

# What is Grid Computing?

*Mike Mineter  
Training Outreach and Education  
National e-Science Centre, UK*

*mjm@nesc.ac.uk*

[www.eu-egee.org](http://www.eu-egee.org)

- **Introduction to**
  - e-Infrastructure
  - e-Research and e-Science
- **Some examples from the EGEE project**
- **Grid concepts**
- **Grids - Where are we now?**

*‘e-Science is about global collaboration in key areas of science, and the **next generation of infrastructure** that will enable it.’*

*John Taylor*

*Director General of Research Councils*

*Office of Science and Technology*

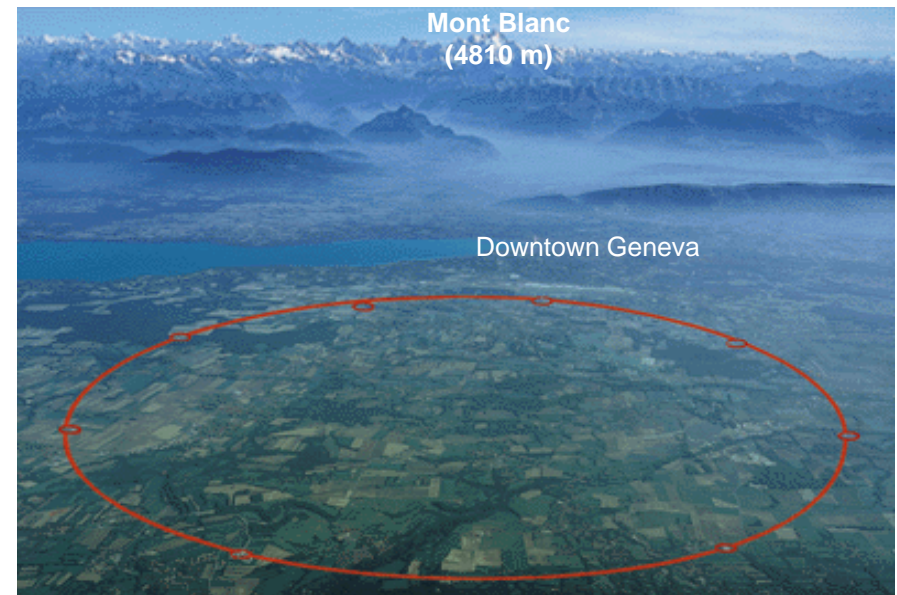
*UK*

***e-Infrastructure = Networks + Grids***

- Networks connect resources*
- Grids enable “virtual computing”*

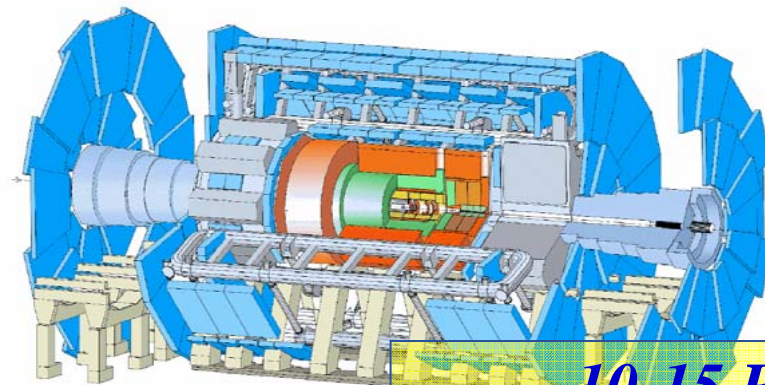
- **Collaborative research that is made possible by the sharing across the Internet of resources (data, instruments, computation, people’s expertise...)**
  - Crosses organisational boundaries
  - Often very compute intensive
  - Often very data intensive
  - Sometimes large-scale collaboration
- **Early examples were in science: “e-science”**
- **Relevance of “e-science technologies” to new user communities (social science, arts, humanities...) led to the term “e-research”**

- Large amount of data
- Large worldwide organized collaborations
- Computing and data management resources distributed world-wide owned and managed by many different entities
- Large Hadron Collider (LHC) at CERN in Geneva Switzerland:
  - One of the most powerful instruments ever built to investigate matter

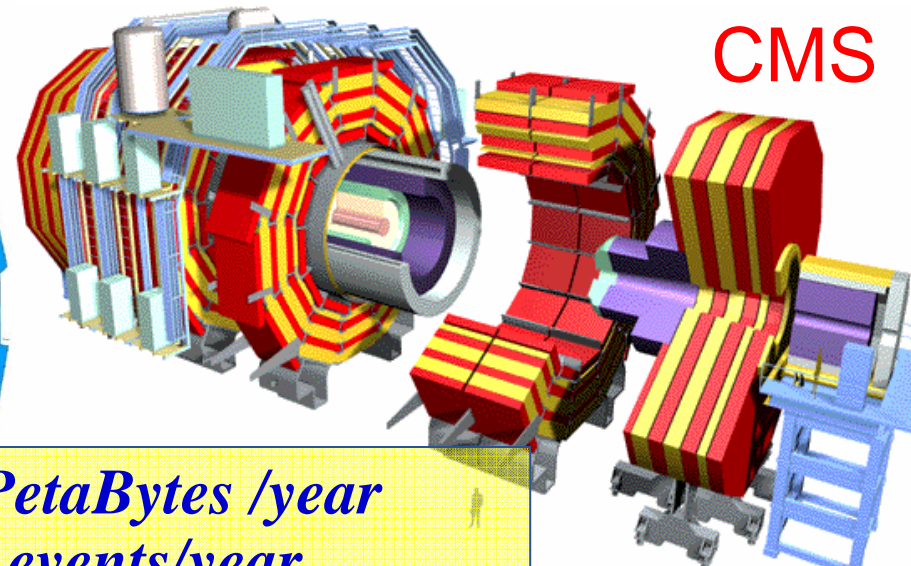




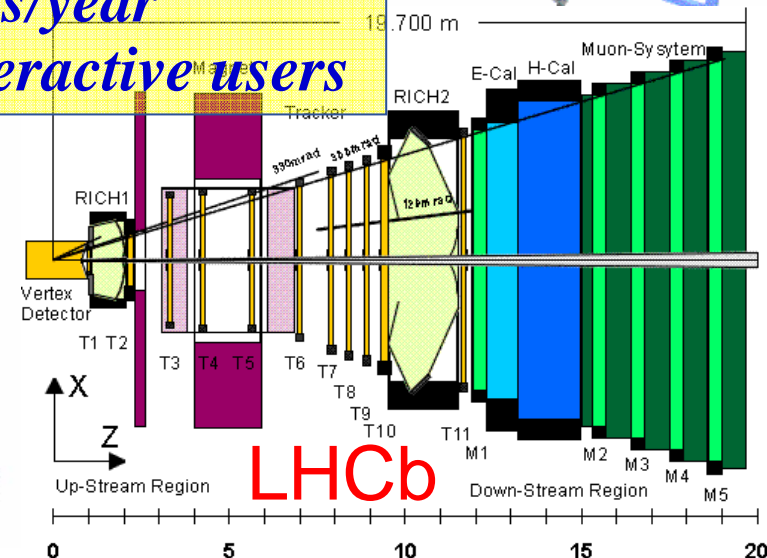
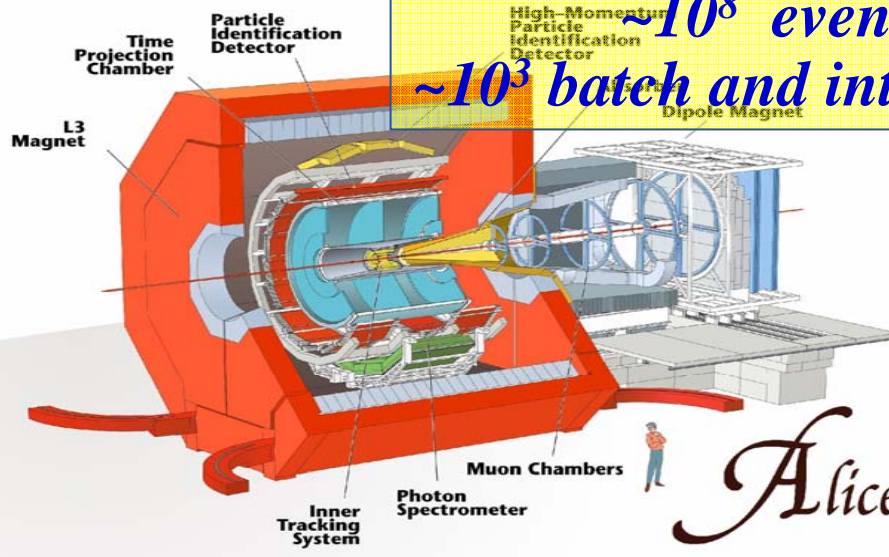
ATLAS



