Edinburgh e-Science TOE
Training Outreach and Education

Grid Computing and the National Grid Service

Mike Mineter
NeSC-TOE
mjm@nesc.ac.uk

This morning…

- Talk 1: Introduction to Grid Computing
- Pause for thought and brief discussion
- Talk 2: Production Grids
  - European – EGEE
  - UK – National Grid Service
- Discussion

All slides will be accessible via the agenda page

Grid Computing

Mike Mineter
Training Outreach and Education
National e-Science Centre

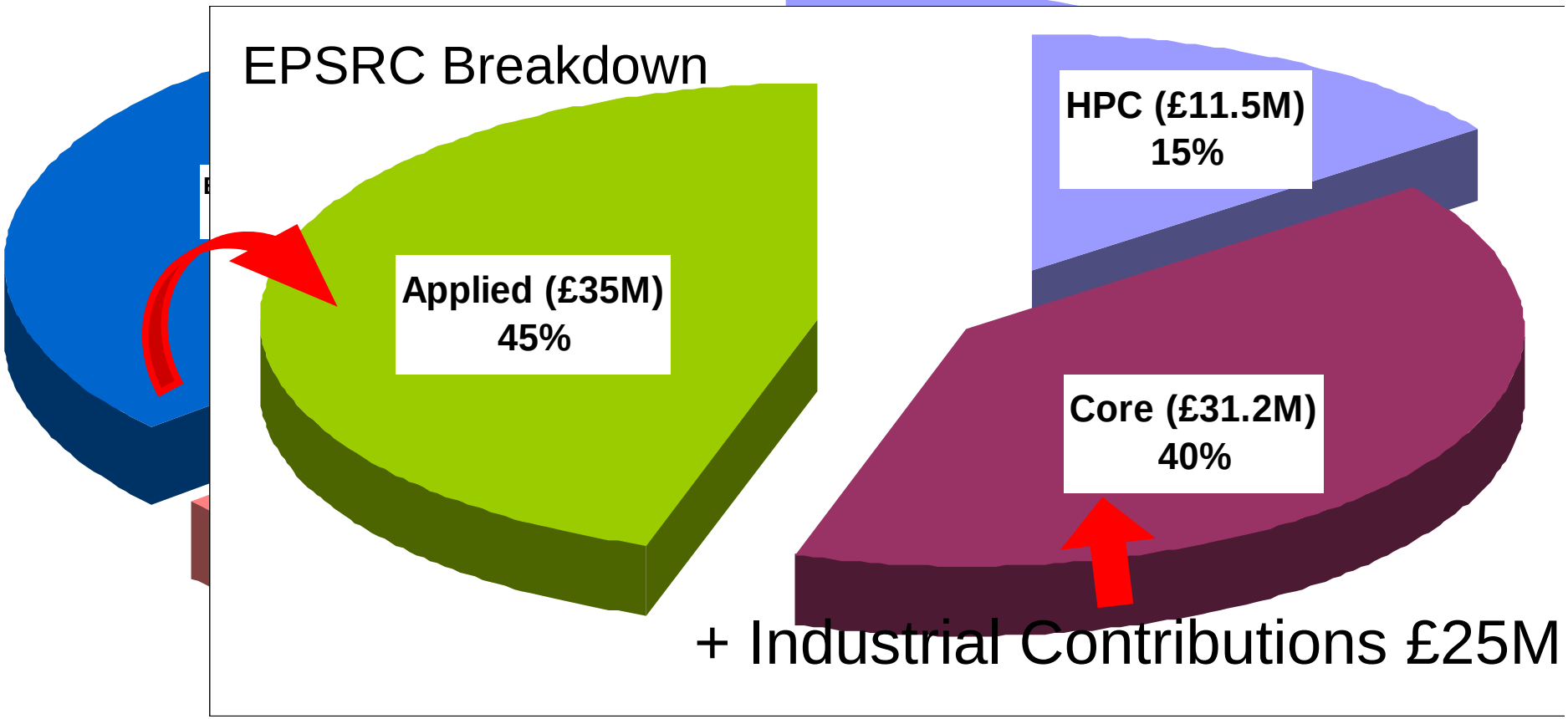
mjm@nesc.ac.uk

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

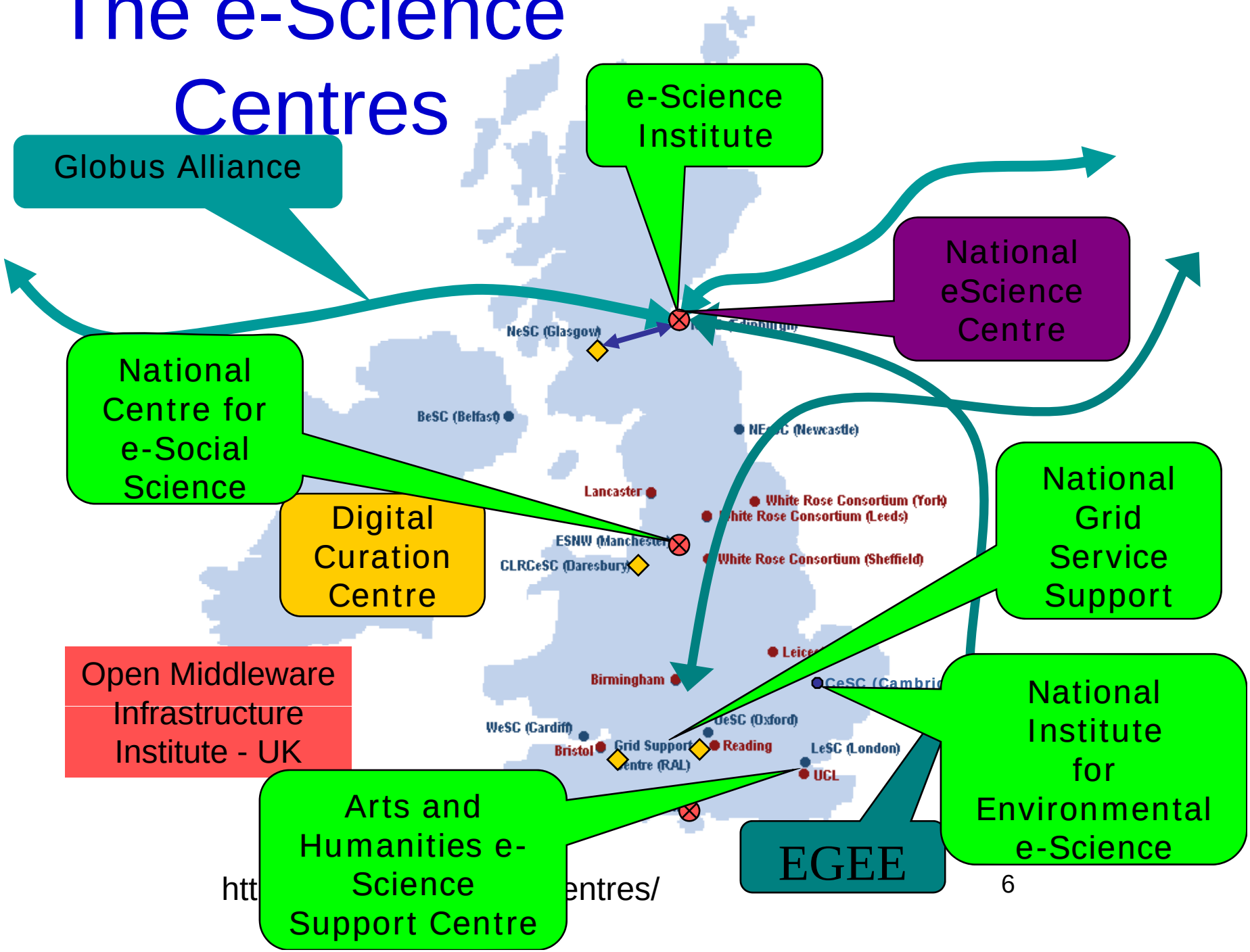
UK e-Science Budget (2001-2006)

Total: £213M + £100M via JISC



Source: Science Budget 2003/4 – 2005/6, DTI(OST)

The e-Science Centres



htt ntres/

Contents

- **Introduction to**
  - e-Science
  - e-Research
  - Grids
  - e-Infrastructure
- **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

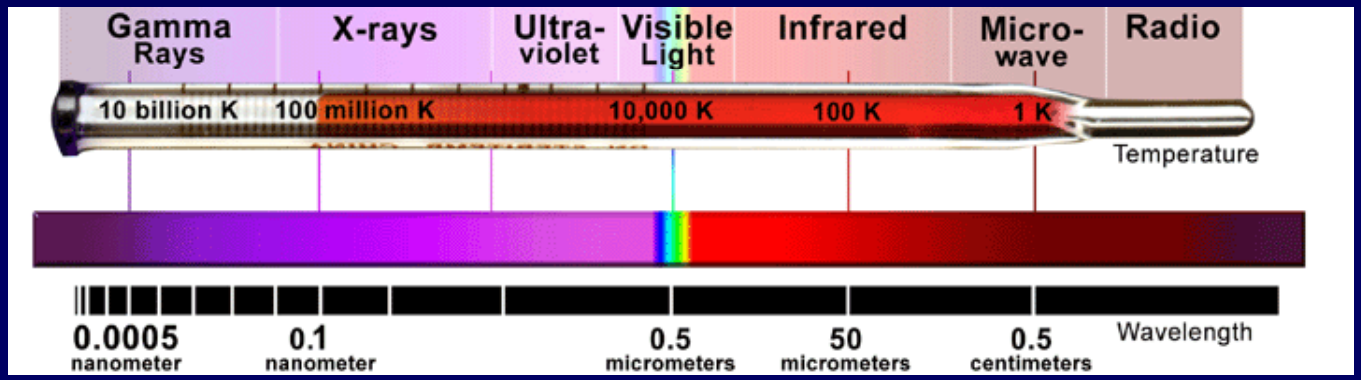
*'e-Science is about global collaboration in key areas of science, and the **next generation of infrastructure** that will enable it.'*

