egee

What is Grid Computing?

Mike Mineter
Training Outreach and Education
National e-Science Centre

mjm@nesc.ac.uk

www.eu-egee.org

Information Society

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

Enabling Grids for E-sciencE

- **Introduction to**
  - e-Research and e-Science
  - 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.'
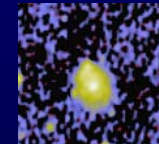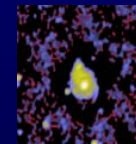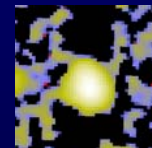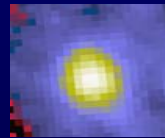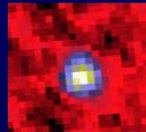
John Taylor

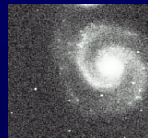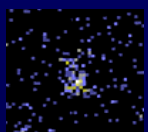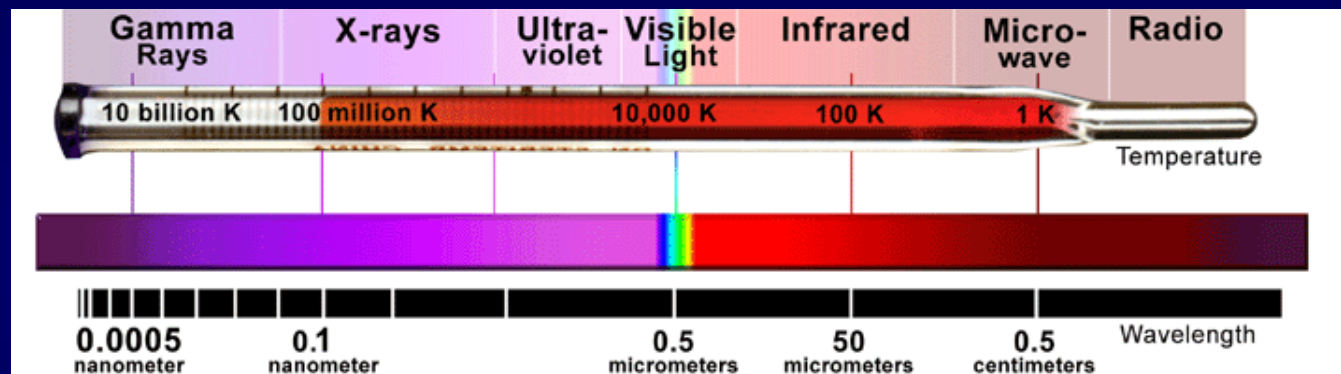Director General of Research Councils

Office of Science and Technology

Virtual Observatories

Observations made across entire electromagnetic spectrum

ROSAT ~keV   DSS Optical   2MASS 2μ   IRAS 25μ   IRAS 100μ   GB 6cm   NVSS 20cm   WENSS 92cm

$\Rightarrow$e.g. different views of a local galaxy

Need all of them to understand physics fully

Databases are located throughout the world

Peter Clarke

Biomedical Research Informatics Delivered by Grid Enabled Services

VO Authorisation

CFG Virtual Organisation

Publically Curated Data

Ensembl
OMIM
SWISS-PROT
MGI
HUGO
RGD
...

Glasgow
Private data

Edinburgh
Private data

Information Integrator

DATA HUB

Leicester
Private data

Oxford
Private data

London
Private data

Netherlands
Private data
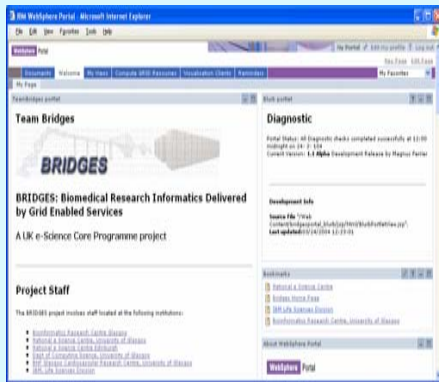
Synteny Grid Service

