What is Grid Computing?

Mike Mineter
Training Outreach and Education
National e-Science Centre

mjm@nesc.ac.uk

You are welcome to re-use these slides. We ask only that you let us know, by email to training-support@nesc.ac.uk

National e-Science Centre

Contents

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

What is e-science about?

'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

e-Infrastructure

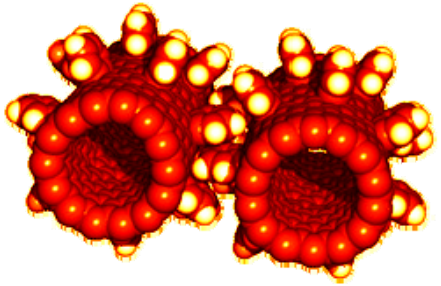
*'e-Science is about global collaboration in key areas of science, and the **next generation of infrastructure** that will enable it.'*

Networks + Grids

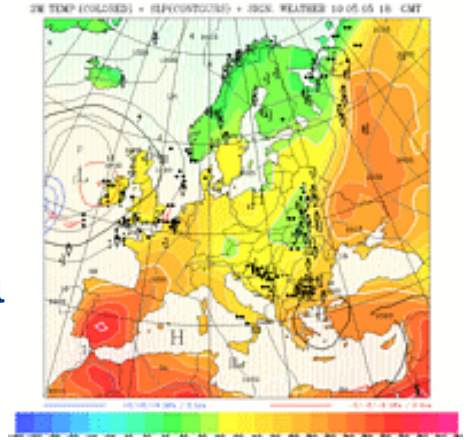
- *Networks connect resources*
- *Grids enable "virtual computing"*

Computing intensive science

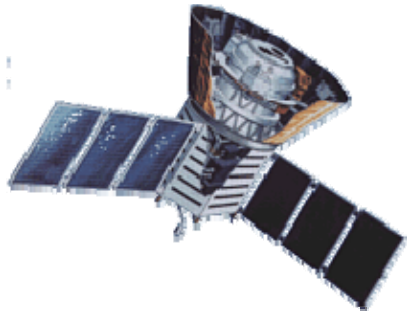
- **Science is becoming increasingly digital and needs to deal with increasing amounts of data**



- **Simulations get ever more detailed**
  - Nanotechnology – design of new materials from the molecular scale
  - Modelling and predicting complex systems (weather forecasting, river floods, earthquake)
  - Decoding the human genome
  - Cross multiple scales: environmental, biologica



- **Experimental Science uses ever more sophisticated sensors to make precise measurements**
  - → Huge amounts of data
  - → Serves user communities around the world

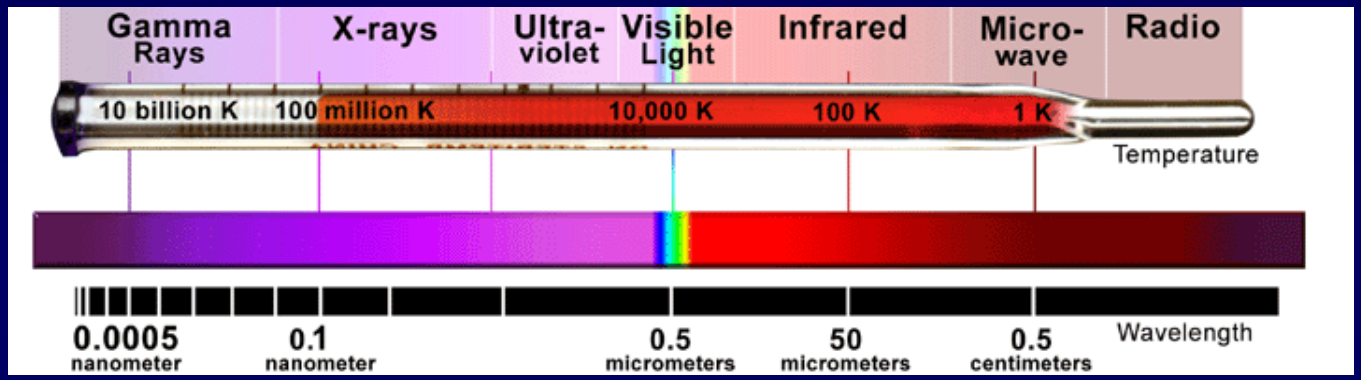


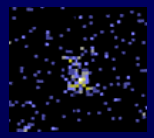
e-Science technologies support…

- **Flexible orchestration of**
  - Computers
  - Data storage
  - Programs
  - Research data
- **Now allow**
  - Collaborative investigation of complex systems
- **Increasing focus within different roles**
  - Researcher: uses data, creates models,……
  - “Service provider” : services take researcher requirements and “provision” these by use of shared resources
  - Resource provider: shared computers, stores, data services,….
  - Compute cycles, storage start to be seen as commodities

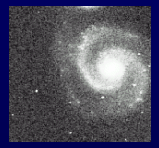
Virtual Observatories

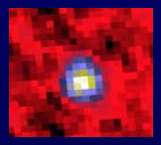
Observations made across entire electromagnetic spectrum

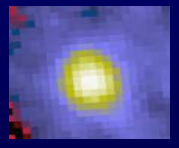


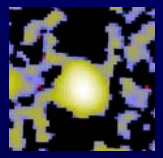

ROSAT ~keV

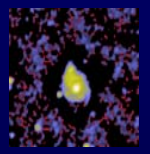

DSS Optical

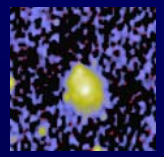

2MASS 2μ


IRAS 25μ


IRAS 100μ


GB 6cm


NVSS 20cm


WENSS 92cm

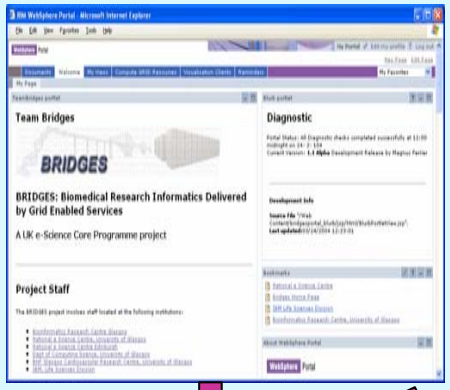
⇒e.g. different views of a local galaxy

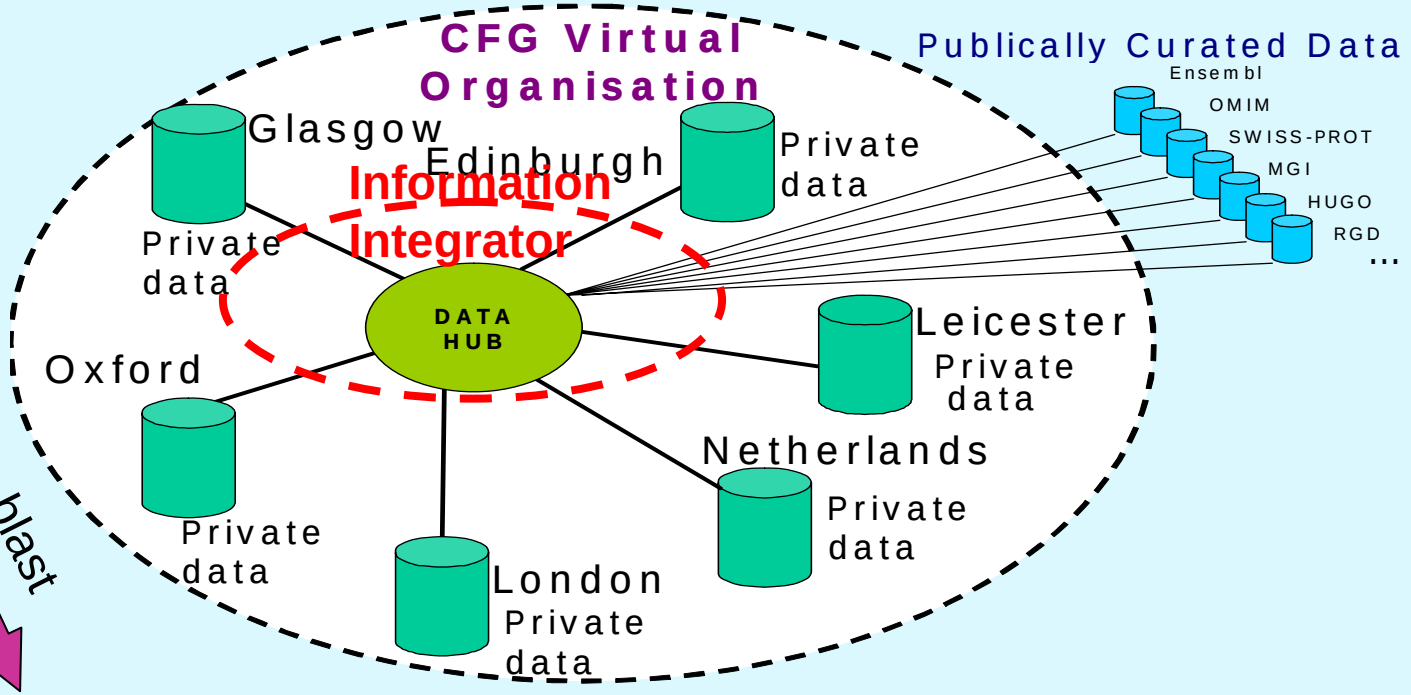
Need all of them to understand physics fully

