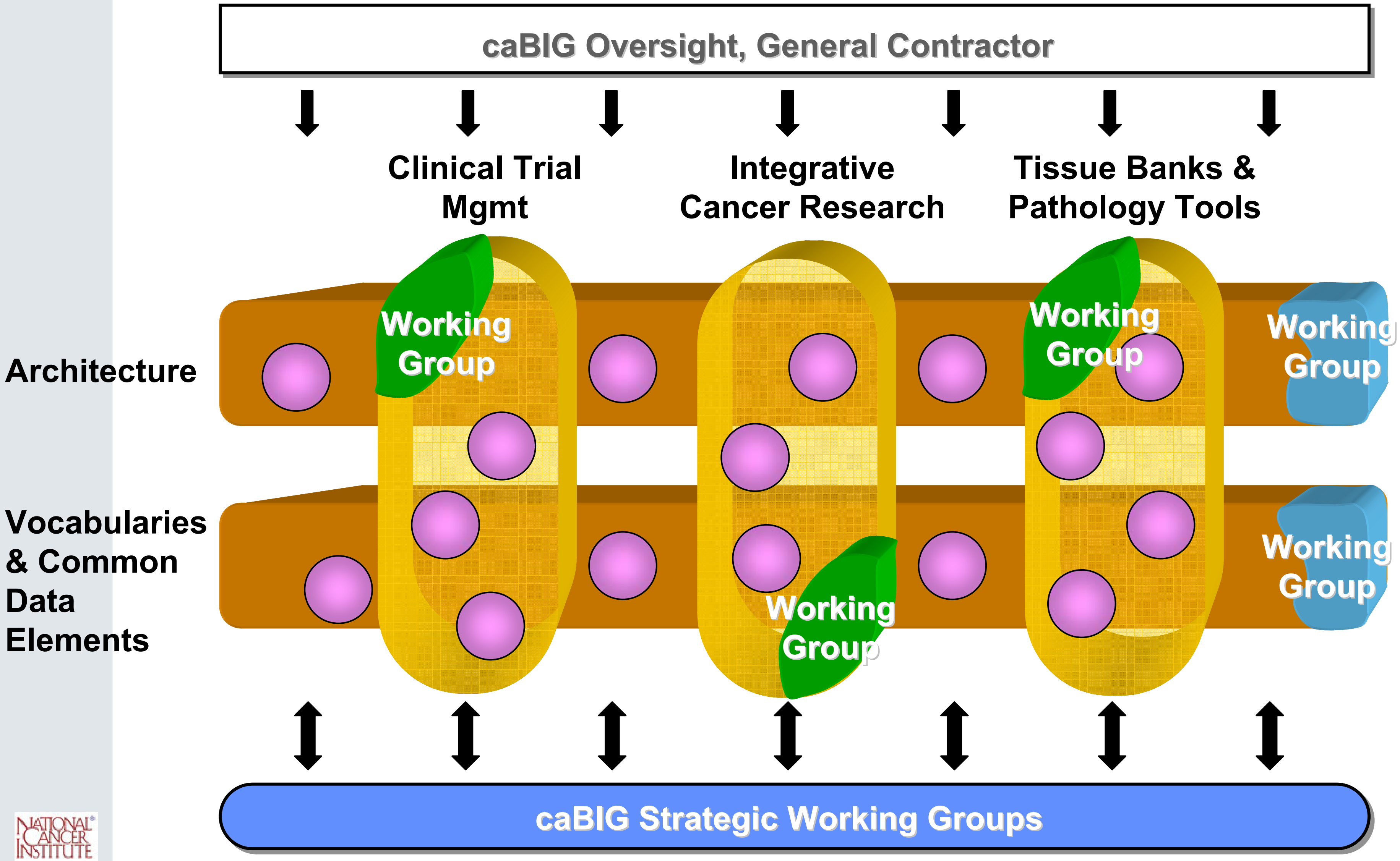


SAP AG R/3 Internet Pricing & Configurator (IPC)

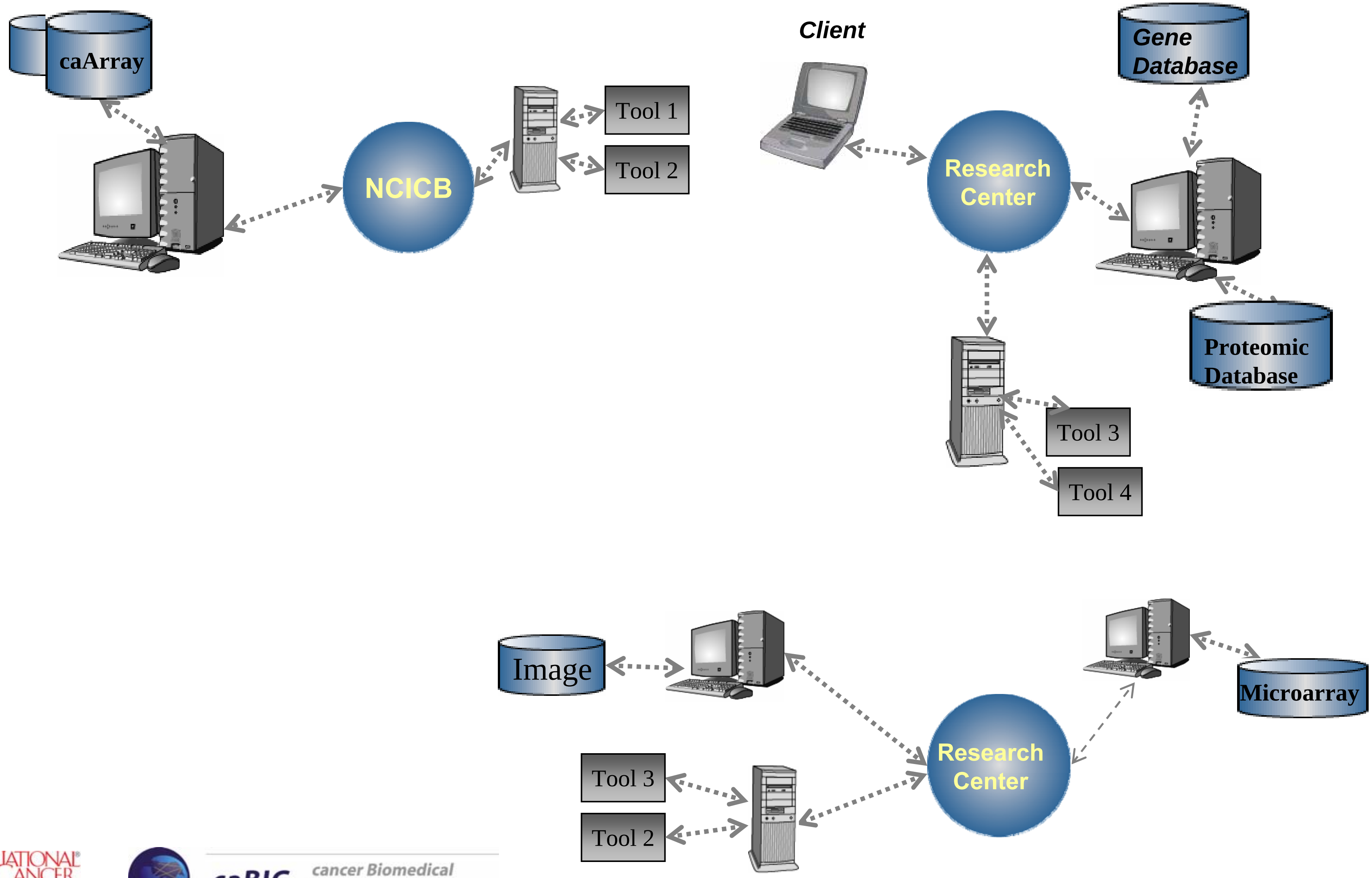
caBIG Motivation

- **In order to better understand disease mechanisms, it is necessary for a researcher to be able to synthesize information from large volumes of data and multiple data types.**
- **Example: UPenn Application Scenario**
  - A research would like to study the error rate in pathological diagnoses of solid tumor samples and compare numerous molecular diagnostic approaches to determine if the molecular diagnostic approach can enhance the accuracy of pathological diagnoses.
- **Query:**
  - I want all solid tumors, specifically for lung cancer, that have a diagnosis based on tumor pathology. Each diagnosis must have an image of the tumor that allows for independent verification of diagnoses. Each record retrieved must also have either proteomics marker data or microarray data (Affy or two-color) included so that different molecular techniques can be correlated to the tumor pathology. In addition, I want all protein annotations for markers and genes associated with the proteomics and microarray data so I can perform meta-analyses.