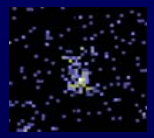
Networks + Grids

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

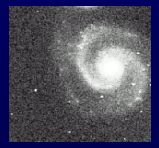
Virtual Observatories

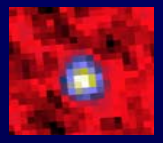
Observations made across entire electromagnetic spectrum

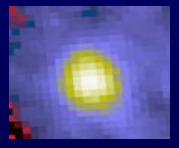


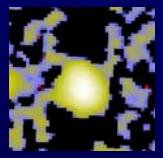

ROSAT ~keV

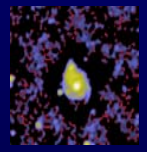

DSS Optical

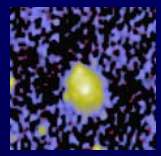

2MASS 2μ


IRAS 25μ


IRAS 100μ


GB 6cm


NVSS 20cm


WENSS 92cm

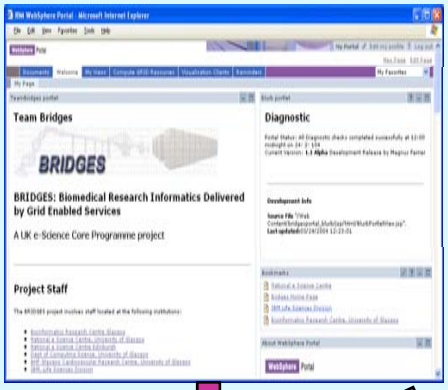
⇒e.g. different views of a local galaxy

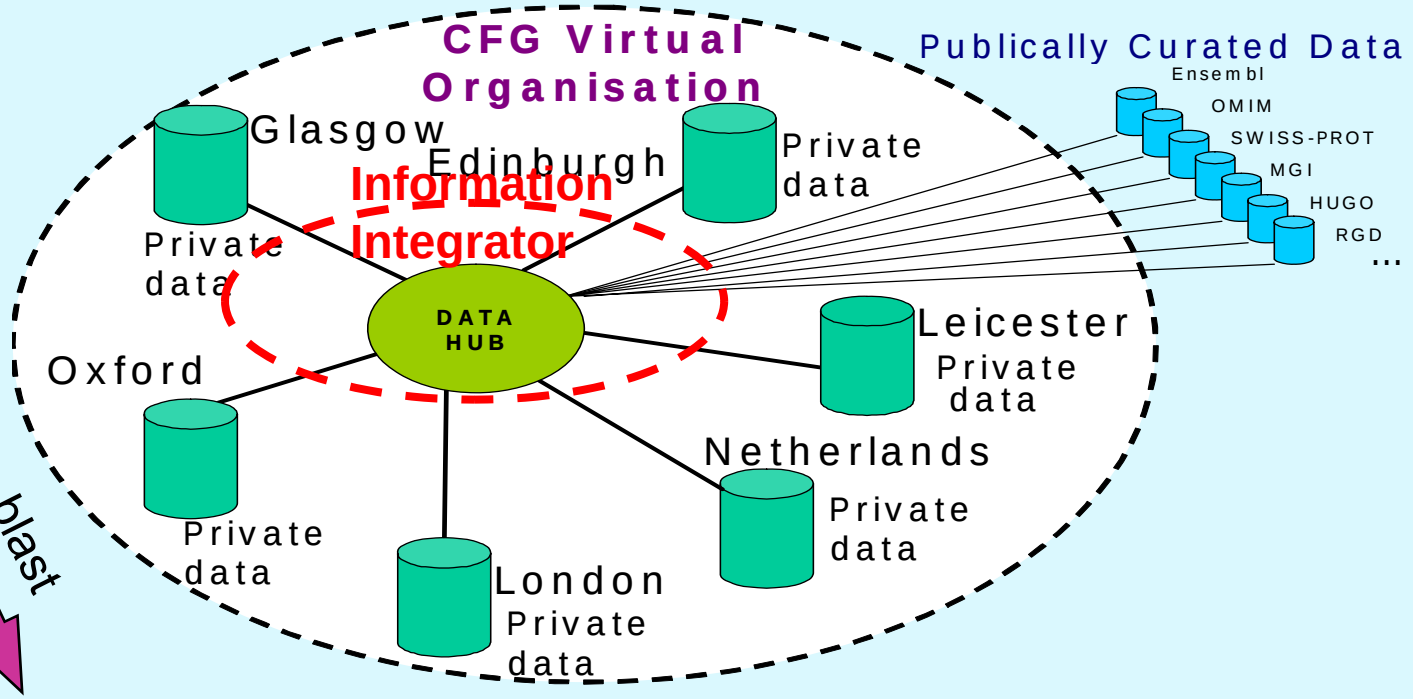
Need all of them to understand physics fully

