

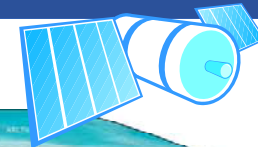
What is Grid Computing?

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

mjm@nesc.ac.uk

With thanks to EGEE colleagues for many of these slides

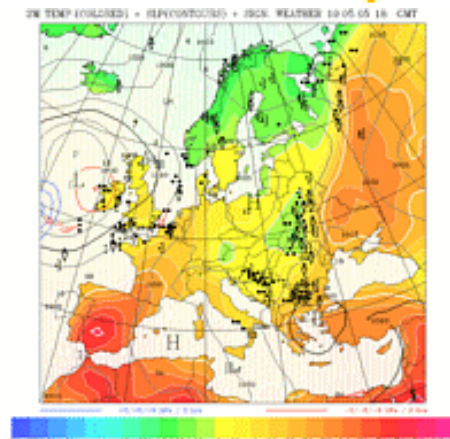
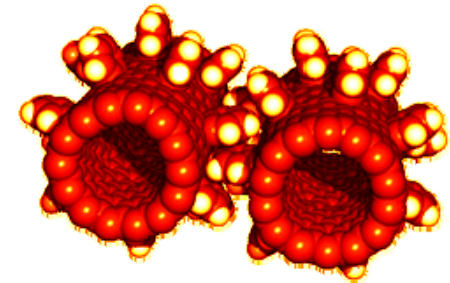
- **Introduction to**
 - e-Infrastructure
 - e-Research and e-Science
- **Some examples from the EGEE project**
 - EGEE: Enabling Grids for E-scienceE
 - EGEE is an EU-funded project that is running the largest international Grid
- **Grid concepts**
- **Grids - Where are we now?**





**Grid Computing “equates to the world's largest computer”
.... dynamic, flexible, distributed computing**

- **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



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

John Taylor

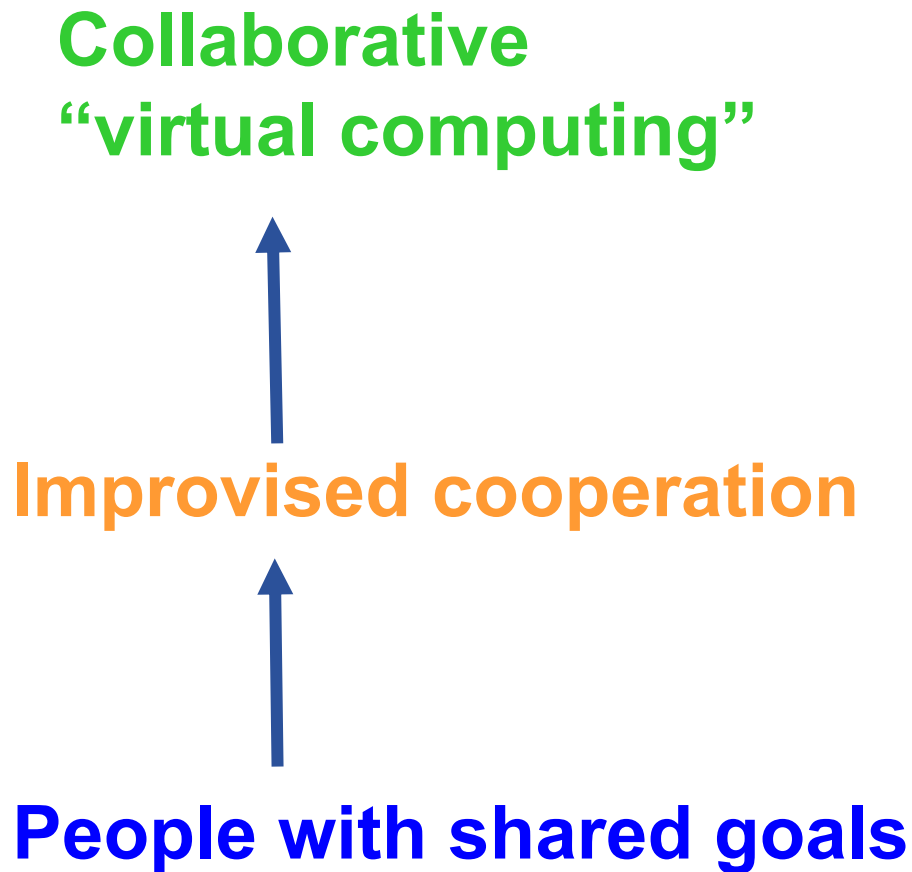
Director General of Research Councils

Office of Science and Technology

UK

e-Infrastructure = Networks + Grids

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



Sharing data, computers, software
Enabled by Grids:

National: BG

International: EGEE grid

Email

File exchange

ssh access to run programs

Enabled by networks:

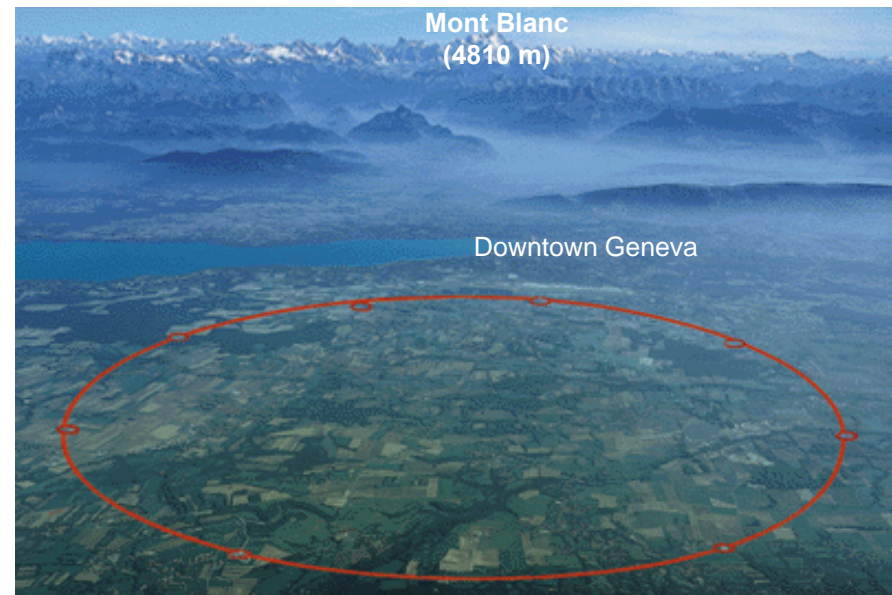
national, regional and

International: GEANT

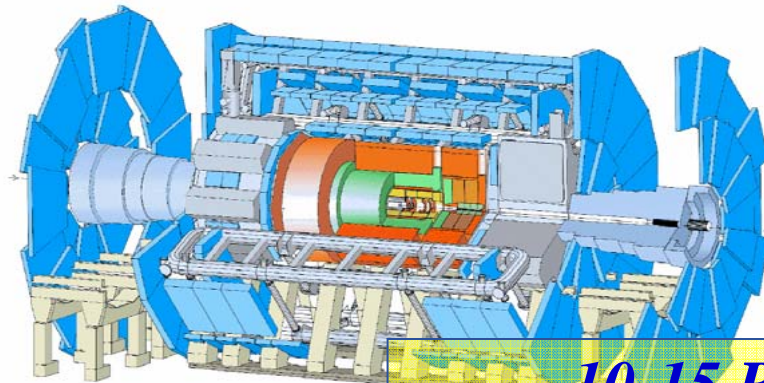
- **Collaborative research that is made possible by the sharing across the Internet of resources (data, instruments, computation, people’s expertise...)**
 - Crosses organisational boundaries
 - Often very compute intensive
 - Often very data intensive
 - Sometimes large-scale collaboration
- **Early 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

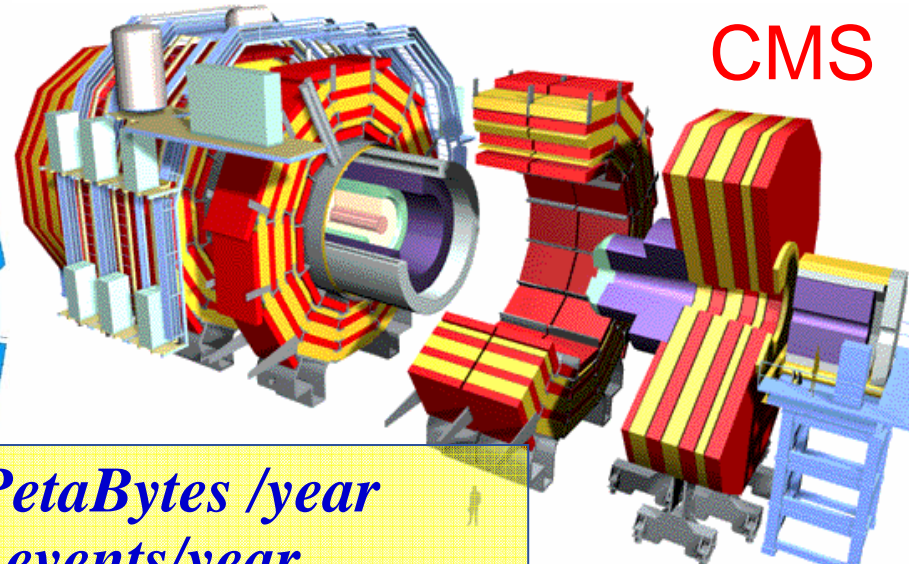
- Large Hadron Collider (LHC) at CERN in Geneva Switzerland:
 - One of the most powerful instruments ever built to investigate matter