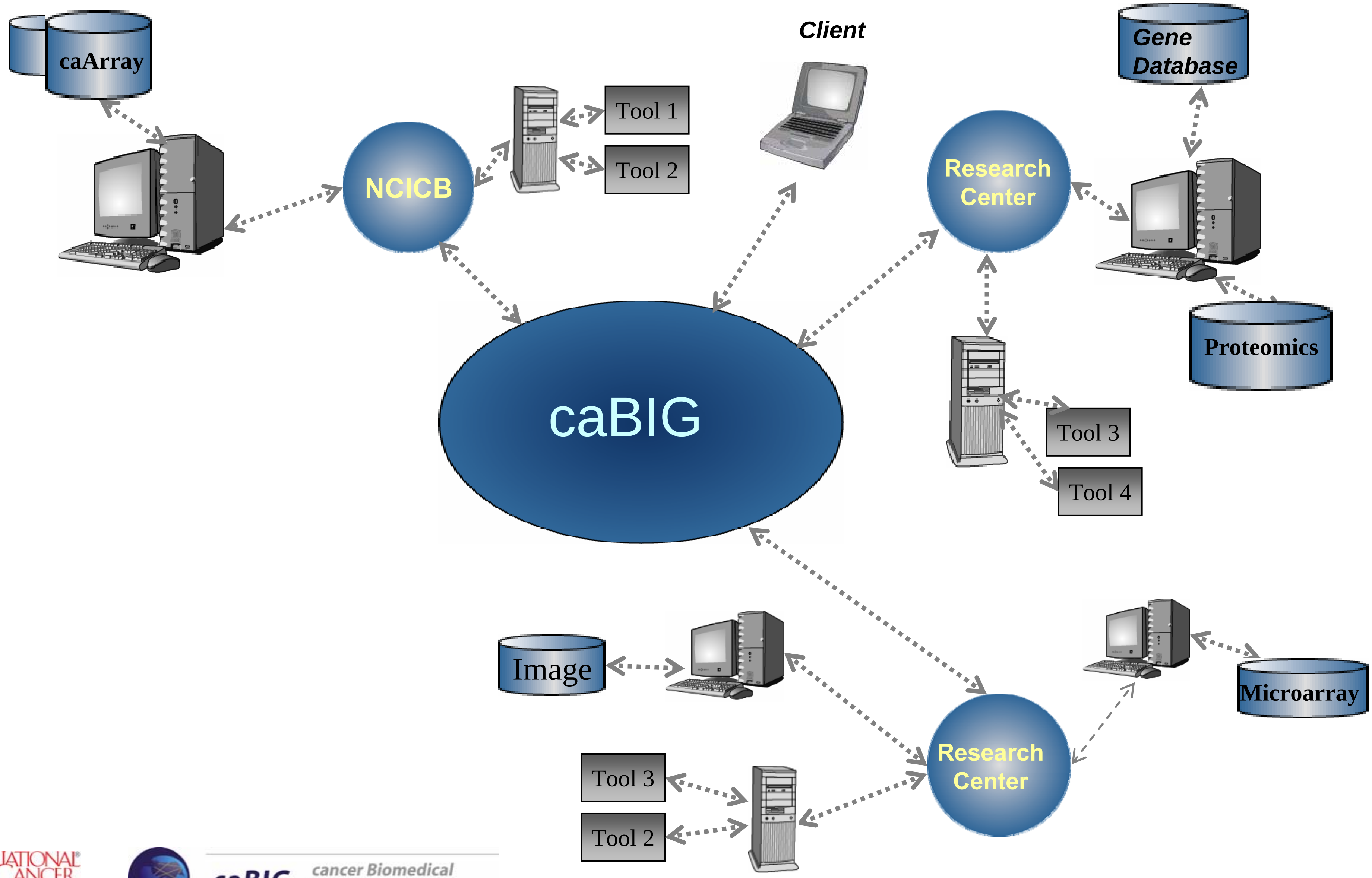
caBIG Organization



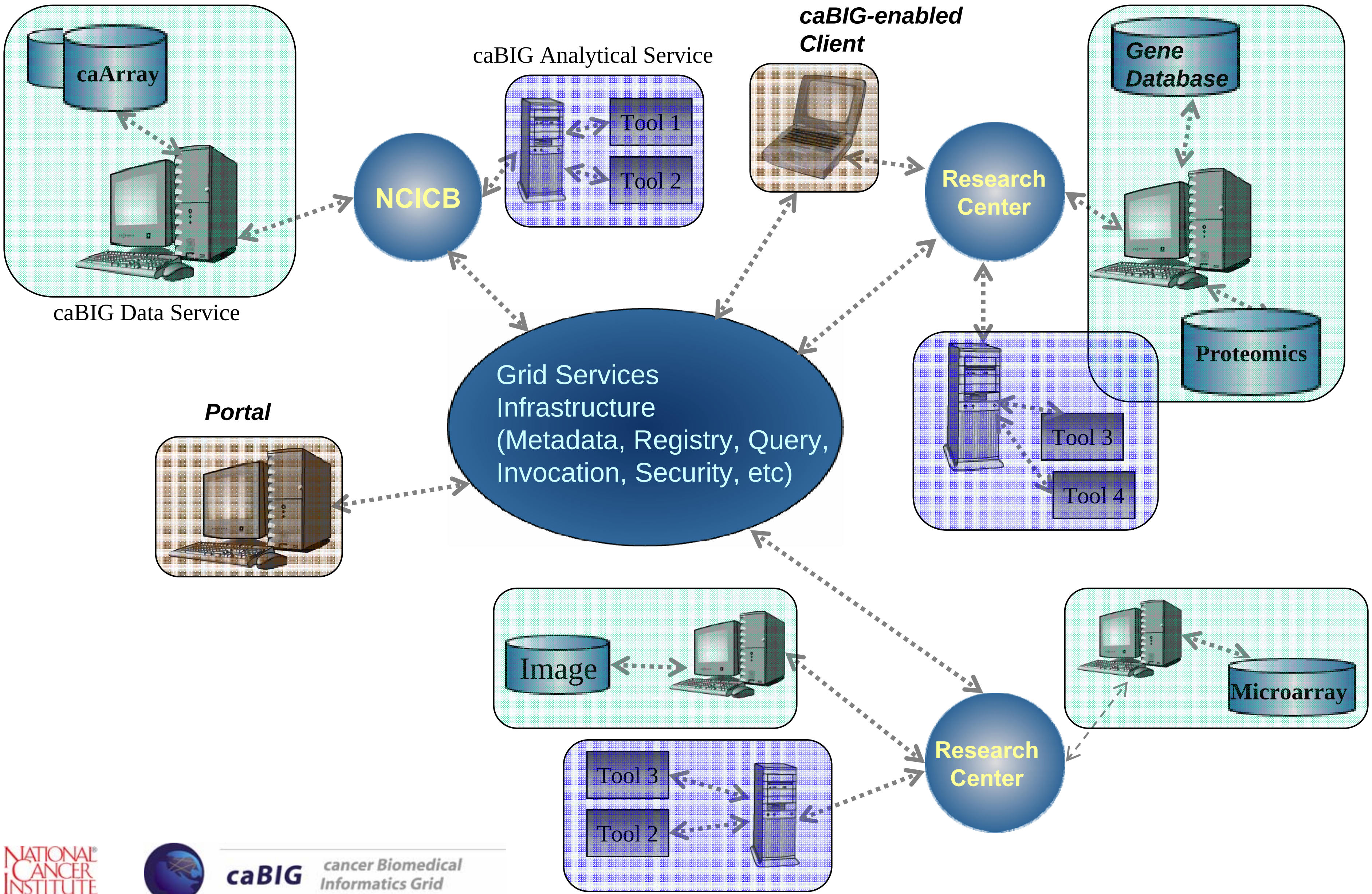
Application Scenario



Application Scenario using caBIG



What is caBIG Architecture (caGRID Infrastructure)?