Databases are located throughout the world

Biomedical Research Informatics Delivered by Grid Enabled Services

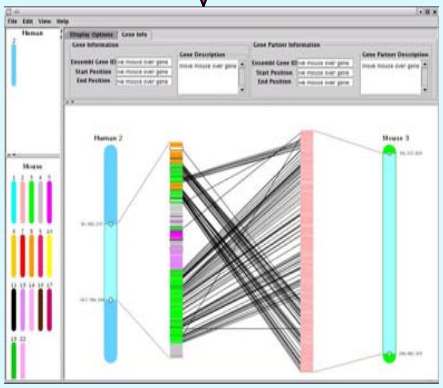
VO Authorisation





Synteny Grid Service

blast

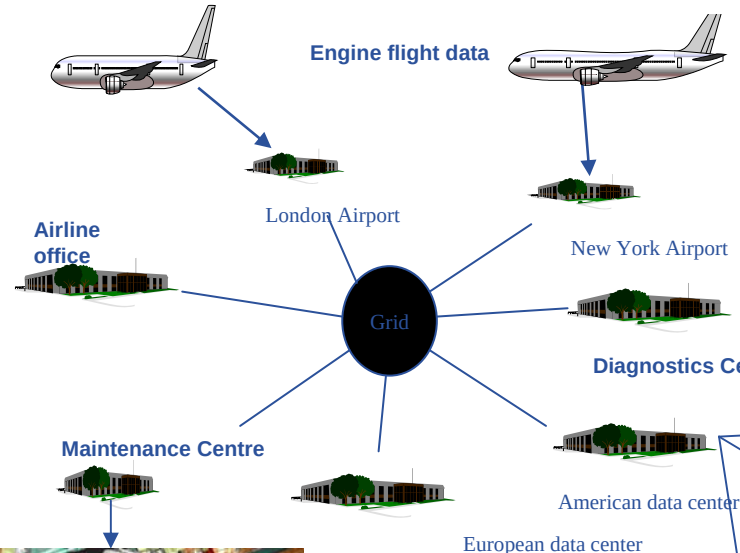


+

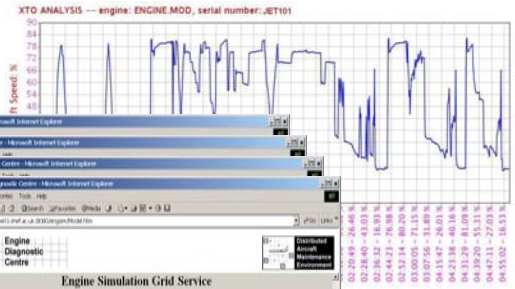
http://www.brc.dcs.gla.ac.uk/projects/bridges/

DAME: Grid based tools and Infer-structure for Aero-Engine Diagnosis and Prognosis

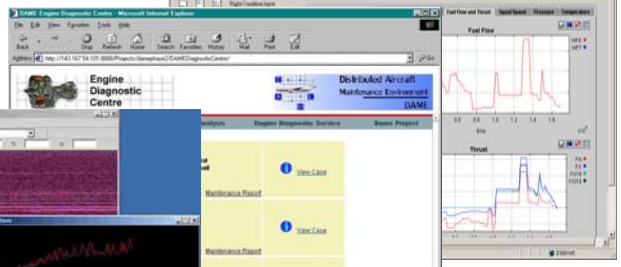
Distributed Aircraft Maintenance Environment
DAME



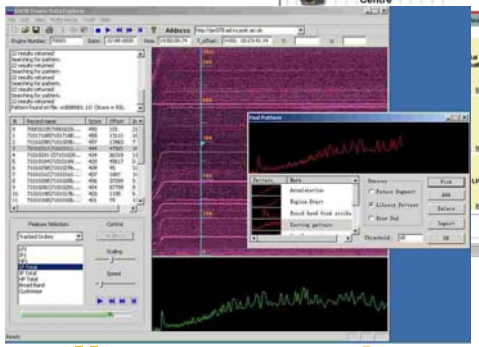
•"A Significant factor in the success of the Rolls-Royce campaign to power the Boeing 7E7 with the Trent 1000 was the emphasis on the new aftermarket support service for the engines provided via DS&S. Boeing personnel were shown DAME as an example of the new ways of gathering and processing the large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



XTO



Engine Model



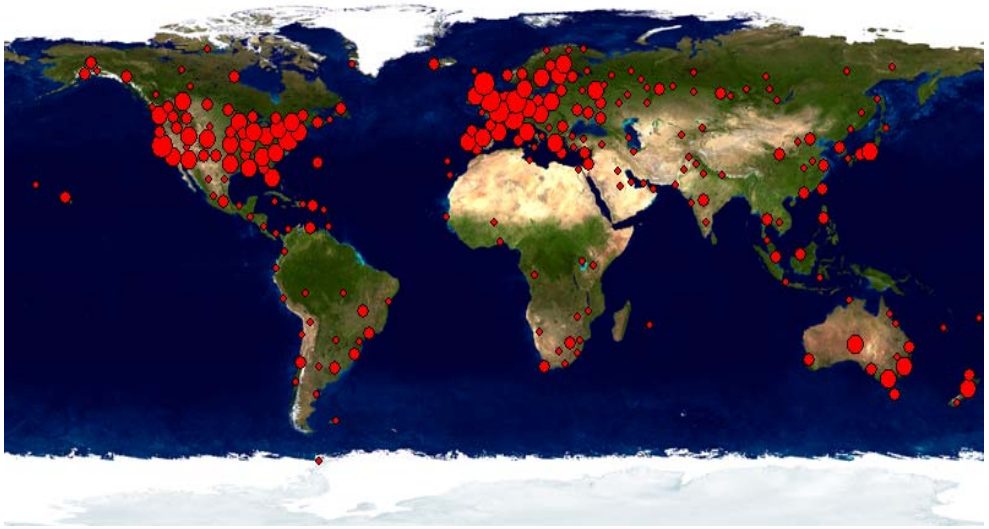
Case Based Reasoning

Companies:
Rolls-Royce
DS&S
Cybula

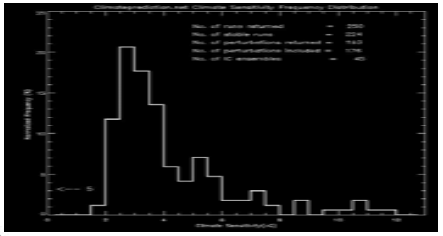
Universities:
York,
Leeds,
Sheffield, Oxford

Follow-on project: BROADEN

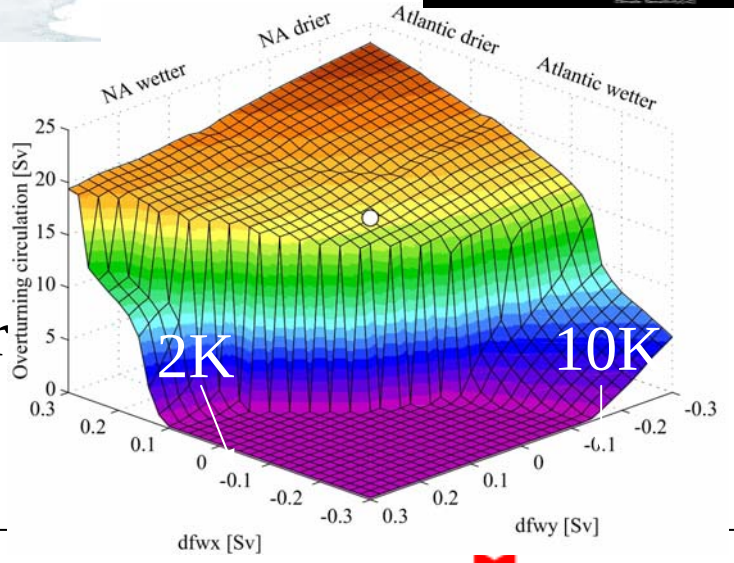
climate*prediction*.net and GENIE



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



Response of Atlantic circulation to freshwater forcing

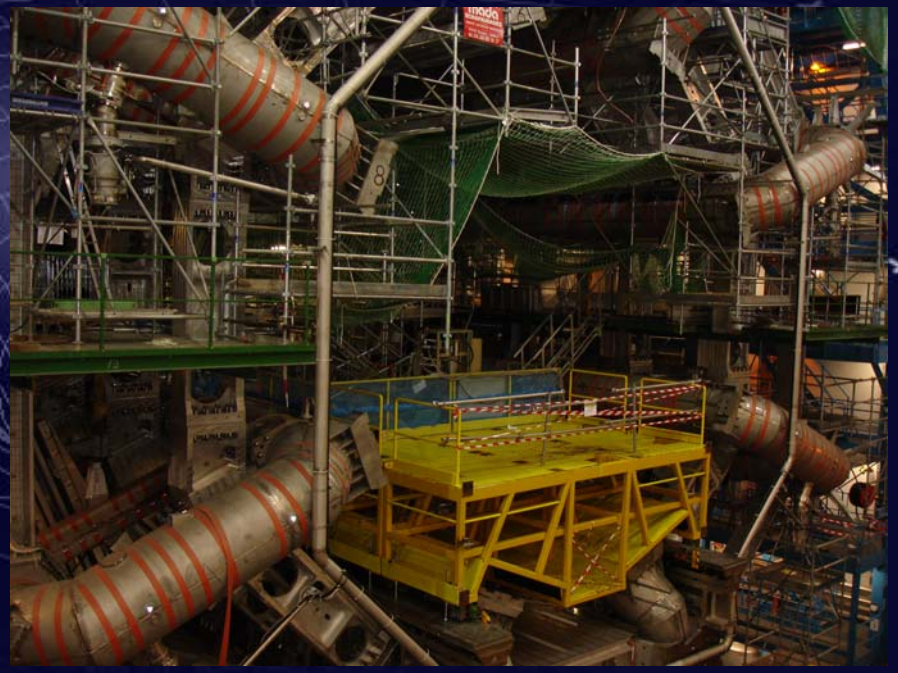


e-Science

NATURAL ENVIRONMENT RESEARCH COUNCIL

PPARC

UK Grid for Particle Physics



GridPP www.gridpp.ac.uk

ATLAS detectors, 2/3/06

Connecting *people*: Access Grid

http://www.accessgrid.org/

What is e-Research?