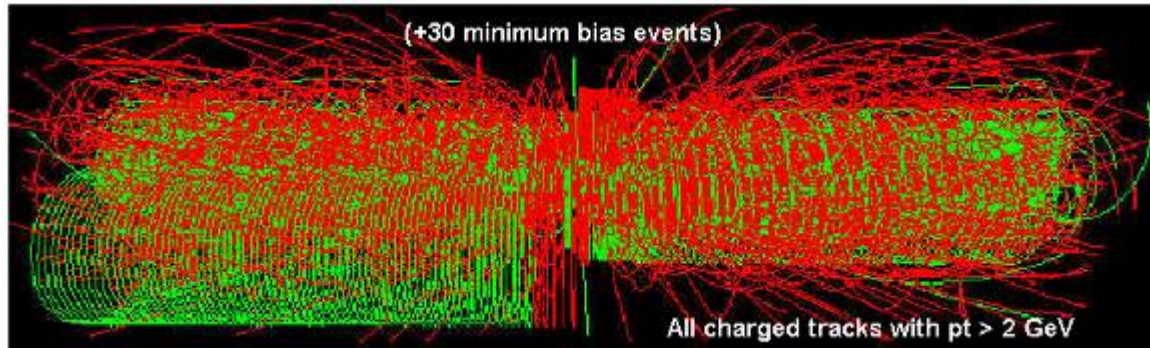
CMS



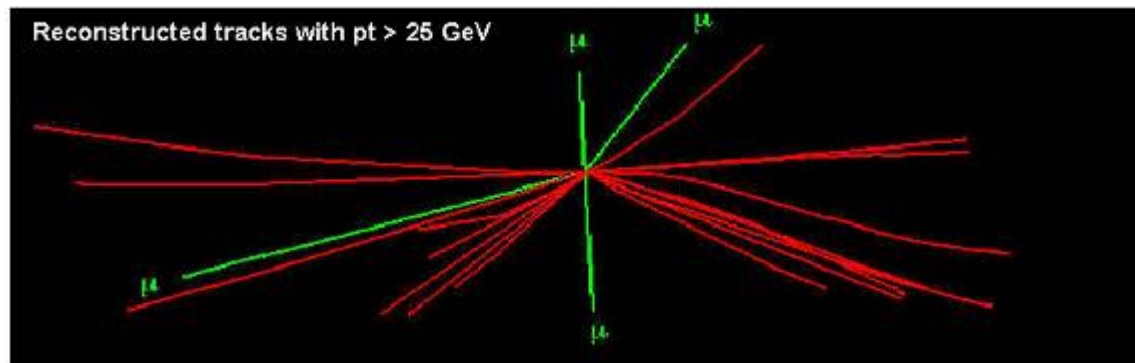
*~10-15 PetaBytes /year*  
*~10<sup>8</sup> events/year*  
*~10<sup>3</sup> batch and interactive users*



Starting from  
this event



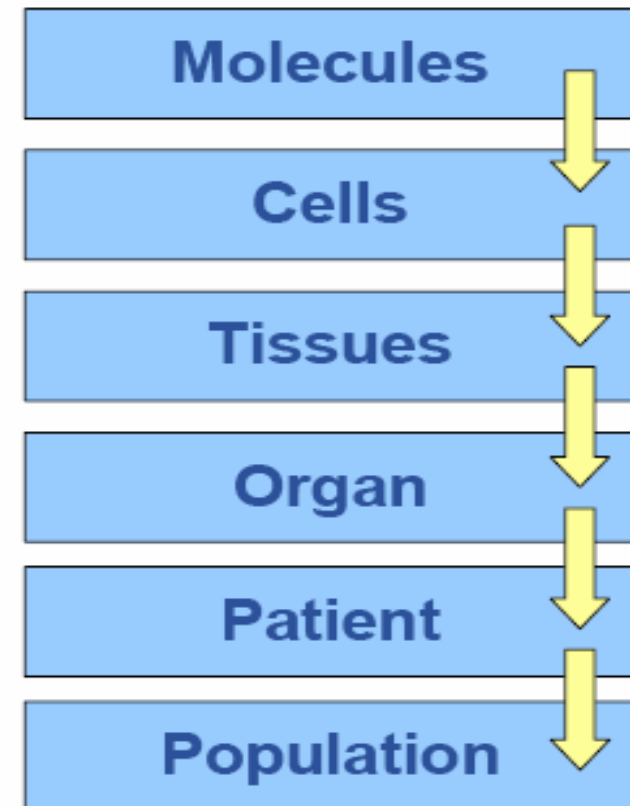
Looking for  
this “signature”