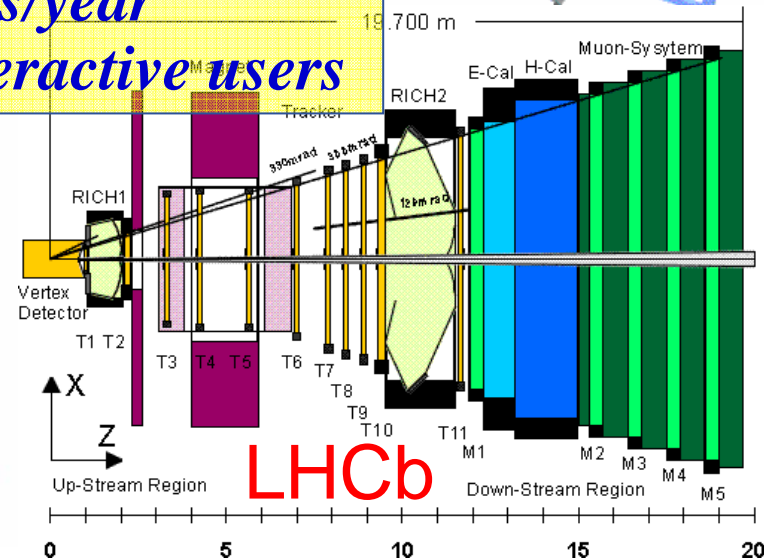
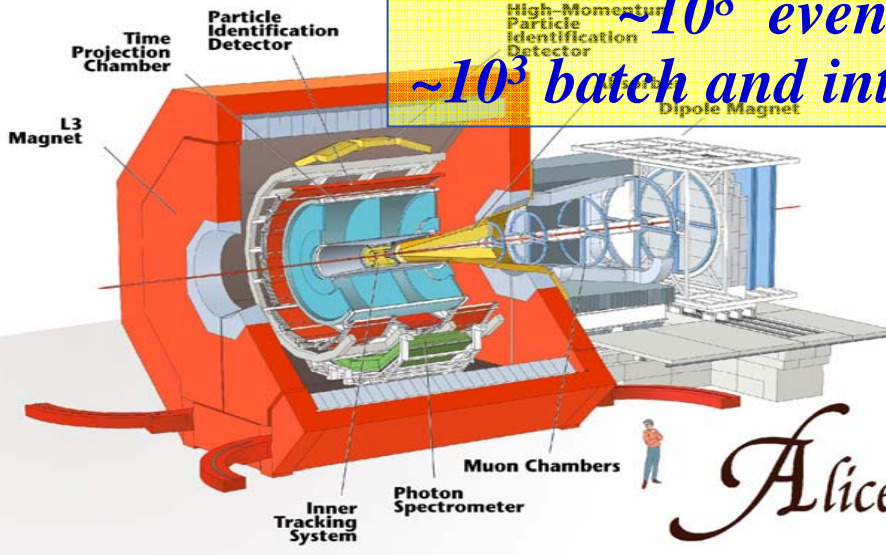
ATLAS



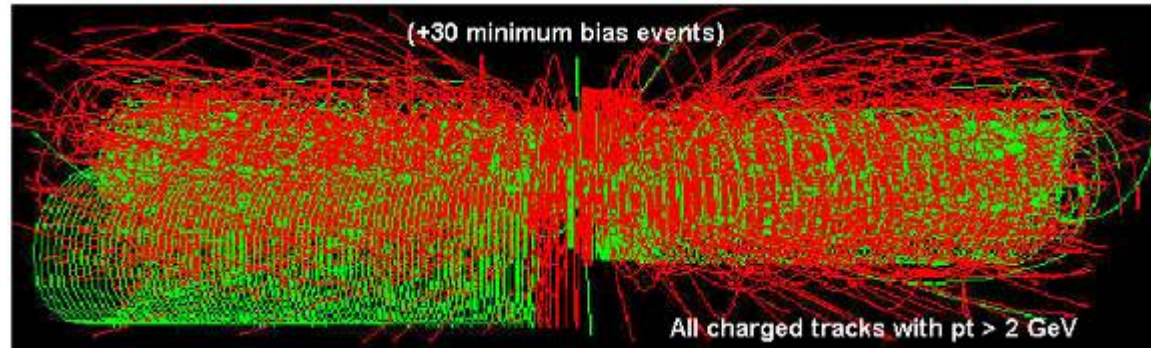
CMS



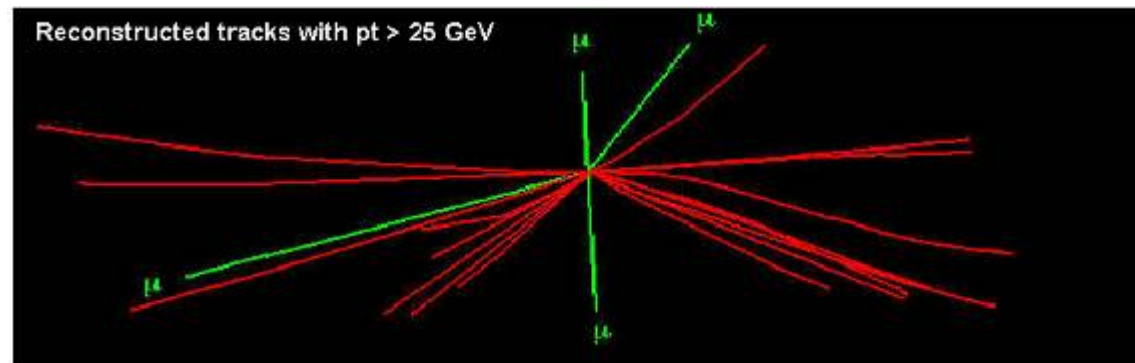
~10-15 PetaBytes /year
~10⁸ events/year
~10³ batch and interactive users



Starting from
this event



Looking for
this “signature”



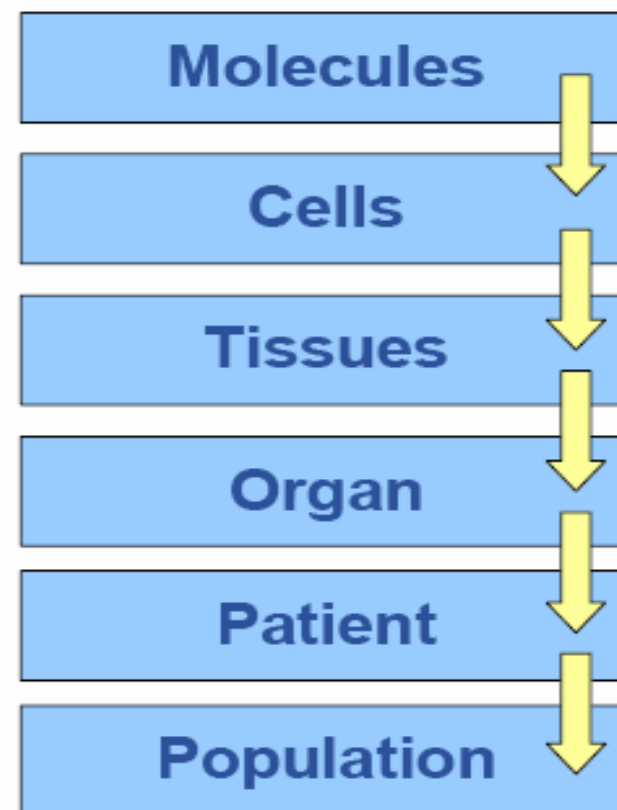
→ **Selectivity: 1 in 10^{13}**

(Like looking for a needle in 20 million haystacks)