blast

+

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

Distributed Aircraft Maintenance Environment — DAME
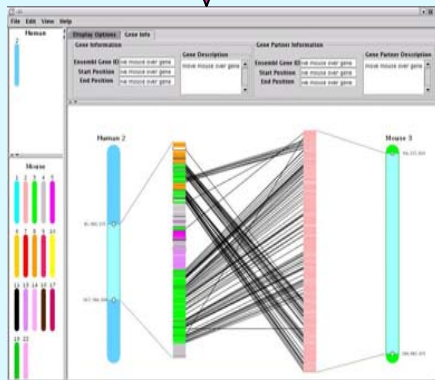
DAME: Grid based tools and Infer-structure for Aero-Engine Diagnosis and Prognosis
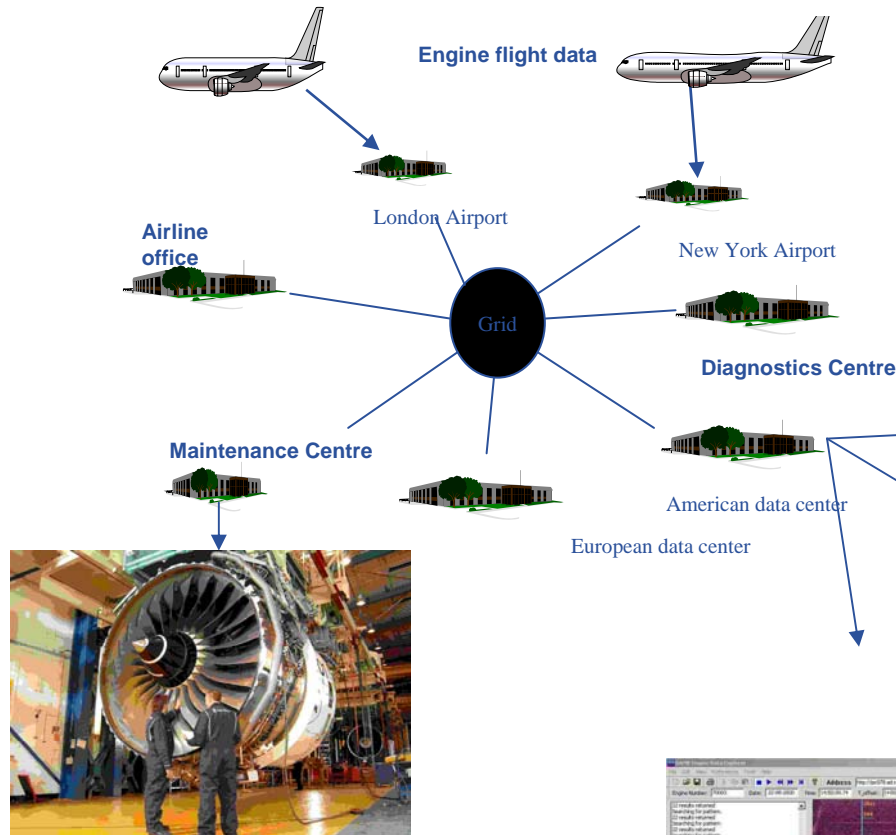
• "A Significant factor in the success of the Rolls-Royce campaign to power the Boeing 7E7 with the