→ **Selectivity: 1 in  $10^{13}$**

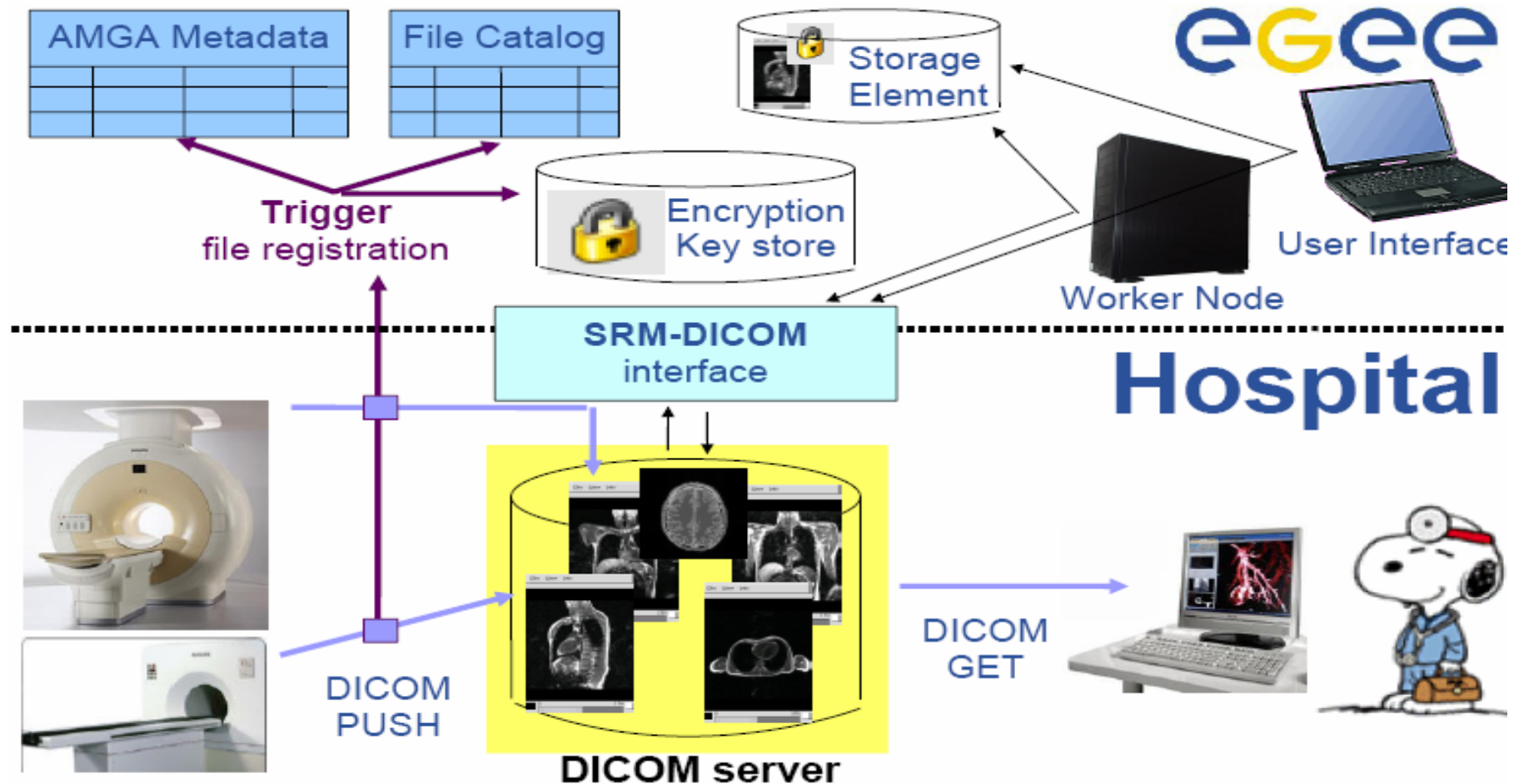
(Like looking for a needle in 20 million haystacks)

- **Bioinformatics**
  - Genomics
  - Proteomics
  - Phylogeny...
- **Medical imaging**
  - Medical imaging
  - Computer Aided Diagnosis
  - Therapy planning
  - Simulation...
- **Life sciences**
  - Drug discovery
  - Epidemiology
  - ...



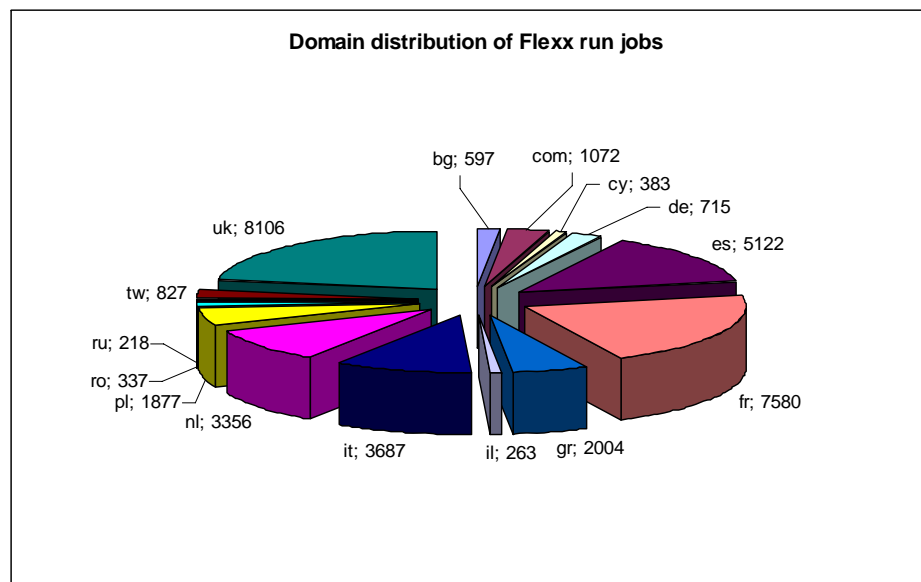
*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*





*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*

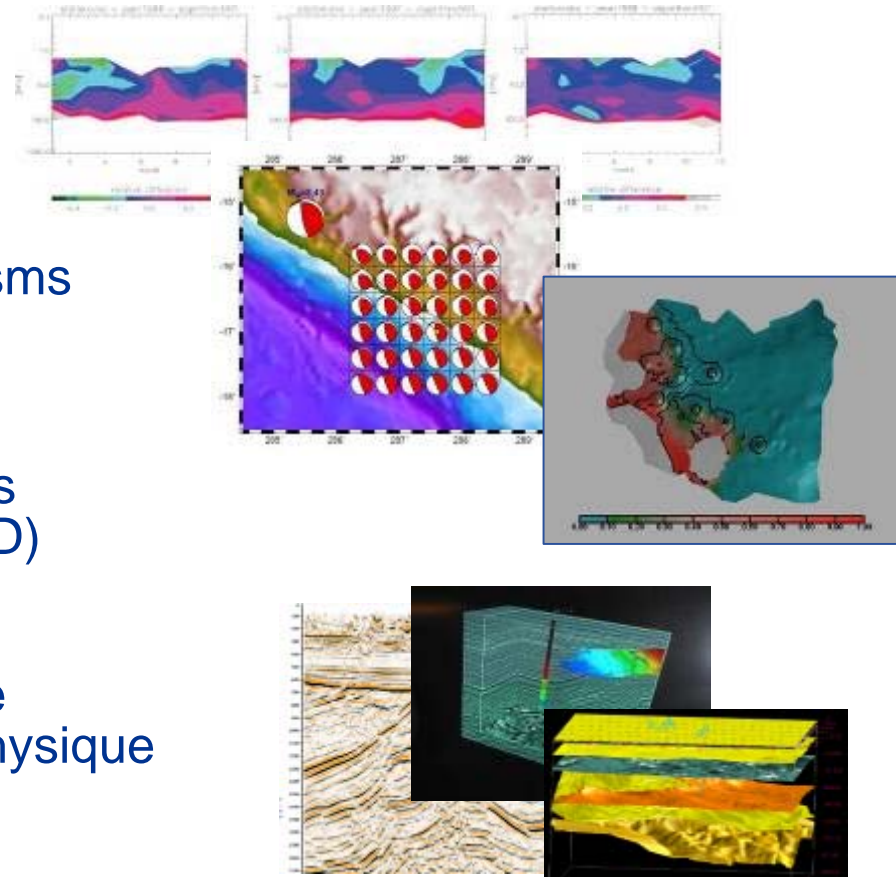
- **Significant biological parameters**
  - two different molecular docking applications (Autodock and FlexX)
  - about one million virtual ligands selected
  - target proteins from the parasite responsible for malaria
- **Significant numbers**
  - Total of about 46 million ligands docked in 6 weeks
  - 1TB of data produced
  - Up to 1000 computers in 15 countries used simultaneously for a total of about 80 CPU years
- **Significant results**
  - Best hits to be re-ranked using Molecular Dynamics



