



Enabling Grids  
for E-science



# What is Grid Computing?

*Richard Hopkins  
Edinburgh e-Science  
Training Outreach and Education*

*National e-Science Centre, UK*

*rph@nesc.ac.uk*

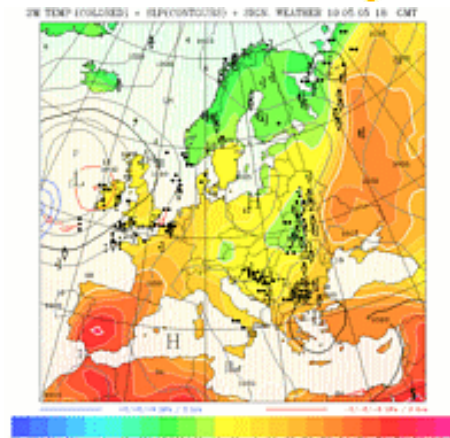
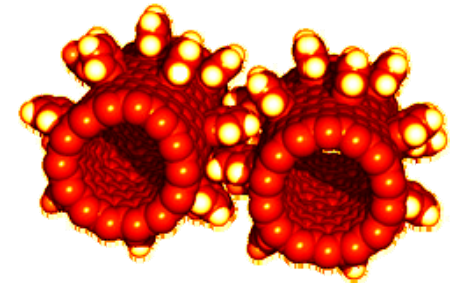
[www.eu-egee.org](http://www.eu-egee.org)



EGEE-II INFOSO-RI-031688

- **Introduction to**
  - e-Research and e-Science
  - Grid Computing
  - e-Infrastructure
- **Some examples**
- **Grid concepts**
- **Grids - Where are we now?**

- **Many vital challenges require community effort**
    - Fundamental properties of matter
    - Genomics
    - Climate change
    - Medical diagnostics
  - **Research is increasingly digital, with increasing amounts of data**
  - **Computation ever more demanding**
- e.g.: experimental science uses ever more sophisticated sensors**
- Huge amounts of data
  - Serves user communities around the world
  - International collaborations



- **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”
  - often “e-Science” really means “e-Research”

**Collaborative  
“virtual computing”**



**Improvised cooperation**



**People with shared goals**

**Sharing data, computers, software  
Enabled by Grids – two main types**

- specific to a project
- supporting many collaborations

**Email**

**File exchange**

**ssh access to run programs**

**Enabled by networks:**

**national, regional and  
International: GEANT**

- **Networks + Grids**
  - *Networks connect resources*
  - *Grids enable “virtual computing” - resource sharing across administrative domains*
    - *“admin. domain”: institute, country where resource is; system management processes;...*
- **+ Operations, Support, Training...**
- **+ Data centres, archives,...**


- **A shared resource**
  - That enables science, research, engineering, medicine, industry, ...
  - It will improve UK / European / ... productivity
    - Lisbon Accord 2000
    - E-Science Vision SR2000 – John Taylor
  - Commitment by UK government
    - Sections 2.23-2.25
  - Always there
    - c.f. telephones, transport, power, internet

## Science & innovation investment framework 2004 - 2014

July 2004



department for  
education and skills



Gordon Brown  
Chancellor of the  
Exchequer

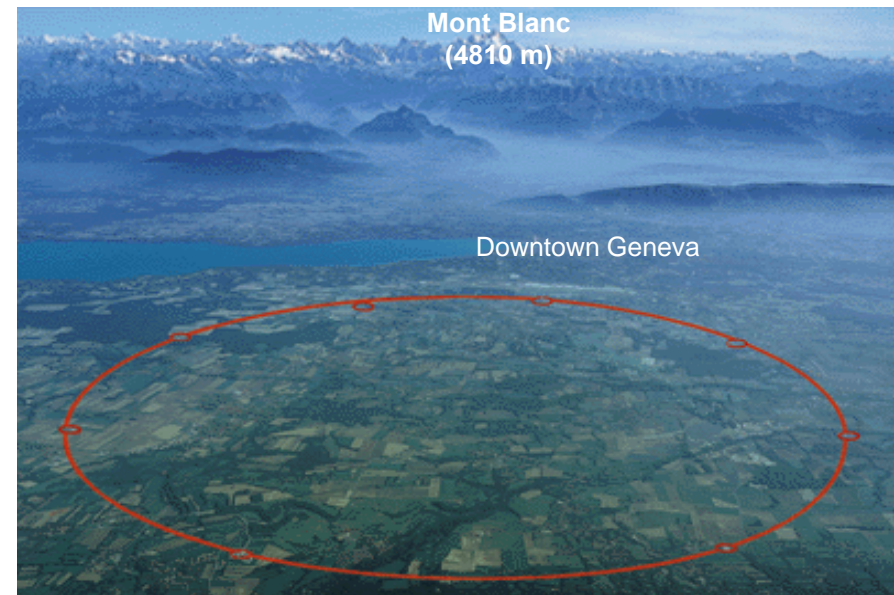
Charles Clarke  
Secretary of State for  
Education and Skills

Patricia Hewitt  
Secretary of State for  
Trade and Industry

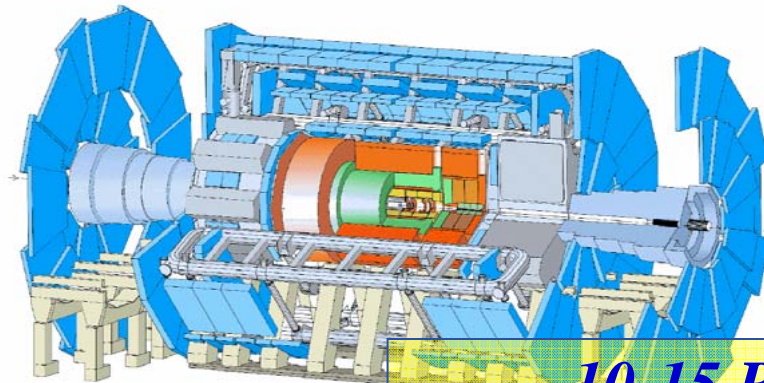
## Some examples of e-science



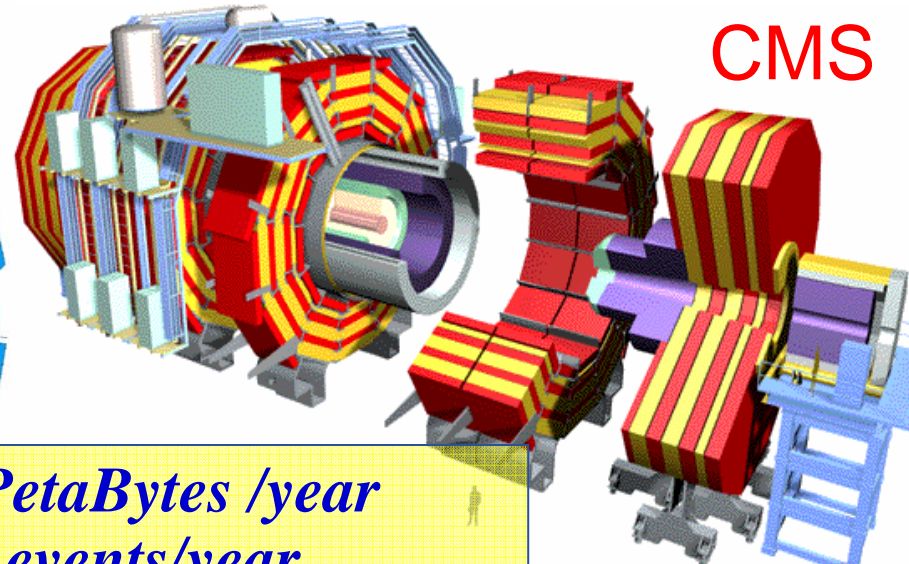
- Large amount of data
- Large worldwide organized collaborations
- Computing and data management resources distributed world-wide owned and managed by many different entities
- Large Hadron Collider (LHC) at CERN in Geneva Switzerland:
  - One of the most powerful instruments ever built to investigate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