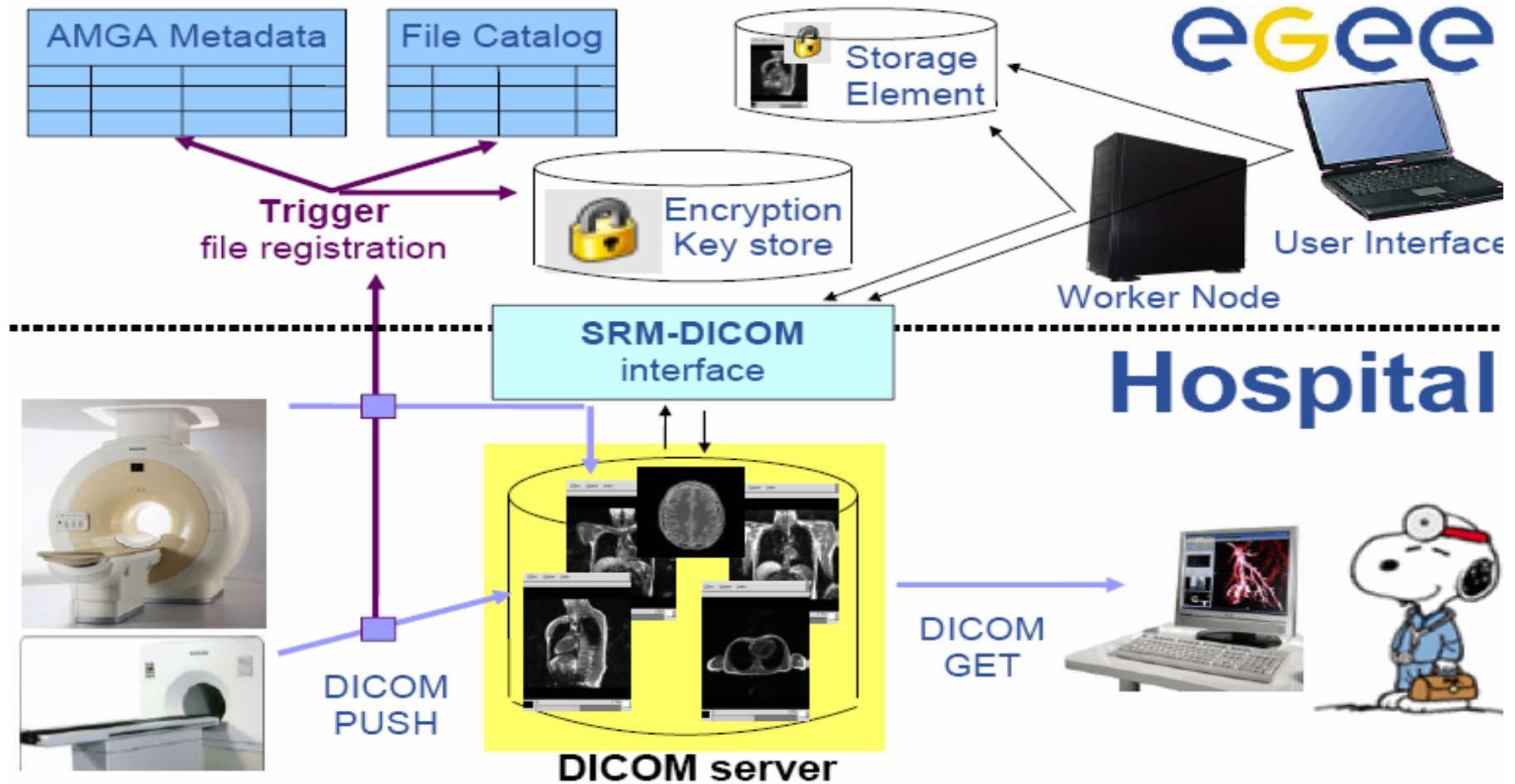
- **Bioinformatics**
 - Genomics
 - Proteomics
 - Phylogeny...

- **Medical imaging**
 - Medical imaging
 - Computer Aided Diagnosis
 - Therapy planning
 - Simulation...

- **Life sciences**
 - Drug discovery
 - Epidemiology
 - ...

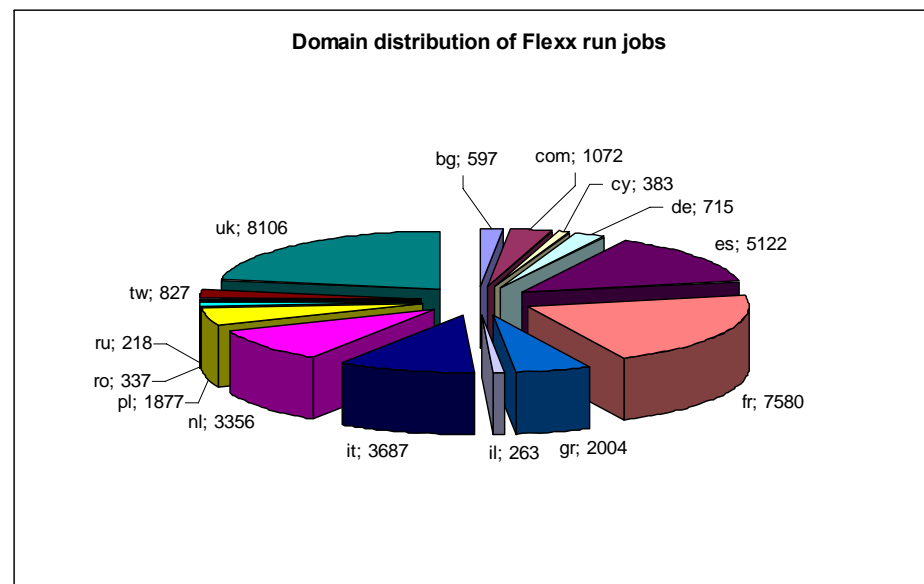


Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin



Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin

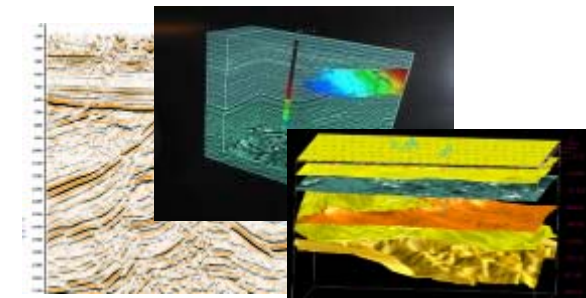
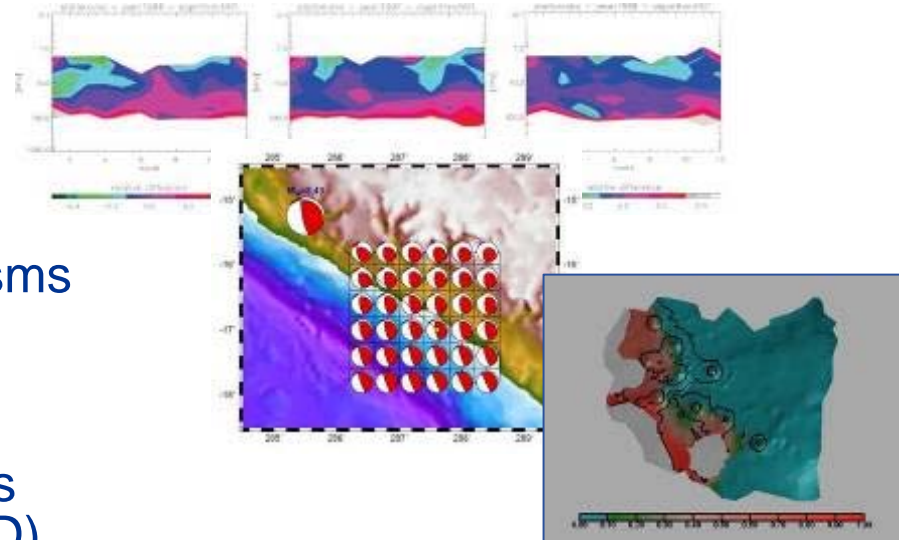
- **Significant biological parameters**
 - two different molecular docking applications (Autodock and FlexX)
 - about one million virtual ligands selected
 - target proteins from the parasite responsible for malaria
- **Significant numbers**
 - Total of about 46 million ligands docked in 6 weeks
 - 1TB 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