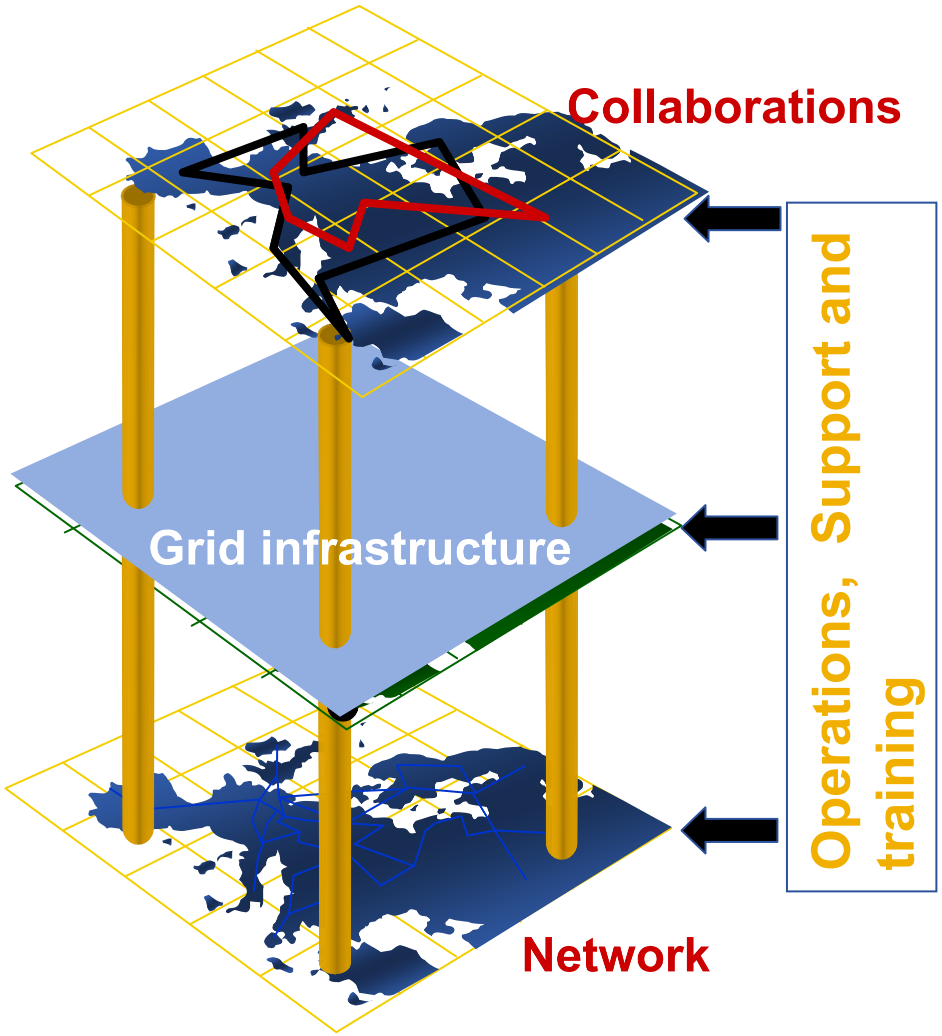


NATIONAL CANCER INSTITUTE caBIG cancer Biomedical Informatics Grid

Why EGEE?

Build a large-scale production grid service to:

- Support science and technology worldwide
- Link with and build on national, regional and international initiatives
- Foster international cooperation both in the creation and the use of the e-infrastructure

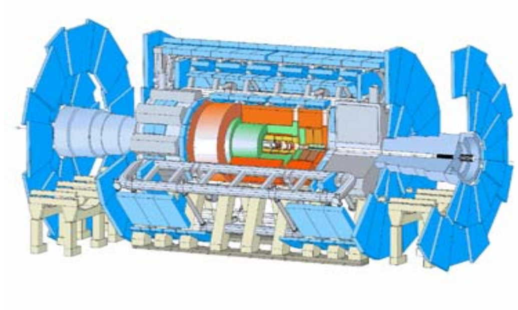


LHC experiments

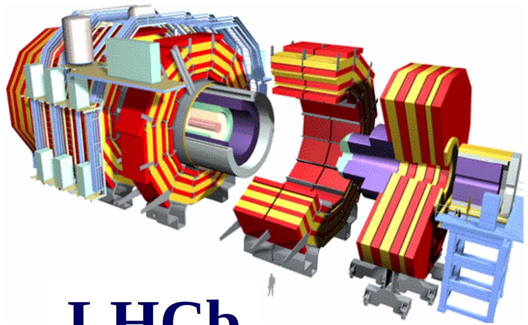
- **Storage**
  – Raw recording rate 0.1 – 1 GByte/s
  – Accumulating at 5-8 PetaByte/year
  – 10 PetaByte of disk