Trent 1000 was the emphasis on the new aftermarket support service for the engines provided via DS&S. Boeing personnel were shown DAME as an example of the new ways of gathering and processing the large amounts of data that could be retrieved from an advanced aircraft such as the 7E7, and they were very impressed", DS&S 2004
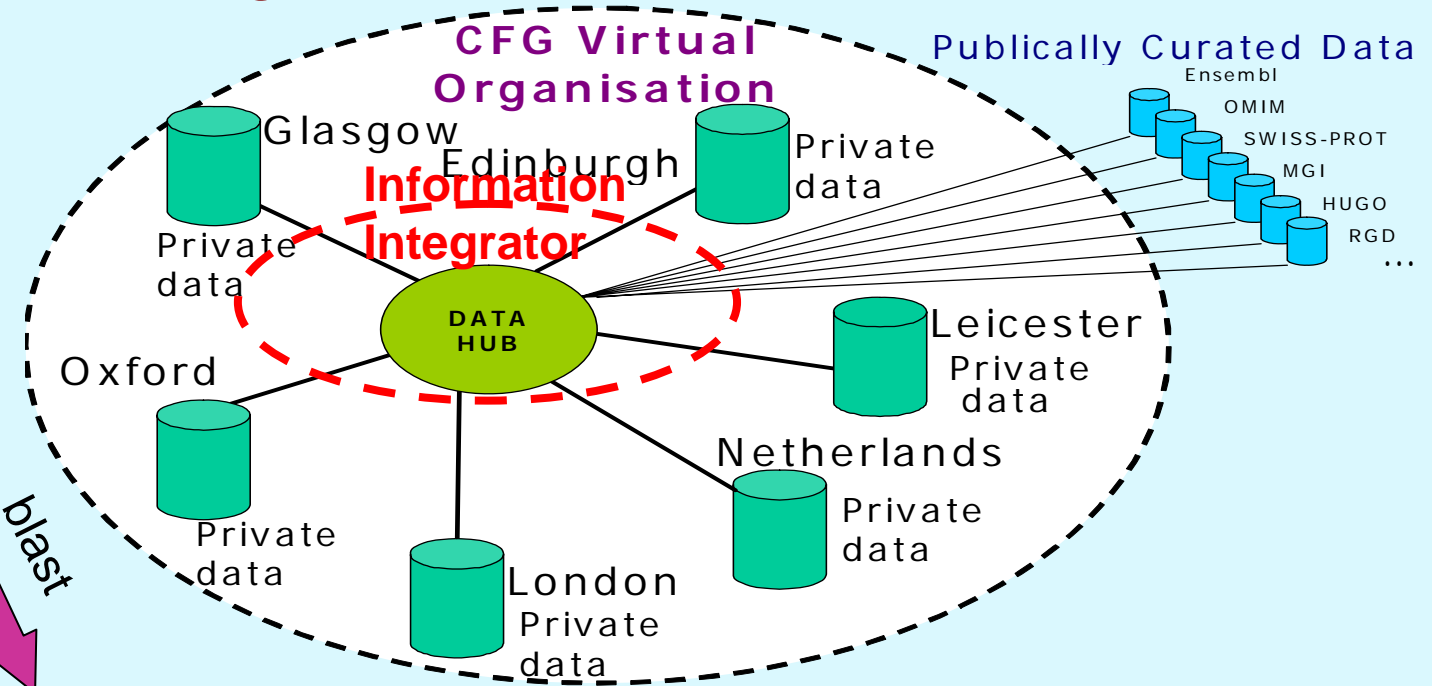
Engine flight data