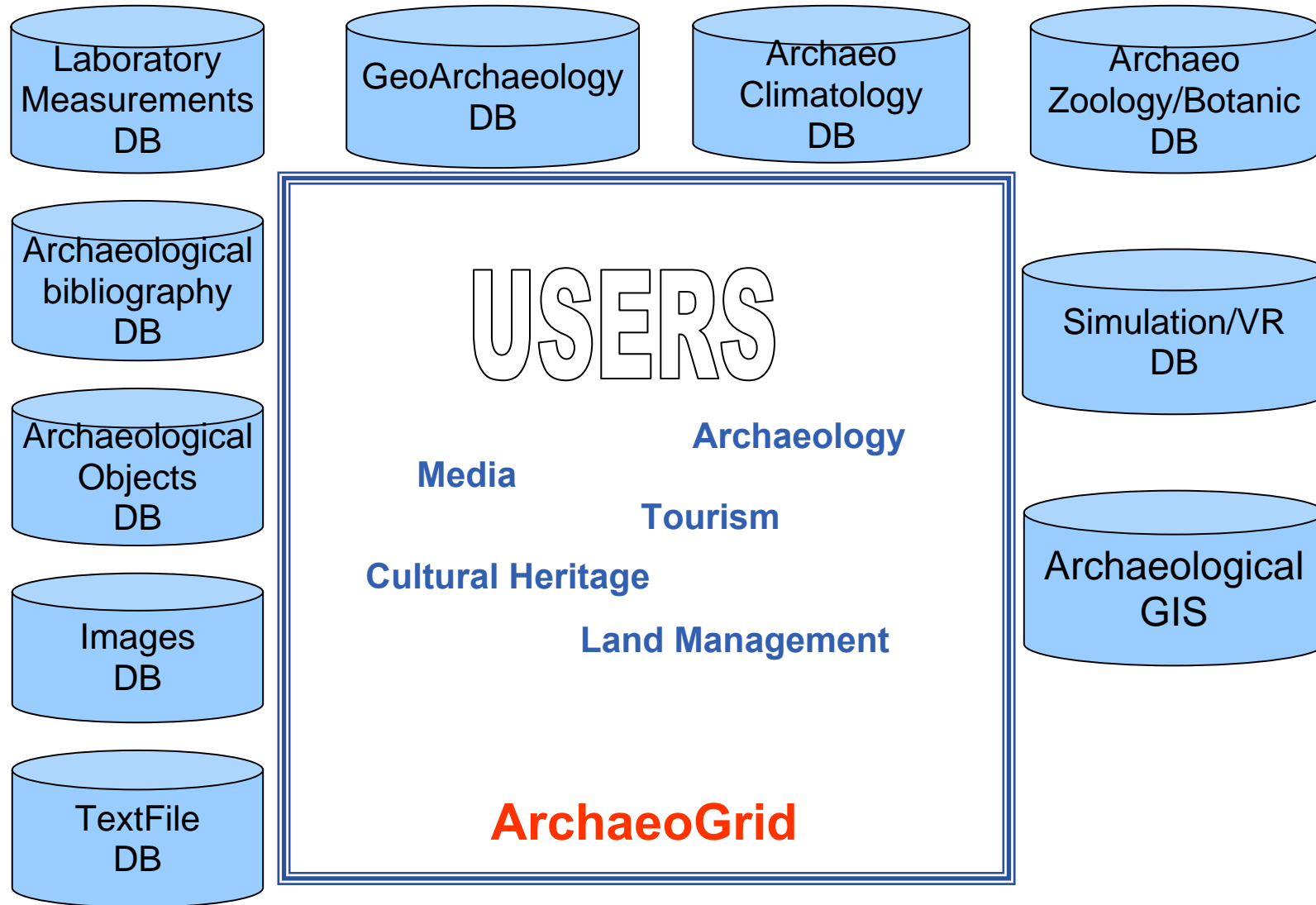
New data challenge in the fall of 2006
 New malaria targets
 Focus on other neglected diseases
 Enlarged collaboration
 (possibly including related projects)

Roberto Barbera, 1st EGEE User Forum, CERN, 1st March 2006

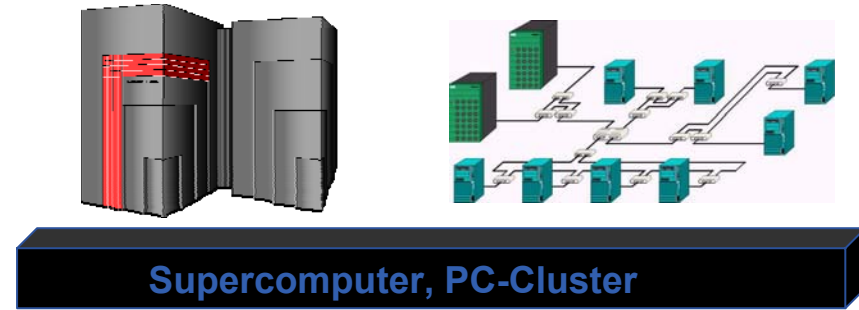
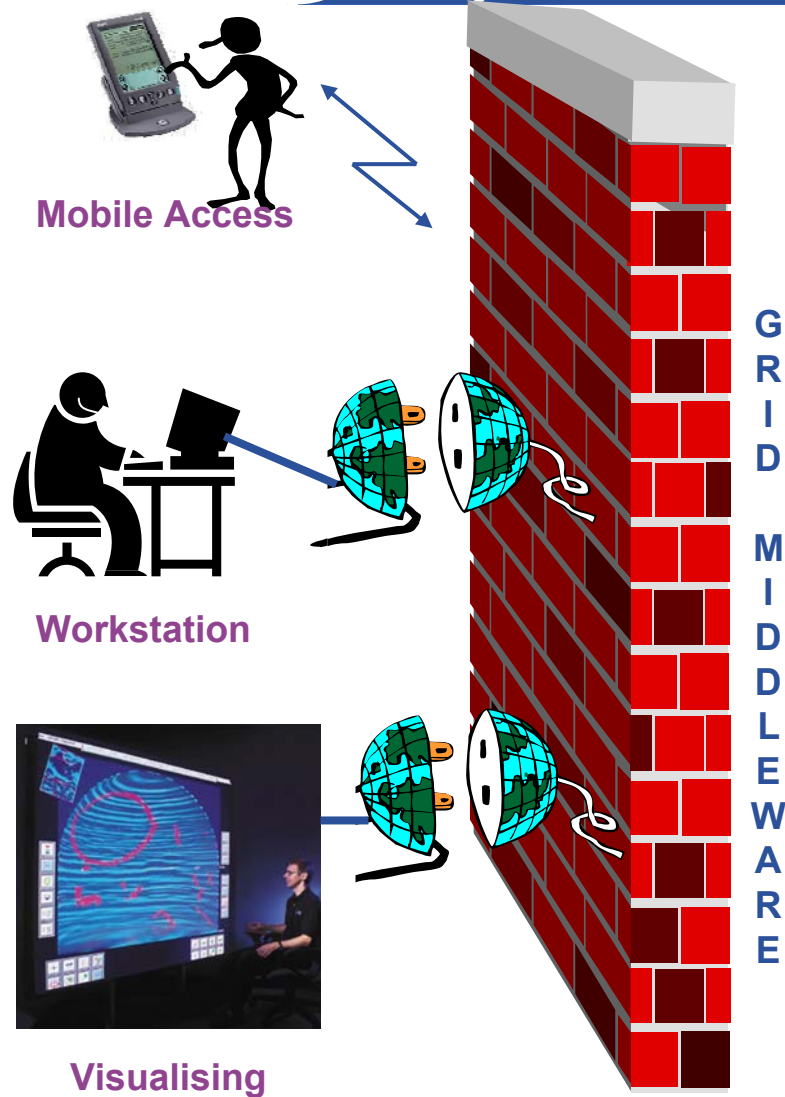
- **Earth Observations by Satellite**
 - Ozone profiles
- **Solid Earth Physics**
 - Fast Determination of mechanisms of important earthquakes
- **Hydrology**
 - Management of water resources in Mediterranean area (SWIMED)
- **Geology**
 - Geocluster: R&D initiative of the Compagnie Générale de Géophysique



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

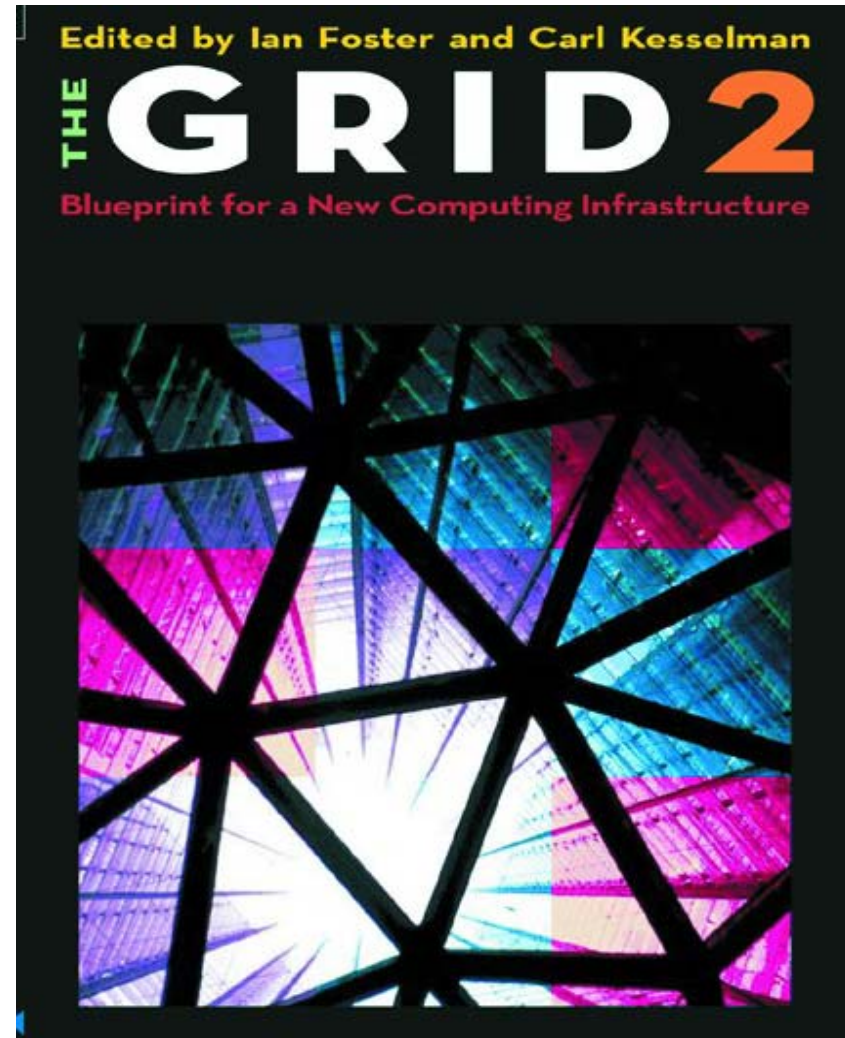


Grid concepts



- 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,...