- **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
- **Began with focus in the "big sciences" hence initiatives are often badged as "e-science"**
- **Relevance of "e-science technologies" to new user communities (social science, arts, humanities…) led to the term "e-research"**

Grids: a foundation for e-Research

- e-Science methodologies will rapidly transform science, engineering, medicine and business
  - driven by exponential growth (×1000/decade)
    - enabling a whole-system approach

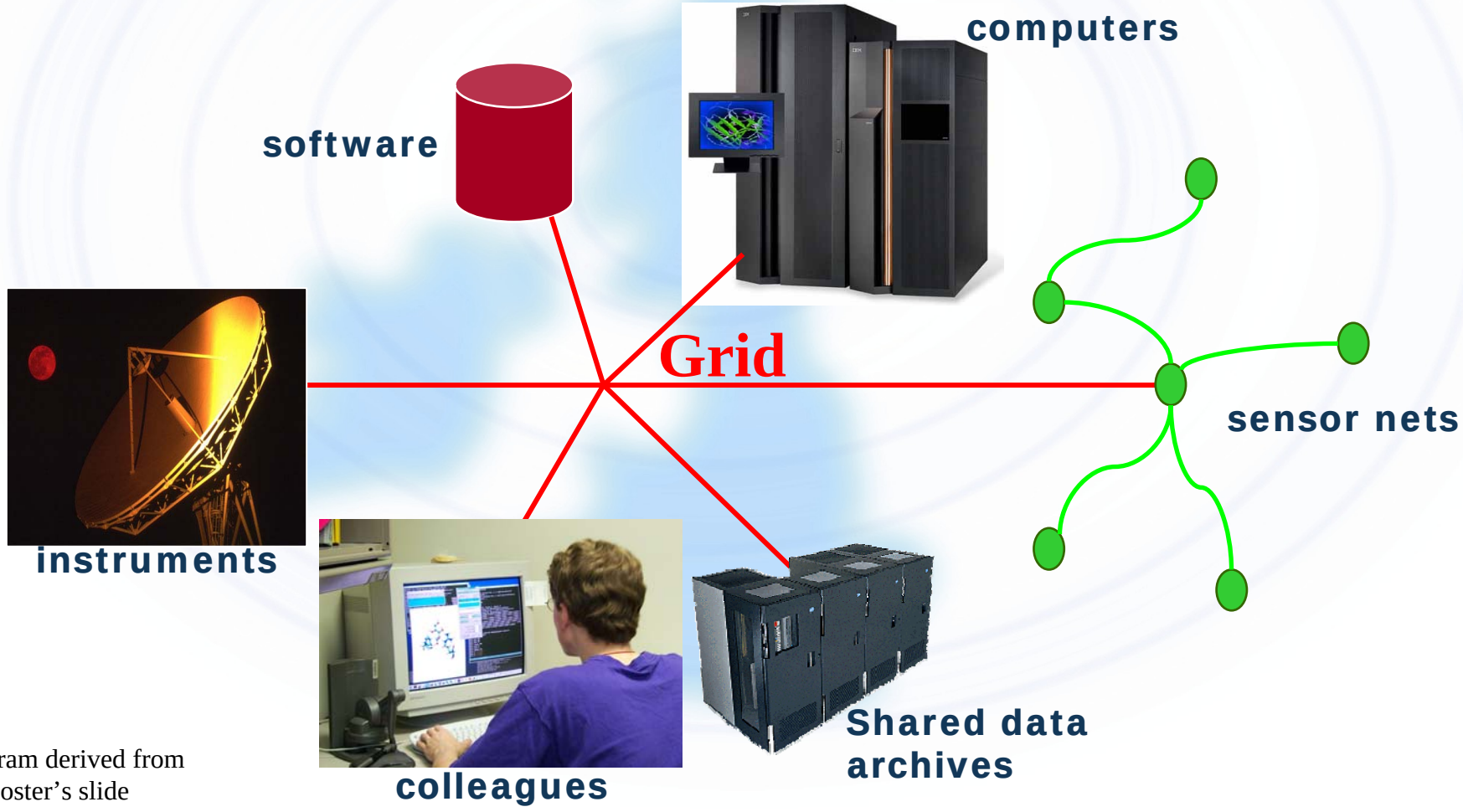
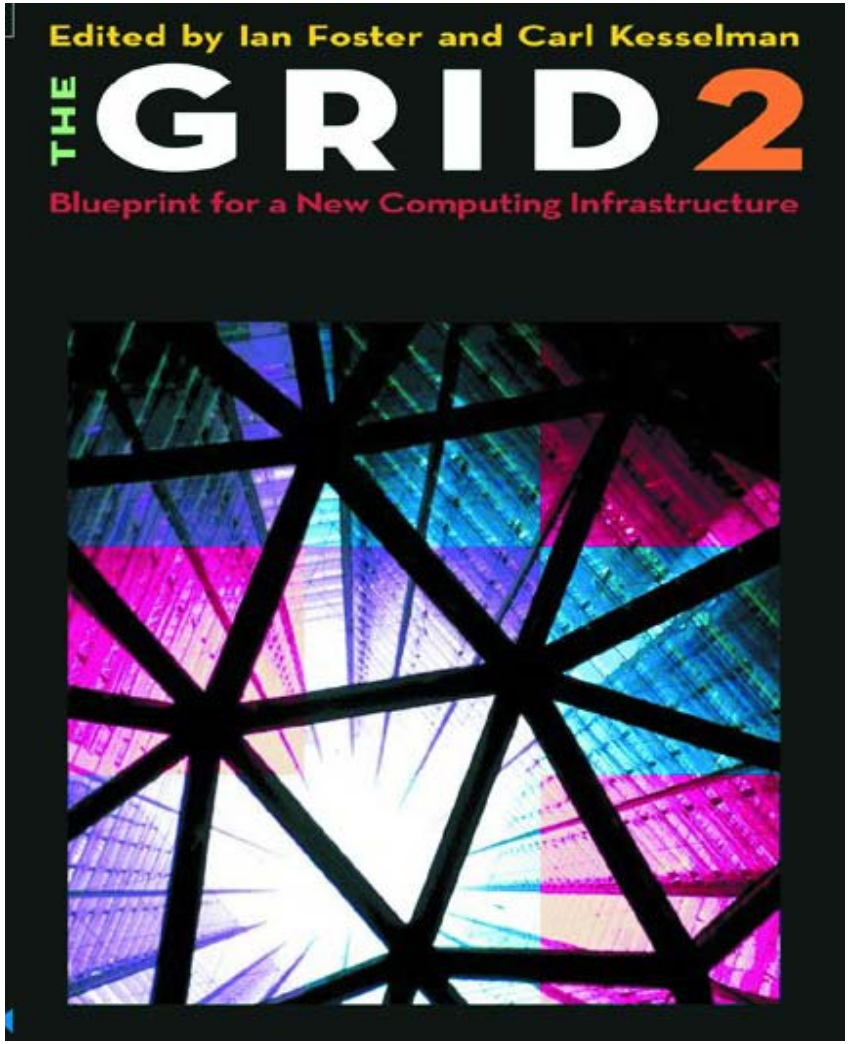


Diagram derived from Ian Foster's slide

What is Grid Computing?