London Airport

New York Airport

Airline office

Grid

Maintenance Centre

Diagnostics Centre

American data center

European data center
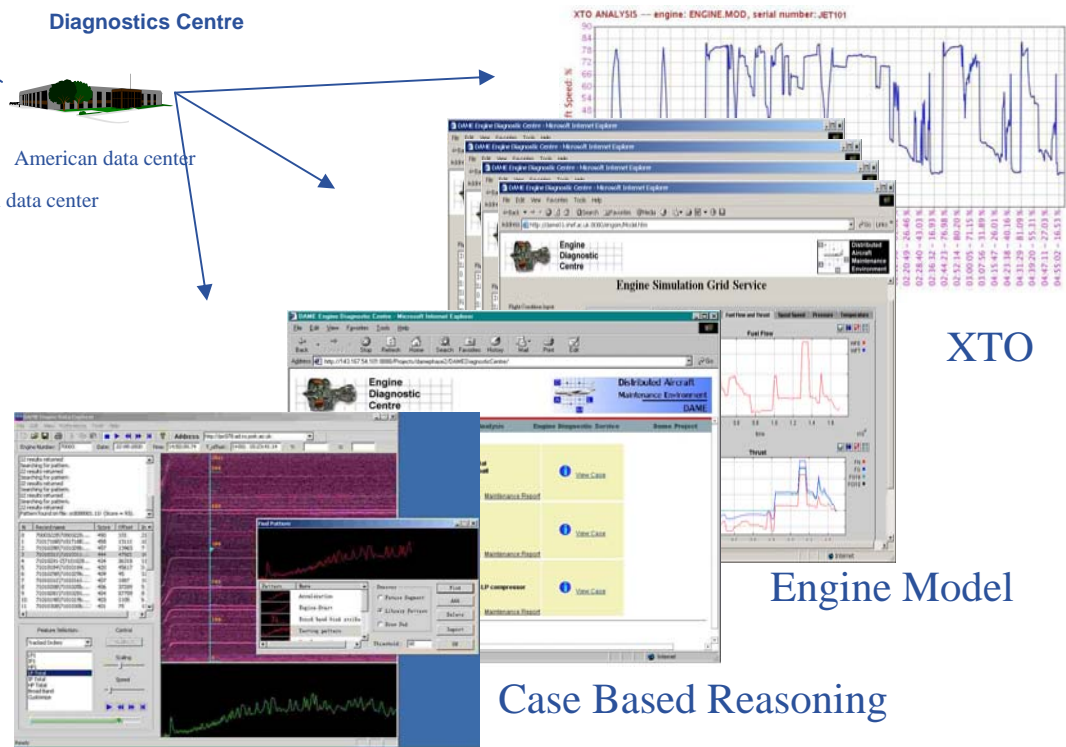
XTO

Engine Model

Case Based Reasoning

Companies:	Universities:
Rolls-Royce	York,
DS&S	Leeds,
Cybula	Sheffield, Oxford

climate*prediction*.net and GENIE