Electrical power Grid	To Grid
<p>The use of electricity is done without worrying about where the electricity you are using comes from and how it is produced. The power grid is transparent.</p>	<p>The user of Grid has access to computing power and storage capacity all over the world. The Grid will be transparent.</p>
<p>An infrastructure, which links together power plants of many different kinds with your home through transmission stations, power stations, transformers, power lines and so forth.</p>	<p>An infrastructure, which links together resources such as PCs, workstations, storage elements, sensors, etc and provides the mechanism needed to access them.</p>
<p>The power grid is pervasive: electricity is available essentially everywhere and you can simply access it through a standard wall socket.</p>	<p>The Grid will be pervasive: remote resources will be accessible from any platform (desktops, laptops, PDAs, mobile phones) through a Web portal.</p>
<p>The power grid is a utility: you ask for electricity and you get it. You also pay for what you get.</p>	<p>The Grid is a utility: you ask for computer power or storage capacity and you get it. For the time being, it is free!</p>

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

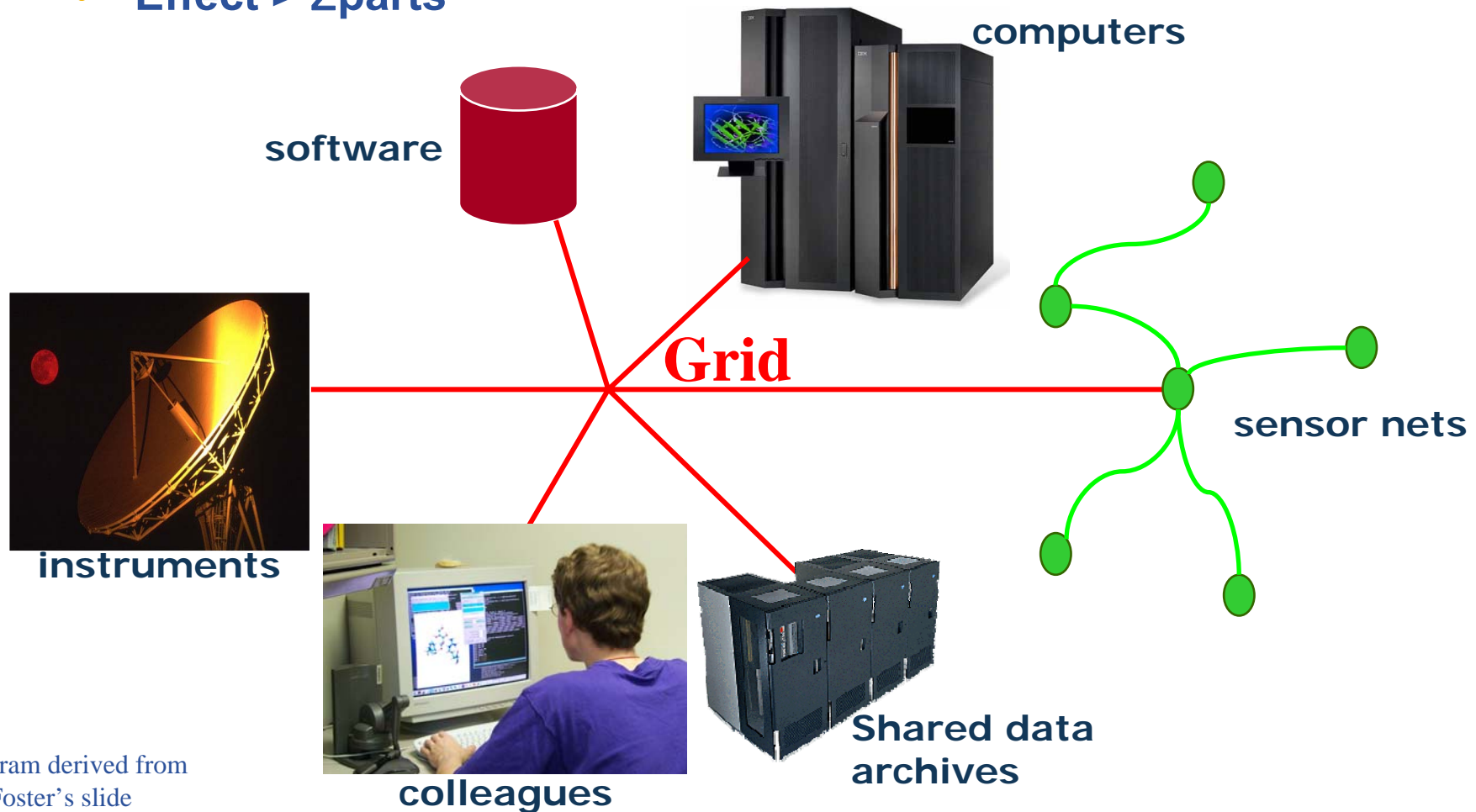


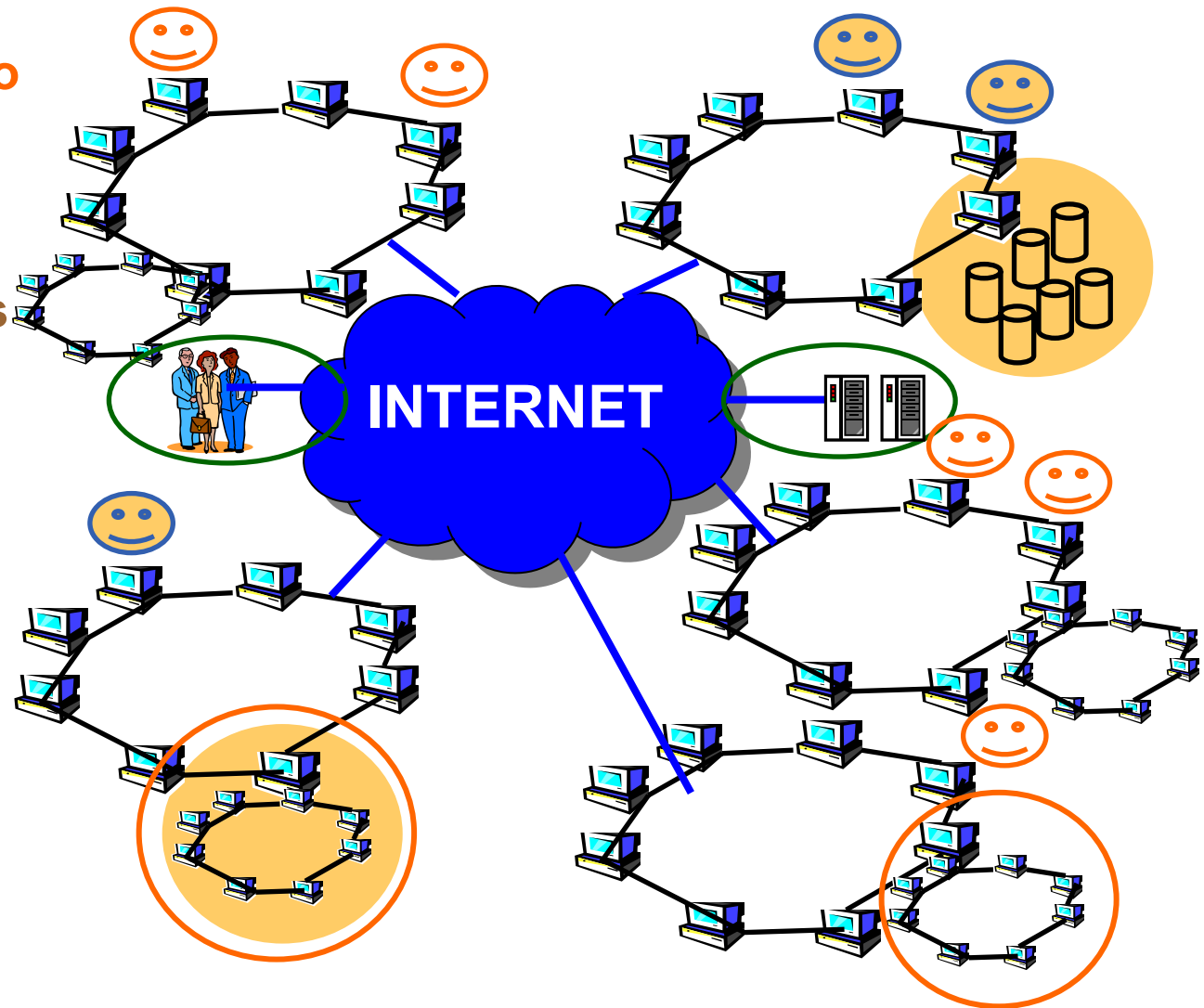
Diagram derived from
Ian Foster's slide

- **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*

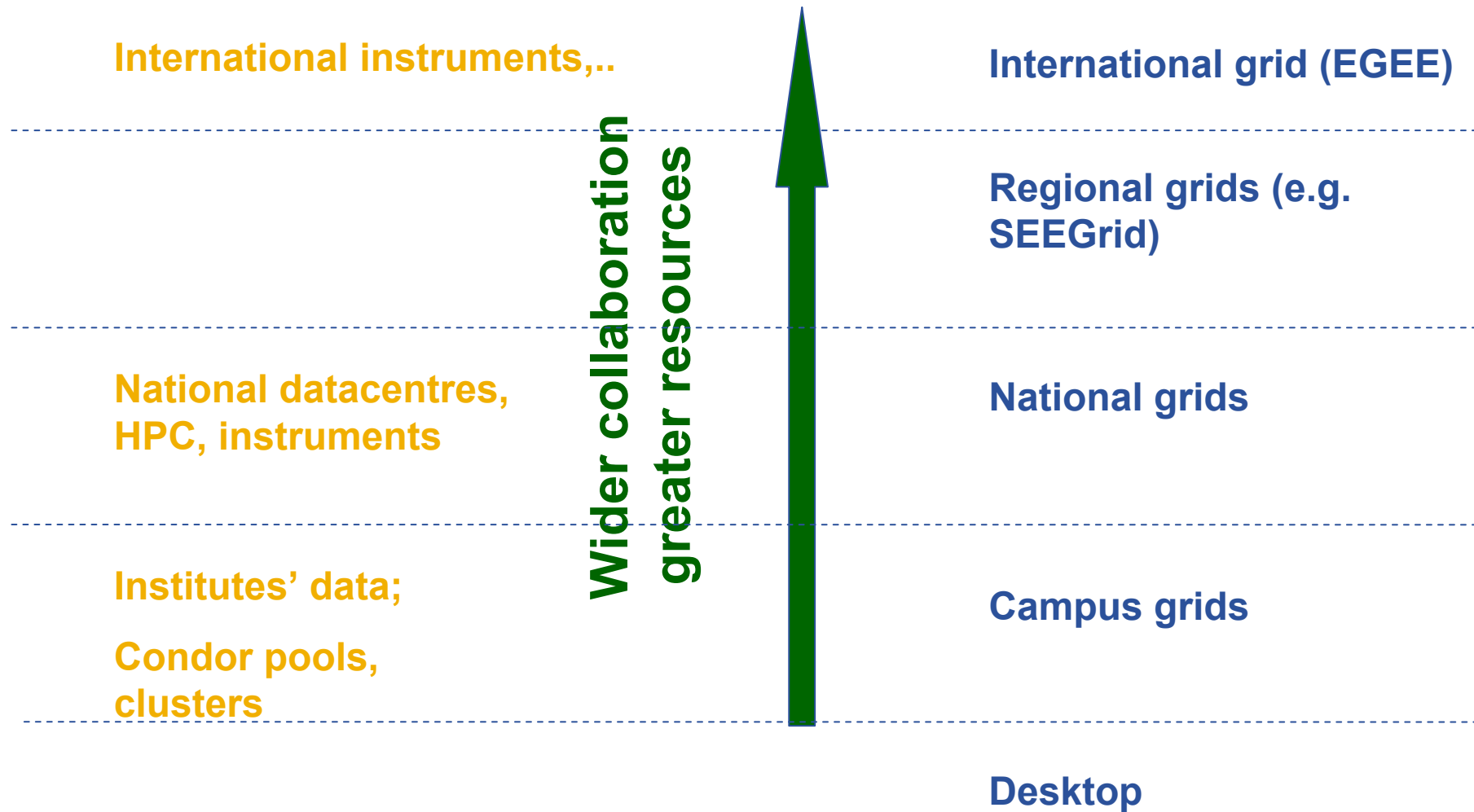
- **What is a Virtual Organisation?**
 - People in different organisations seeking to cooperate and share resources across their organisational boundaries
 - E.g. A research collaboration
- **Each grid is an infrastructure enabling one or more “virtual organisations” to share and access resources**
-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

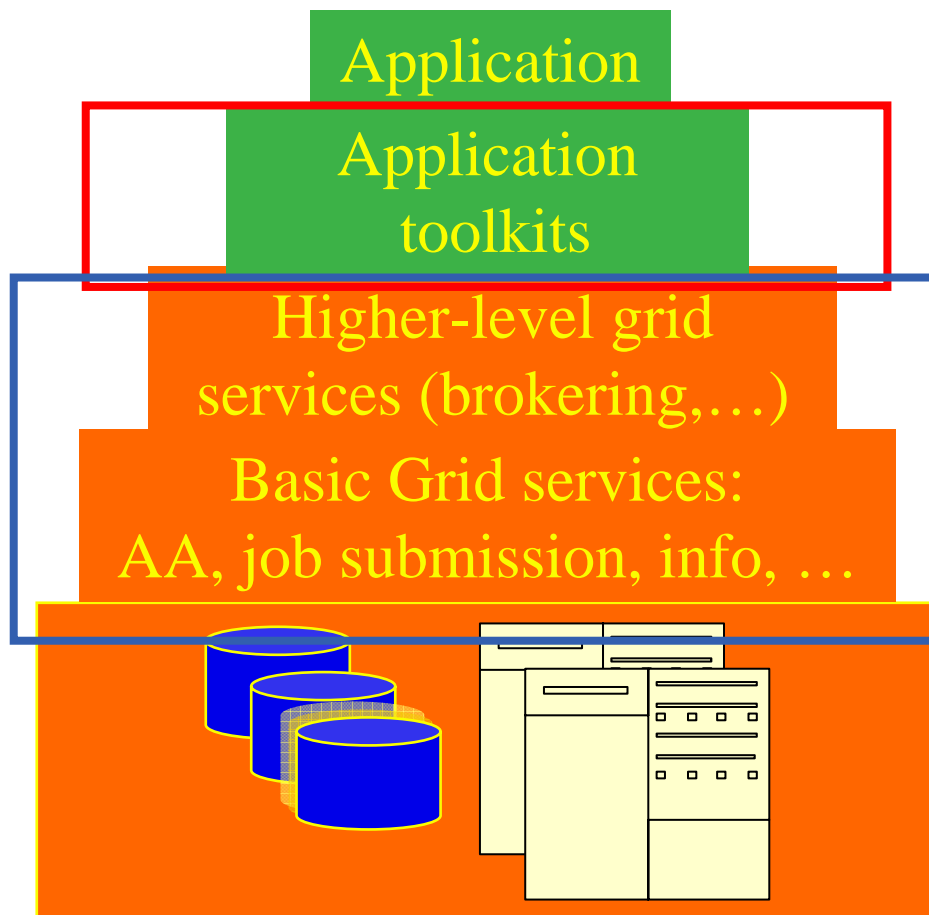
- **Virtual organisations negotiate with sites to agree access to resources**
- **Grid middleware runs on each shared resource to provide**
 - Data services
 - Computation services
 - Single sign-on
- **Distributed services (both people and middleware) enable the grid**