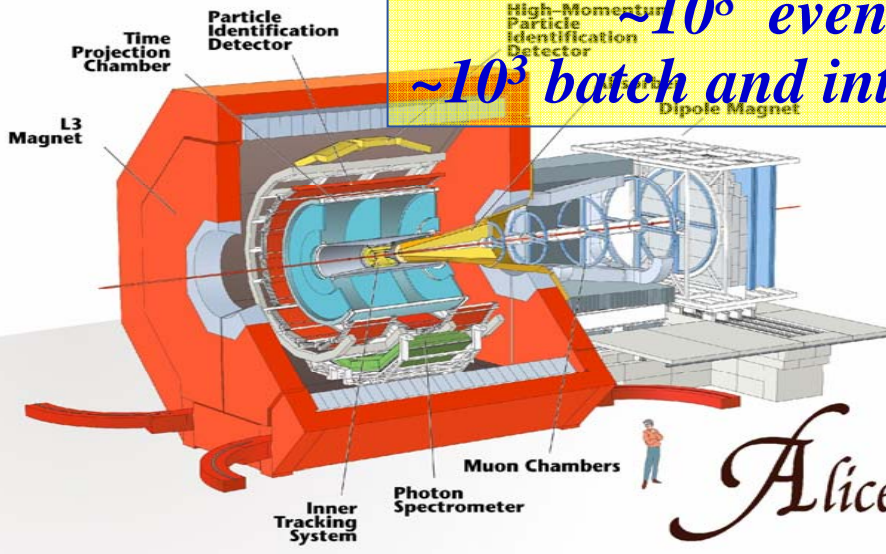
ATLAS



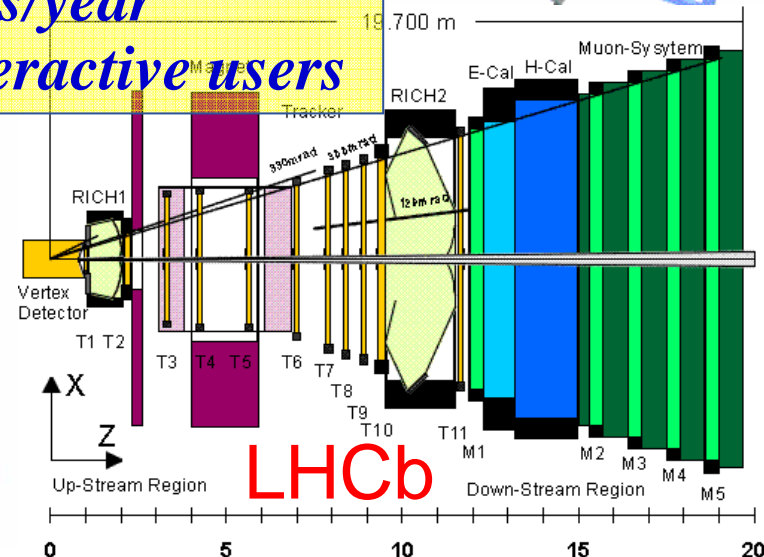
CMS



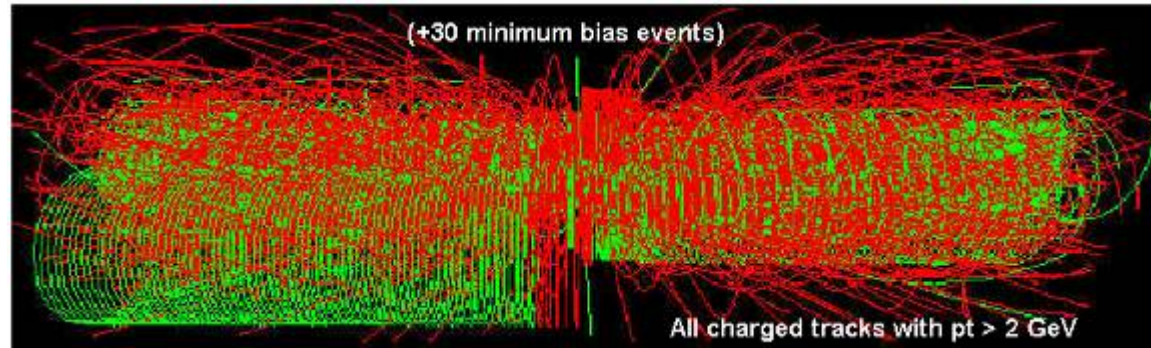
*~10-15 PetaBytes /year*  
*~10<sup>8</sup> events/year*  
*~10<sup>3</sup> batch and interactive users*



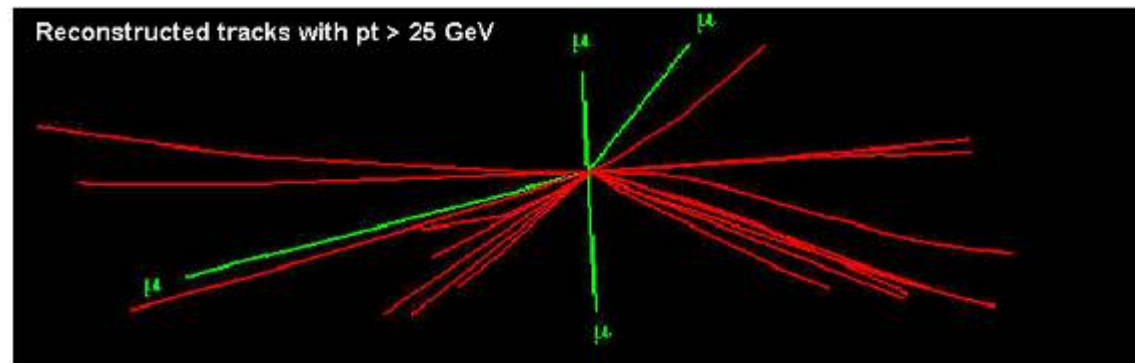
*Alice*



Starting from  
this event



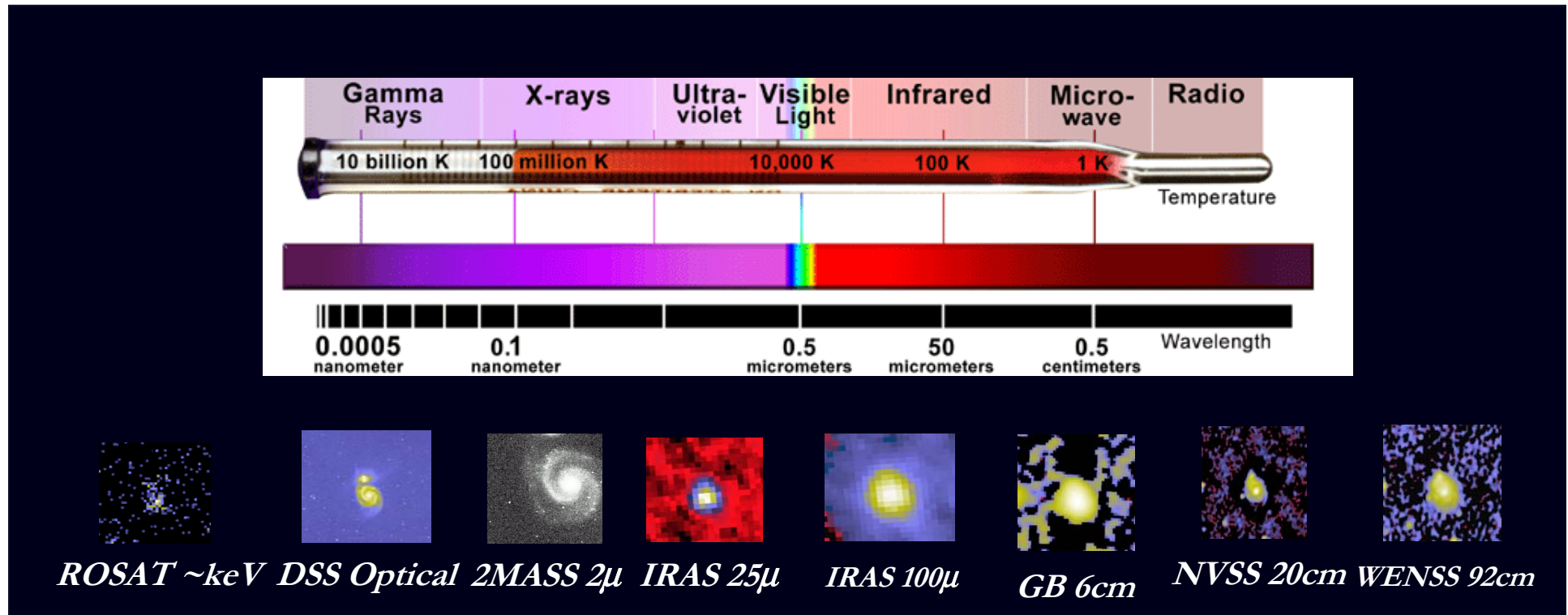
Looking for  
this “signature”



→ **Selectivity: 1 in  $10^{13}$**

(Like looking for a needle in 20 million haystacks)