- Largest climate model ensemble
- >45,000 users, >1,000,000 model years
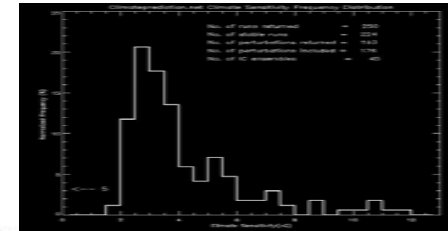
Response of Atlantic circulation to freshwater forcing
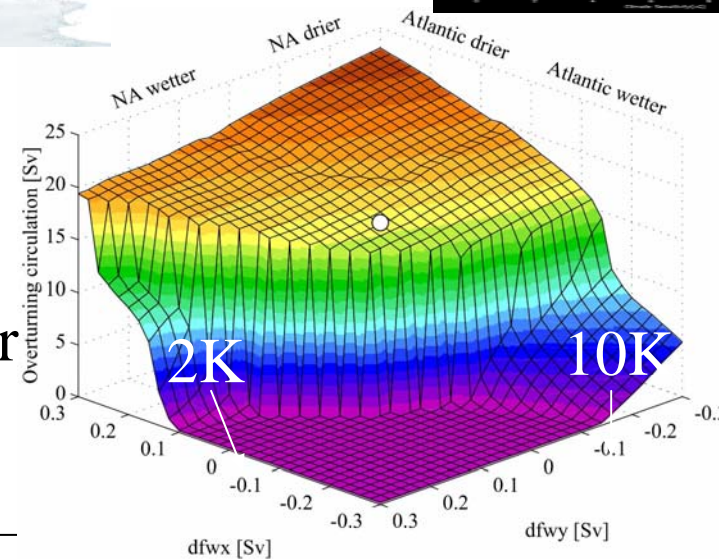
UK Grid for Particle Physics
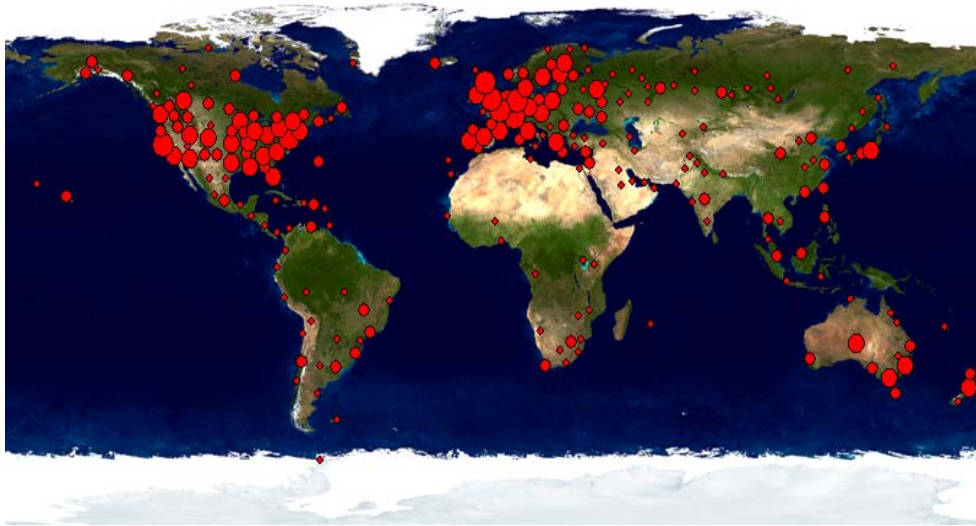
GridPP www.gridpp.ac.uk
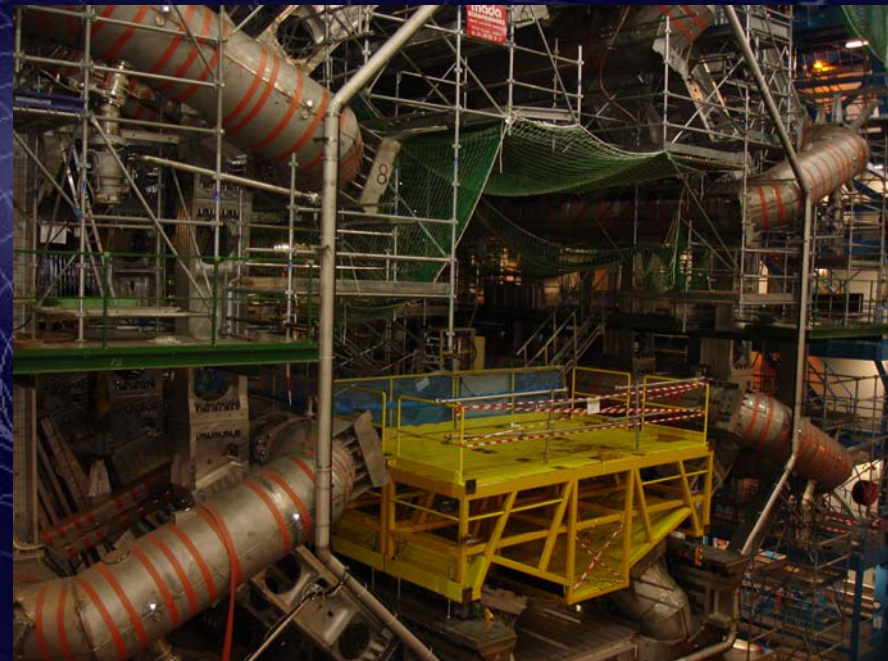
ATLAS detectors, 2/3/06