New data challenge in the fall of 2006  
 New malaria targets  
 Focus on other neglected diseases  
 Enlarged collaboration  
 (possibly including related projects)

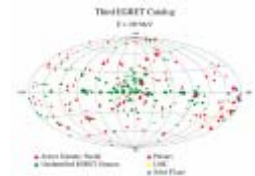
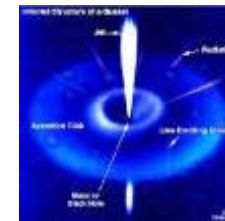
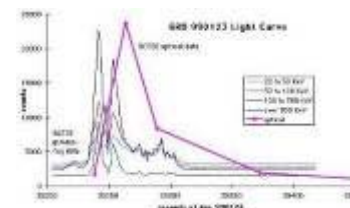
*Roberto Barbera, 1<sup>st</sup> EGEE User Forum, CERN, 1<sup>st</sup> March 2006*

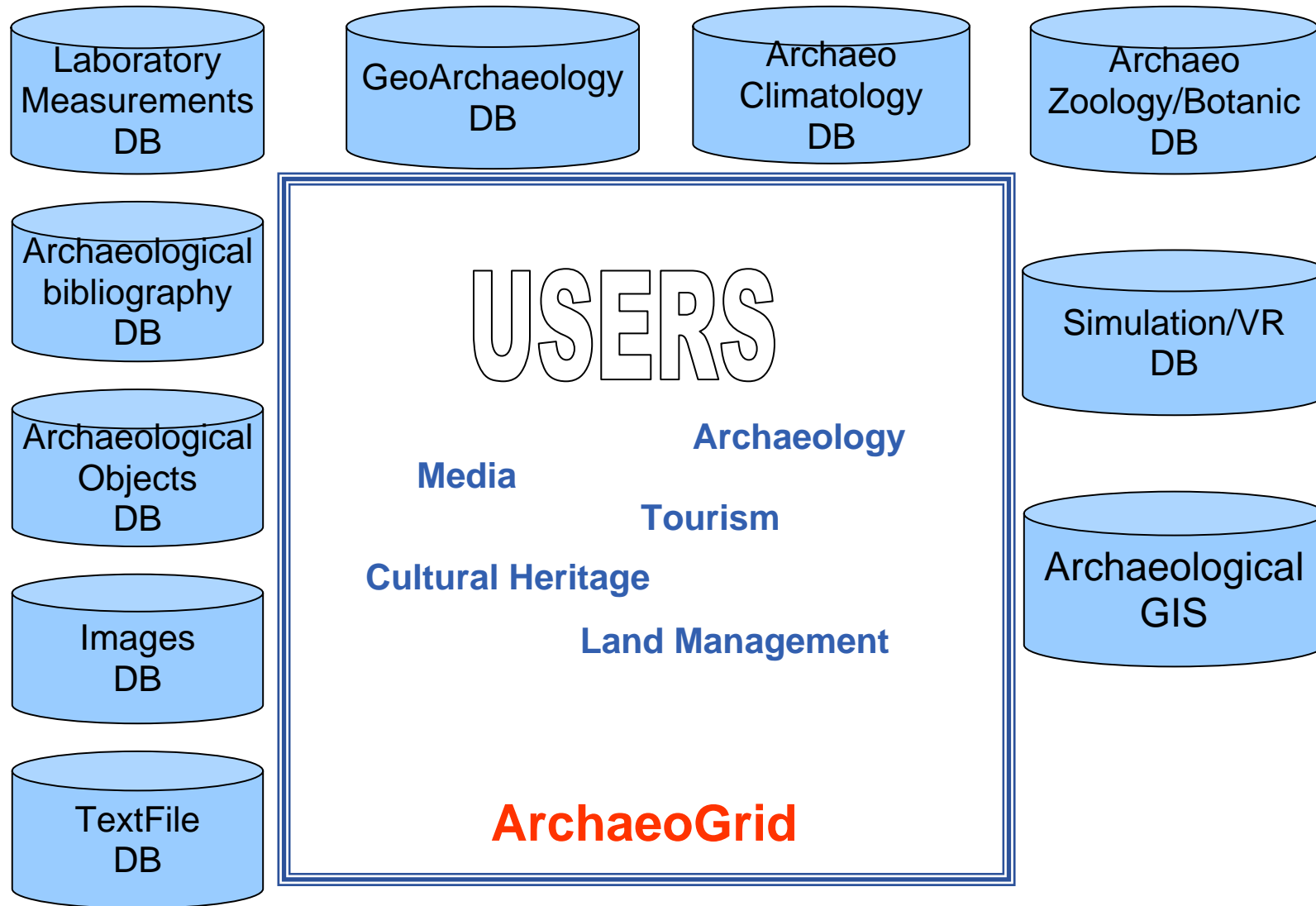
- **Earth Observations by Satellite**
  - Ozone profiles
- **Solid Earth Physics**
  - Fast Determination of mechanisms of important earthquakes
- **Hydrology**
  - Management of water resources in Mediterranean area (SWIMED)
- **Geology**
  - Geocluster: R&D initiative of the Compagnie Générale de Géophysique



➤ **A large variety of applications ported on EGEE**

- **Ground based Air Cerenkov Telescope 17 m diameter**
- **Physics Goals:**
  - Origin of VHE Gamma rays
  - Active Galactic Nuclei
  - Supernova Remnants
  - Unidentified EGRET sources
  - Gamma Ray Burst
- **MAGIC II will come 2007**
- **Grid added value**
  - Enable “(e-)scientific” collaboration between partners
  - Enable the cooperation between different experiments
  - Enable the participation on Virtual Observatories







# Grids: a foundation for e-Research

- e-Science methodologies will **rapidly transform** science, engineering, medicine and business
  - driven by exponential growth ( $\times 1000/\text{decade}$ )
    - ▶ enabling a whole-system approach