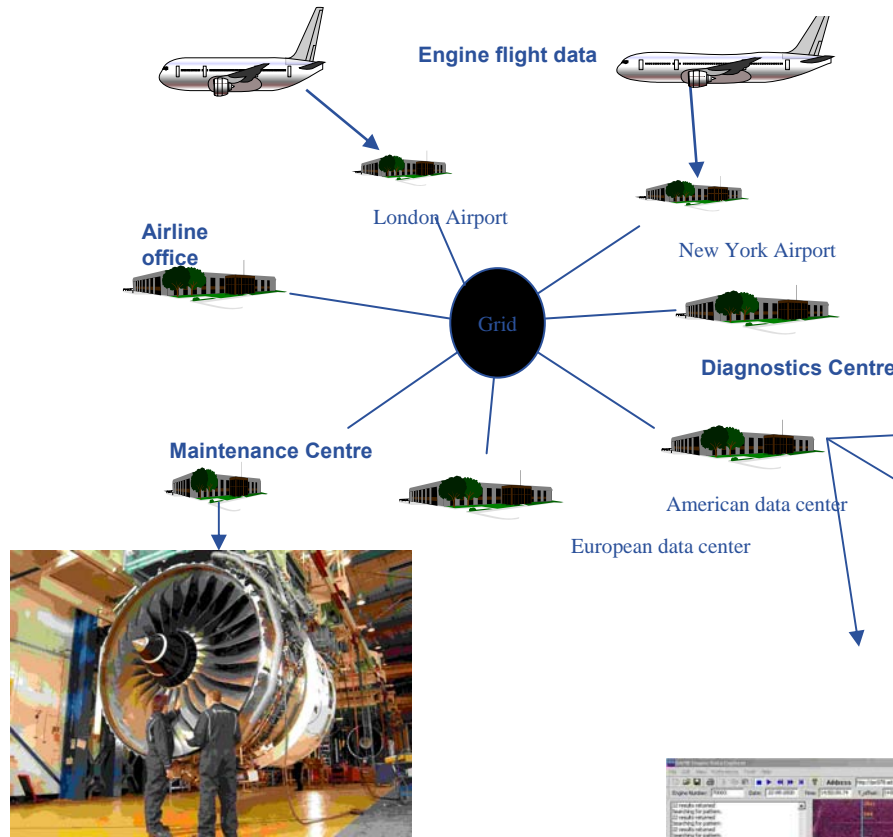
## Observations made across entire electromagnetic spectrum



⇒ e.g. different views of a local galaxy  
 Need all of them to understand physics fully  
 Databases are located throughout the world



# DAME: Grid based tools and Infrastructure for Aero-Engine Diagnosis and Prognosis

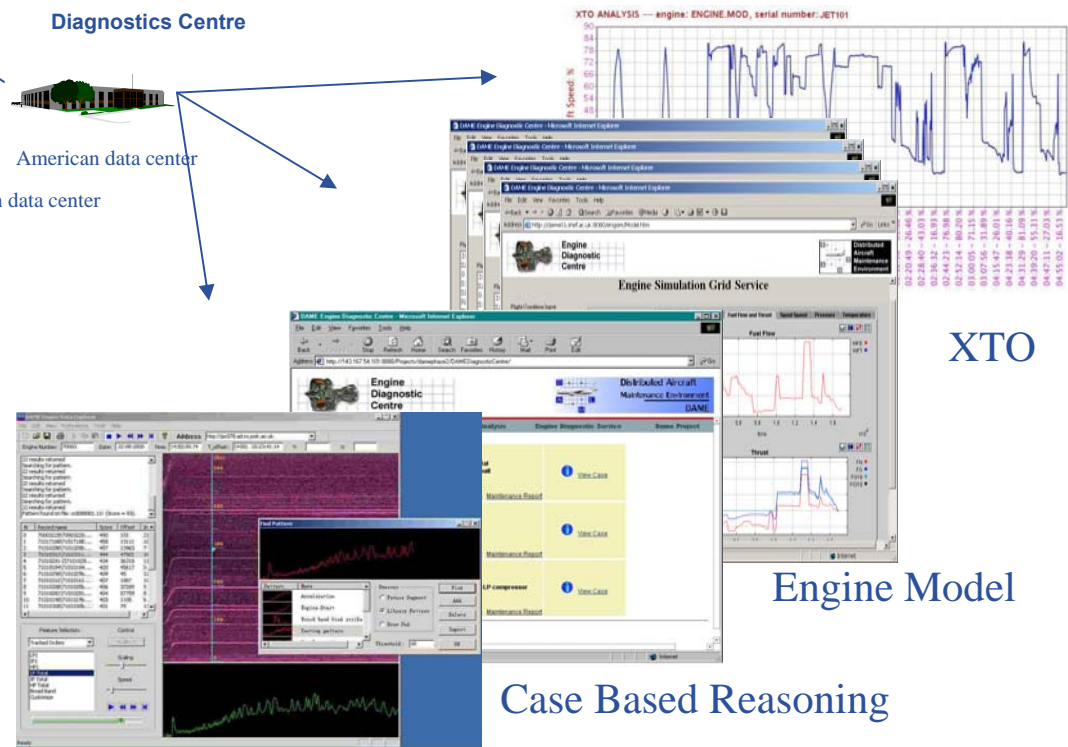


“A Significant factor in the success of the Rolls-Royce campaign to power the Boeing 7E7 with the Trent 1000 was the emphasis on the new aftermarket support service for the engines provided via DS&S. Boeing personnel were shown DAME as an example of the new ways of gathering and processing the large amounts of data that could be retrieved from an advanced aircraft such as the 7E7, and they were very impressed”, DS&S 2004