Databases are located throughout the world

Biomedical Research Informatics Delivered by Grid Enabled Services

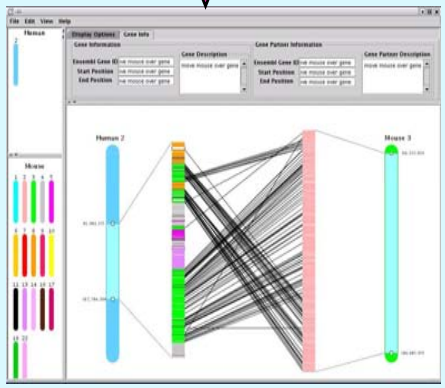
VO Authorisation





Synteny Grid Service

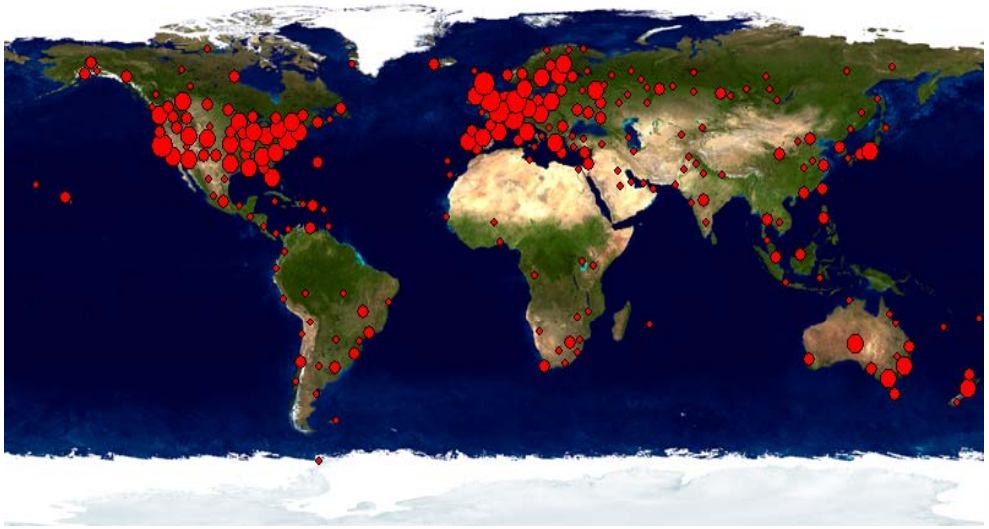
blast



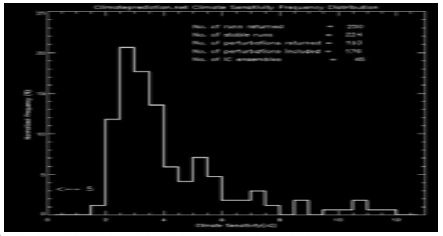
+

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

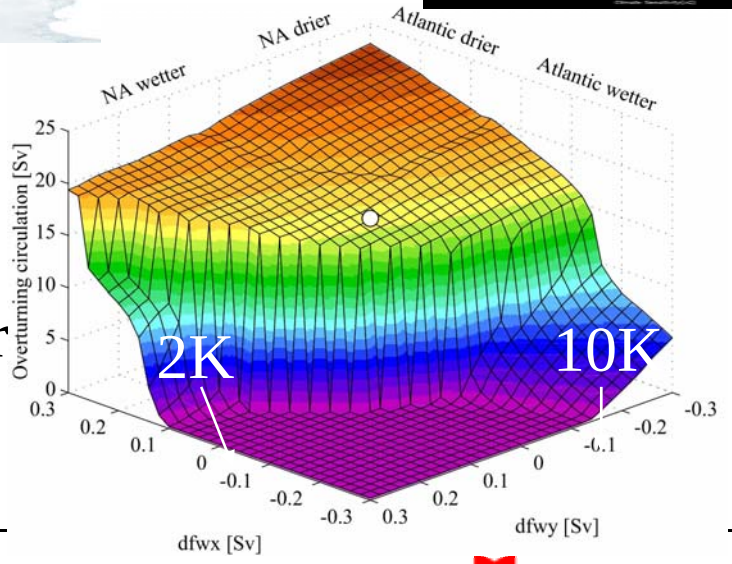
climate*prediction*.net and GENIE



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



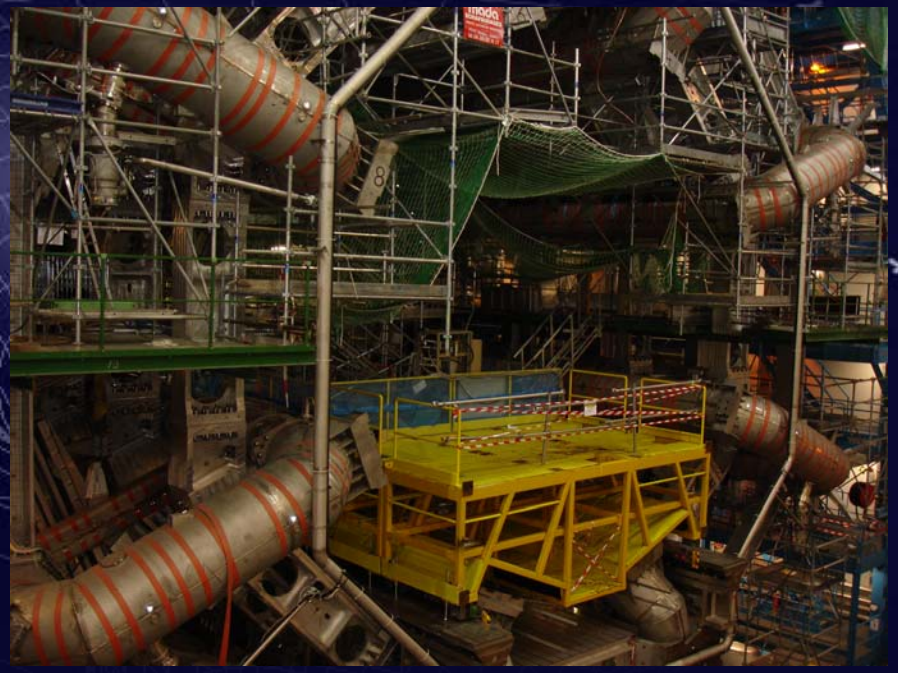
Response of Atlantic circulation to freshwater forcing



NATURAL ENVIRONMENT RESEARCH COUNCIL

PPARC

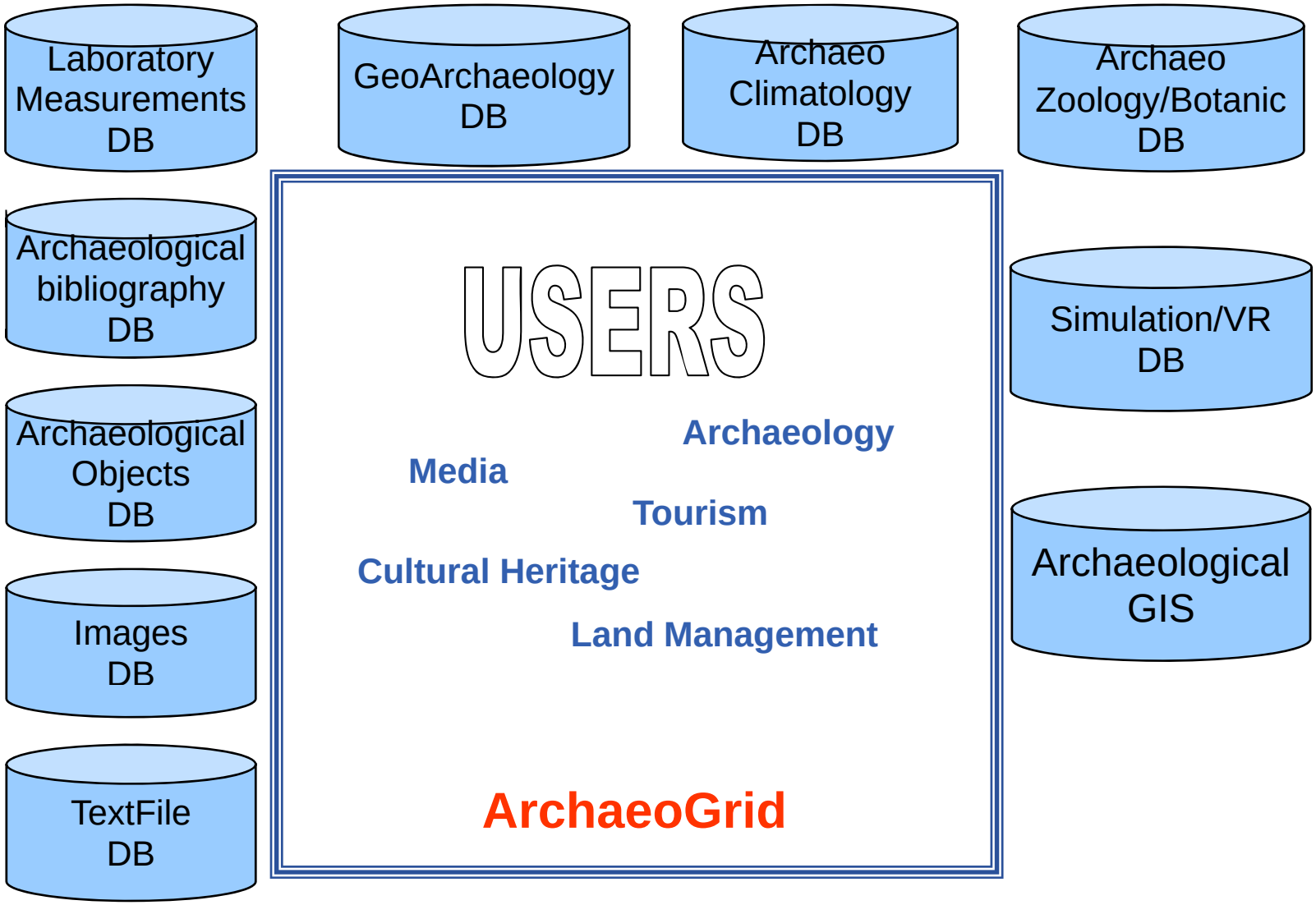
UK Grid for Particle Physics



GridPP www.gridpp.ac.uk

ATLAS detectors, 2/3/06

ArchaeoGrid



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 on enhancing scientific research - hence "e-science"**
- **Relevance of "e-science technologies" to new user communities (social science, arts, humanities…) led to the term "e-research"**

e-Research and Grids

Collaborative
"virtual computing"

Sharing data, computers, software
Enabled by Grids:
Campus, National, regional
International: EGEE grid

Improvised cooperation

Email
File exchange
ssh access to run programs
Enabled by networks:
national, regional and
International: GEANT

People with shared goals

Grids: a foundation for e-Research

- Enabling a whole-system approach
- Facilitating collaboration: data, program, storage, computer sharing