- **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**

The many scales of grids



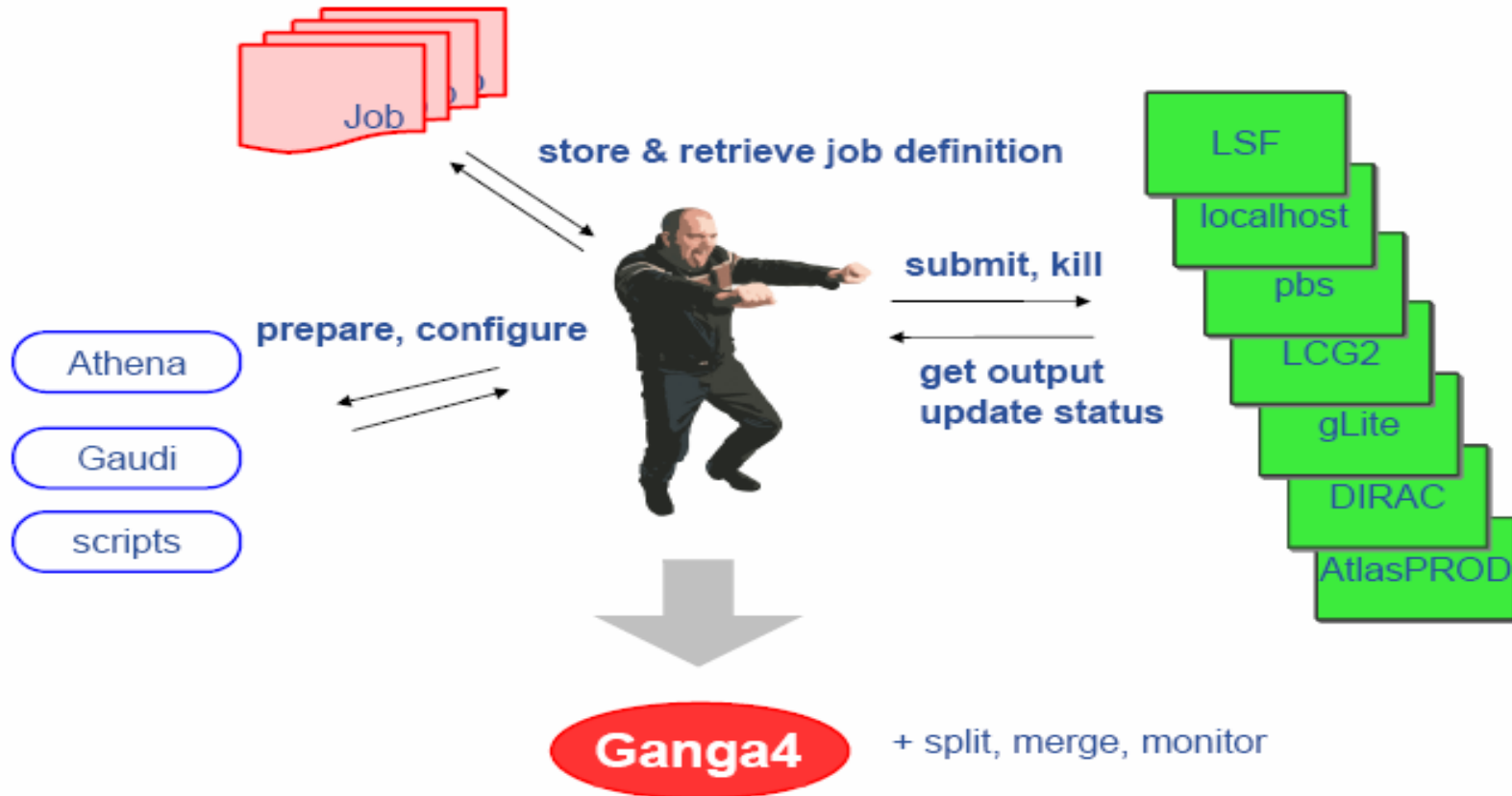


Where computer science meets the application communities!

High level tools and VO-specific developments:

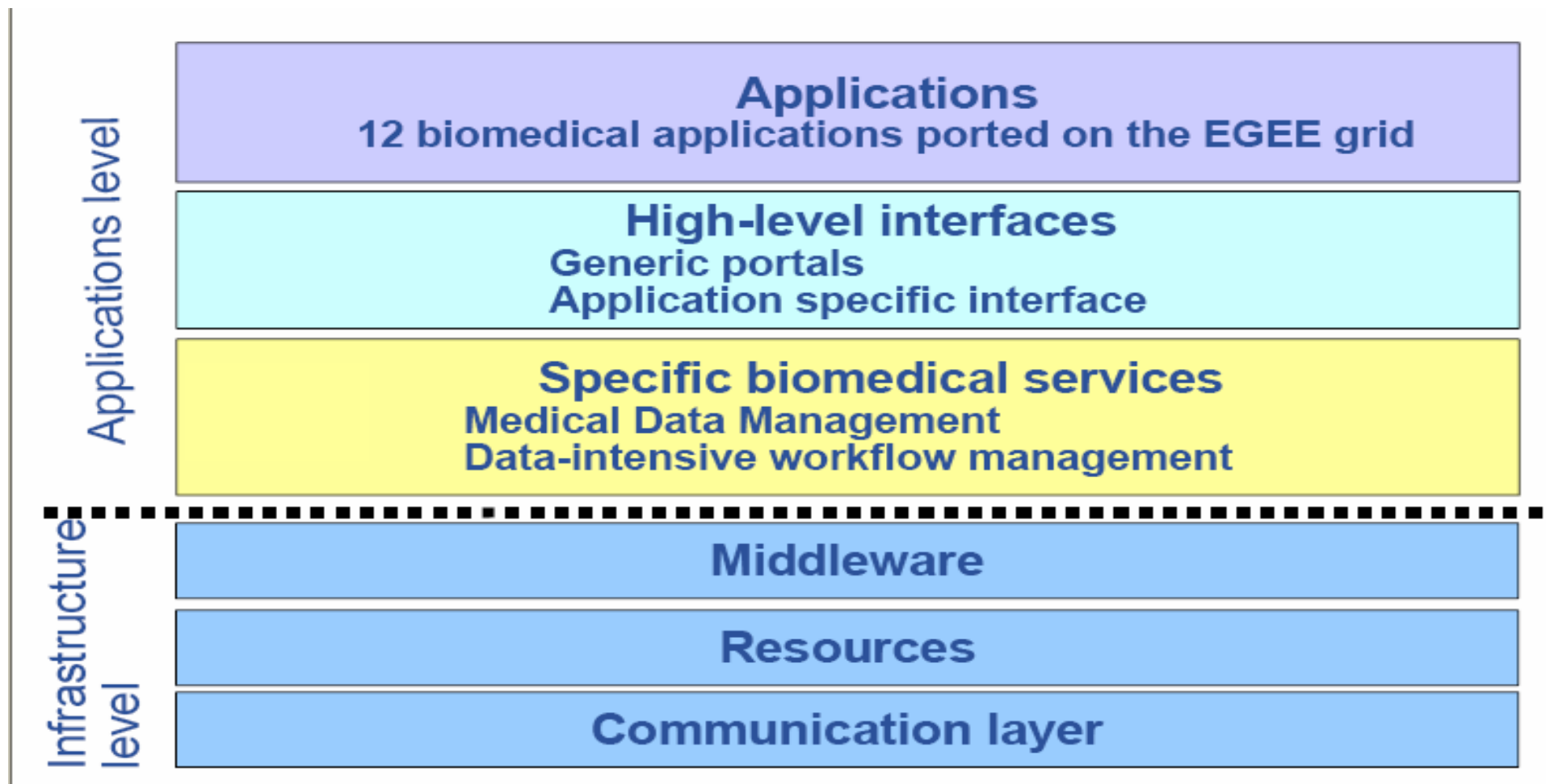
- Portals
- Virtual Research Environments
- Semantics, ontologies
- Workflow
- Registries of VO services

Production grids provide these services.

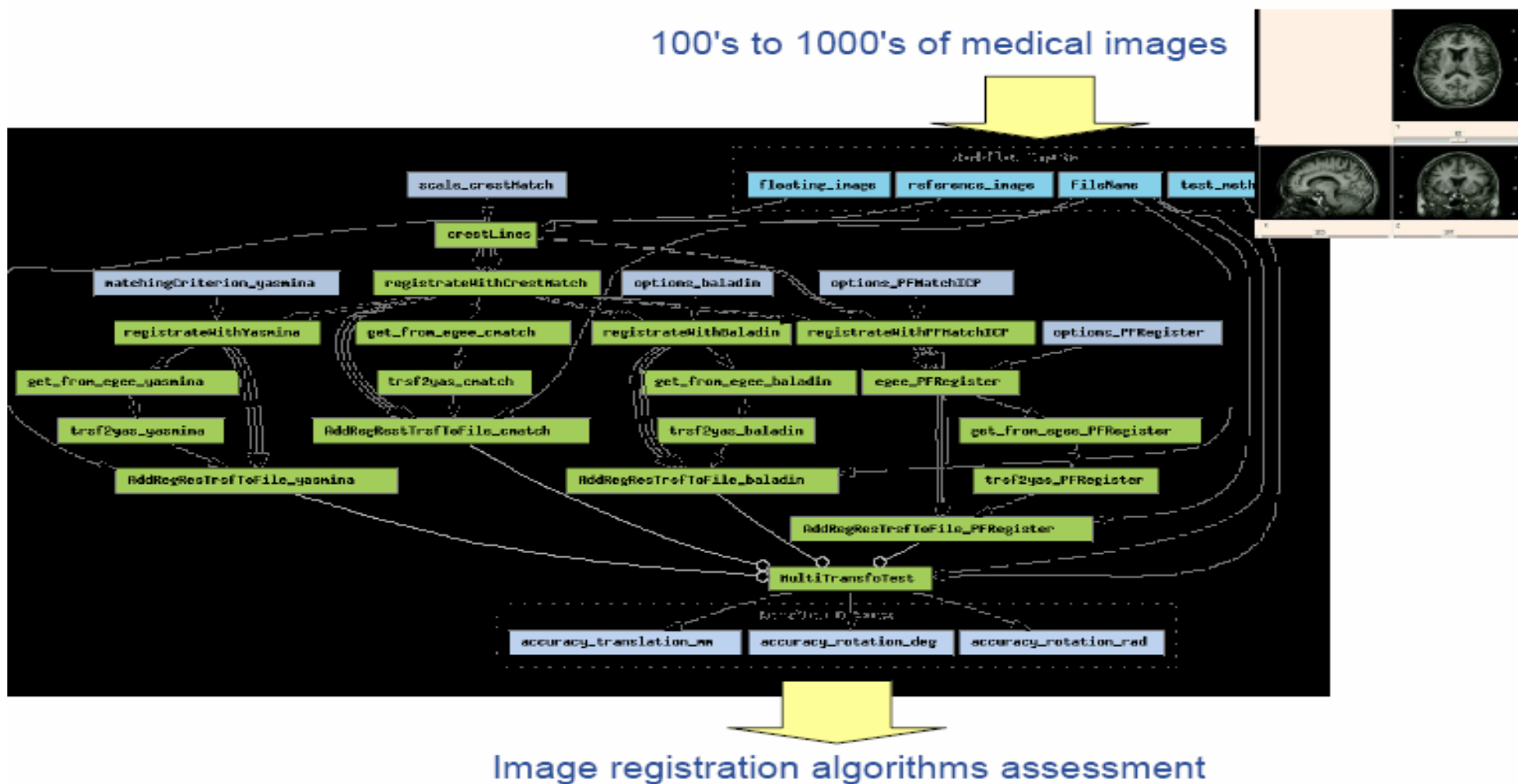


- **Ganga is a lightweight user tool**
ganga.web.cern.ch/
- **But also: Ganga is a developer framework**