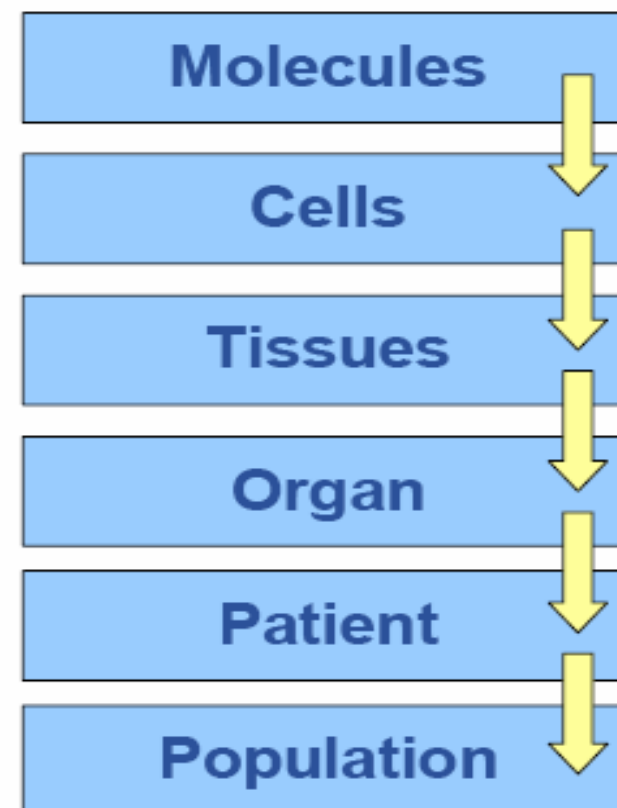


**Companies:**  
Rolls-Royce  
DS&S  
Cybula

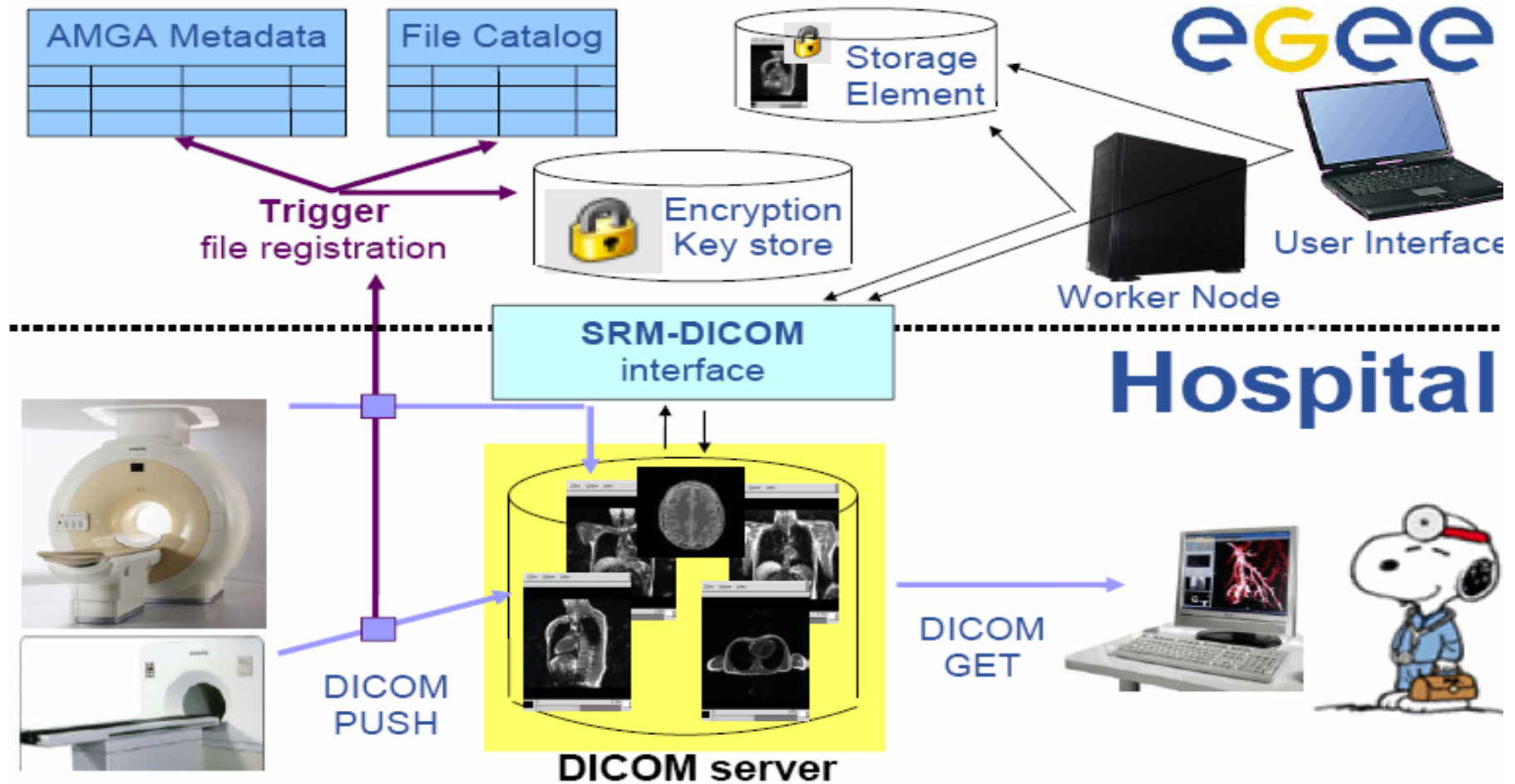
**Universities:**  
York,  
Leeds,  
Sheffield, Oxford



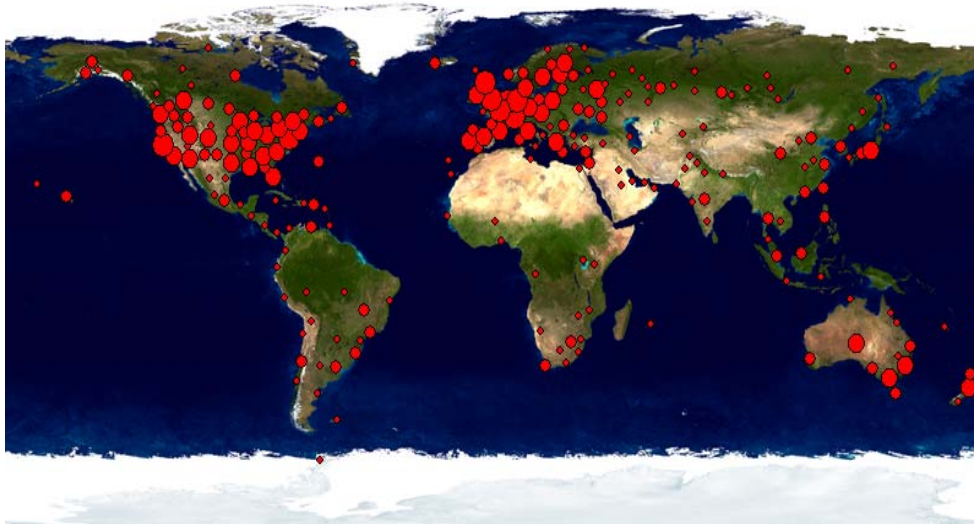
- **Bioinformatics**
  - Genomics
  - Proteomics
  - Phylogeny...
  
- **Medical imaging**
  - Medical imaging
  - Computer Aided Diagnosis
  - Therapy planning
  - Simulation...
  
- **Life sciences**
  - Drug discovery
  - Epidemiology
  - ...



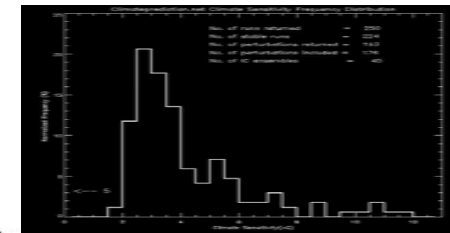
*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*



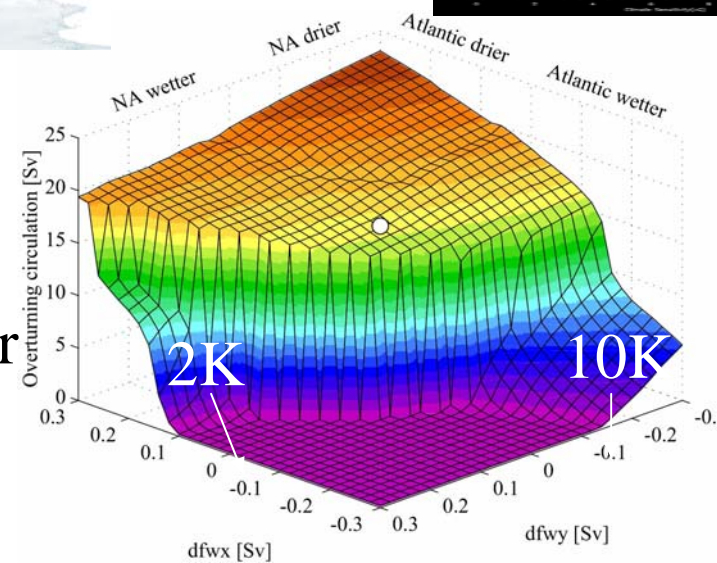
*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*



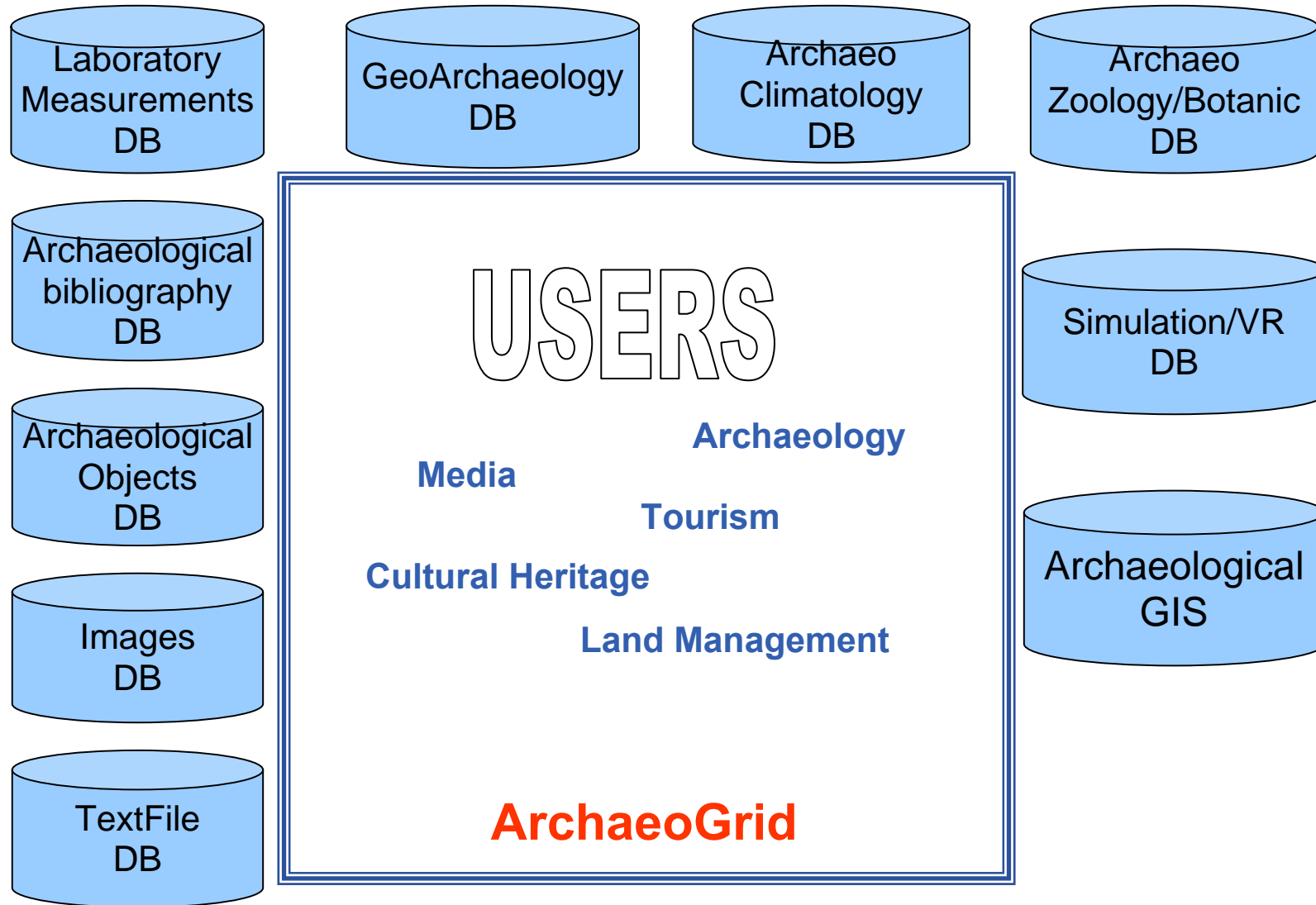
- Largest climate model ensemble
- >45,000 users,  
>1,000,000 model years