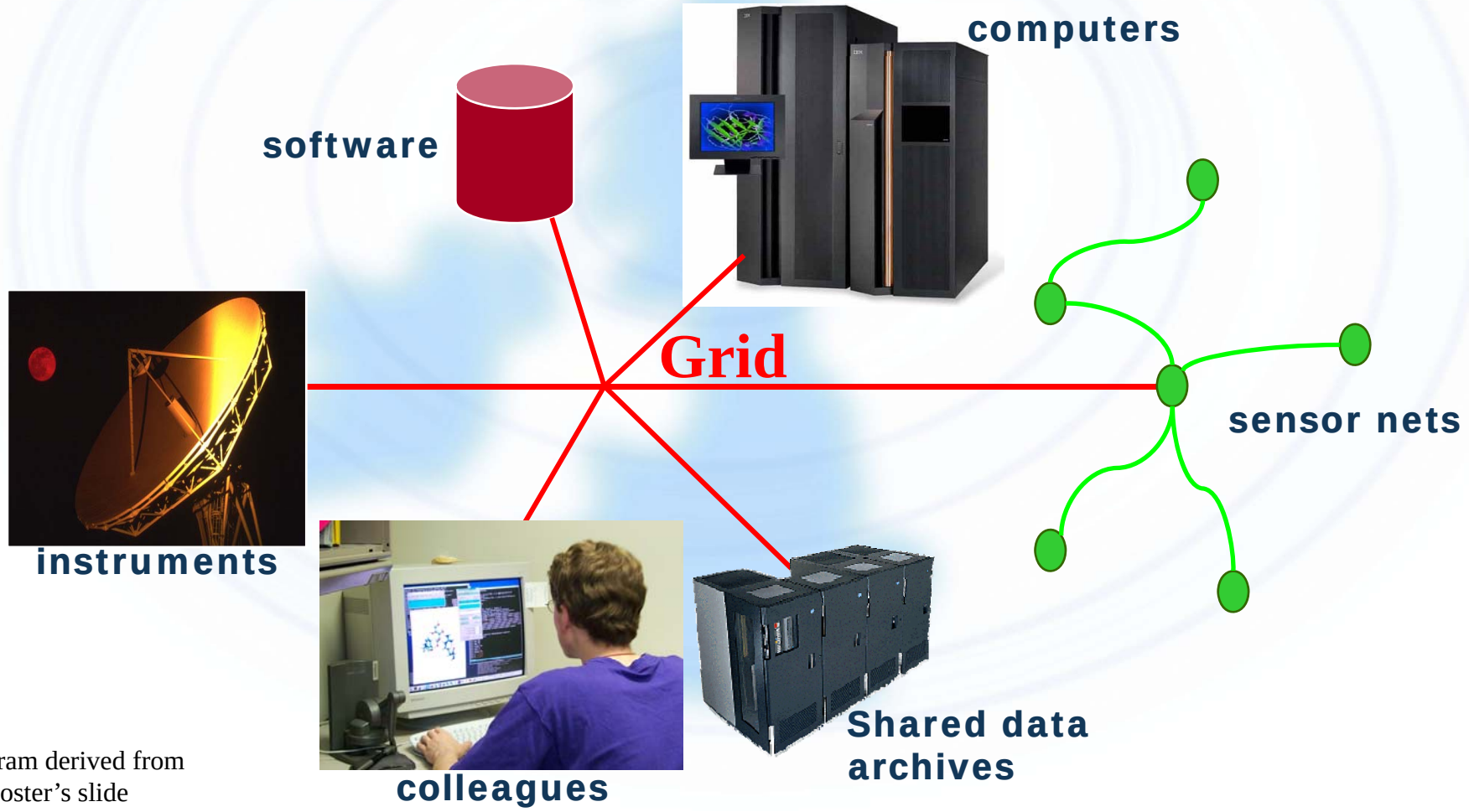
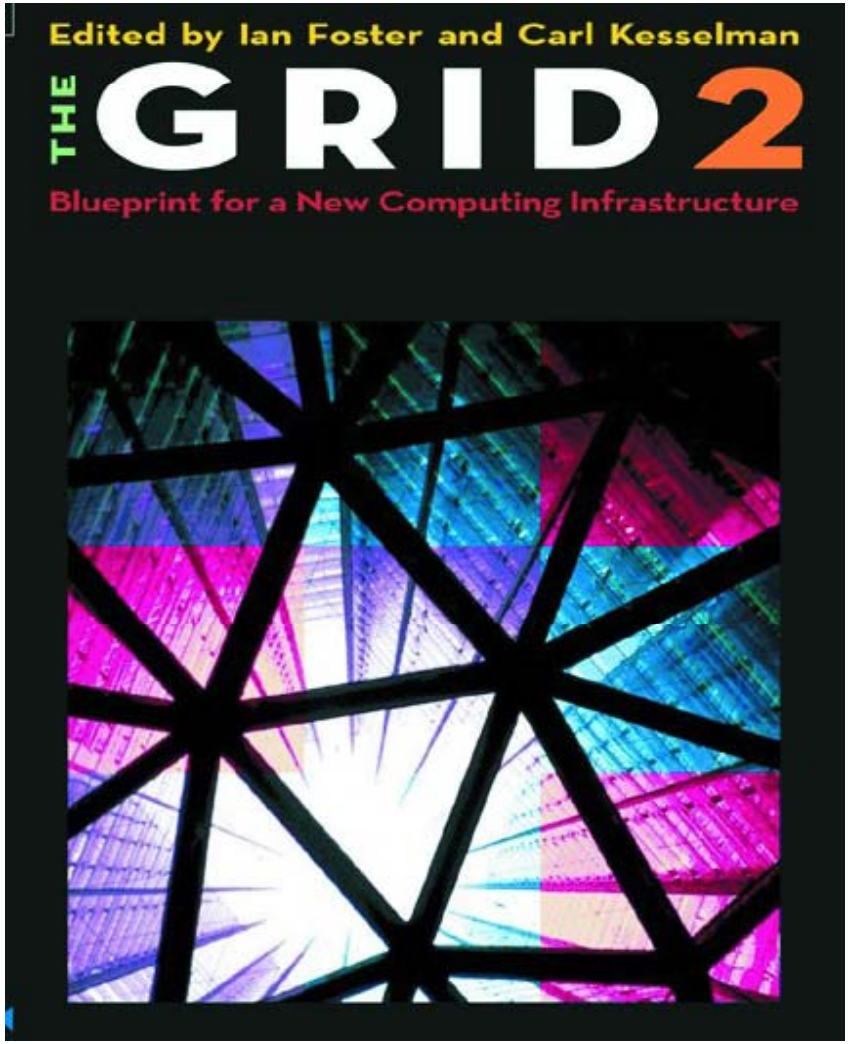


Diagram derived from
Ian Foster's slide

What is Grid Computing?

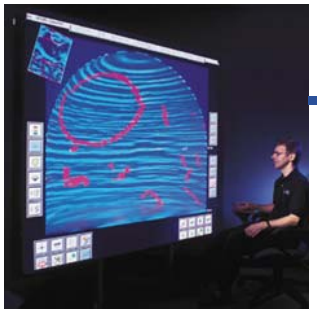
- **The grid vision is of "Virtual computing" (+ information services to locate computation, storage resources)**
  - Compare: The web: "virtual documents" (+ search engine to locate them)
- **MOTIVATION: collaboration through sharing resources (and expertise) to expand horizons of**
  - Research
  -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