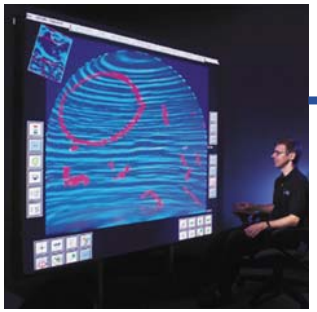
- **The grid vision is of "Virtual computing" (+ information services to locate computation, storage resources)**
  - Compare: The web: "virtual documents" (+ search engine to locate them)
- **MOTIVATION: collaboration through sharing resources (and expertise) to expand horizons of**
  - Research
  -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



The Grid Metaphor

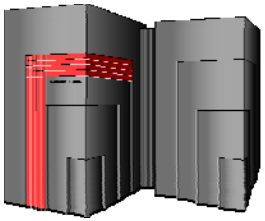
Mobile Access

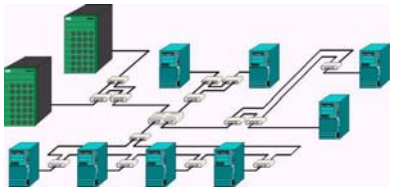
Workstation



Visualising

GRID MIDDLEWARE





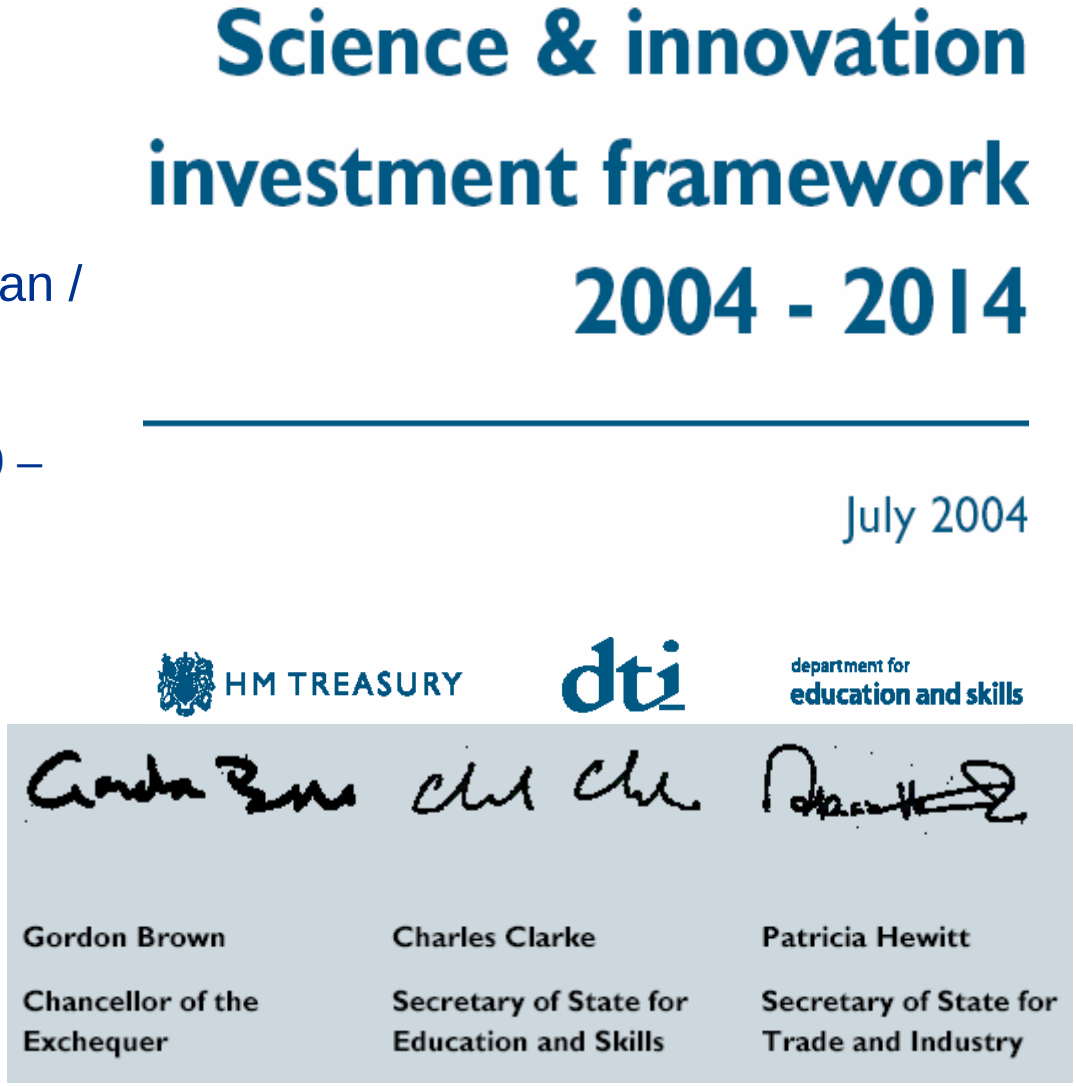
Supercomputer, PC-Cluster

Data-storage, Sensors, Experiments

Internet, networks

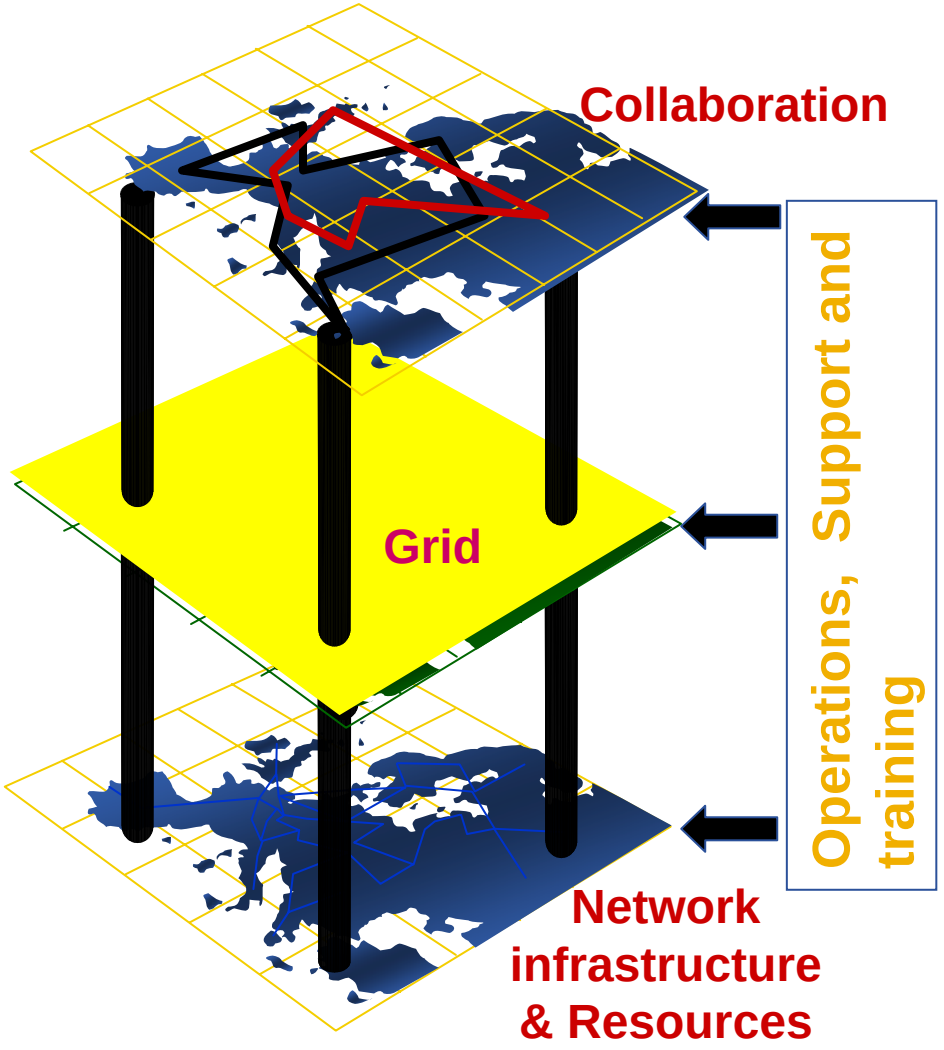
What is e-Infrastructure? – Political view

- **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



What is e-Infrastructure?

- **Grids: permit resource sharing across administrative domains**
- **Networks: permit communication across geographical distance**
- **Supporting organisations**
  - Operations for grids, networks
- **Resources**
  - Computers
  - Digital libraries
  - Research data
  - Instruments
- **Middleware**
  - Authentication, Authorisation
  - Registries, search engines
  - Toolkits, environments
    - E.g. for collaboration



What is Grid computing?