Example – Biomedical applications



Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin



Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin



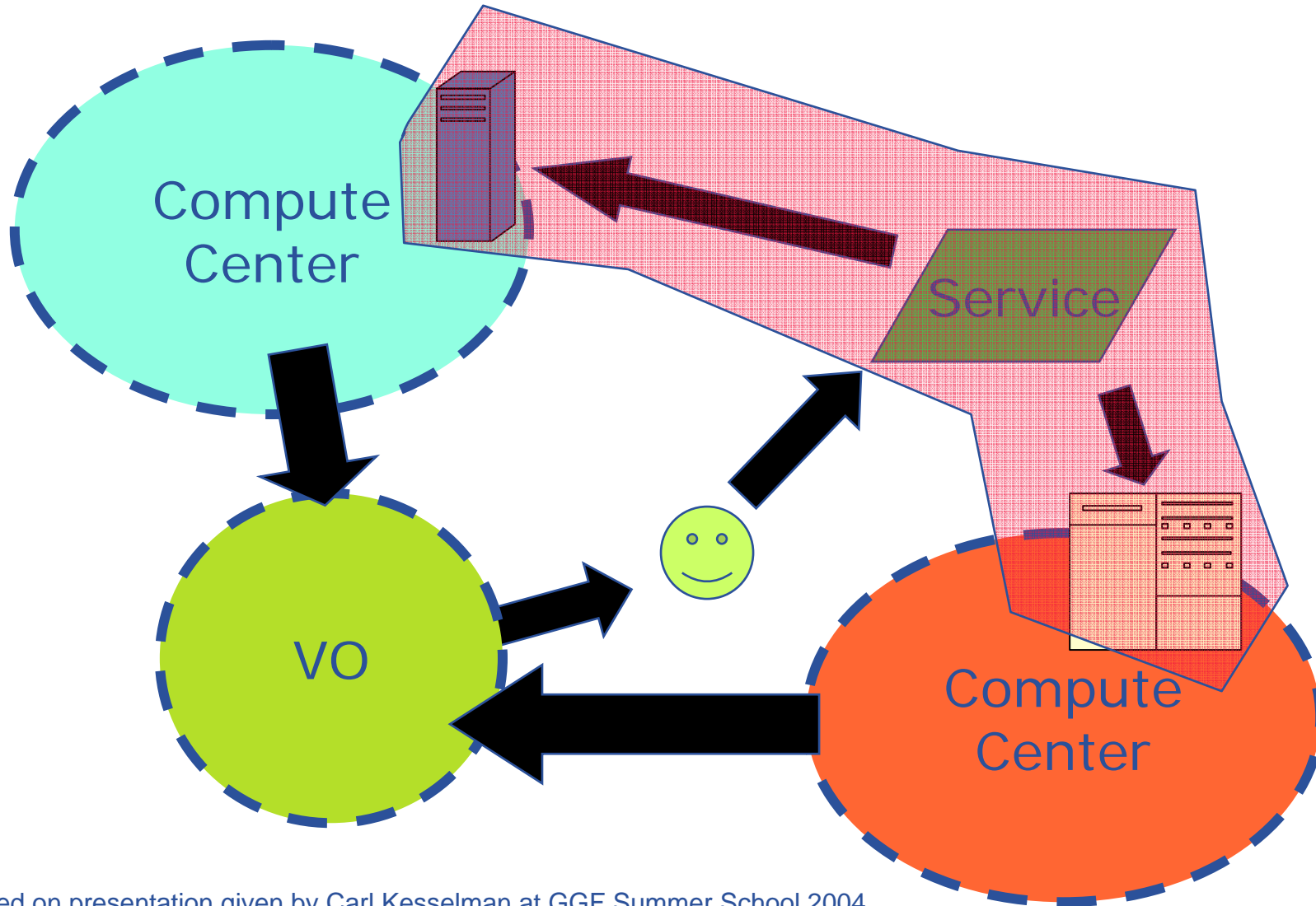
If "The Grid"
vision leads us
here...

... then where are
we now?

- Many key concepts identified and known
- Many grid projects have tested, and benefit from, these
 - 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 ...

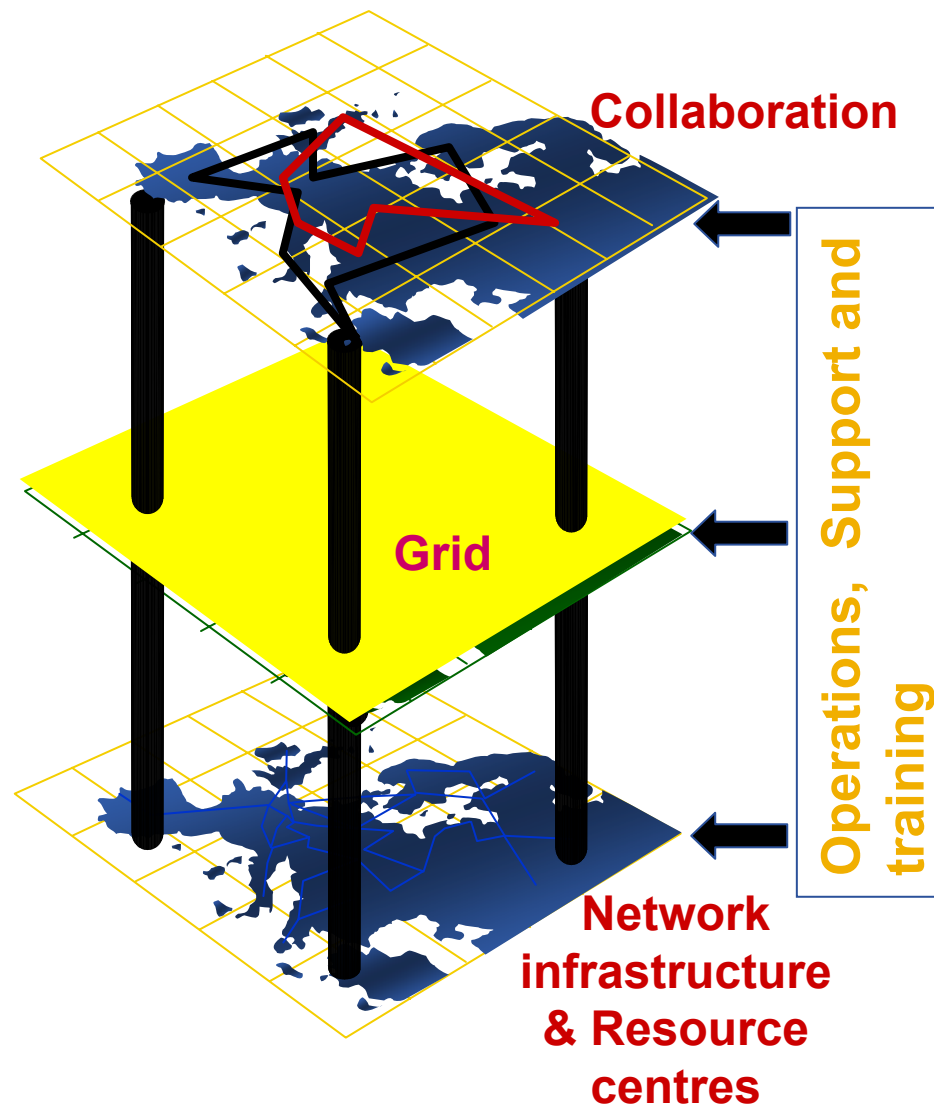


- **Providers of resources (computers, databases,...) need risks to be controlled: they are asked to trust users they do not know**
- **User's need**
 - single sign-on: to be able to logon to a machine that can pass the user's identity to other resources
 - To trust owners of the resources they are using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



slide based on presentation given by Carl Kesselman at GGF Summer School 2004

- **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



- **Open Grid Forum** <http://www.ggf.org/>
- **The Grid Cafe** www.gridcafe.org
- **Grid Today** <http://www.gridtoday.com/>
- **Globus Alliance** <http://www.globus.org/>