The Grid Metaphor

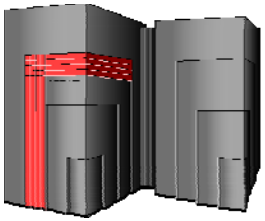
Mobile Access

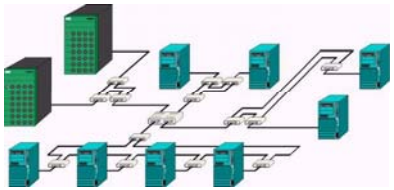
Workstation



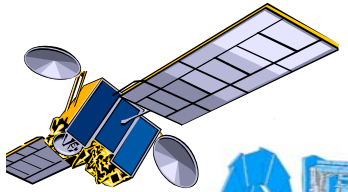
Visualising

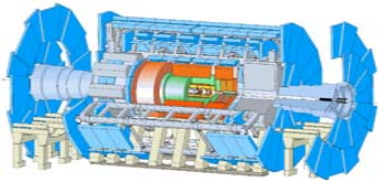
GRID MIDDLEWARE



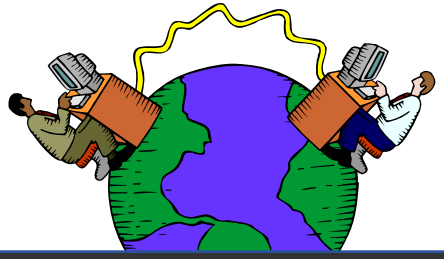


Supercomputer, PC-Cluster





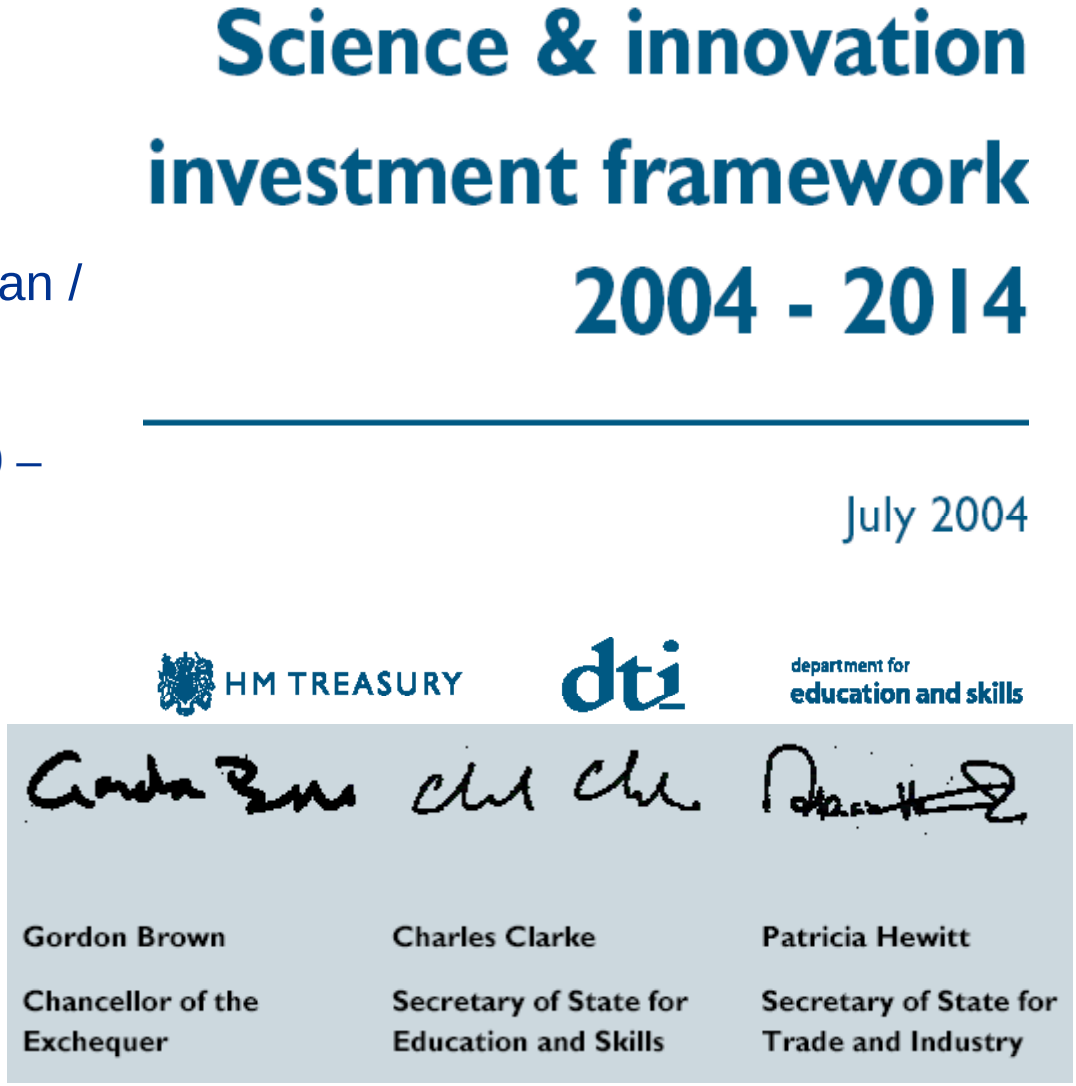
Data-storage, Sensors, Experiments



Internet, networks

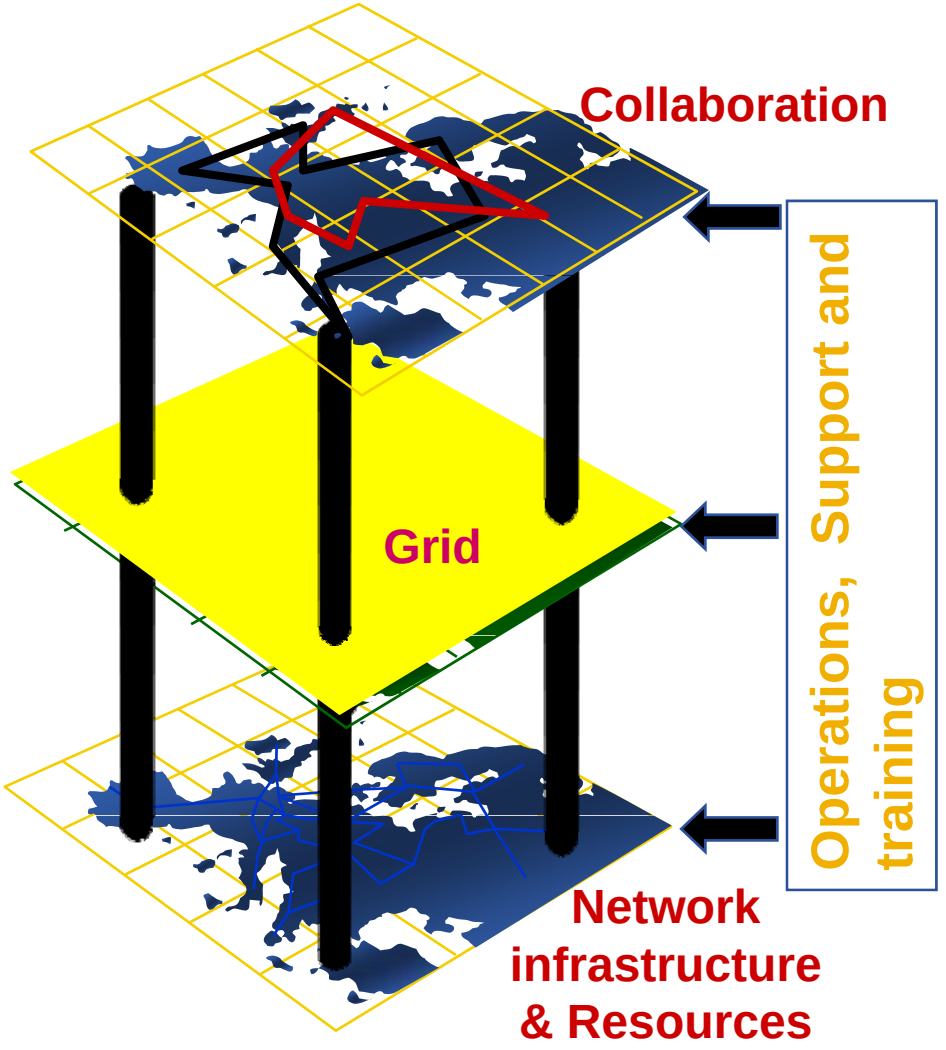
What is e-Infrastructure? – Political view

- **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



Global Drivers of e-Research

- **Digital technology – exponential growth - e.g. bandwidth**
- **Opportunities for e-Infrastructure to support faster, better, different research**
  - Sharing expertise
    - Support for cooperation and communication
  - Sharing computation services
    - E.g. to serve occasional peaks of high demand for computation (especially trivially parallelisable ones)
  - Sharing data
    - New sensors and instruments
    - Databases
- **Based on an infrastructure that requires *and enables* multidisciplinary research**
    - Requires: IT + domain specialists
    - Enables: New interdisciplinary research

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!

- **Grid computing: virtual computing across administrative domains**

Story so far

- **“e-science” and “e-research”**
  - Collaborative
  - Sharing resources
  - Often data, often compute intensive
- **Grids**
  - Permit resources in different administrative domains to be viewed as if one computer
  - Single sign-on
- **e-Infrastructure**
  - Networks + grids + resources + operations, support, training
  - Sustainable services that underpin e-science, commerce…..

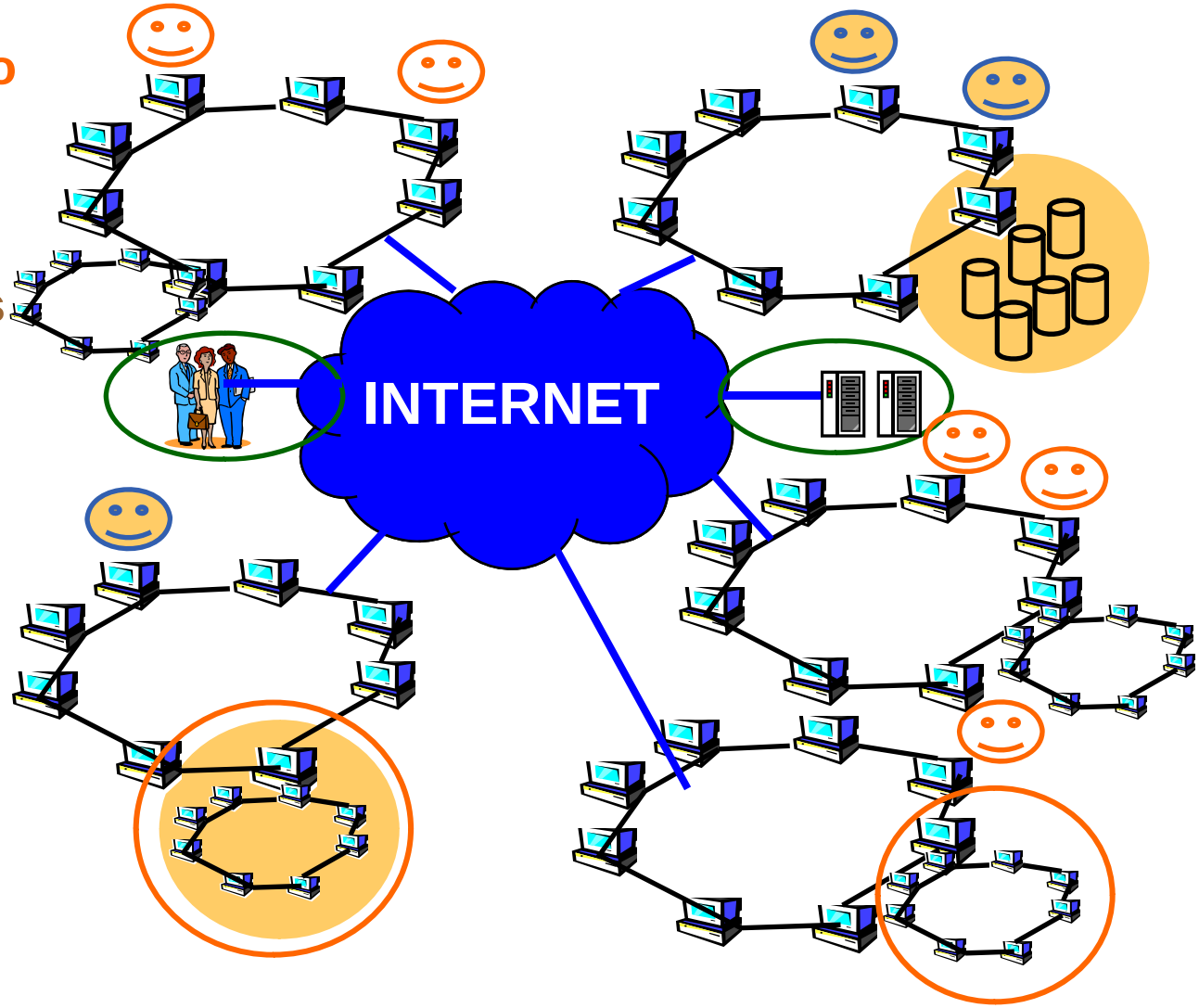
Grid concepts

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**
- **Grid computing is much more than "just software":**
  - 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)