- **Processing**
  – 200,000 of today's fastest PCs

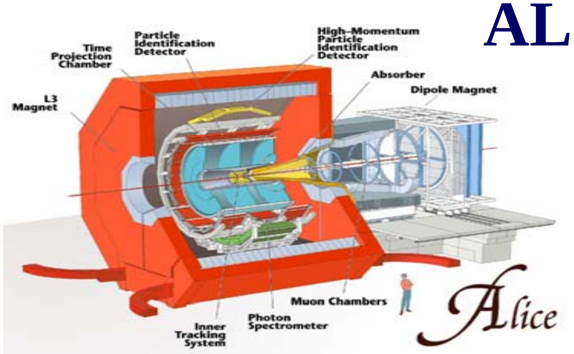
ATLAS



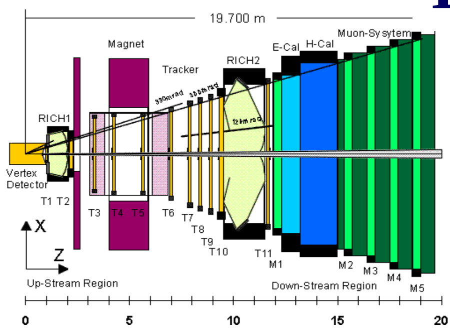
CMS



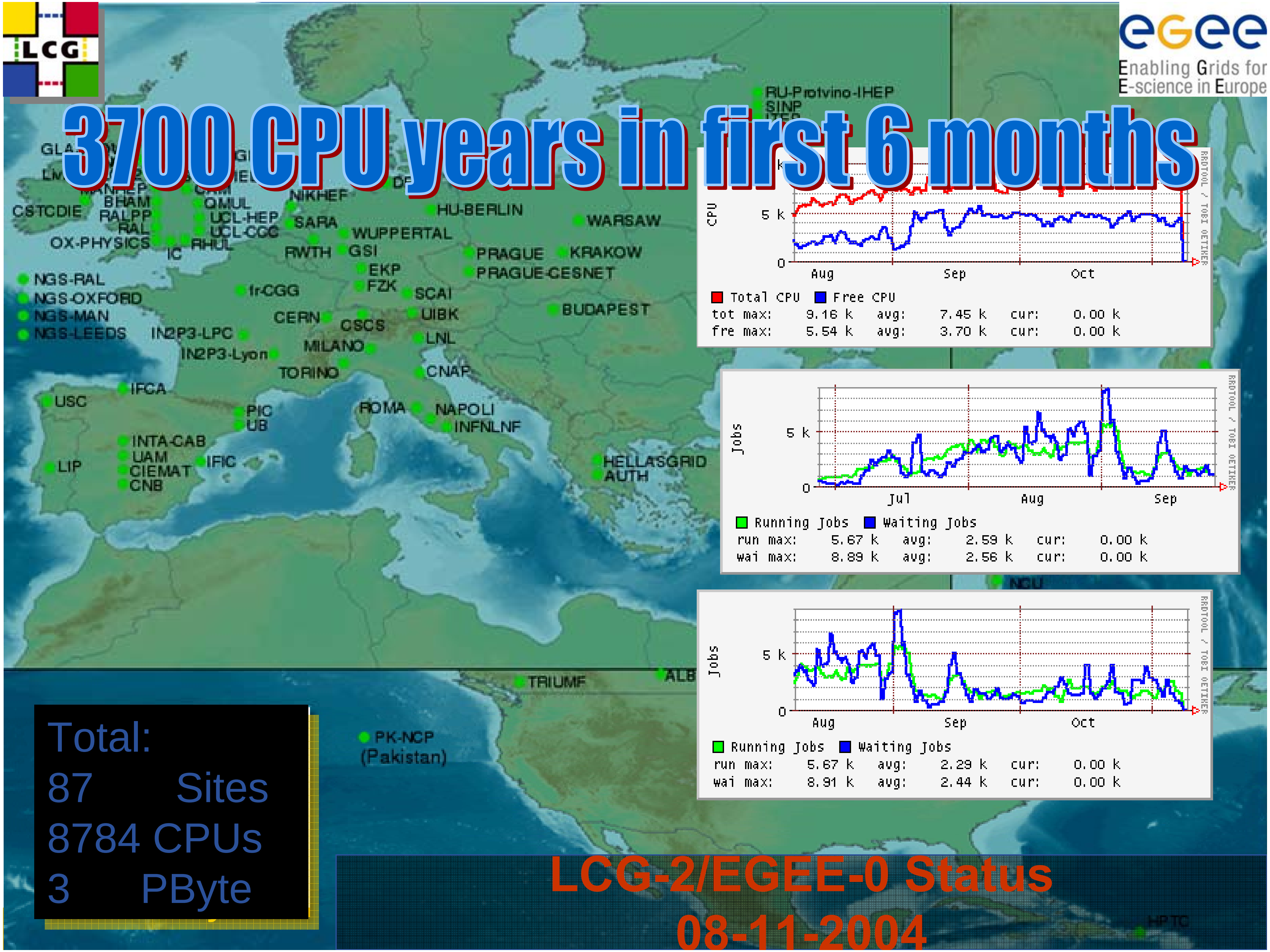
ALICE



LHCb



3700 CPU years in first 6 months

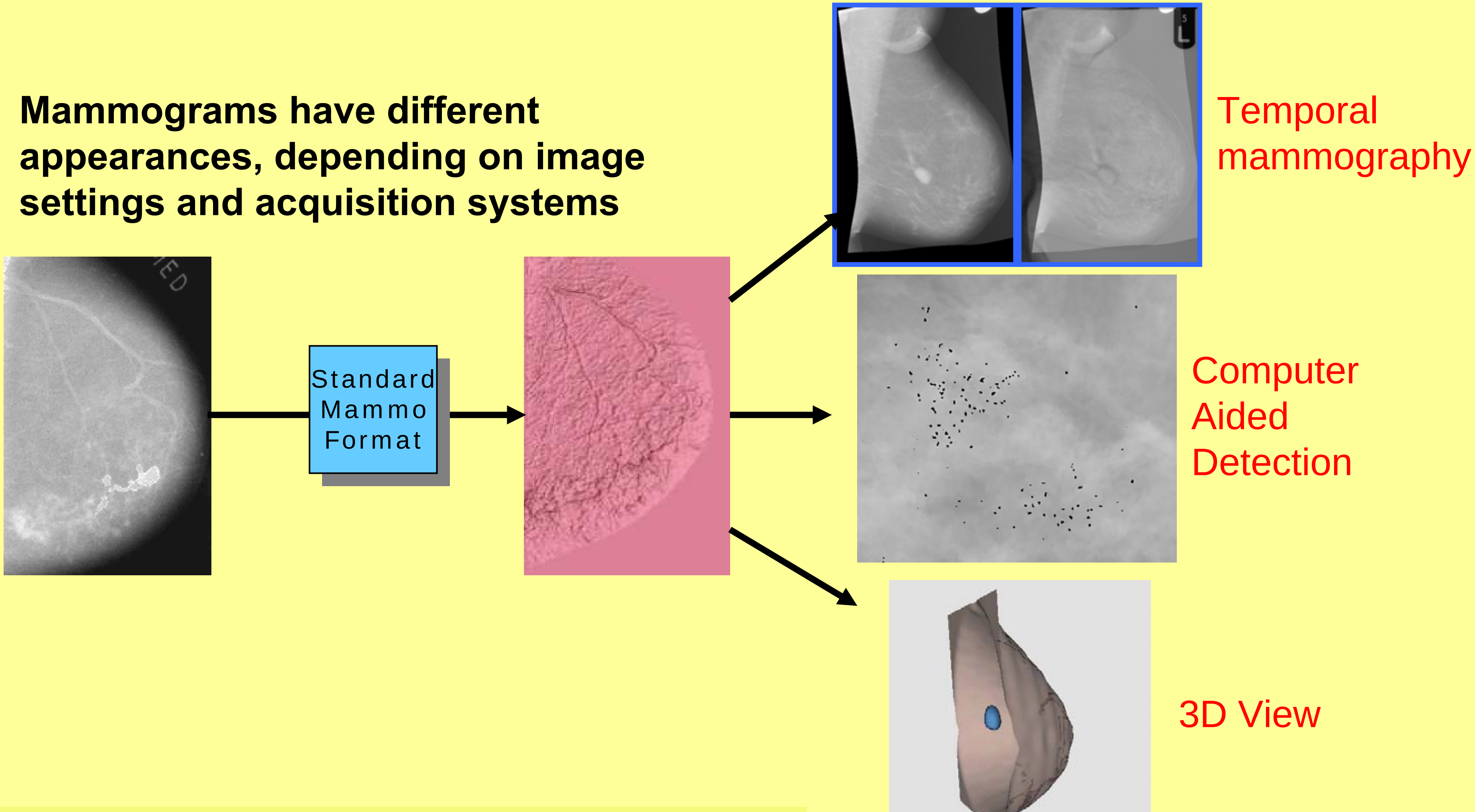


Total:
87 Sites
8784 CPUs
3 PByte

LCG-2/EGEE-0 Status
08-11-2004

eDiaMoND - Compute

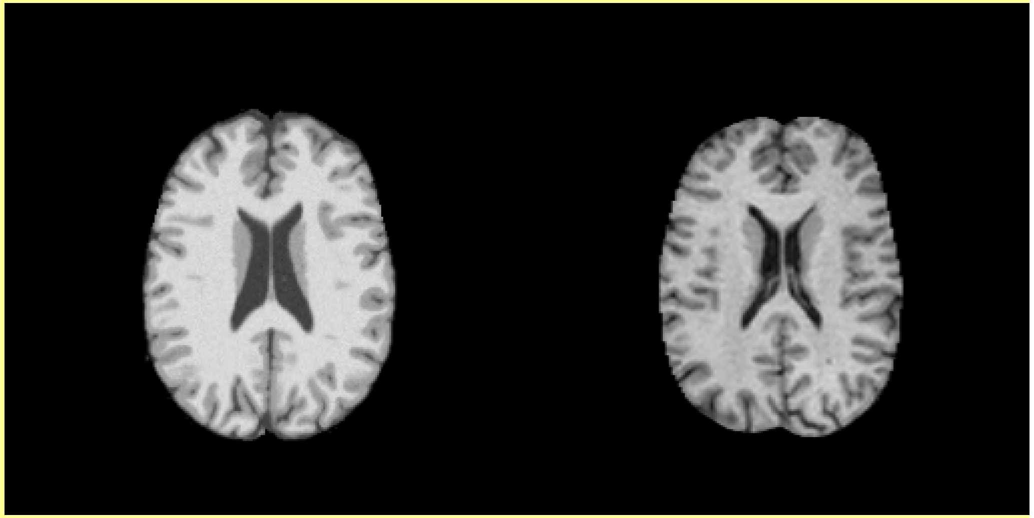
Mammograms have different appearances, depending on image settings and acquisition systems