- **The term “Grid” has become popular!**
  - Sometimes in Industry : “Grids” = clusters
    - Motivations: better use of resources; scope for commercial services
  - Also used to refer to the harvesting of donated, unused compute cycles
    - (SETI@home, Climateprediction.net)
  - These are e-Infrastructure but are not “grids” from the e-Research viewpoint!
- **Our (EGEE and e-research) use of “Grid”**
  - Virtual computing for collaboration across administrative domains
  - Includes resource utilisation and sharing
  - Always: across administrative domains

Grid concepts

Virtual organisations and grids

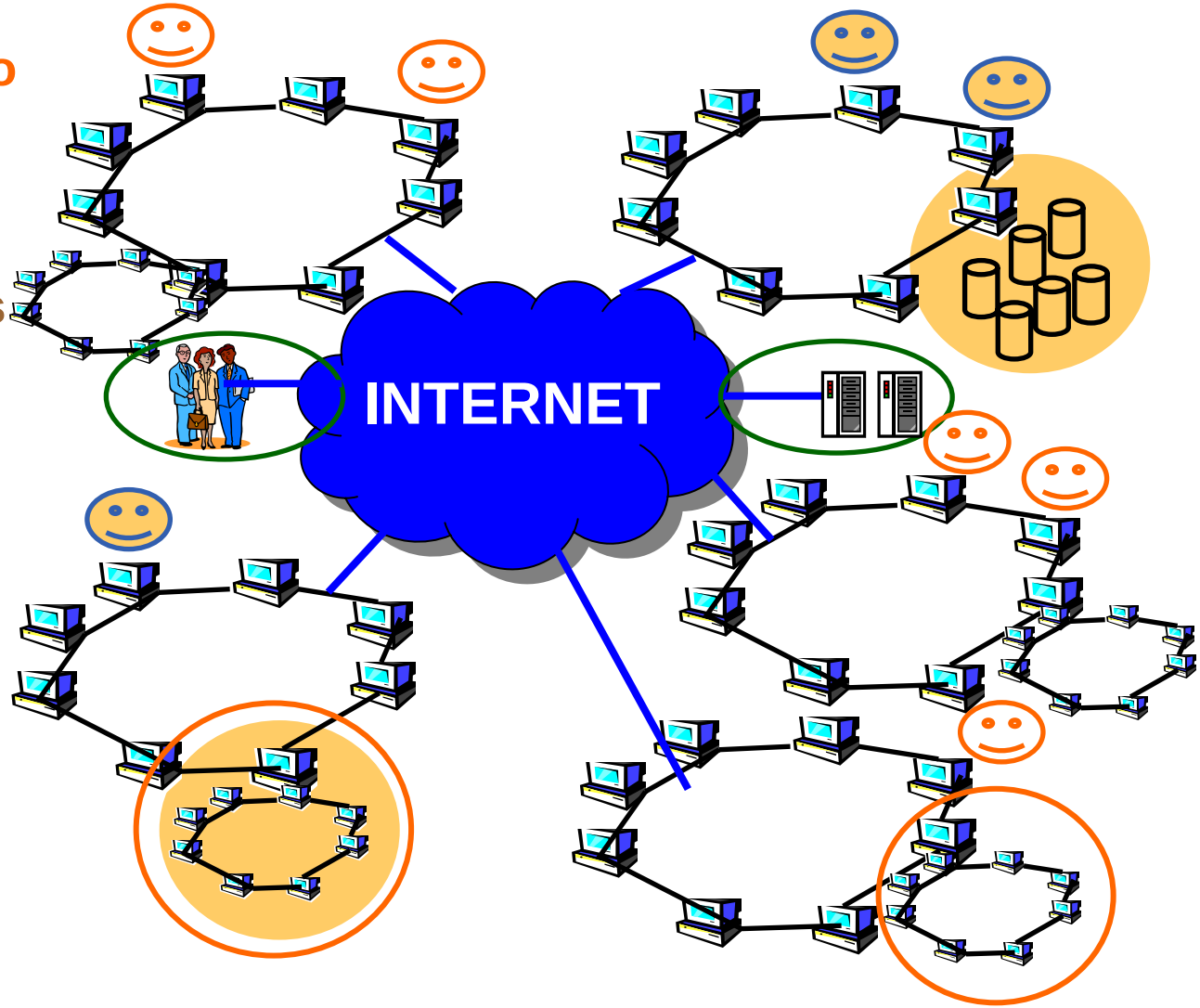
- **What's a Virtual Organisation?**
  - People in different organisations seeking to cooperate and share resources across their organisational boundaries E.g. A research collaboration

- **Each grid is an infrastructure enabling one or more "virtual organisations" to share and access resources**

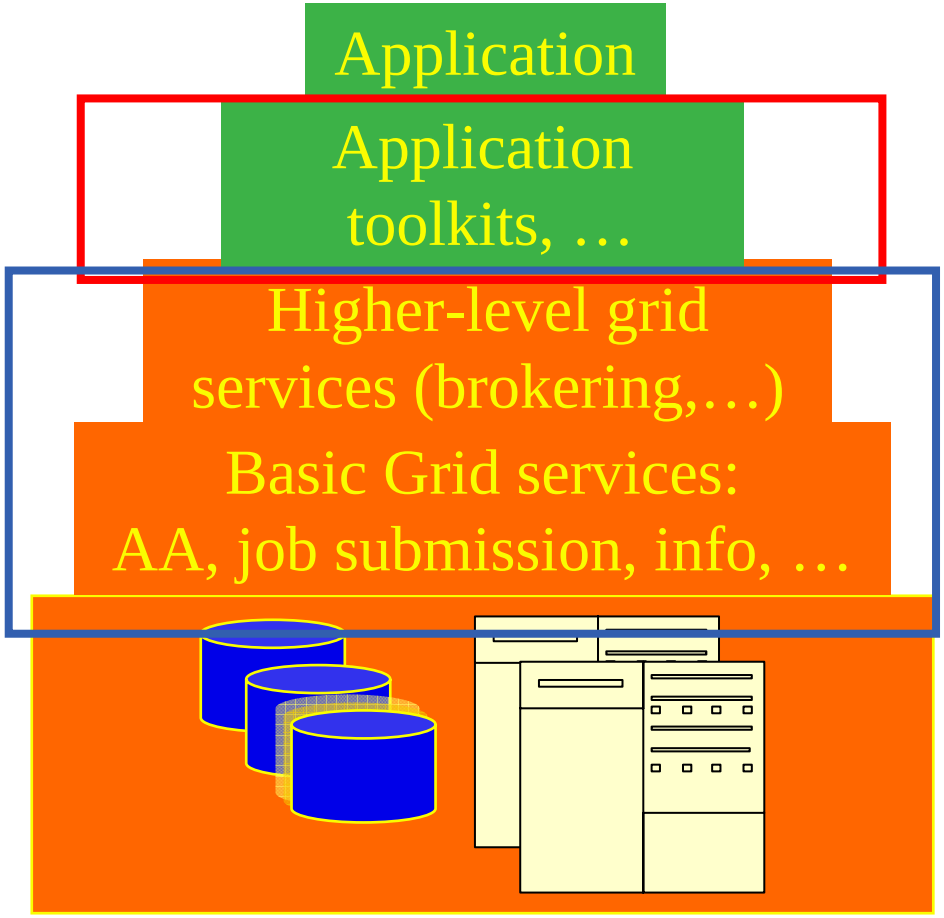
- **Key concept: The ability to negotiate resource-sharing arrangements among a set of participating parties (providers and consumers) and then to use the resulting resource pool for some purpose. (Ian Foster)**

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**



Empowering VO's



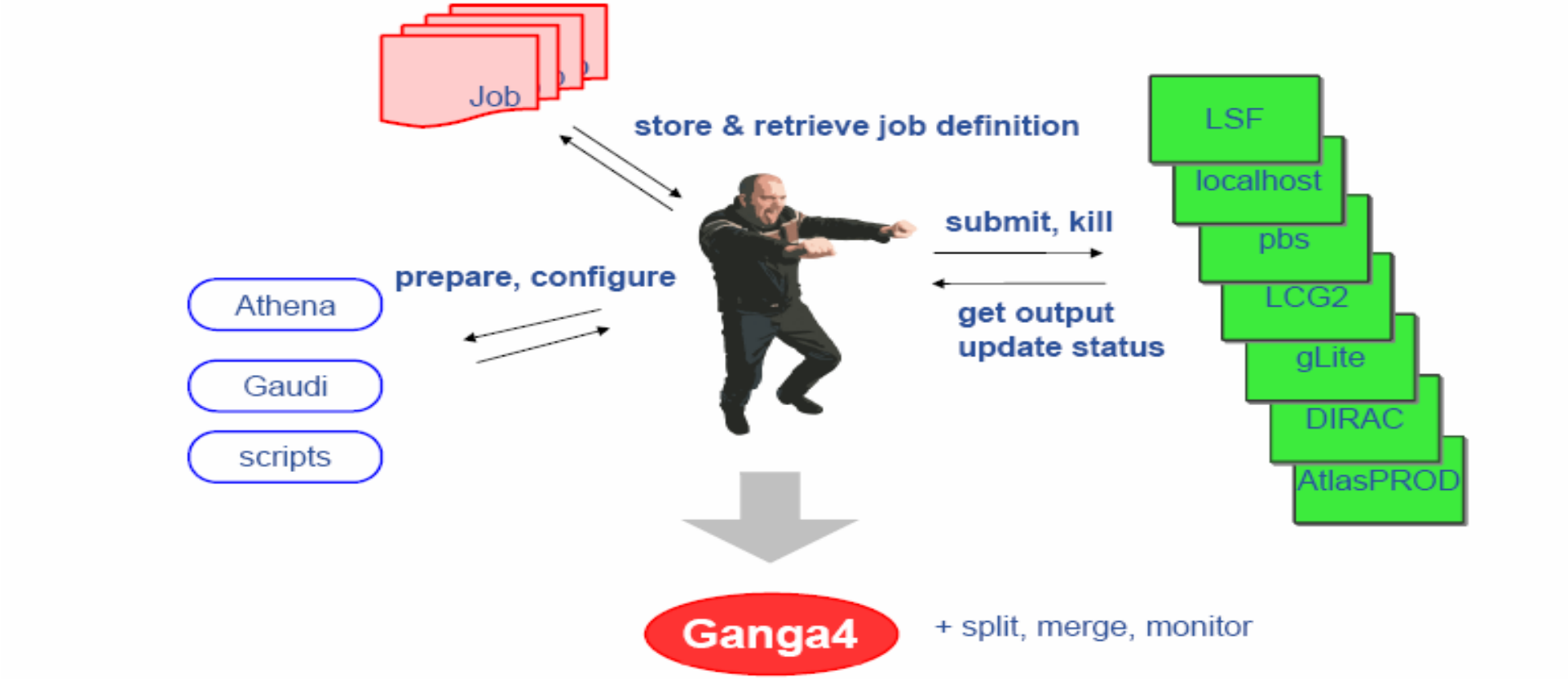
Where computer science meets the application communities!