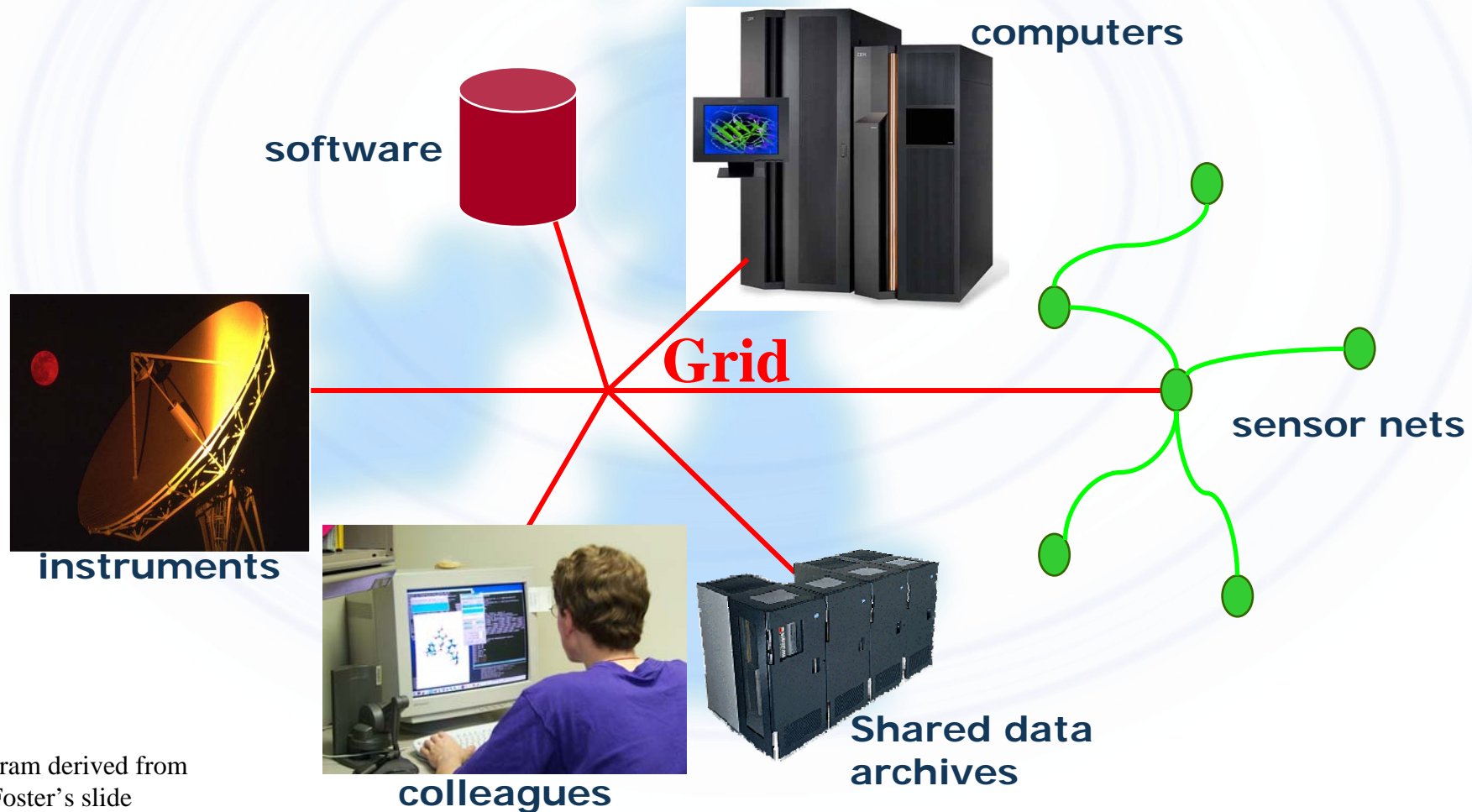
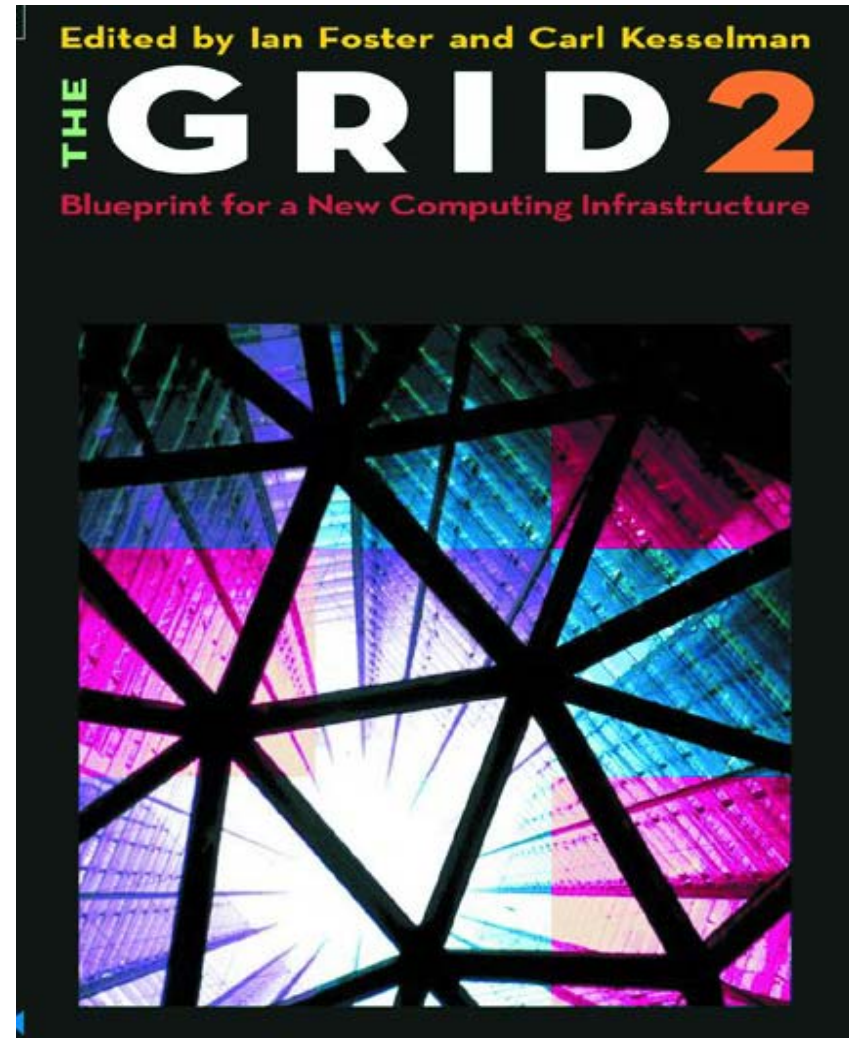


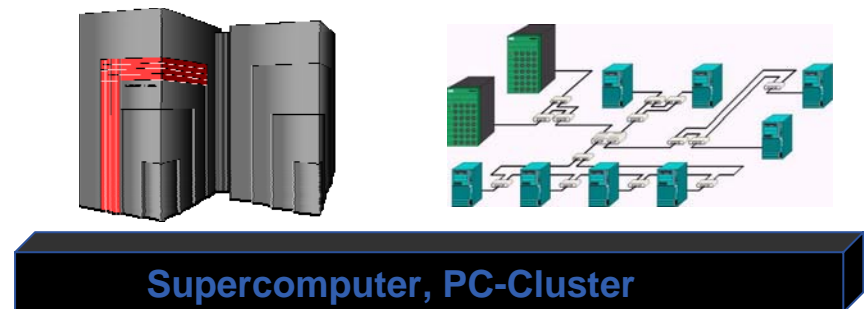
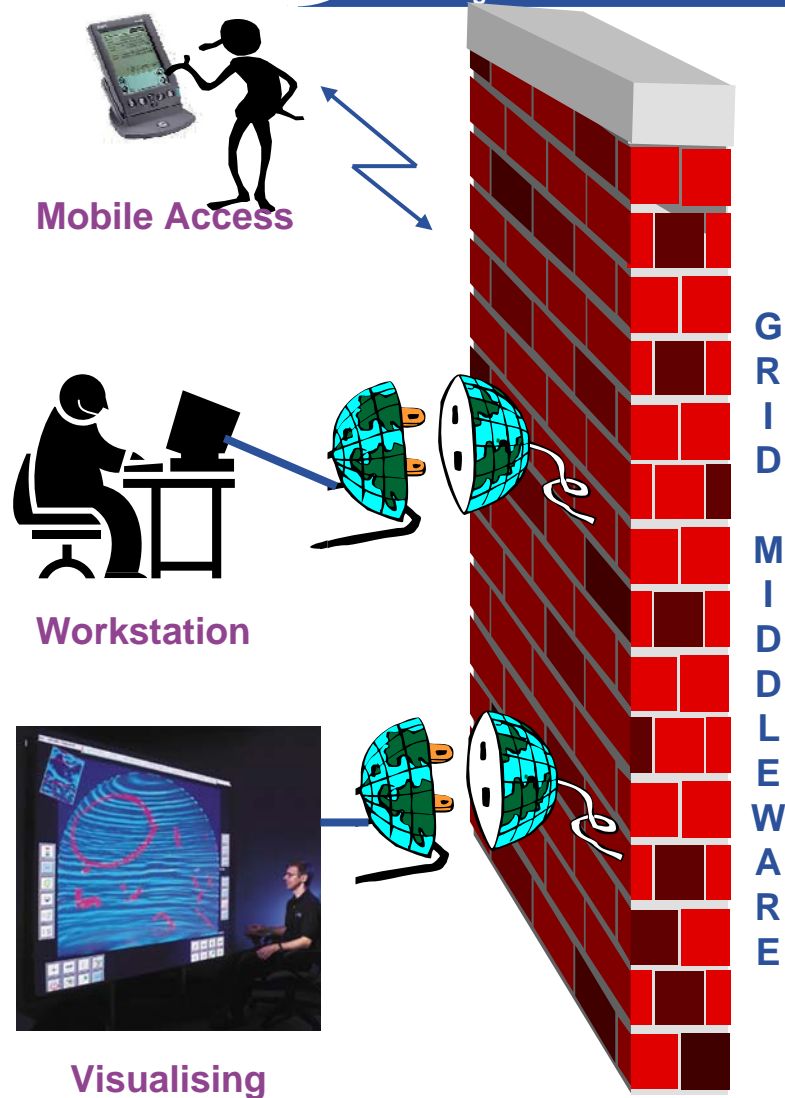
Diagram derived from  
Ian Foster's slide