Provided by eDiamond project: Prof. sir Mike Brady *et al.*

Automatic registration technology

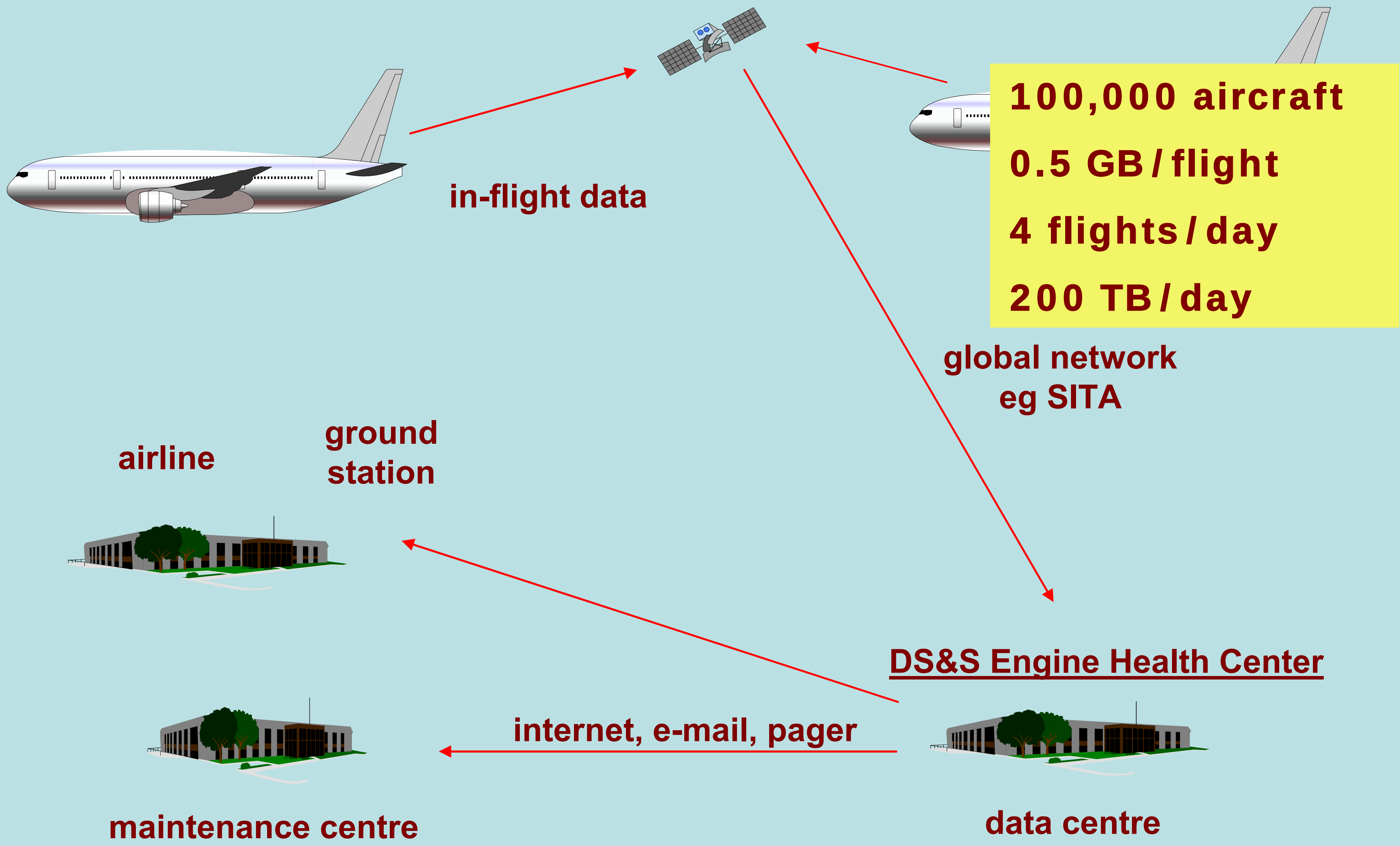
Rigid registration of MR and CT images of the head



Inter-subject image warping

Provided by IXI project: Prof. Derek Hill *et al.*

global in-flight engine diagnostics



Distributed Aircraft Maintenance Environment: Leeds, Oxford, Sheffield &York, Jim Austin

The Growth of Data

- Much faster than simple accumulation
- Typical Doubling Times
  - 9 months to 4 years
- Compare with technology doubling times
  - ~18 Months
- Driven by a composition of forces
  - Moore's Law Squared
    - Devices double in speed
    - Devices halve in cost, size & power consumption
    - Less components – more rugged or disposable – speckled computing
    - Every 18 months
  - Collaborative Investment
    - Human Genome project → functional genomics → proteomics
    - Laboratory automation & Human concern → multiple species, pathology & development
  - Recognition of the value of sharing
  - Annotation as a means of communicating and publishing
  - Derivation as a means of communicating and publishing

Tera → Peta Bytes

Enabling Grids for E-science in Europe

- RAM time to move
  - 15 minutes
- 1GB WAN move time
  - 10 hours ($1000)
- Disk Cost
  - 7 di[illegible] = $50[illegible]
- Disk P[illegible]
  - 10[illegible]
- Disk [illegible]
  - 5.6 Kg
- Disk Footprint
  - Inside machine

- RAM time to move
  - 2 months
- 1TB WAN move time
  - 14 month[illegible]
- [illegible]C[illegible]st
  - [illegible] units +
  - [illegible]cks = $7 million
- Disk Power
  - 100 Kilowatts
- Disk Weight
  - 33 Tonnes
- Disk Footprint
  - 60 m^2

Now make it secure & reliable!