VO-specific developments:

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

Production grids provide these services.

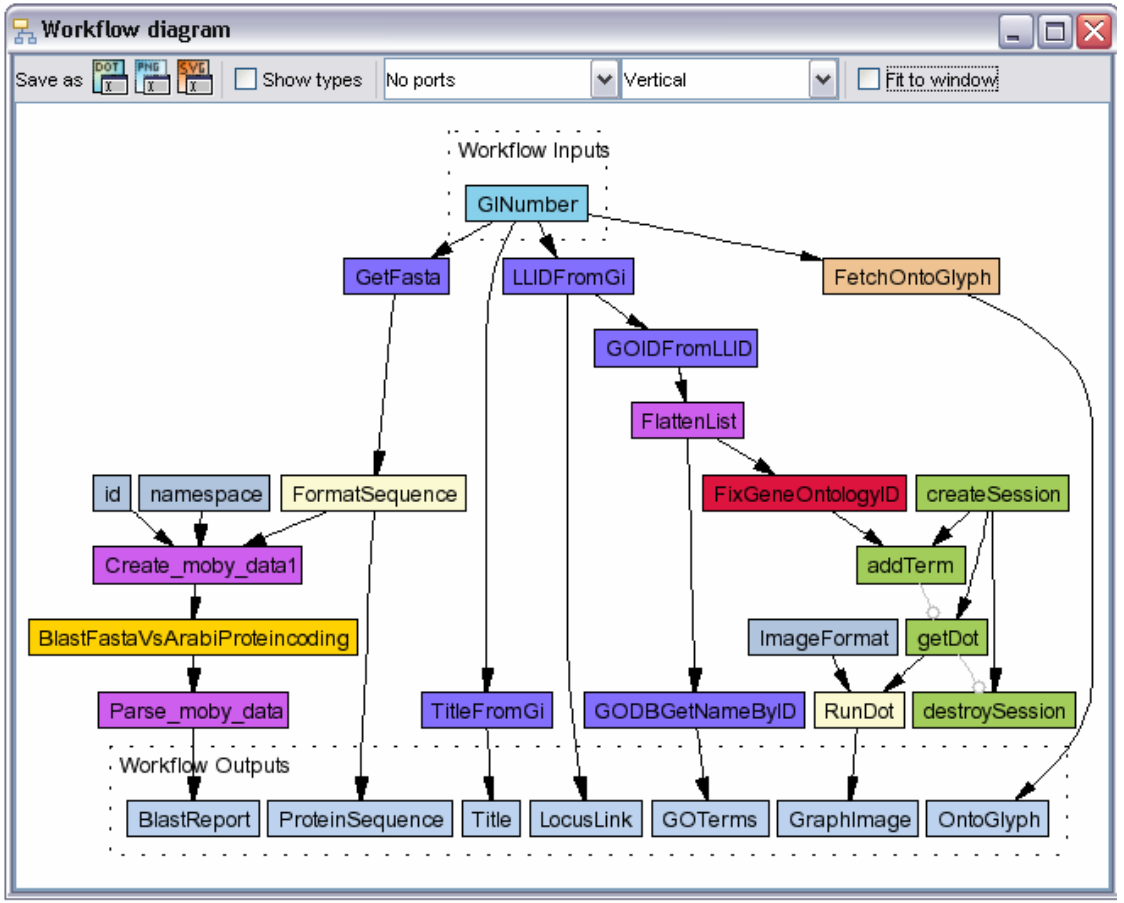
Example of higher-level service -1: GANGA



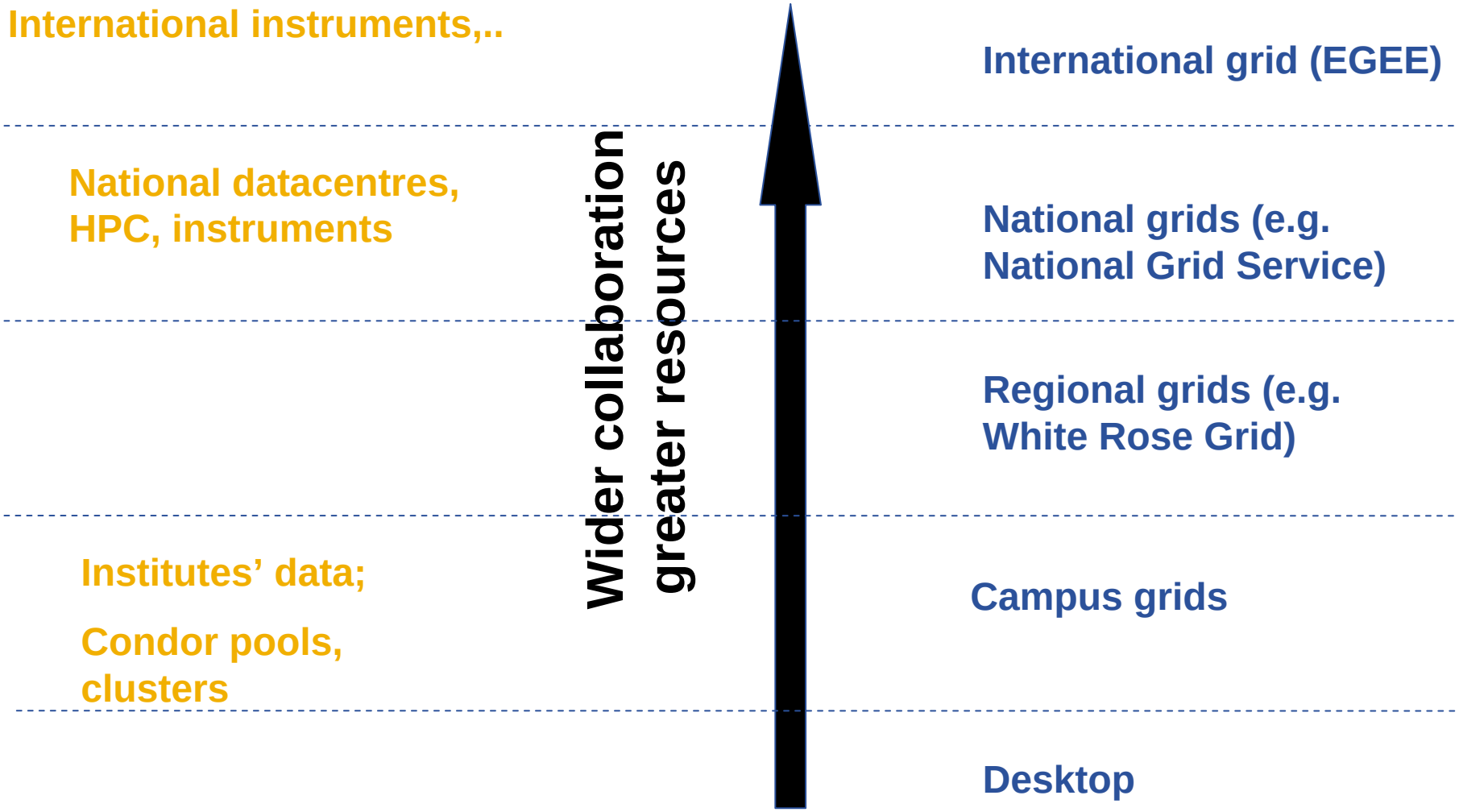
- **Ganga is a lightweight user tool**
  ganga.web.cern.ch/
- **But also: Ganga is a developer framework**

Example 2: Workflow

- Taverna in MyGrid http://www.mygrid.org.uk/
- “allows the e-Scientist to describe and enact their experimental processes in a structured, repeatable and verifiable way”
- GUI
- Workflow language
- enactment engine



The many scales of grids

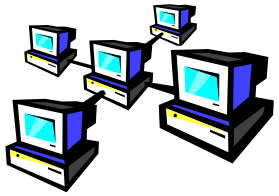


Main components

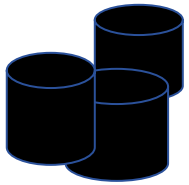
Access service How users logon to a Grid

Resource Broker (RB): Service that matches the user's requirements with the available resources on a Grid

Information System: Characteristics and status of resources



Computing Element (CE): A batch queue on a site's computers where the user's job is executed