Response of Atlantic  
circulation to freshwater  
forcing

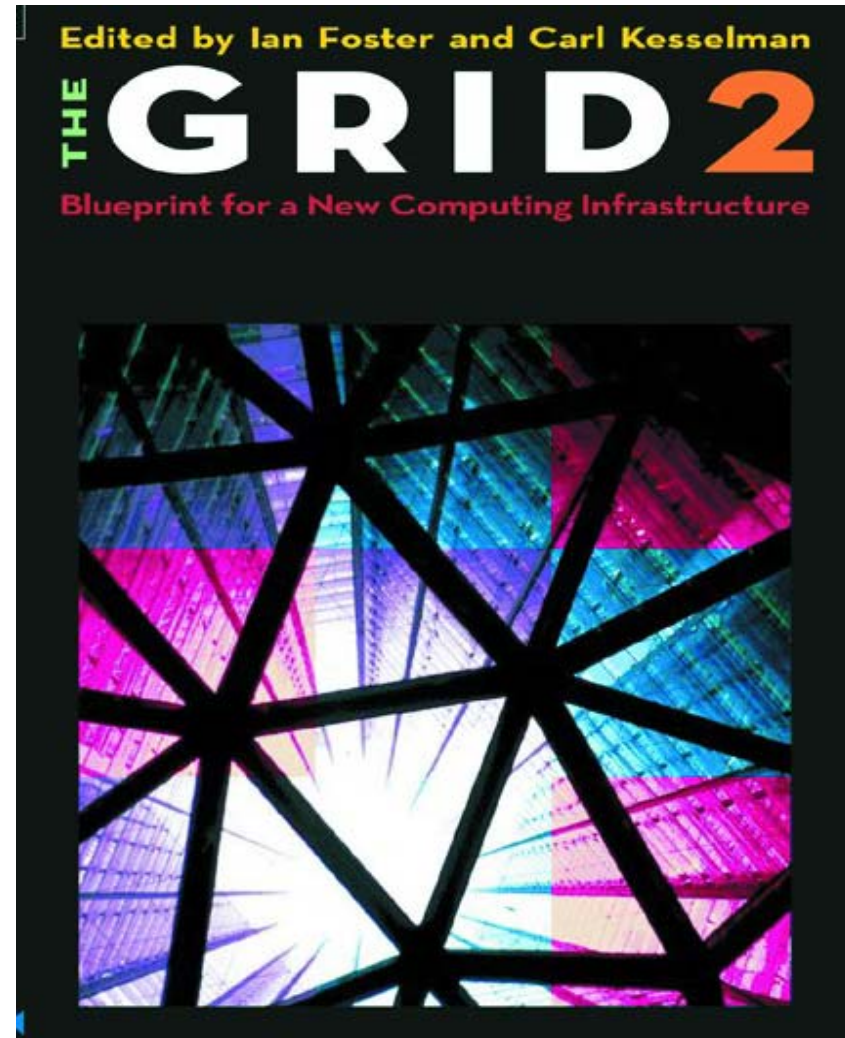


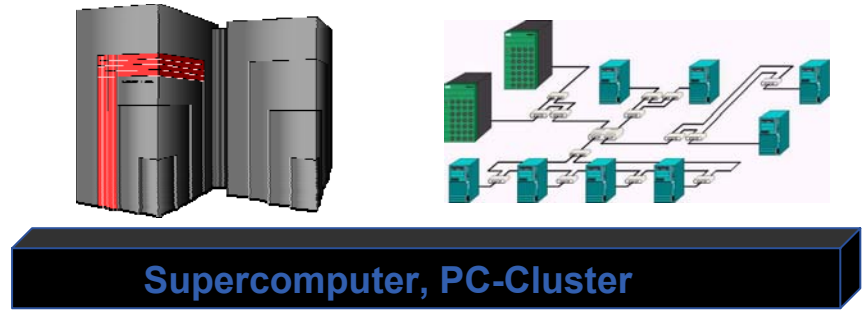
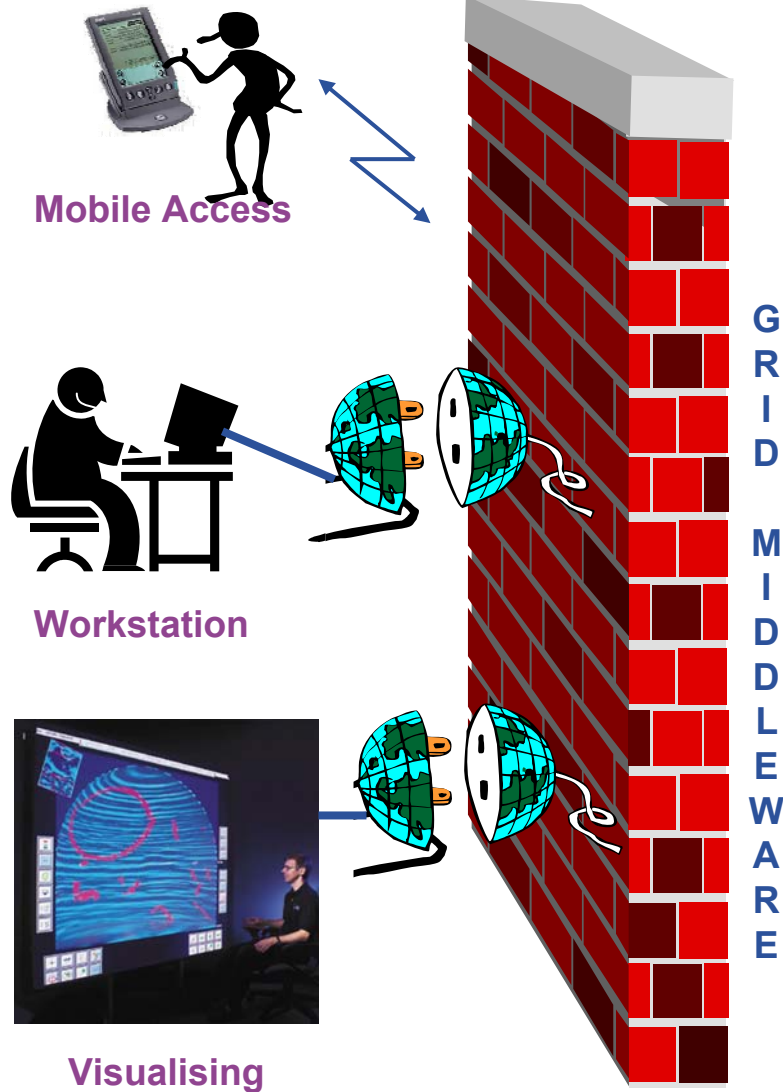




# Grid concepts

- 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,...





- Enabling a whole-system approach
- A challenge to the imagination
- Effect > Σparts