May 2003 Approximately Correct *Distributed Computing Economics*

Jim Gray, Microsoft Research, MSR-TR-2003-24

Challenges for Data Users

Enabling Grids for E-science in Europe

- Finding (all) *relevant* data
- *Understanding* the data
- *Inventing* ways of forming interconnections
- Coping with *distributed* and *autonomous* data sources
- Coping with format, representation and policy *variations*
- Obtaining *relevant* parts of the data
- *Inventing* and developing algorithms to form derivatives
- *Organising* the processing to produce derivatives
- *Storing*, *managing* and *describing* the derivatives
  - Other people's primary data
  - Your primary data next year
- *Assessing* the reliability of results
- *Interpreting* the results

Challenges for Data Users

Enabling Grids for E-science in Europe

- Finding (all) *relevant* data
- *Understanding* the data
- *Inventing* ways of forming interconnections
- Coping with *distributed* and *autonomous* data sources
- Coping with format, representation and policy *variations*
- Obtaining *relevant* parts of the data
- *Inventing* and developing algorithms to form derivatives
- *Organising* the processing to produce derivatives
- *Storing*, *managing* and *describing* the derivatives
  - Oth r people's rimary data
  - Yo primary dat ext year
- *Asses g* the reliab f results
- *Inter*

Depends on

Depends on metadata and

Depends on good

Requires good query systems + data extractors /

Requires creative

Requires persistence, resources & organising skills

Requires

Requires knowledge, insight & courage

Requires relia storage syste

Requires data – see data challenges

Requires standards & metadata

The Challenge for Data Creators

- Coping with data rates & volume
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
- Established & adopted standards
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

- Coping with data rates & volume
  - Parallel channels
  - Buffers
  - Fast feature extraction $\Rightarrow$ discarding other information
  - Compression
  - Coordinated distributed systems – parcelling out the work
  - Good engineering
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
- Established & adopted standards
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

- Coping with data rates & volume
- Reliable & affordable persistent storage
  - Redundant storage – replication & distribution
  - Standard technology – Hardware & Software
    - E.g. SRB and SRM
  - Minimal but sufficient redundancy
  - Separation of data ingest from data publishing / access
  - Compact & efficient representations
  - Flexible structure for future requirements & expansion
  - Smart Engineering
- Reliable & sufficient metadata
- Established & adopted standards
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

- Coping with data rates & volume
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
  - Agree community requirements & standards
  - *As understanding advances these change*
  - Automate as much as possible
    - Instruments and derivation processes generate it
    - Repetitive procedures require only changed data to be input
    - Generate by mining from “primary” data
  - But humans must produce some of it
    - Error prone & poor motivation initially
    - Good tools & community policies / pressure
  - Good engineering
- Established & adopted standards
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

- Coping with data rates & volume
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
- Established & adopted standards
  - For Data formats, Data Access, Data Storage, Data Management
  - For metadata, privacy, security, …
  - Industry & academia – collaborate when common goals
  - Huge investments required to agree and adopt standards
  - Huge costs if you don't
  - Technology trade offs change & engineering advances
  - Requirements change
  - Too many standards & not enough standards
  - Compliance & interoperability illusive
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

- Coping with data rates & volume
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
- Established & adopted standards
- Reliable & sufficient data protection
  - To satisfy data owners – do they trust your systems?
  - To overcome excuses of data huggers!
  - To meet statutory requirements
  - What do they mean in international collaborations?
  - Authentication, Authorisation & Accounting
    - Complex policies
      - e.g. all medical images composites of at least 5 individuals
    - No one (external organisation?) may access more than 10%
    - Kinship, residence & occupation needed but must pseudonymise
  - Encryption – storage & transmission
    - Applies to derivatives & metadata too!
- Agreed publication, recognition & discard policies

The Challenge for Data Creators

Enabling Grids for E-science in Europe

- Coping with data rates & volume
- Reliable & affordable persistent storage
- Reliable & sufficient metadata
- Established & adopted standards
- Reliable & sufficient data protection
- Agreed publication, recognition & discard policies
  - Some disciplines have agreements on the first 2
  - Some funders require the first but don't understand its costs
  - Some researchers accept the first but don't understand …
  - Publication – poorly specified notion at present – with exceptions
  - Recognition
    - Huge creative & valuable efforts unrecognised
    - No lasting record of credit or responsibility in most cases!
  - Discard – chosen and fixed at data collection time ☹
    - Or haphazard later ☹
      - next project , next student ⇒ don't maintain old data
    - Observed funding behaviour
      - Invest in new projects don't maintain access to accumulated data wealth ☹
    - Buy more storage and put off the evil day ☹

Requires urgent attention – talk to archivists? Can we do better than haphazard?

Challenge for the Engineers

Enabling Grids for E-science in Europe

- Building Storage systems
  - File systems – 10^9 files
  - Accession & cataloguing rates – 10^6 per day
    - Commercial systems or DIY?
  - Database systems
    - Commercial systems or DIY?
  - Combining & interfacing these to meet research community requirements
    - See User Challenges & Creator Challenges combined
  - Operating & sustaining these
    - For decades
    - As requirements change
    - As technology is superseded
  - Recovering cost of storage & operations

Once you start you accept a responsibility for your community's data – SLA? SLD at least?

Challenge for the Engineers

- Providing fast data access
  - With audit, security, diagnostics, accounting, …
  - Move computation not data
    - Programs grow slowly compared with data
    - But safety & trust big issues
    - Provide powerful query languages
      - Plus judicious use of stored procedures
      - Beware the QFH
    - Provide powerful encapsulated data derivation tools
      - Datacutter, FFTs, R, …
    - Web service interfaces close to data
      - But operational service costs
  - Alternative – Replicate data (near) where computation will happen
    - Based on planning & high-level work(flow) description – VDT
    - Based on assumption of repeated patterns – caching
    - Based on user-specified change propagation requirements
      - BioDBs with high update rates & InfoD
  - Move data fast & reliably

Be Direct

- Double Handling costs too much
  - Memory cycles, bus capacity, cache (
- Double Handling via discs pathol
- Data translation expensive
  - Avoid or compose
- Main memory is not big enough
  - Nor is it linear and uniform
  - Streaming algorithms essential
- Couple generator & consumer di
  - Data pipe from RAM to RAM
  - Requires coupled computation execu
  - Requires new standards and technol

Breaks down boundaries and merges data, execution & transport requirements.

Demands smart workflow enactment service & foundation services

Architectural Requirement: Dynamically
Move computation to the data

Enabling Grids for E-science in Europe

- Assumption: code size << data size
- Develop the database philosophy for this?
  - Queries are dynamically re-organised & bound
- Develop the storage architecture for this?
  - Compute closer to disk?
    - System on a Chip using free space in the on-disk controller
- Data Cutter a step in this direction
- Develop experiment, sensor & simulation architectures
  - That take code to select and digest data as an output control
- Safe hosting of arbitrary computation
  - *Proof-carrying code* for data and compute intensive tasks + *robust hosting* environments
- Provision combined storage & compute resources
- Decomposition of applications
  - To ship behaviour-bounded sub-computations to data
- Co-scheduling & co-optimisation
  - Data & Code (movement), Code execution
  - Recovery and compensation