http://www.accessgrid.org/

Cameras

Microphones

Enabling Grids for E-sciencE

- **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

computers

software

Grid

instruments

sensor nets

colleagues

Shared data archives
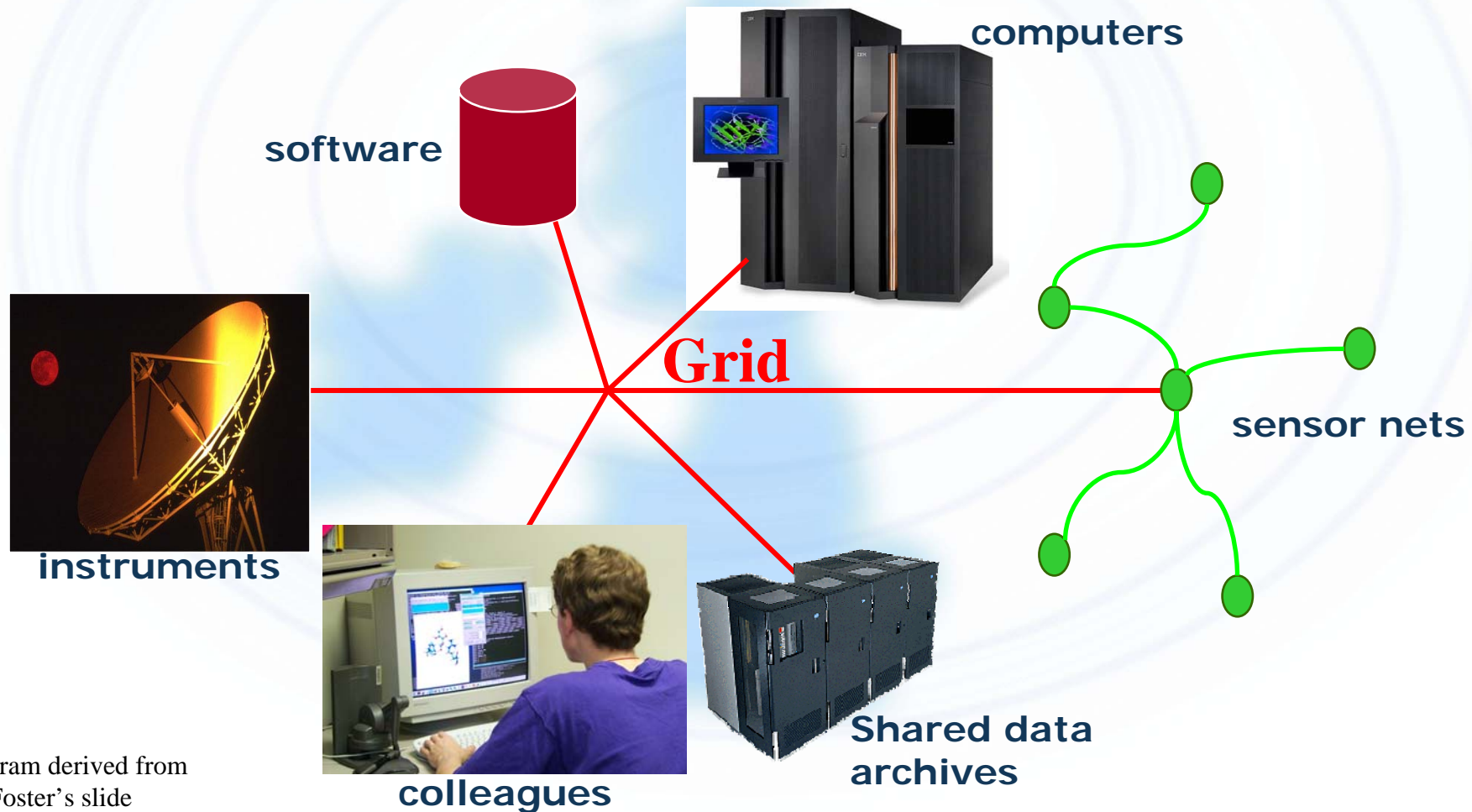
- **The grid vision is of "Virtual computing" (+ information services to locate computation, storage resources)**