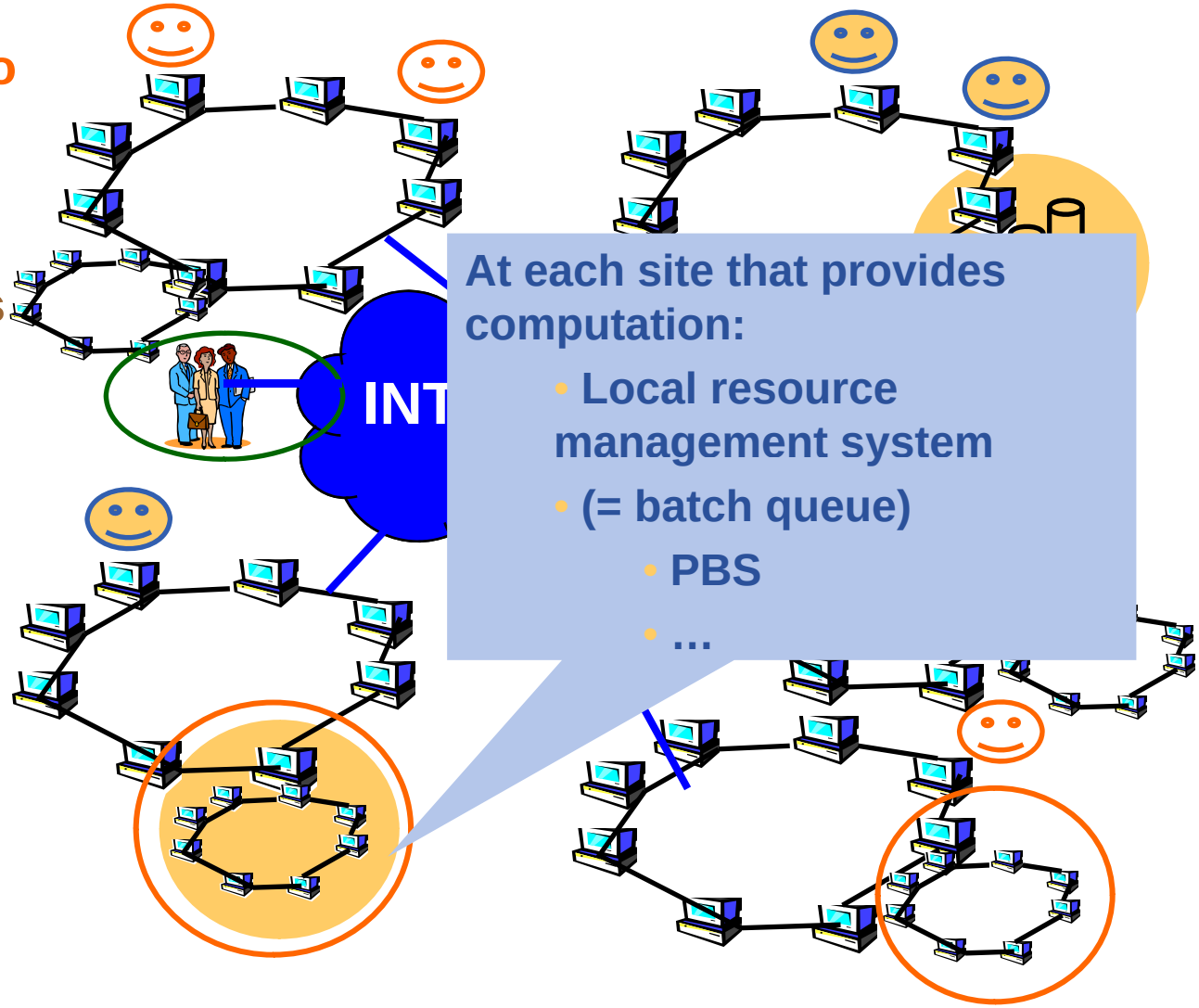
Typical current grid

- **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**



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**

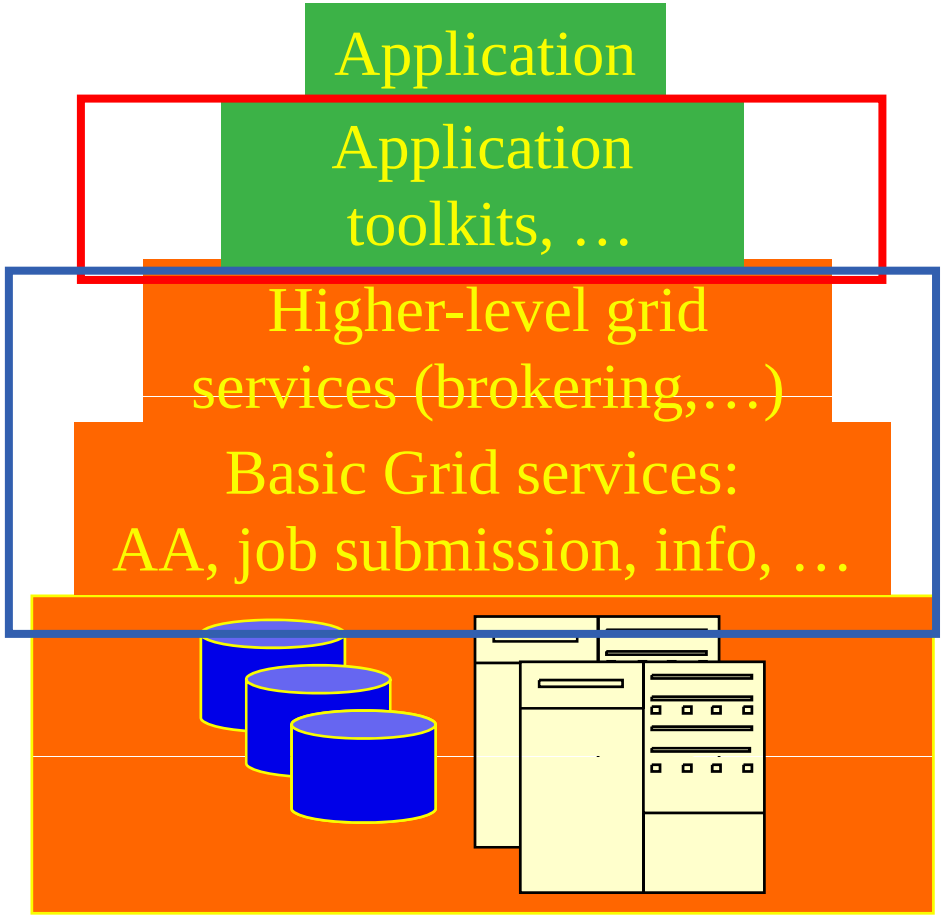


Grid Middleware

- **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**

Empowering VO's



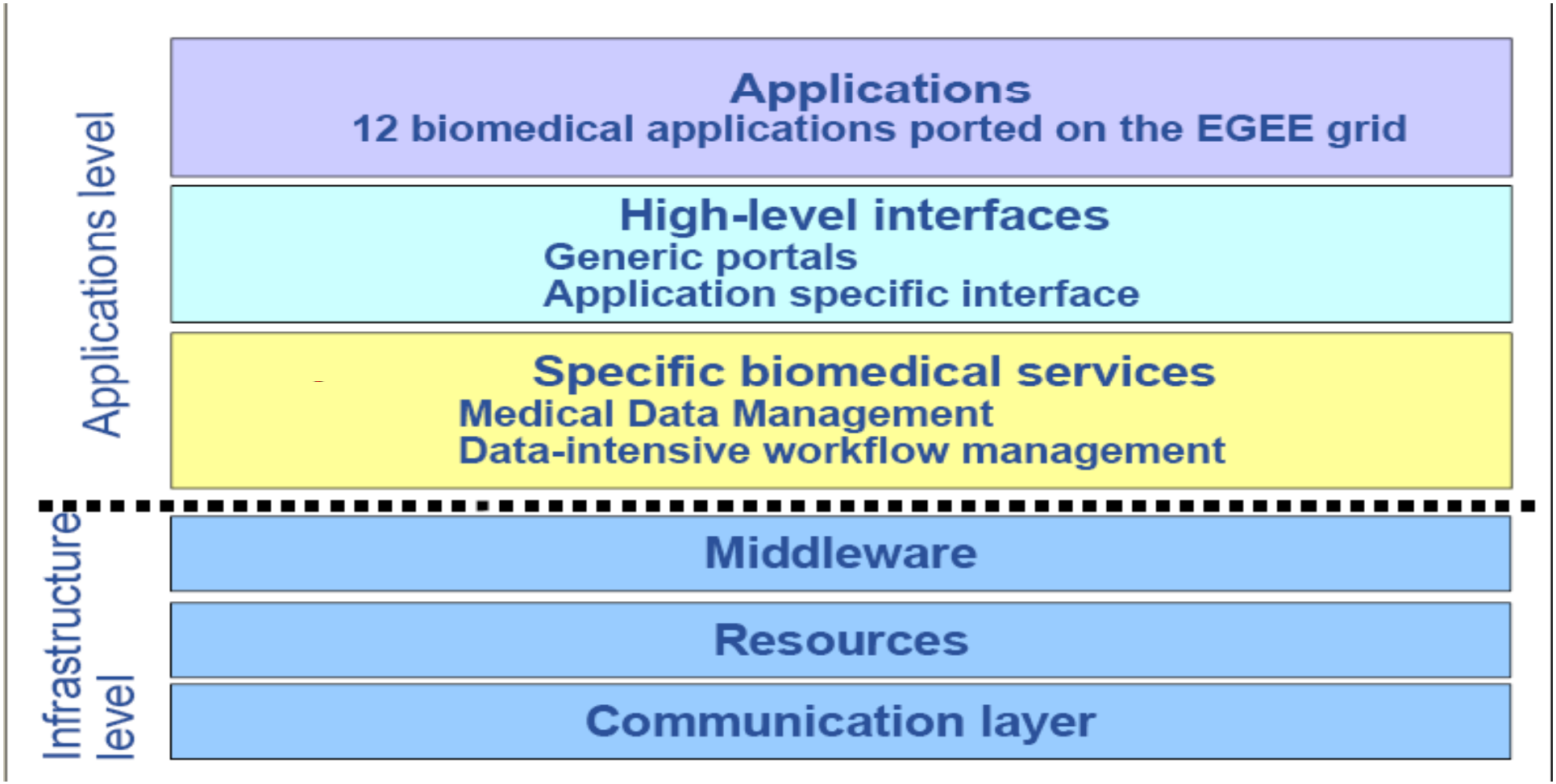
Where computer science meets the application communities!

VO-specific developments:

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

Production grids provide these services.