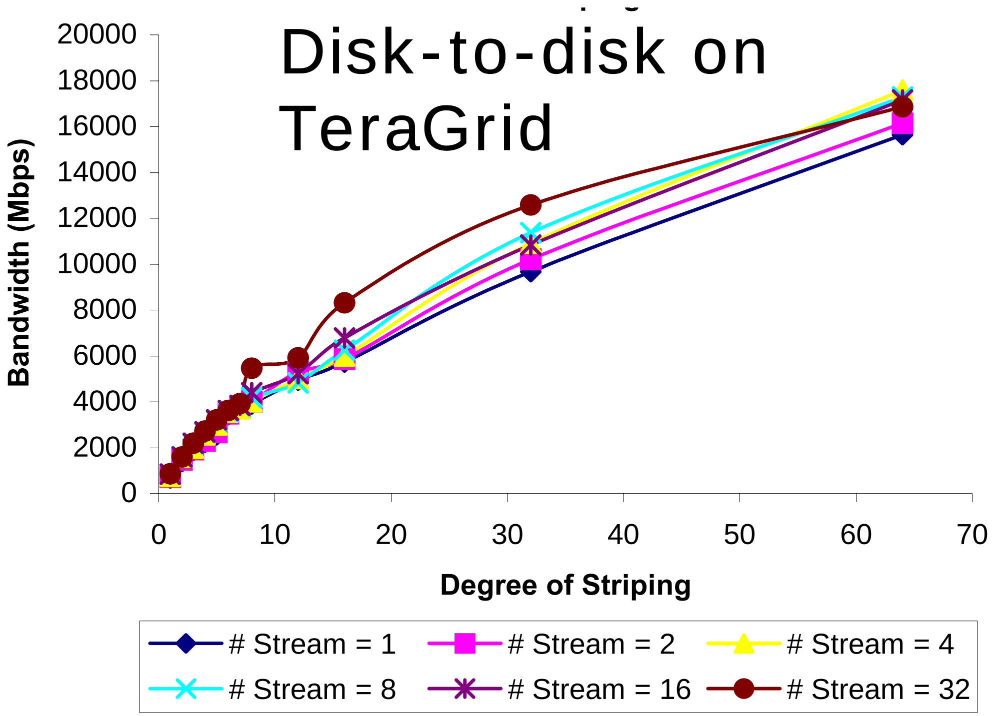
Dave Patterson
Seattle
SIGMOD 98

Little is done yet – requires much R&D and a Grid infrastructure

GT4 Data Management

- **Stage / move** large data to/from nodes
  - GridFTP, Reliable File Transfer (RFT)
  - Alone, and integrated with GRAM
- **Locate** data of interest
  - Replica Location Service (RLS)
- **Replicate** data for performance/reliability
  - Distributed Replication Service (DRS)
- Provide **access** to diverse data sources
  - File systems, parallel file systems, hierarchical storage: GridFTP
  - Databases: OGSA-DAI

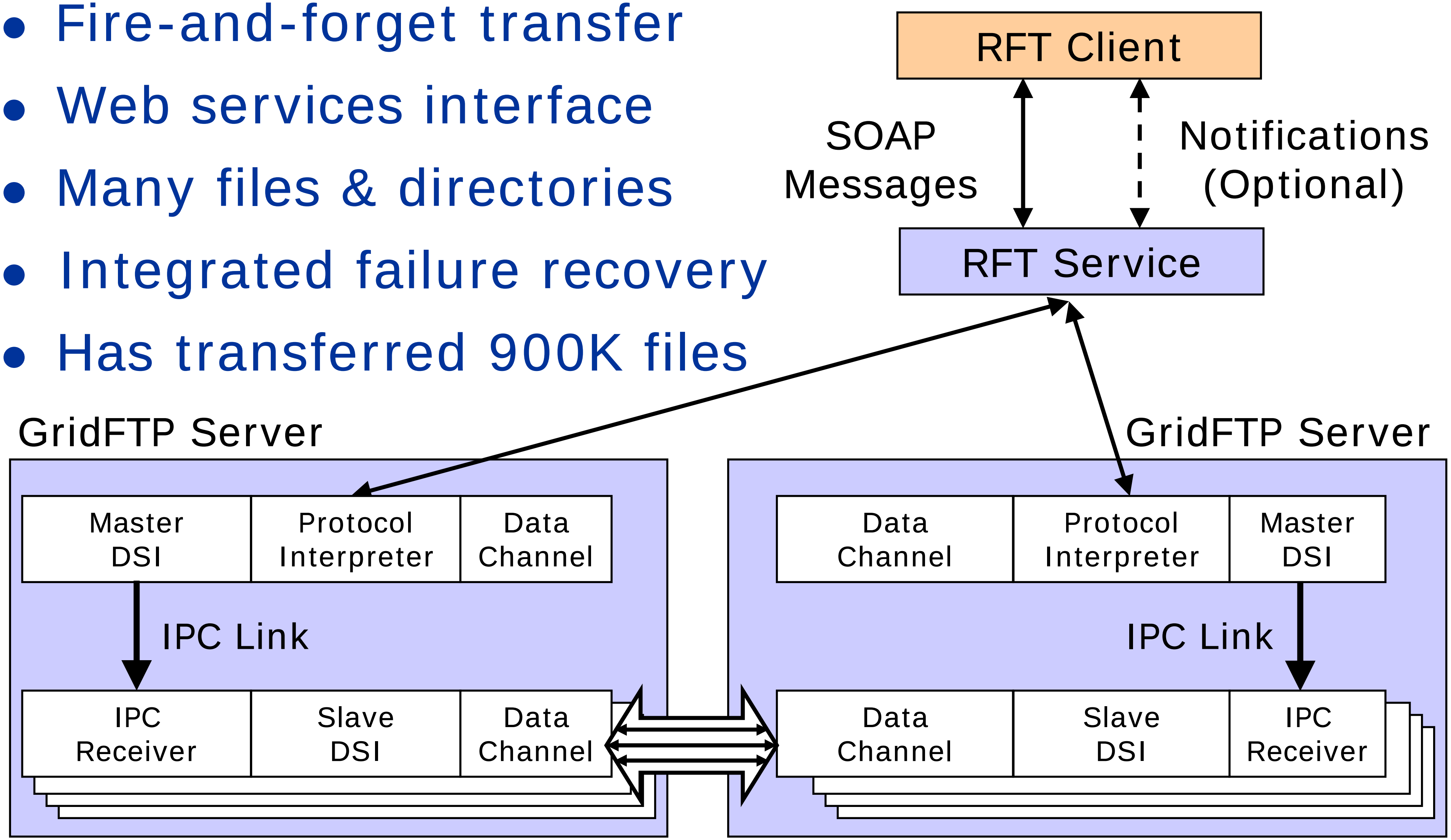
GridFTP in GT4



- 100% Globus code
  - No licensing issues
  - Stable, extensible
- IPv6 Support
- XIO for different transports
- Striping → multi-Gb/sec wide area transport
  - 27 Gbit/s on 30 Gbit/s link
- Pluggable
  - Front-end: e.g., future WS control channel
  - Back-end: e.g., HPSS, cluster file systems
  - Transfer: e.g., UDP, NetBLT transport

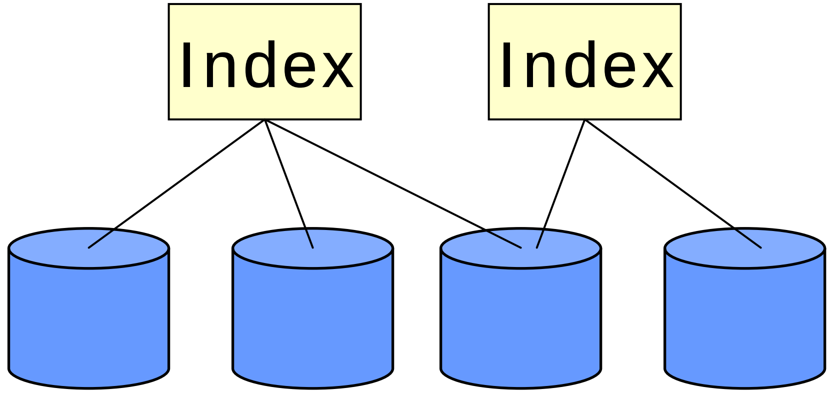
Reliable File Transfer: Third Party Transfer

- Fire-and-forget transfer
- Web services interface
- Many files & directories
- Integrated failure recovery
- Has transferred 900K files



Replica Location Service

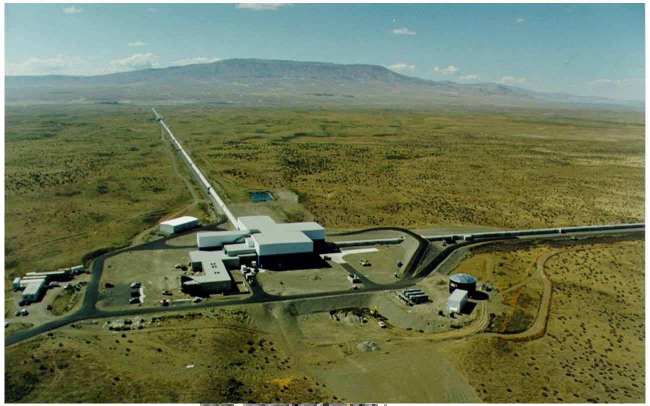
- Identify location of files via logical to physical name map
- Distributed indexing of names, fault tolerant update protocols
- GT4 version scalable & stable
- Managing ~40 million files across ~10 sites