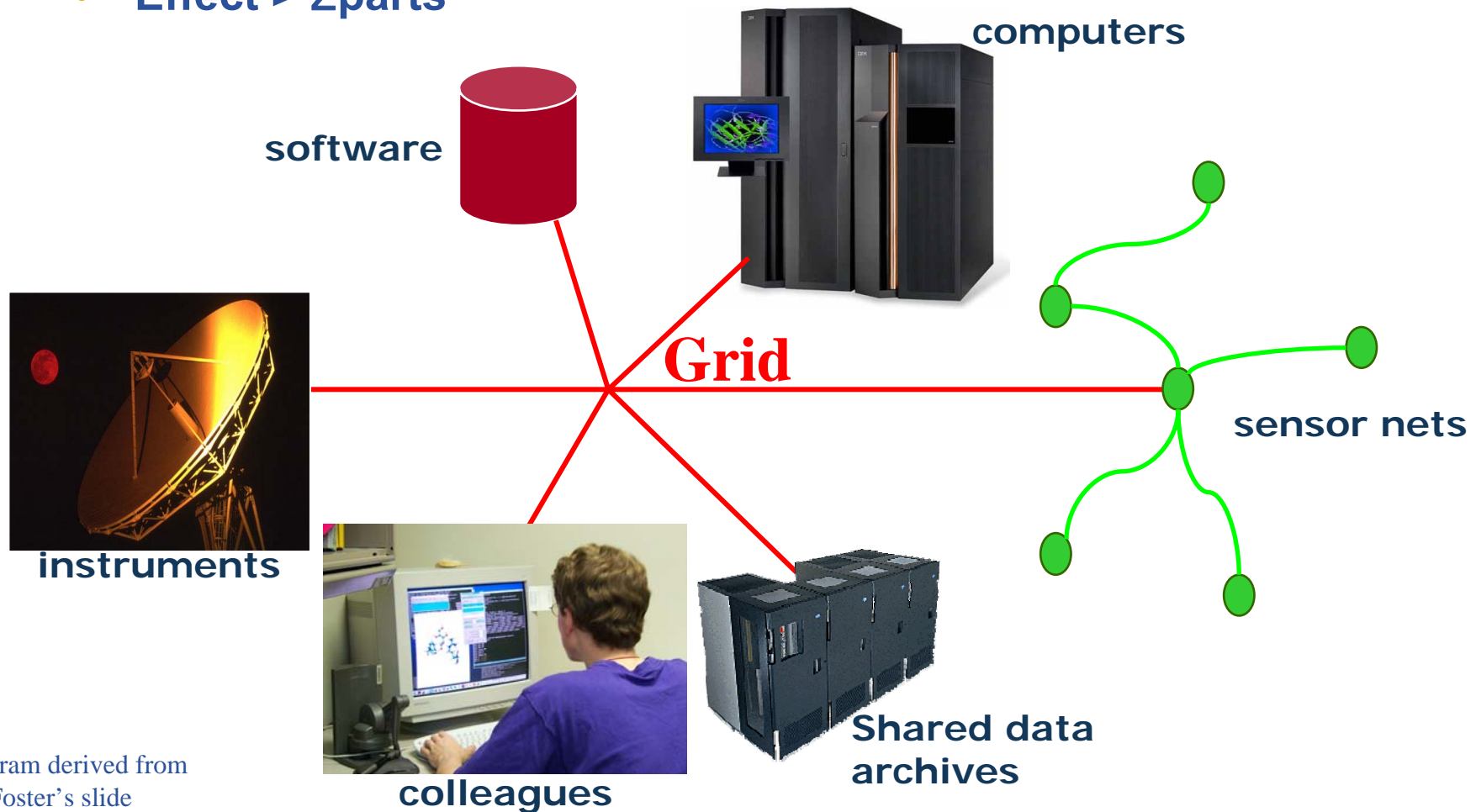


Diagram derived from  
Ian Foster's slide

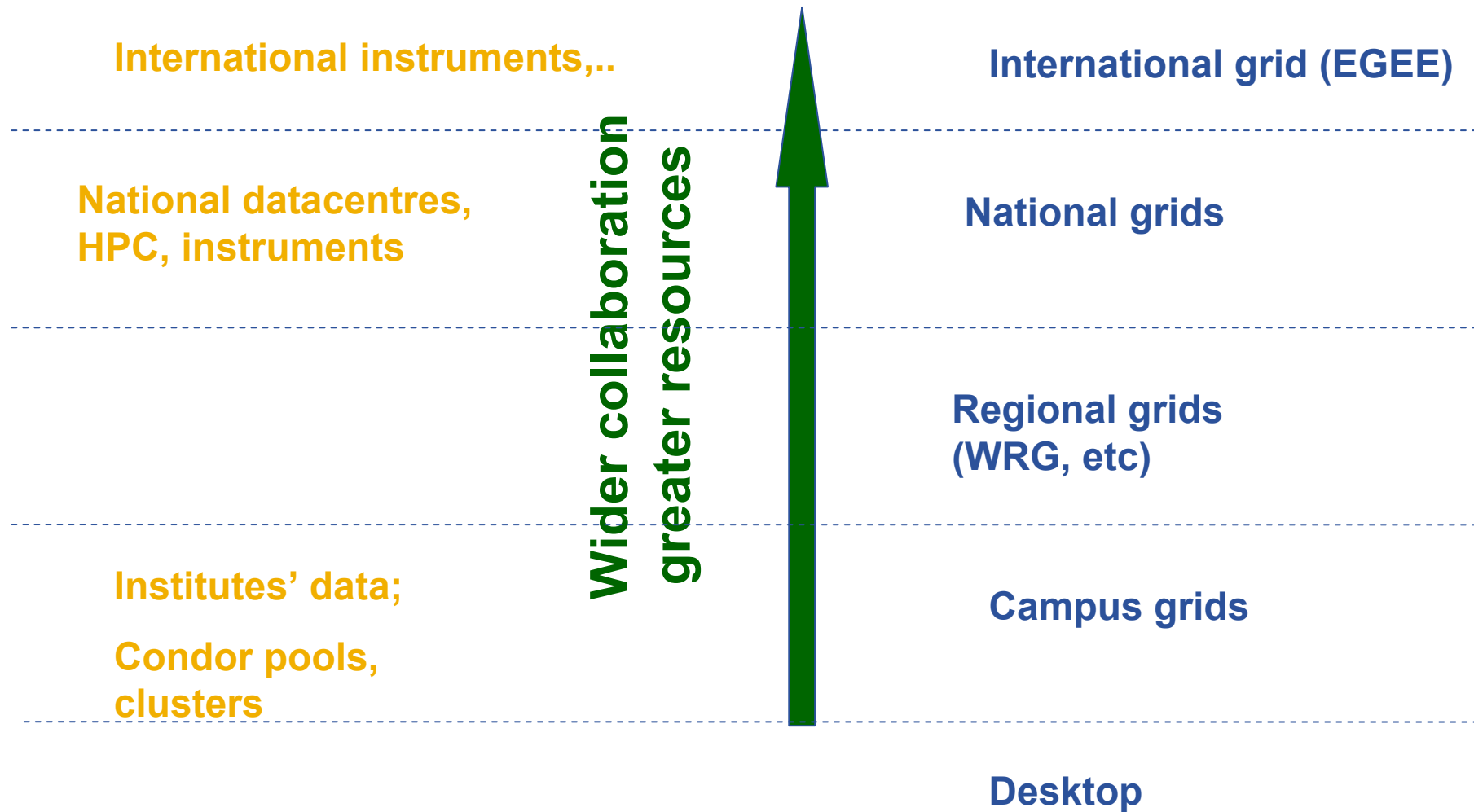
- **Flexible, simplified orchestration of resources available to a collaboration**
  - Across administrative domains
  - Abstractions hide detail of individual resources
    - Conform to Grid’s procedures to gain benefit
  - Operations services (people and software)
  
- **Increased utilisation**
  - A collaboration shares its resources building on Grid services
  - Collaborations share resources
    - Each contributes average requirements (cpus, storage)
    - Each can benefit from
      - *Heterogeneity*
      - *Scale*

- **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**
- **Each resource is exposed to the grid through an abstraction that masks heterogeneity, e.g.**
  - Multiple diverse computational platforms
  - Multiple data resources
- **Resources are usually owned by VO members.**
- **Negotiations lead to VOs sharing resources**
  - Usage organised at level of VO, not individuals

- **When using a PC or workstation you**
  - Login with a username and password (“Authentication”)
  - Use rights given to you (“Authorisation”)
  - Run jobs
  - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **Operating system**
- **One admin. domain**
- **When using a Grid you**
  - Login with digital credentials – single sign-on (“Authentication”)
  - Use rights given you (“Authorisation”)
  - Run jobs
  - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**