# What is Grid Computing?

- The grid vision is of “Virtual computing” (+ information services to locate computation, storage resources)
  - Compare: The web: “virtual documents” (+ search engine to locate them)
- **MOTIVATION:** collaboration through sharing resources (and expertise) to expand horizons of
  - Research
  - Commerce – engineering, ...
  - Public service – health, environment,...



# The Grid Metaphor



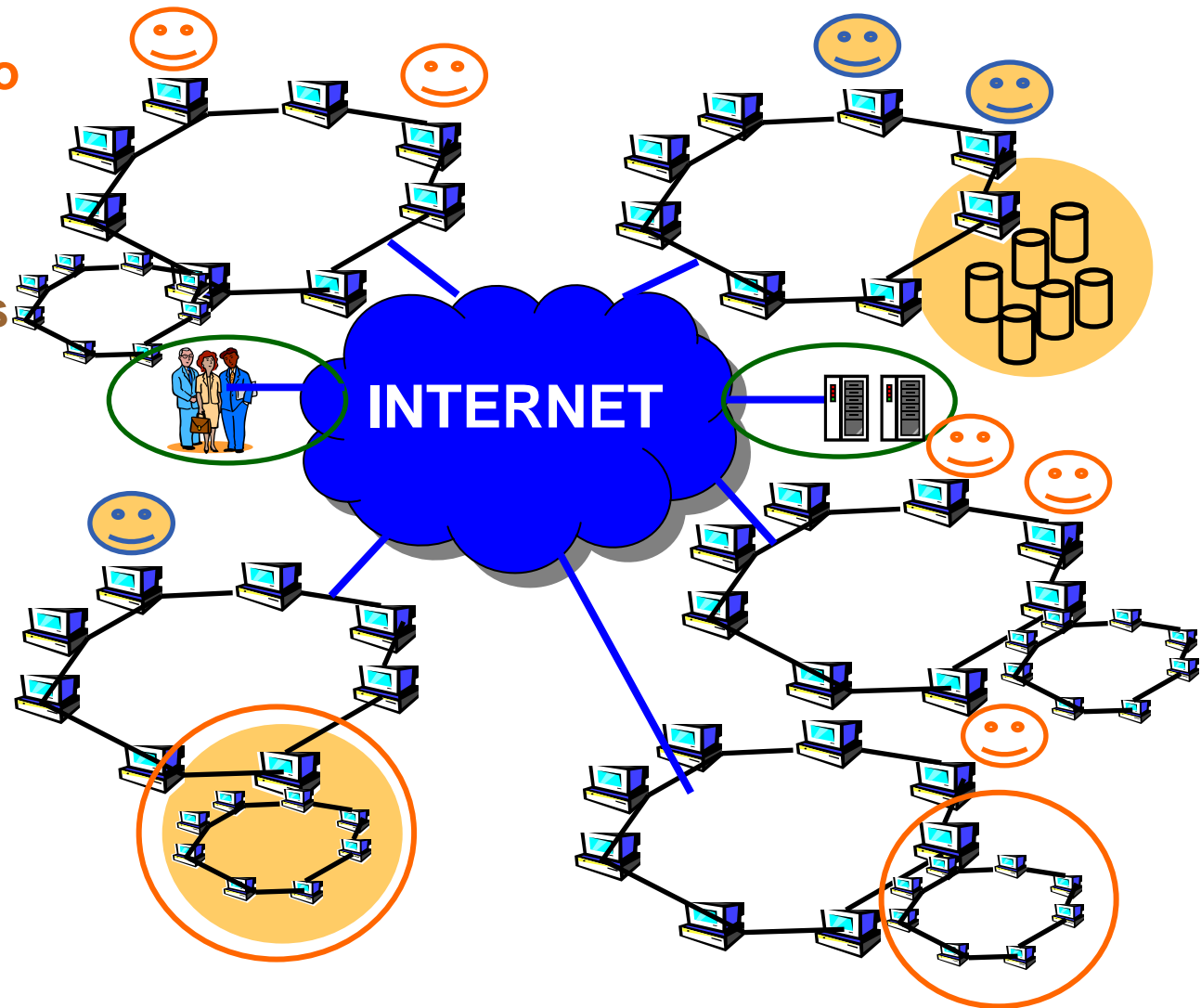
# Grid concepts

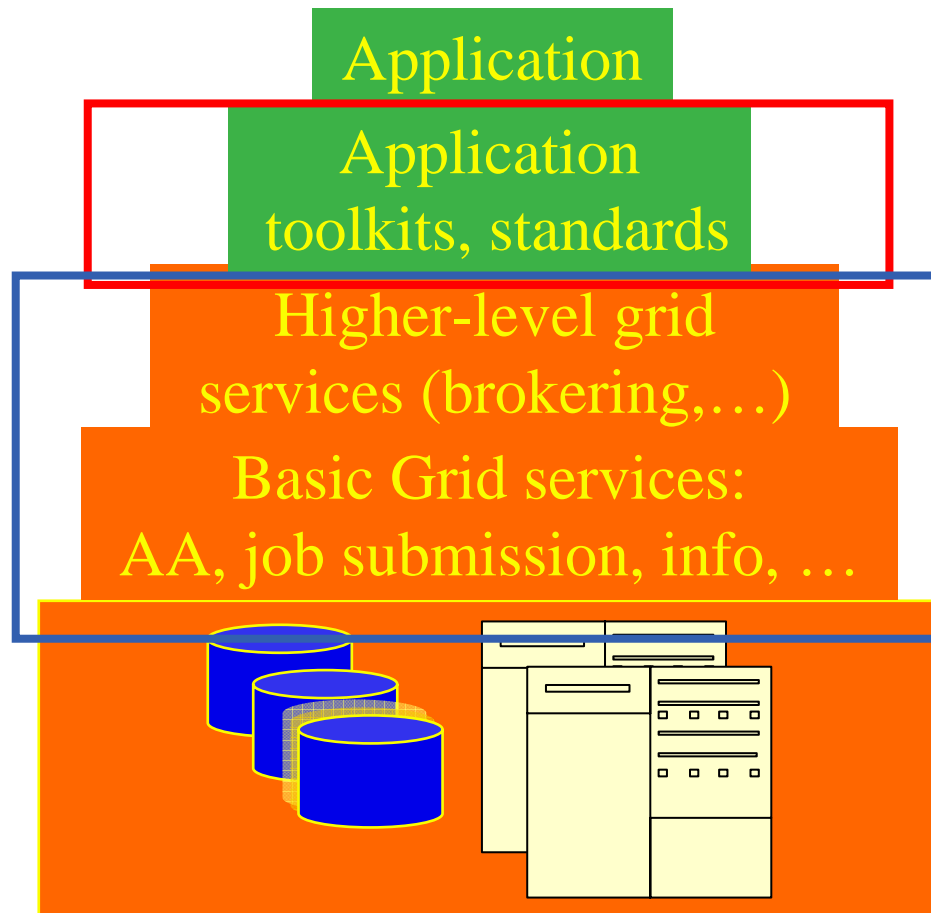
- **What is a Virtual Organisation?**
  - People in different organisations seeking to cooperate and share resources across their organisational boundaries