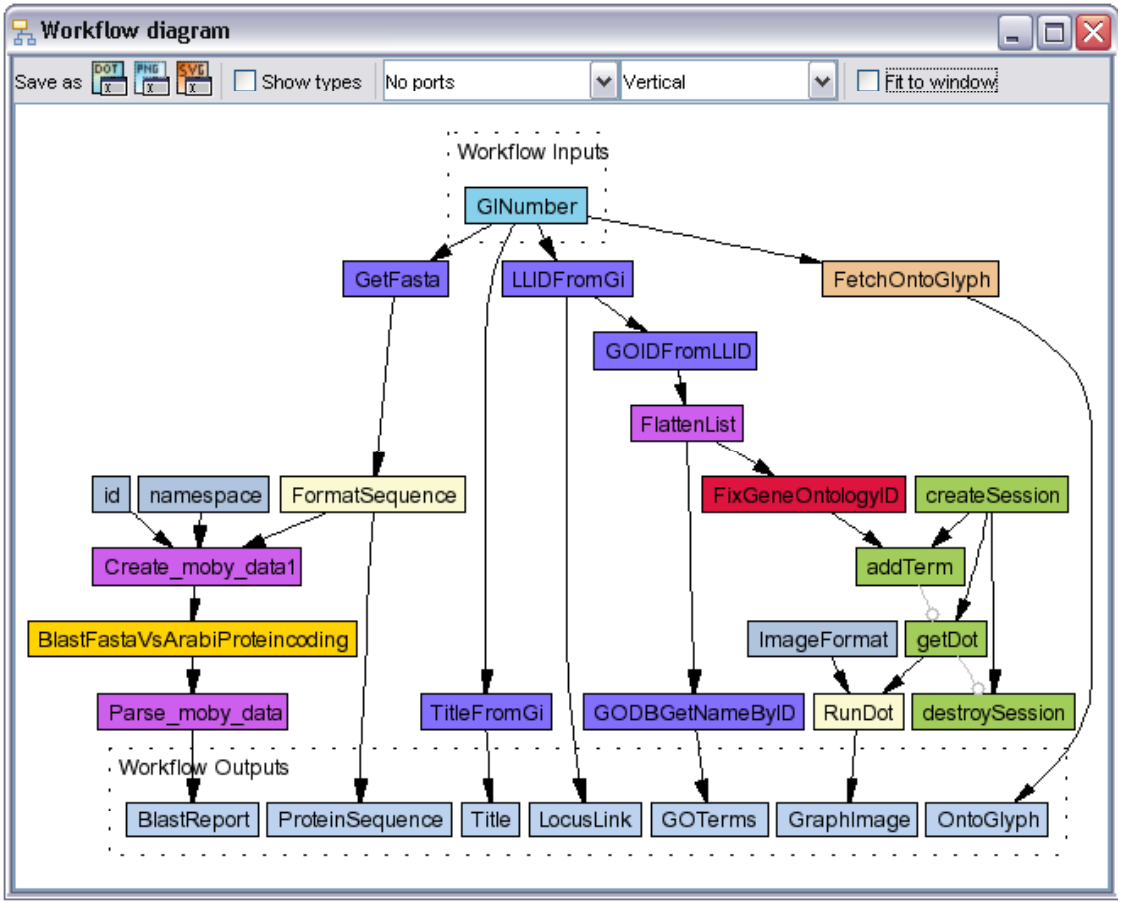
Example – Biomedical applications



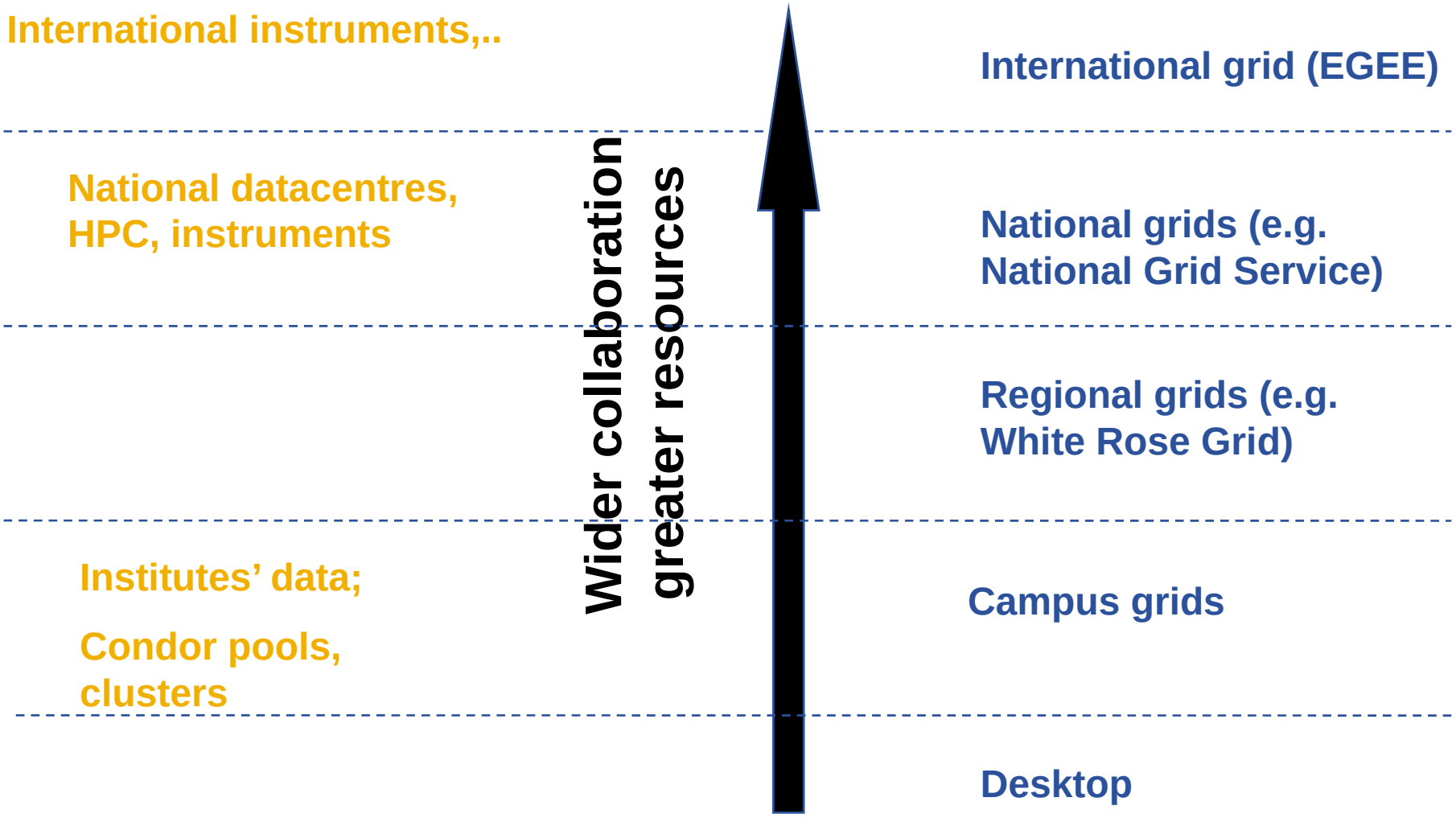
Biomedical community and the Grid, EGEE User Forum, March 1st 2006, I. Magnin