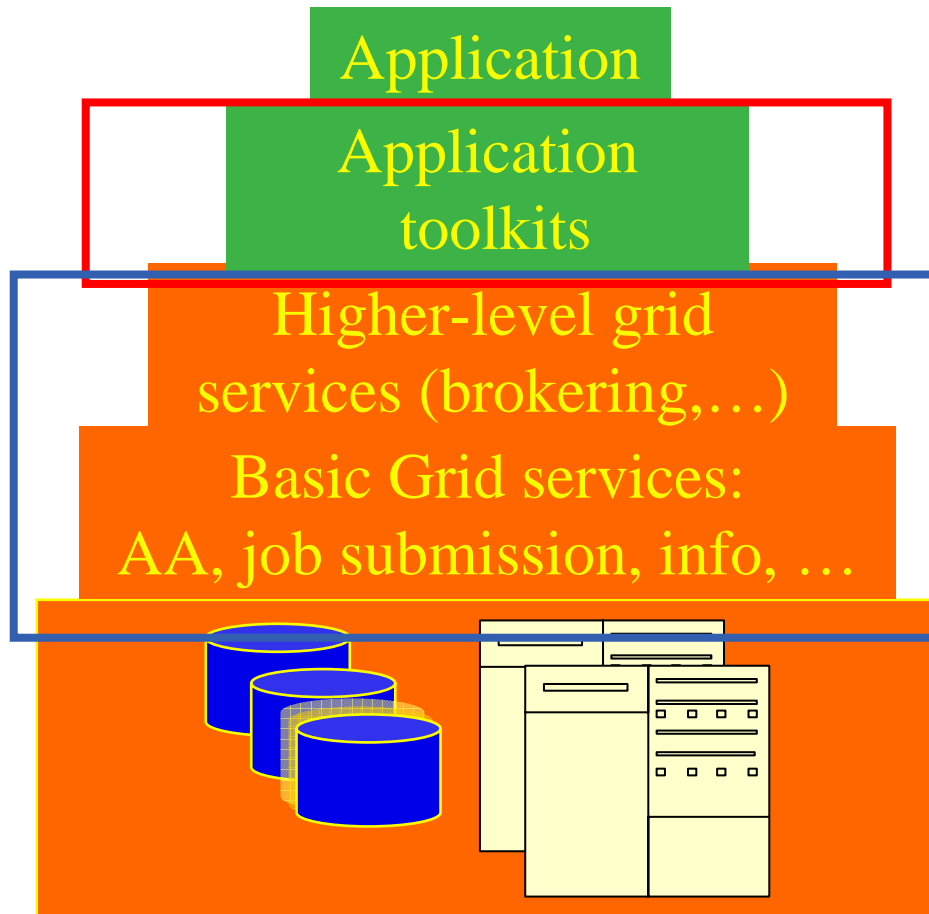
- **Providers of resources (computers, databases,...)**
  - need risks to be controlled:
  - they are asked to trust users they do not know
- **User's need**
  - single sign-on: to be able to logon to a machine that can pass the user's identity to other resources
  - To trust owners of the resources 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**



- **I need resources for my research**
  - I need richer functionality
    - MPI, parametric sweeps,...
    - Data and compute services together...
  
- **I provide an application for (y)our research**
  - How!?
    - Pre-install executables ?
    - Hosting environment?
    - Share data
    - Use it via portal?
  
- **We provide applications for (y)our research**
  - Also need:
    - Coordination of development
    - Standards
    - ...



**Engineering challenges increasing**

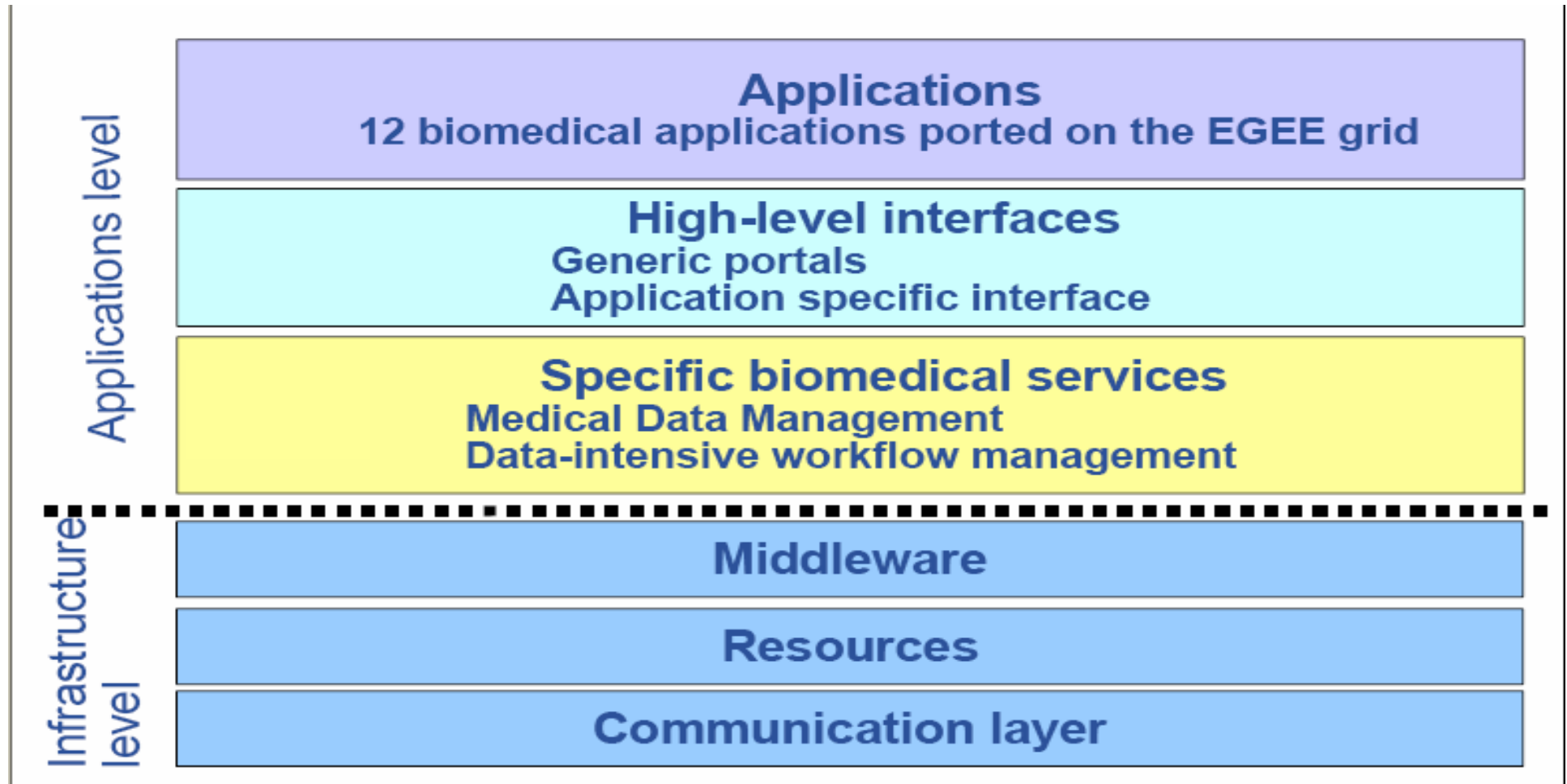


**Where computer science meets the application communities!**

**High level tools and VO-specific developments:**

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

**Production grids provide these services.**

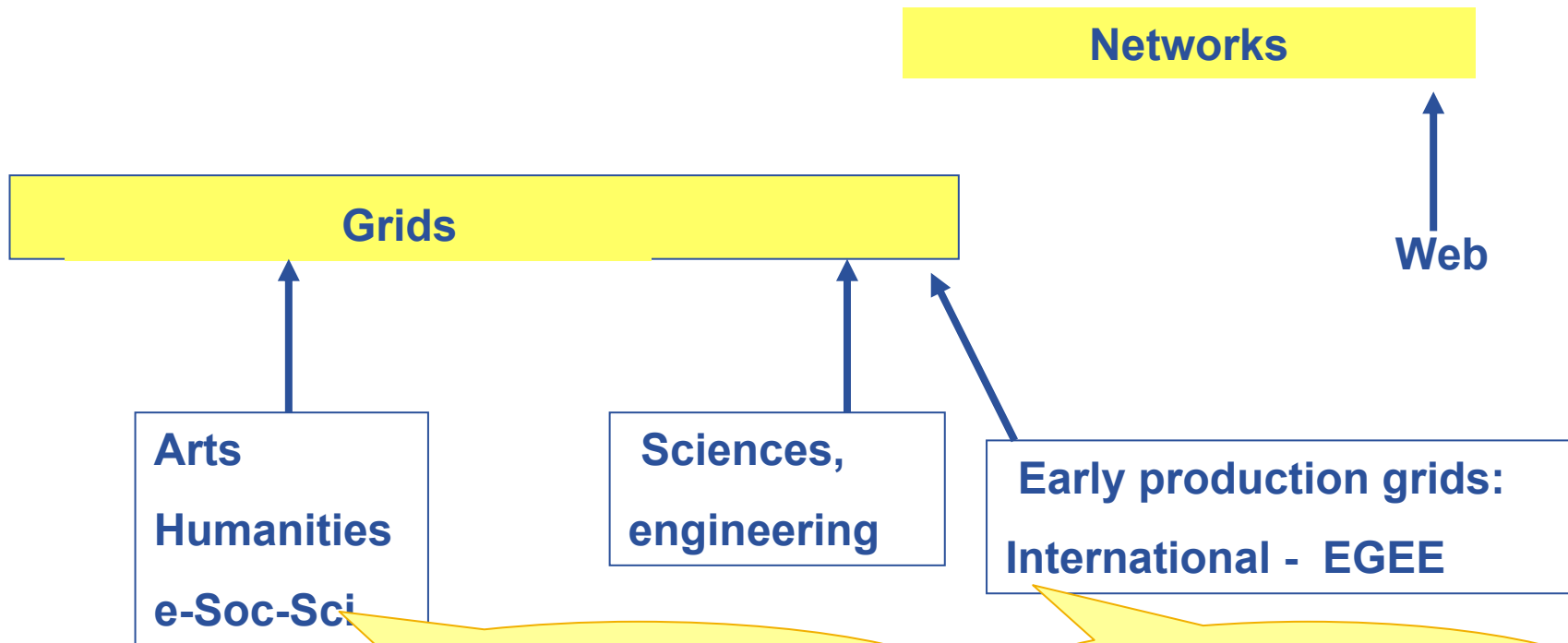


*Biomedical community and the Grid, EGEE User Forum, March 1<sup>st</sup> 2006, I. Magnin*



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

... then where are  
we now?