Storage Element (SE): provides (large-scale) storage for files

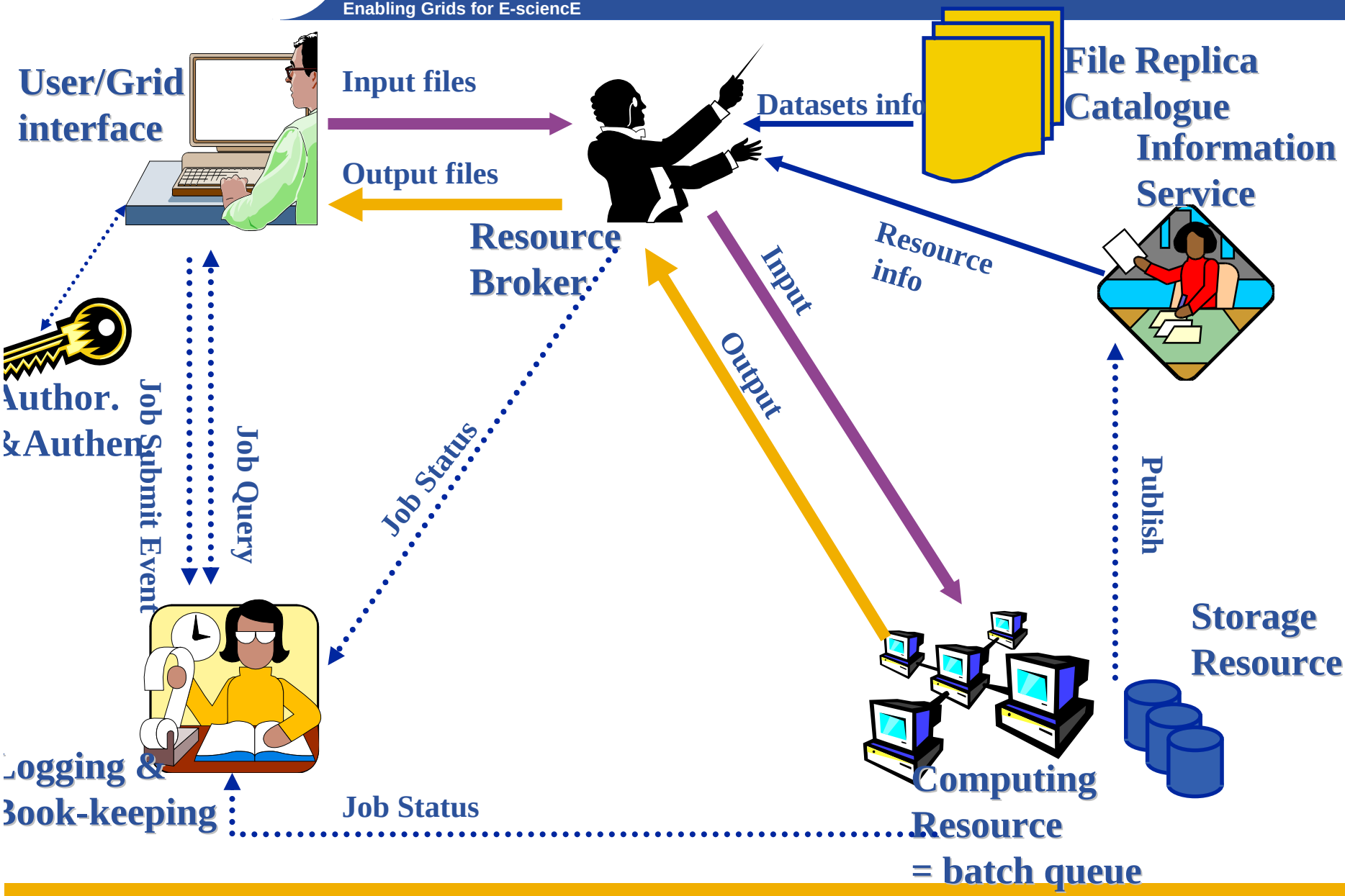
Who provides the resources?!

Service	Provider	Note
Access service	User / institute/ VO / grid operations	Computer with client software
Resource Broker (RB)	VO / grid operations	(No NGS-wide RB exists)
Information System:	Grid operations	
Computing Element (CE)	VO / sometimes centralised provision also	Scalability requires that VOs provide resources to match average need
Storage Element (SE)	ditto	ditto