  - E.g. A research collaboration
- **Each grid is an infrastructure enabling one or more “virtual organisations” to share and access resources**



# Typical current grid

- **Virtual organisations negotiate with sites to agree access to resources**
- **Grid middleware runs on each shared resource to provide**
  - Data services
  - Computation services
  - Single sign-on
- **Distributed services (both people and middleware) enable the grid**





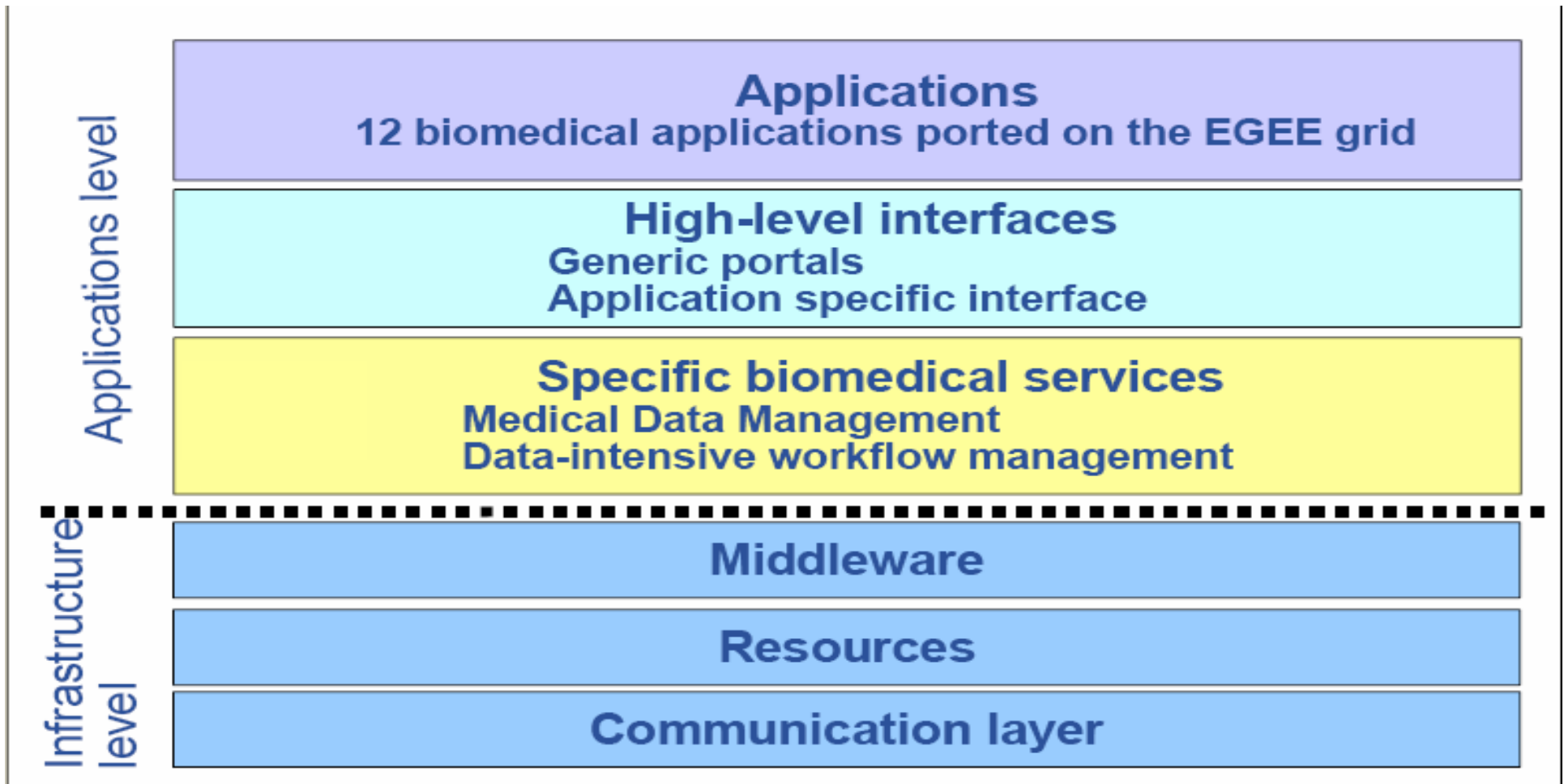
**Where computer science meets the application communities!**

**VO-specific developments:**

- Portals
- Virtual Research Environments
- Semantics, ontologies
- Workflow
- Registries of VO services