Workflow example

- 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

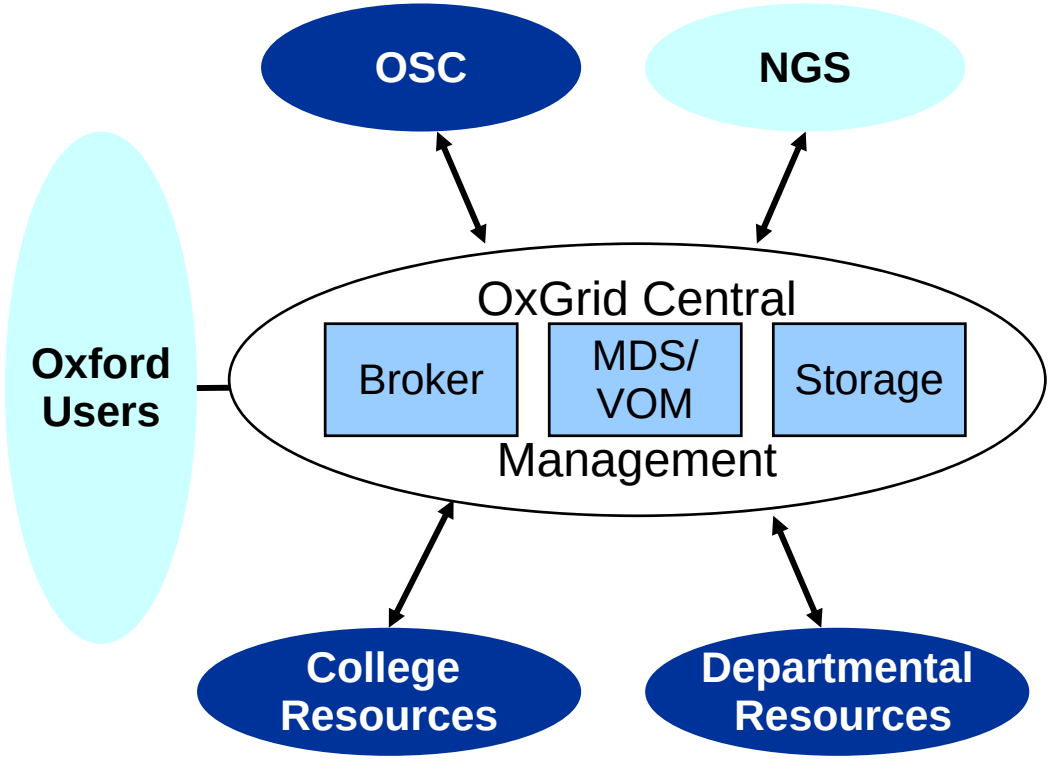


The many scales of grids



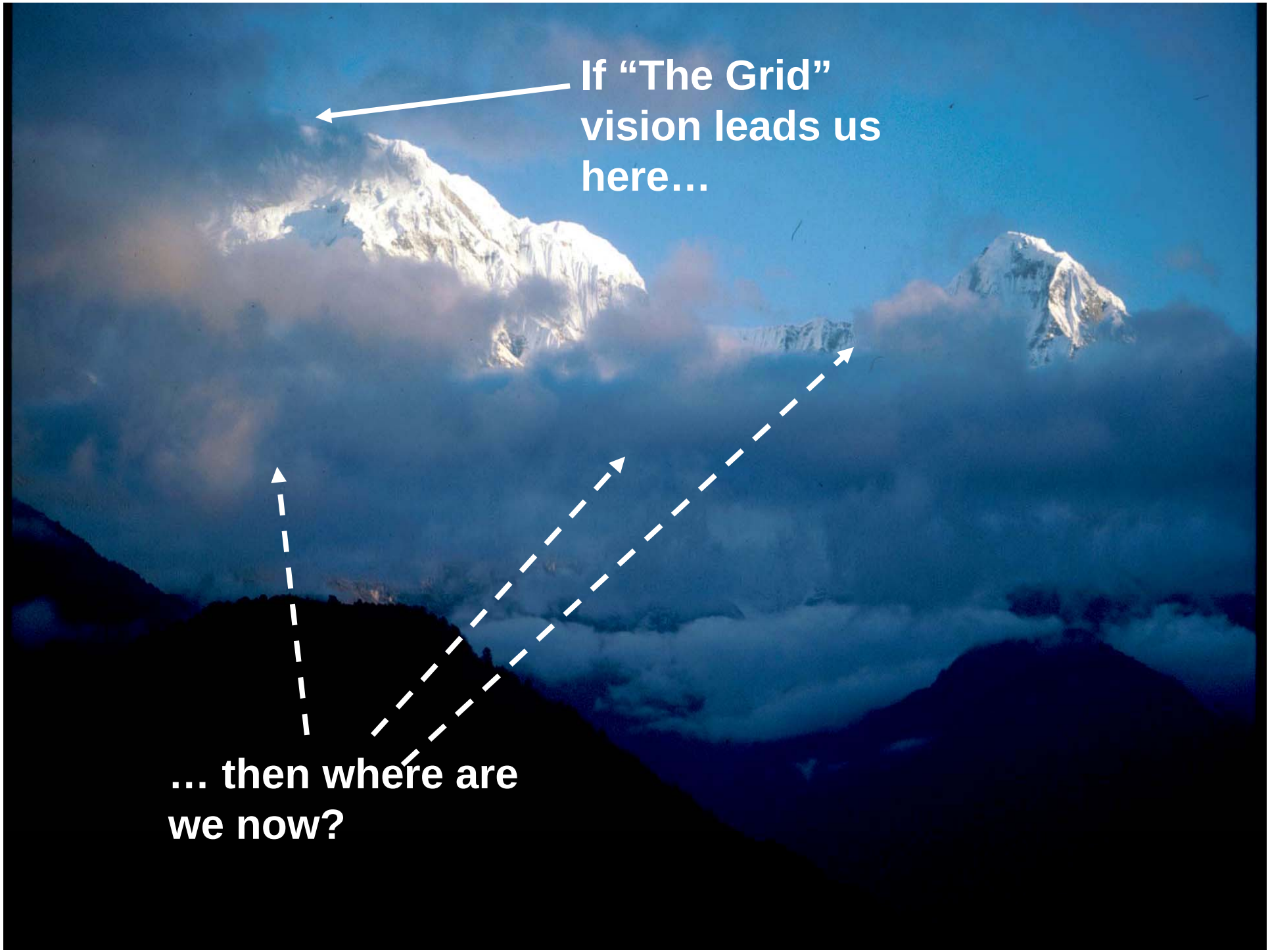
Example: OxGrid, Oxford University Campus Grid

- Single entry point for Oxford users to shared and dedicated resources
- Seamless access to National Grid Service and Oxford Supercomputing Centre for registered users
- Single sign-on



David Wallom

National e-Science Centre


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 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. Open Grid Forum, http://www.gridforum.org/)
  - **Sustainable infrastructure**
    - Move from research funding to GEANT-like model
    - European Grid Infrastructure federating National Grid Infrastructures

Where are we now? –user's view

Networks

Grids

Web

Arts
Humanities
e-Soc-Sci

Sciences,
engineering

Early production grids:
International - EGEE

Types of use:

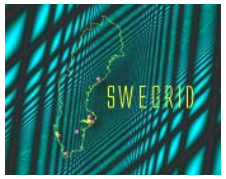
Service-oriented, workflow, "legacy" data

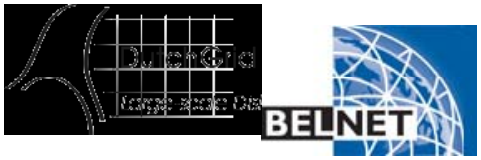
High throughput, new data

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

National grid initiatives now include…

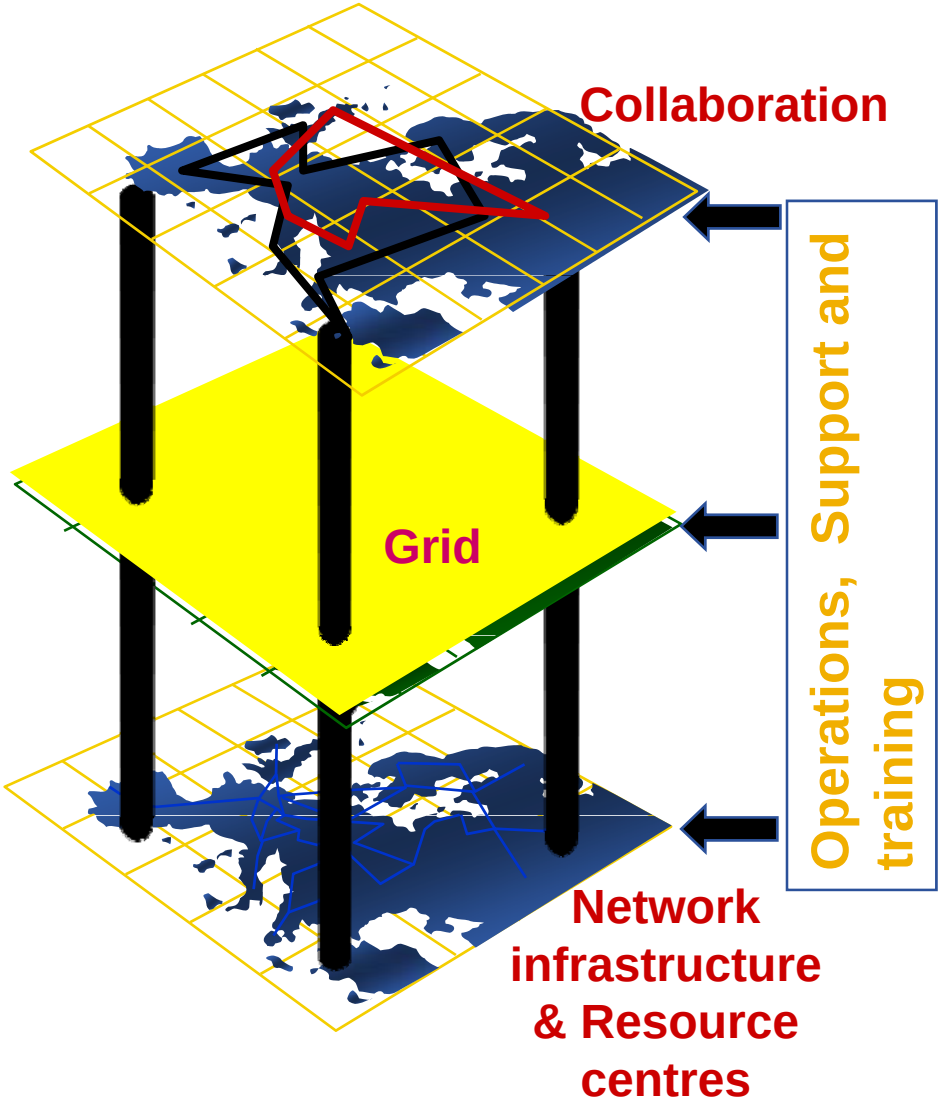




CroGrid

Grid concepts - Summary

- **Grids enable virtual computing across administrative domains**
  - Resources share authorisation and authentication
  - Social as well as technical challenges
- **Motivations:**
  - Collaborative research, diagnostics, engineering, public service,..
  - Resource utilisation and orchestration



Further reading

- **Open Grid Forum** http://www.ggf.org/
- **The Grid Cafe** www.gridcafe.org
- **Grid Today** http://www.gridtoday.com/
- **International Science Grid This Week** http://www.isgtw.org/
- **Training material and events** http://www.nesc.ac.uk/training
- UK All Hands Meeting 2007 – Nottingham this week

Pause for thought

- "Does Grid Computing open new horizons for my research, my School, my University?"
  - With whom do you collaborate?
  - What resources (data, programs, computers, people) can be orchestrated to benefit your research?
  - Benefit of a campus or regional grid??
- Any questions so far??
- Next talk: glance at grids at national and international scales