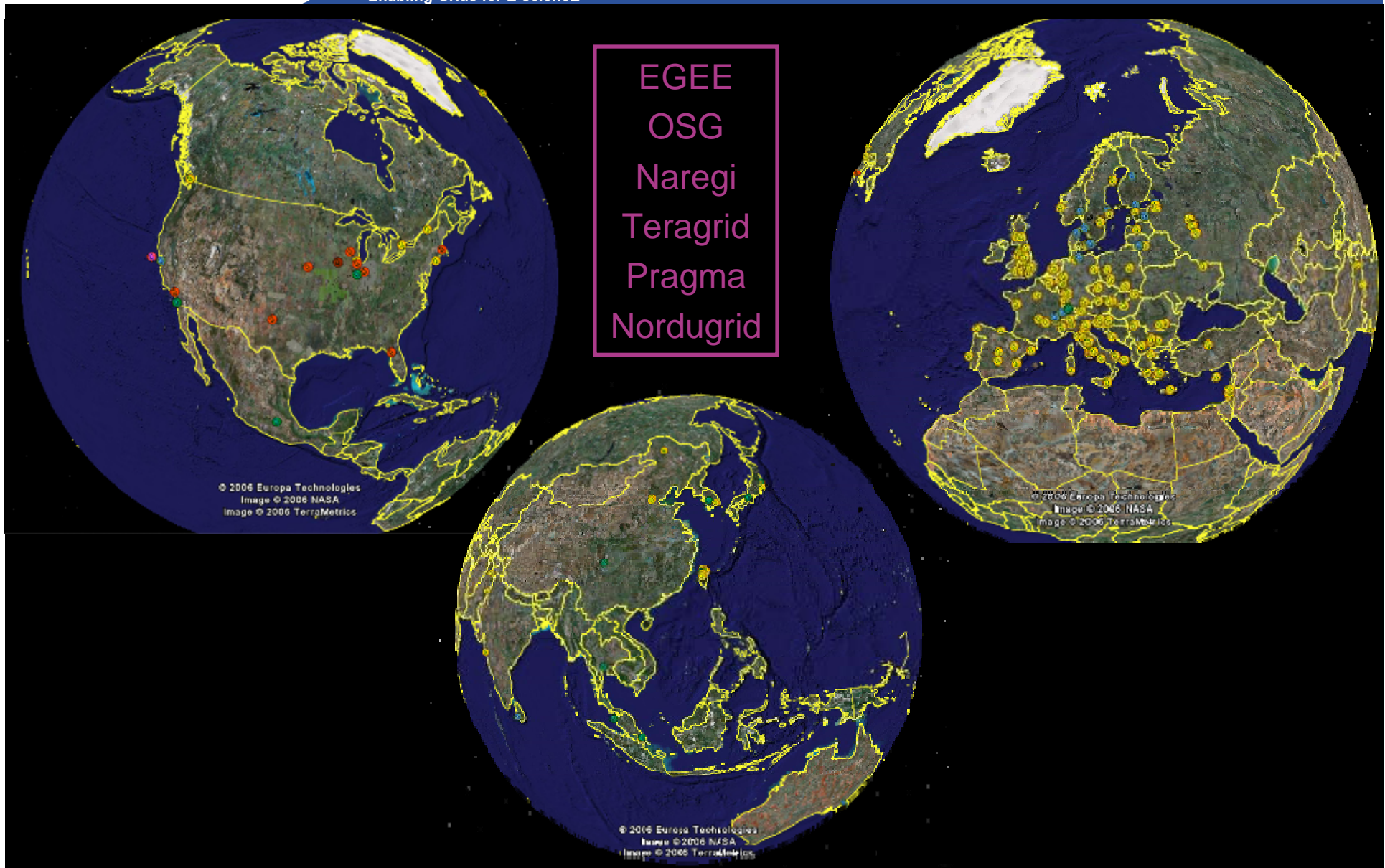
Types of use:

Service-oriented, workflow, “legacy” data

High throughput, new data

- Many key concepts identified and known
- Many grid projects have tested, and benefit from, these
  - Empowering collaborations
  - Resource-sharing
- 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. Open (formerly Global) Grid Forum, <http://www.gridforum.org/>
    - Extending web services
  - **Broadening range of research communities**
    - arts and humanities, social science ...



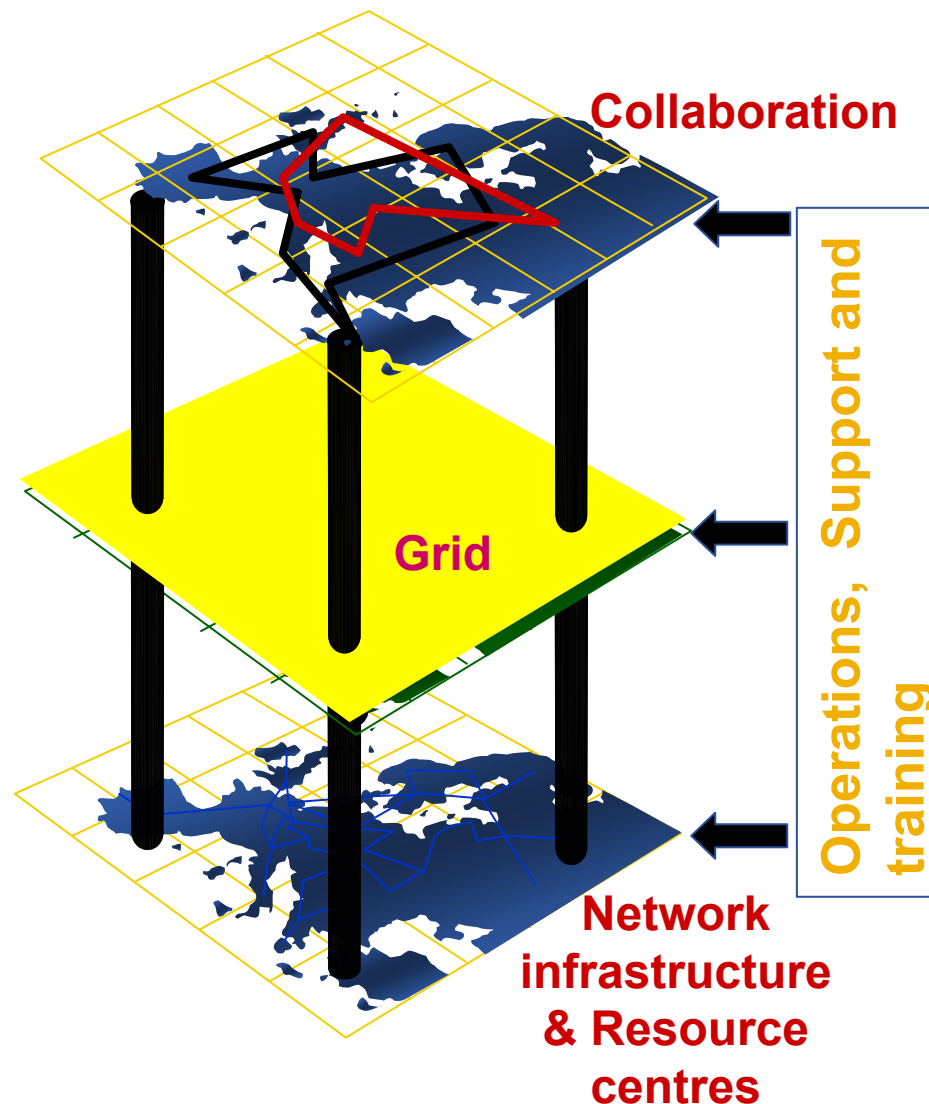


- To obtain a Google map of the Grids in the Globus Interoperability Now (GIN) initiative go to:

<http://www.pparc.ac.uk/Nw/GIN.asp>

- (You will need to install GoogleEarth)

- **Grids enable virtual computing across administrative domains**
  - Resources share authorisation and authentication
  - Resources accessed thru abstractions
- **Motivations:**
  - Collaborative research, diagnostics, engineering, public service,..
  - Resource utilisation and sharing



- The End