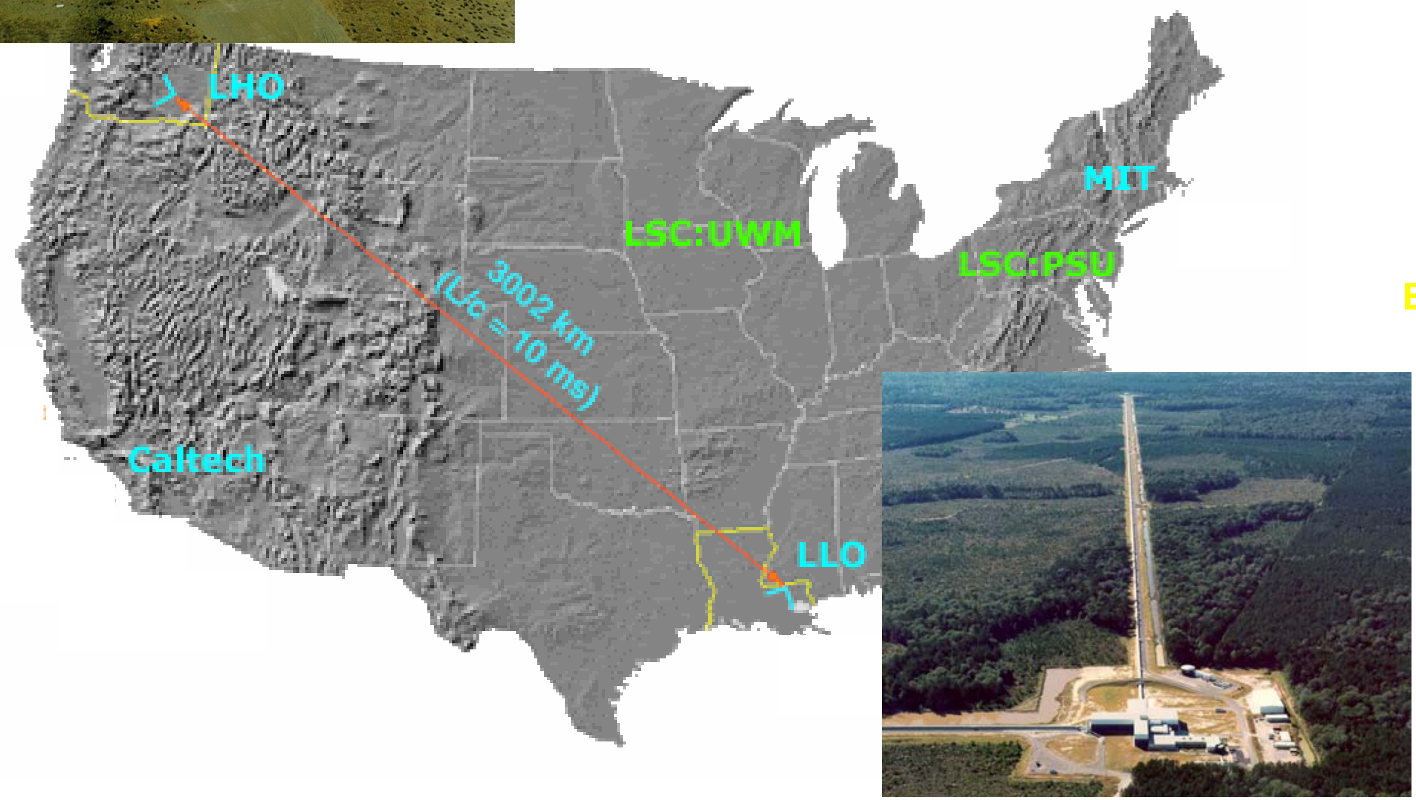


Local DB	Update send (secs)	Bloom filter (secs)	Bloom filter (bits)
10K	<1	2	1 M
1 M	2	24	10 M
5 M	7	175	50 M

Reliable Wide Area Data Replication

LIGO Gravitational Wave Observatory





Replicating >1 Terabyte/day to 8 sites

>30 million replicas so far

MTBF = 1 month

www.globus.org/solutions

e-Science & Data

What is e-Science?

- **Goal: to enable better research**
- **Method: Invention and exploitation of advanced computational methods**
  - to generate, curate and analyse research data
    - From experiments, observations and simulations
    - Quality management, preservation and reliable evidence
  - to develop and explore models and simulations
    - Computation and data at extreme scales
    - Trustworthy, economic, timely and relevant results
  - to enable *dynamic* distributed virtual organisations
    - Facilitating collaboration with information and resource sharing
    - Security, reliability, accountability, manageability and *agility*

Multiple, independently managed sources of data – each with own time-varying structure

Creative researchers discover new knowledge by combining data from multiple sources

The Primary Requirement …

Enabling Grids for E-science in Europe

Enabling *People* to Work Together on Challenging Projects: Science, Engineering & Medicine

Data Access and Integration: Scientific discovery

- **Choosing data sources**
  - **How do *you* find them?**
  - **How do *they* describe and advertise them?**
  - **Is the equivalent of Google possible?**
- **Obtaining access to that data**
  - **Overcoming administrative barriers**
  - **Overcoming technical barriers**
- **Understanding that data**
  - **The parts *you* care about for *your* research**
- **Extracting nuggets from multiple sources**
  - **Pieces of *your* jigsaw puzzle**
- **Combing them using sophisticated models**
  - **The *picture* of reality in *your* head**
- **Analysis on scales required by statistics**
  - **Coupling data access with computation**
- **Repeated Processes**
  - **Examining variations, covering a set of candidates**
  - **Monitoring the emerging details**
  - **Coupling with scientific workflows**

You're an innovator

$\therefore$ Your model $\neq$ their model

$\Rightarrow$ Negotiation & patience needed from *both* sides

Scientific Data: Opportunities and Challenges

Enabling Grids for E-science in Europe

- **Opportunities**
  - Global Production of *Published* Data
  - Volume↑ Diversity↑
  - Combination ⇒ Analysis ⇒ Discovery

- **Challenges**
  - Data Huggers
  - Meagre metadata
  - Ease of Use
  - …ntegration
  - Depen…

A Cornucopia of Research Challenges

- **Opportunities**
  - Sp… …xing
  - New Data Organisation
  - New Algorithms
  - Varied Replication
  - Shared Annotation
  - Intensive Data & Computation

- **…allenges**
  - Fundamental Principles
  - Approximate Matching
  - Multi-scale optimisation
  - Autonomous Change
  - Legacy structures
  - Scale and Longevity
  - Privacy and Mobility
  - Sustained Support / Funding

Summary & Conclusions

Take Home Message

- **There are plenty of Research Challenges**
  - Workflow & DB integration, co-optimised
  - Distributed Queries on a global scale
  - Heterogeneity on a global scale
  - Dynamic variability
    - Authorisation, Resources, Data & Schema
    - Performance
  - *Some* Massive Data
  - Metadata for discovery, automation, repetition, …
  - Catalogues
  - Optimising Data replication & Data movement
  - Provenance tracking
- **Grasp the theoretical & practical challenges**
  - Working in Open & Dynamic systems
  - Incorporate all computation
  - Welcome "code" visiting your data

Take Home Message (2)

Enabling Grids for E-science in Europe

- Information Grids
  - Support for collaboration
  - Support for computation and data grids
  - Structured data fundamental
    - Relations, XML, semi-structured, (described) files, …
  - Integrated strategies & technologies needed
- OGSA-DAI is here now
  - We'll talk about it on Saturday
- You live in Exciting Times
  - Can you ride the wave?

Comments & Questions Please

www.ogsadai.org.uk

www.nesc.ac.uk