"VO": virtual organisation

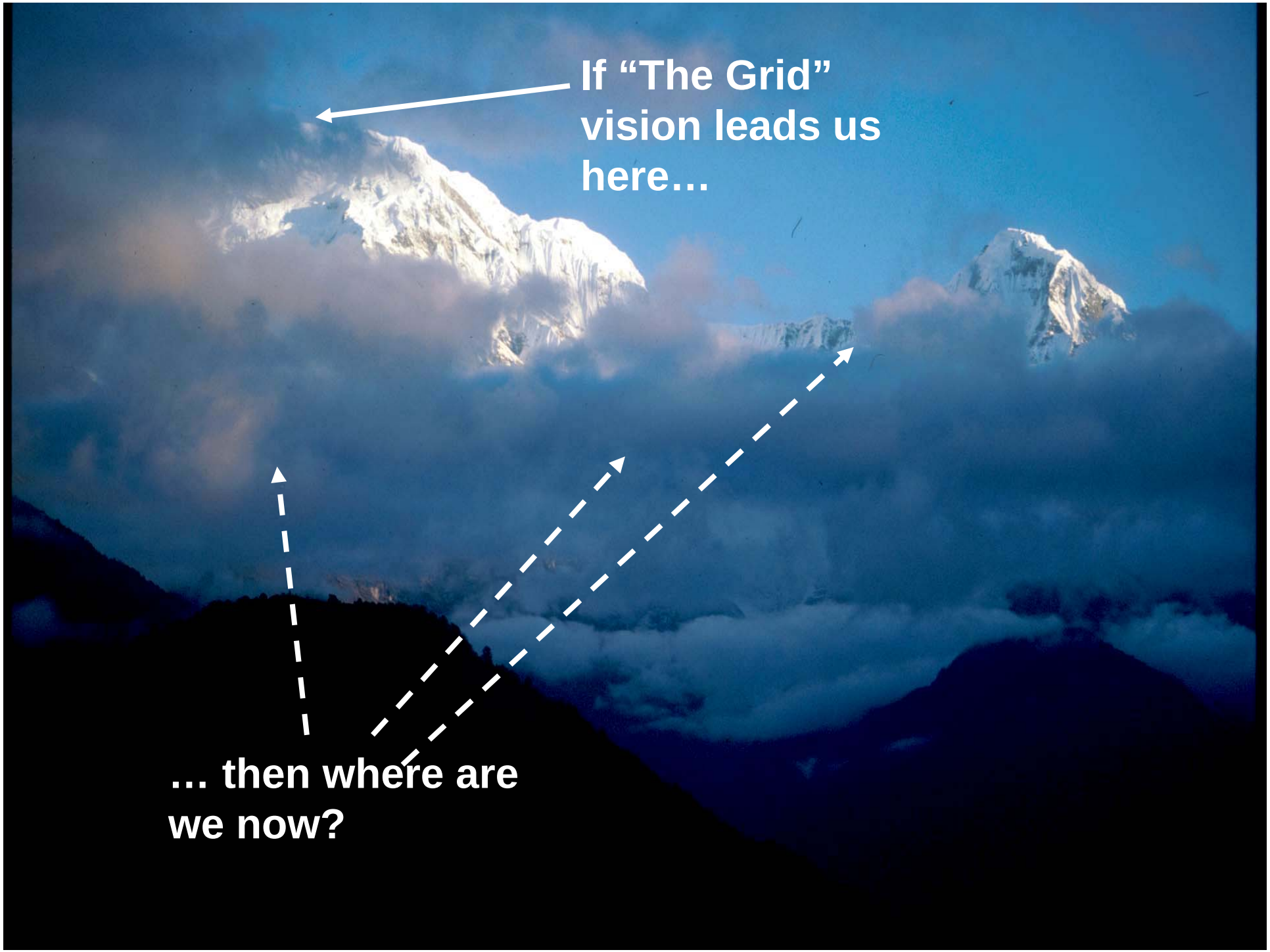
"Grid operations": funded effort

Current EGEE grid



EGEE is …

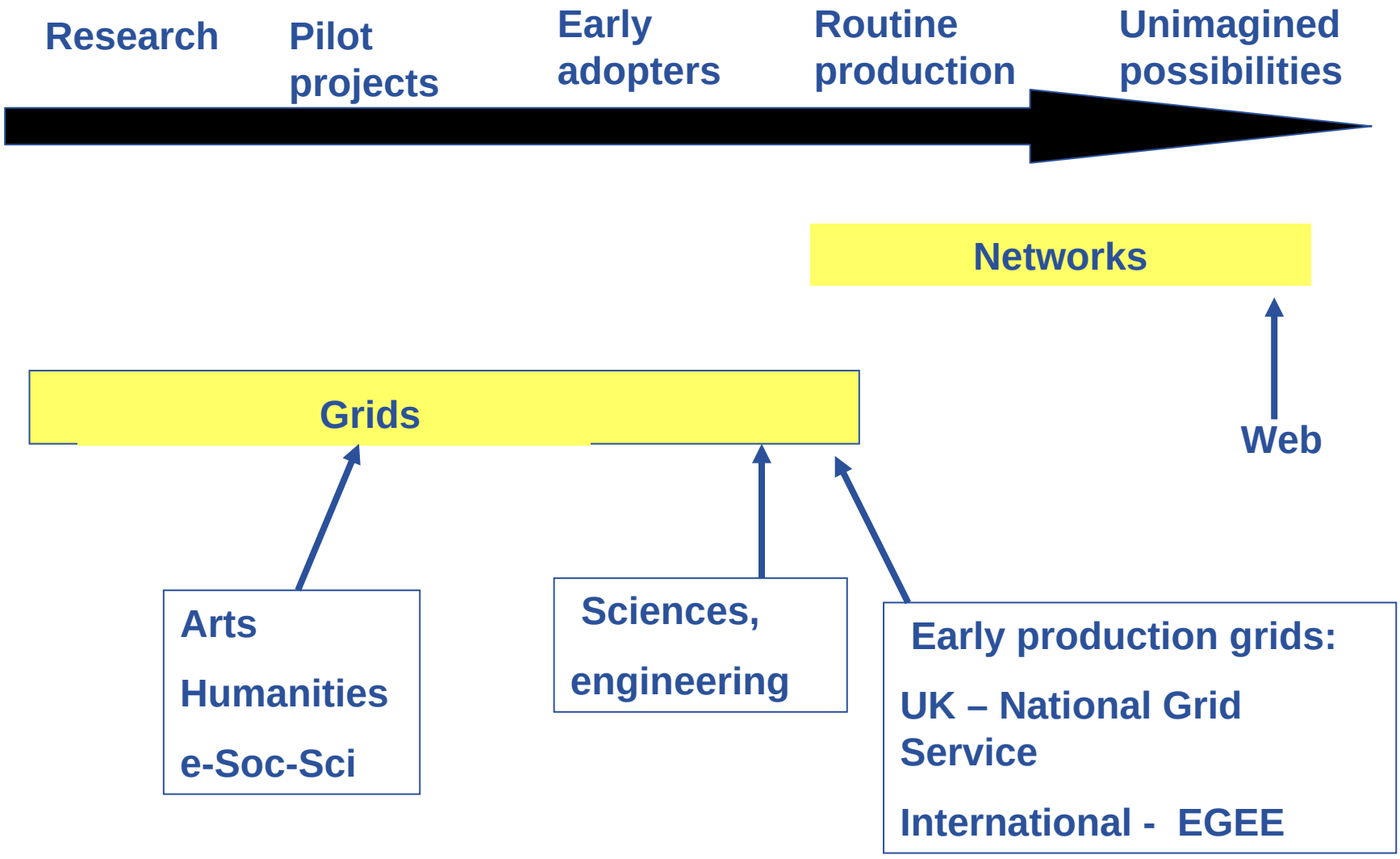
- **EU-funded project that has established the largest multi-VO production grid in the world!**
- **What's happening now?**
  http://gridportal.hep.ph.ic.ac.uk/rtm/
- **What resources are connected?**
  http://goc.grid-support.ac.uk/gridsite/monitoring/


If “The Grid” vision leads us here…
… then where are we now?

Grids: 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
      - ***In Europe, EGEE***
      - ***In UK, National Grid Service***
      - *In US, Teragrid and OSG*
    - One stack of middleware that serves many research communities
    - Establishing operational procedures and organisation
  - **Standards** (a slow process)
    (e.g. Global Grid Forum, http://www.gridforum.org/)
- **Service orientation - "*the* way to build grids"**

Where are we now? –user's view



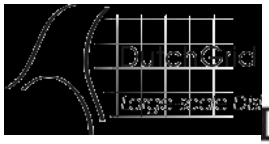
Where are we now?!

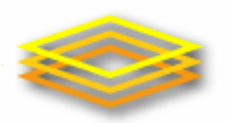
- **Standards are emerging… some near acceptance and some being discarded**
  - Standards bodies:
    - W3C http://www.w3c.org/
    - GGF http://www.ggf.org/
    - OASIS http://www.oasis-open.org/home/index.php
    - IETF http://www.ietf.org/
  - For a summary see http://www.innoq.com/soa/ws-standards/poster/

- **Production grids are based on de-facto standards at present**
  - Inevitably!
  - GT2 especially
  - But locks a grid into one middleware stack unable to benefit from the diverse developments of new services

- **Globus Toolkit 4 has been released**

National grid initiatives now include…

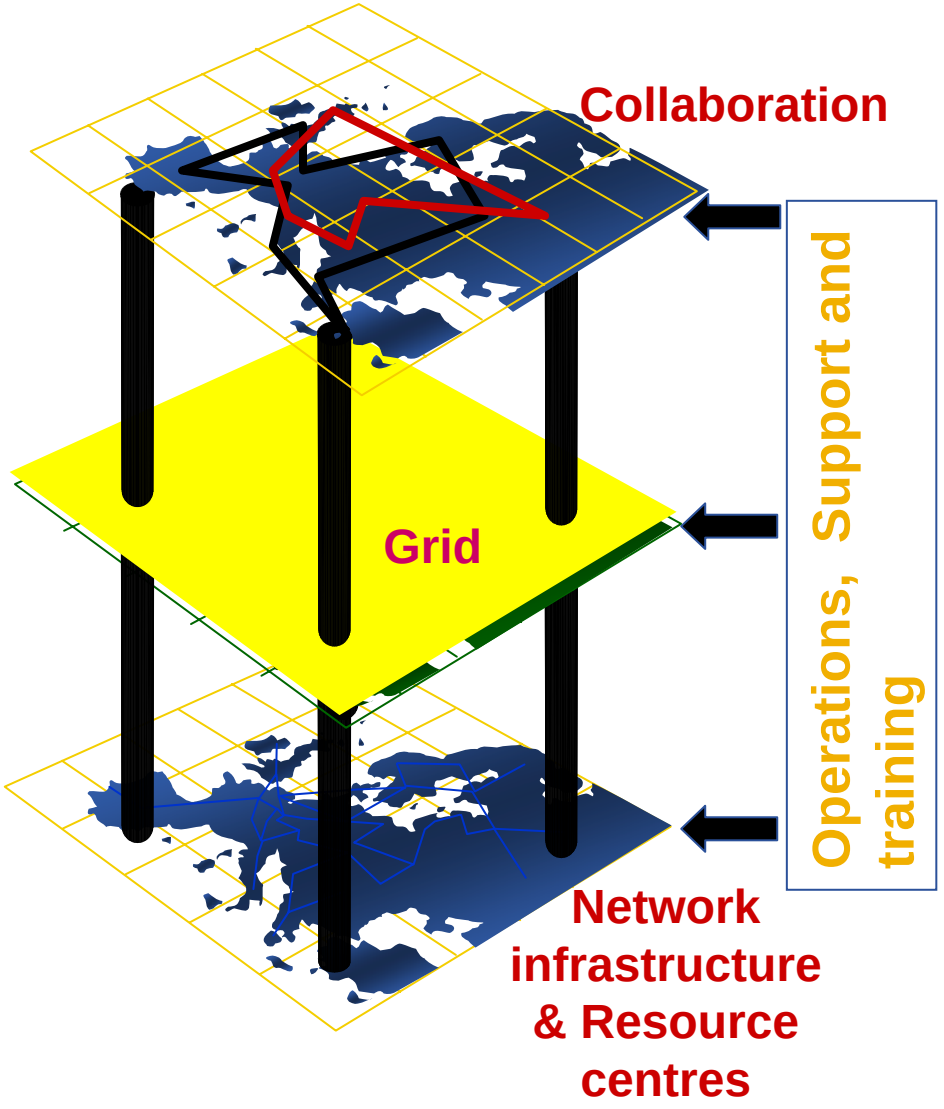




CroGrid

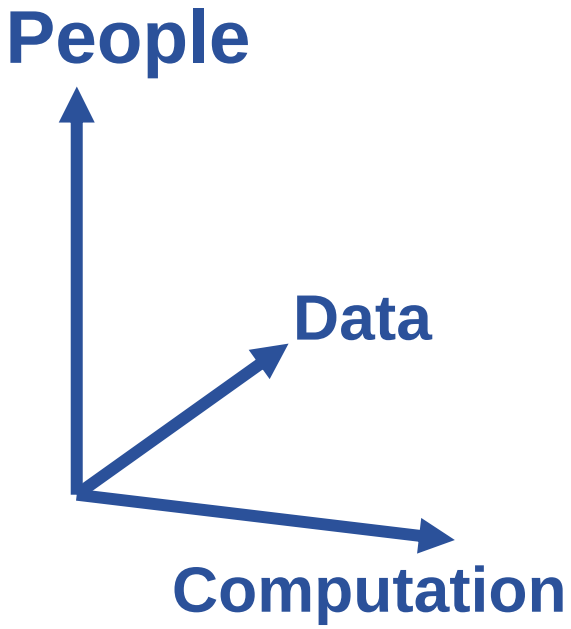
Summary -1: enabling collaboration

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



Summary -2

- Ask not what “the Grid” can do for *you*
- BUT
- With whom do you collaborate?
- What resources / services can you provide?
- What resources would empower your research?



Further reading

- Global Grid Forum http://www.ggf.org/ (see GGF16)
- National e-Science Centre http://www.nesc.ac.uk
- UK All Hands Meeting - see 2005 proceedings and go to 2006! http://www.allhands.org.uk/
- National Grid Service http://www.ngs.ac.uk
- EGEE www.eu-egee.org
- The Grid Cafe www.gridcafe.or

- The Grid Core Technologies, Maozhen Li and Mark Baker, Wiley, 2005