**Production grids provide these services.**

# Example – Biomedical applications



*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*

# Example of Workflow

100's to 1000's of medical images

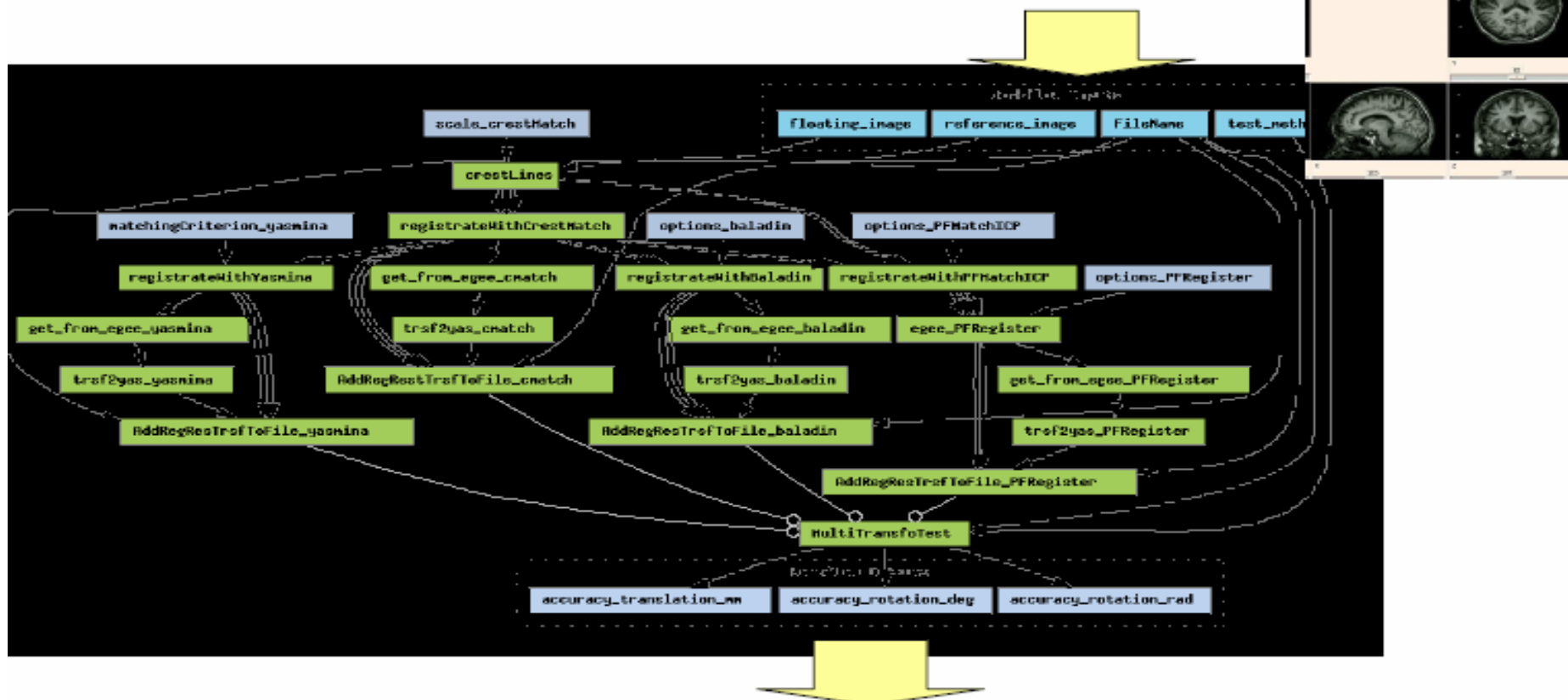


Image registration algorithms assessment

*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*

- **Providers of resources (computers, databases,...) need risks to be controlled: they are asked to trust users they do not know**
  - They trust a VO
  - The VO trusts its members
- **User's need**
  - single sign-on: to be able to logon to a machine that can pass the user's identity to other resources
  - To trust owners of the resources they are using
- **Build middleware on layer providing:**
  - *Authentication*: know who wants to use resource
  - *Authorisation*: know what the user is allowed to do
  - *Security*: reduce vulnerability, e.g. from outside the firewall
  - *Non-repudiation*: knowing who did what
- **The “Grid Security Infrastructure” middleware is the basis of (most) production grids**





If "The Grid"  
vision leads us  
here...

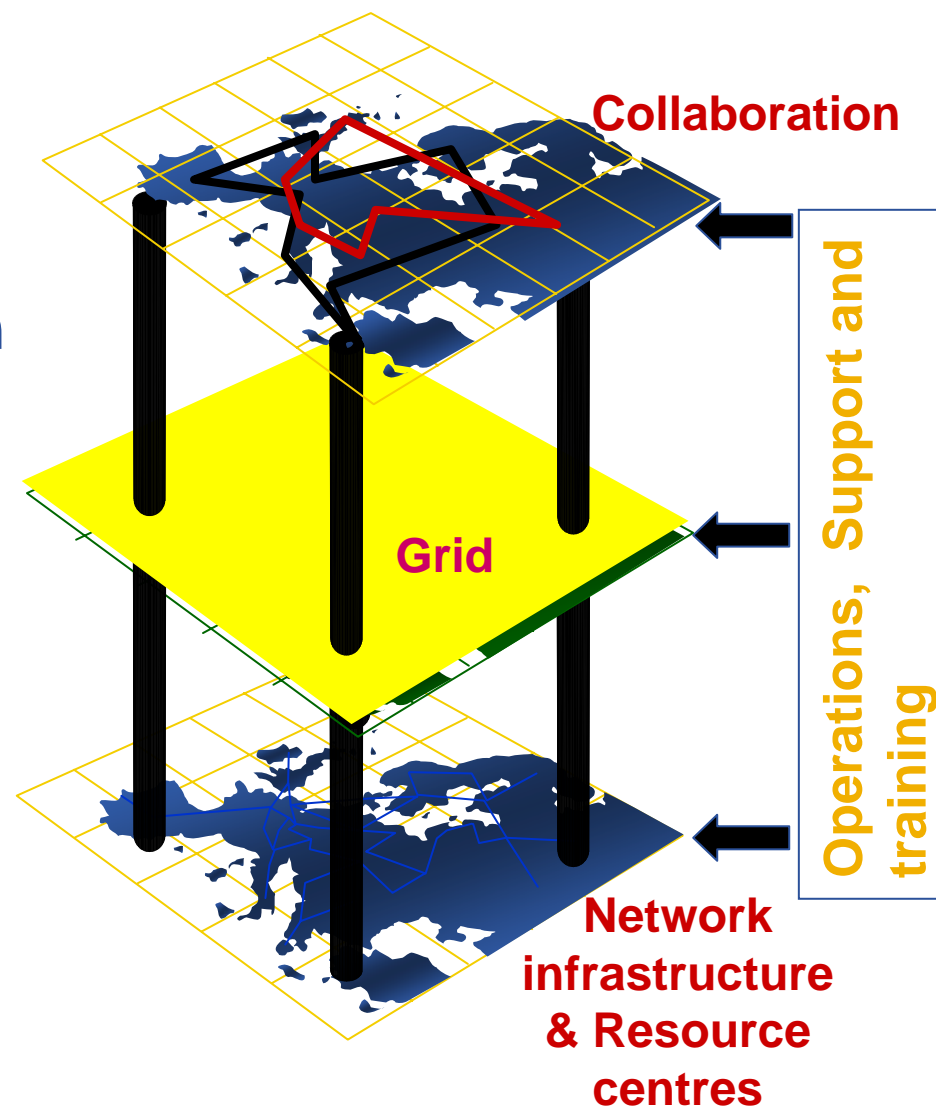
... then where are  
we now?

- Many key concepts identified and known
- Many grid projects have tested, and benefit from, these
- Major efforts now on establishing:
  - **Production Grids for multiple VO's**
    - “Production” = Reliable, sustainable, with commitments to quality of service
    - Each has
      - *One stack of middleware that serves many research communities*
      - *Establishing operational procedures and organisation*
    - Challenge for EGEE-II: federate these!
  - **Standards** (a slow process)
    - e.g. Global Grid Forum, <http://www.gridforum.org/>
    - Extending web services
  - **Broadening range of user communities**
    - arts and humanities, social science ...

# National grid initiatives now include...



- **Grids:** collaboration across administrative domains
- **Networks:** collaboration across geographical distance
- **Semantics, ontologies:** collaboration across disciplines / groups
- **Storage, (“curation”):** collaboration across time



- **Global Grid Forum** <http://www.ggf.org/> (see GGF16)
- **EGEE** [www.eu-egee.org](http://www.eu-egee.org)
- **The Grid Cafe** [www.gridcafe.org](http://www.gridcafe.org)