    – 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,…

Edited by Ian Foster and Carl Kesselman

THE GRID 2

Blueprint for a New Computing Infrastructure

Mobile Access

Workstation

Visualising

GRID MIDDLEWARE

Supercomputer, PC-Cluster

Data-storage, Sensors, Experiments

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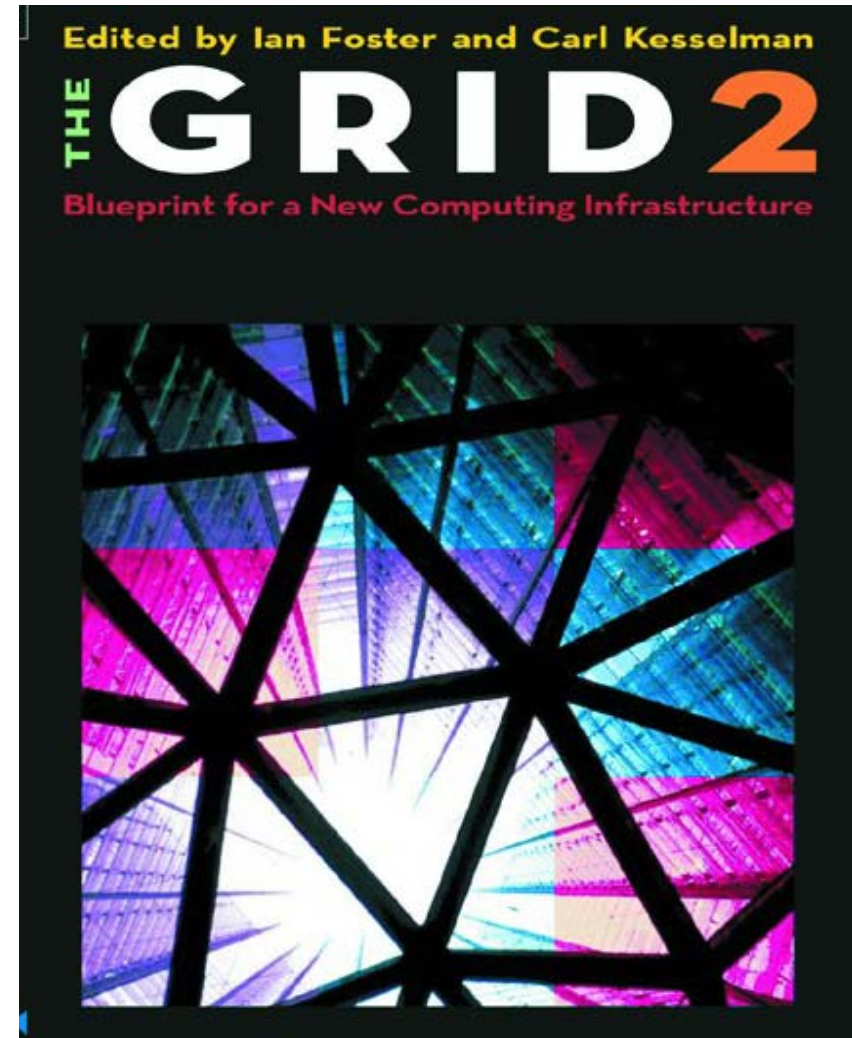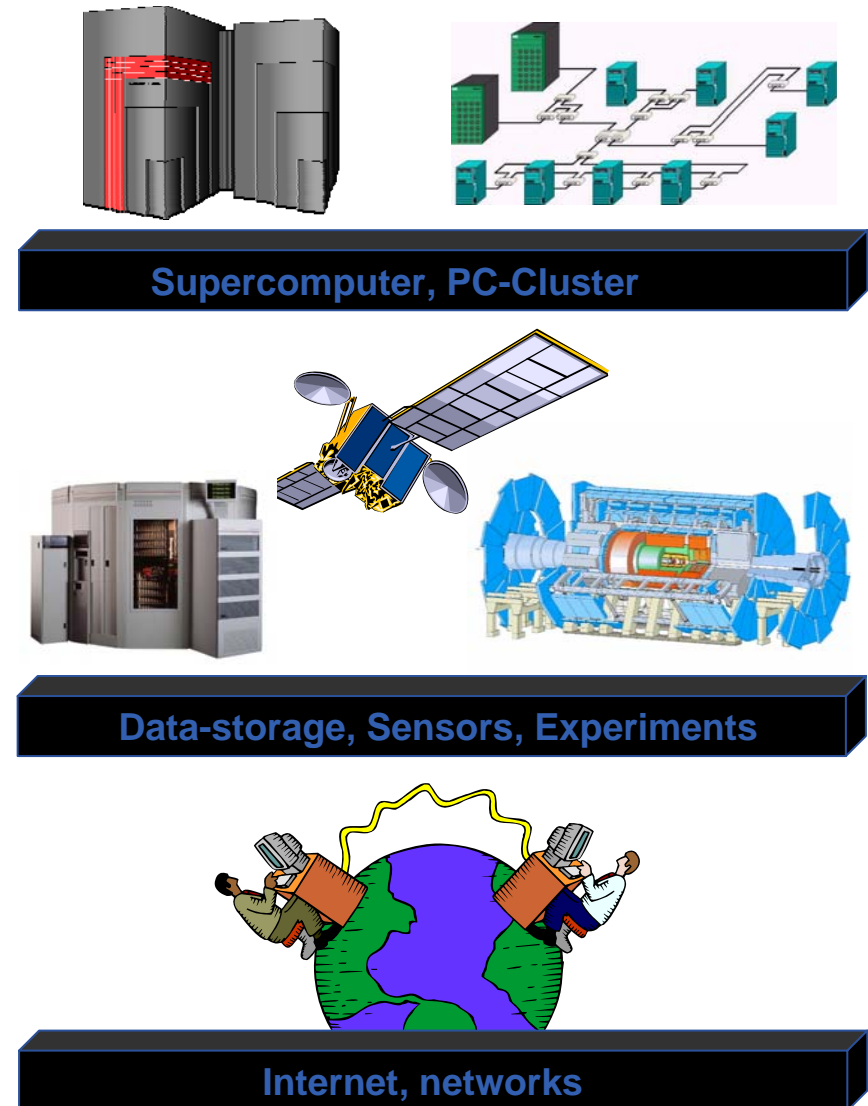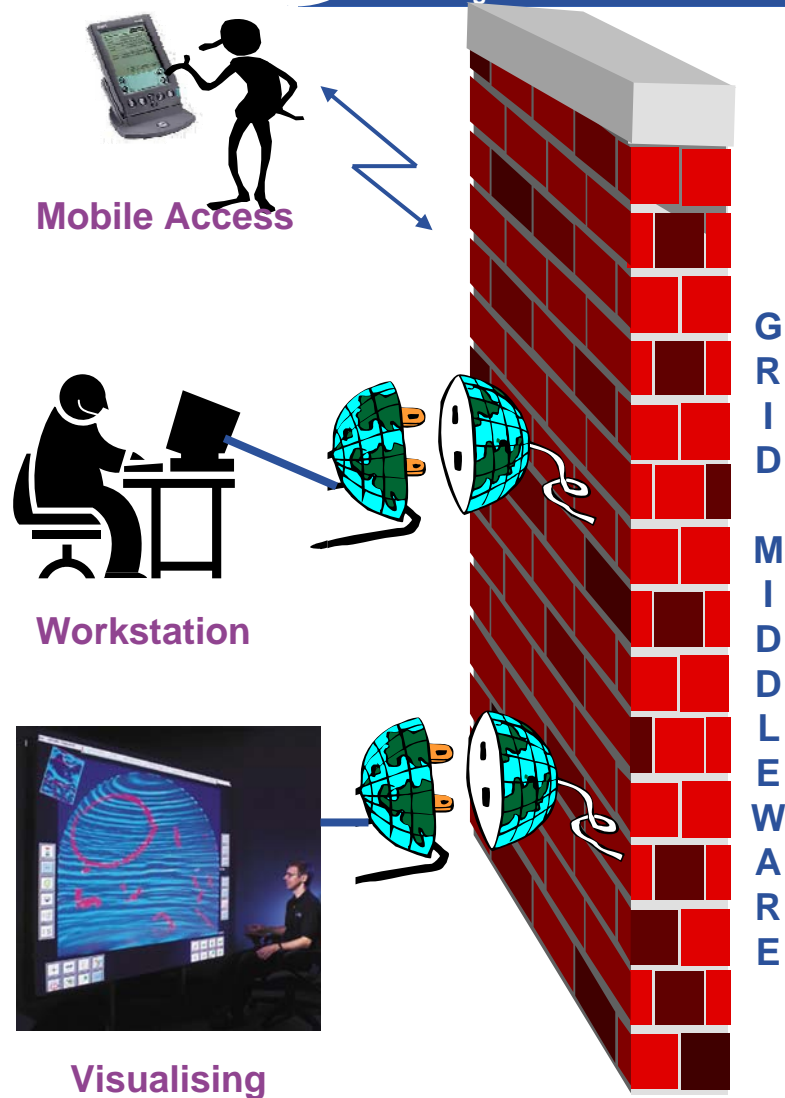
- **A shared resource**
  - That enables science, research, engineering, medicine, industry, …
  - It will improve UK / European / … productivity
    - Lisbon Accord 2000
    - E-Science Vision SR2000 – John Taylor
  - Commitment by UK government
    - Sections 2.23-2.25
  - Always there
    - c.f. telephones,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

Science & innovation investment framework

2004 - 2014

July 2004

HM TREASURY   dti   department for education and skills

Gordon Brown	Charles Clarke	Patricia Hewitt
Chancellor of the Exchequer	Secretary of State for Education and Skills	Secretary of State for Trade and Industry

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

Collaboration

Grid

Operations, Support and training

Network infrastructure & Resources
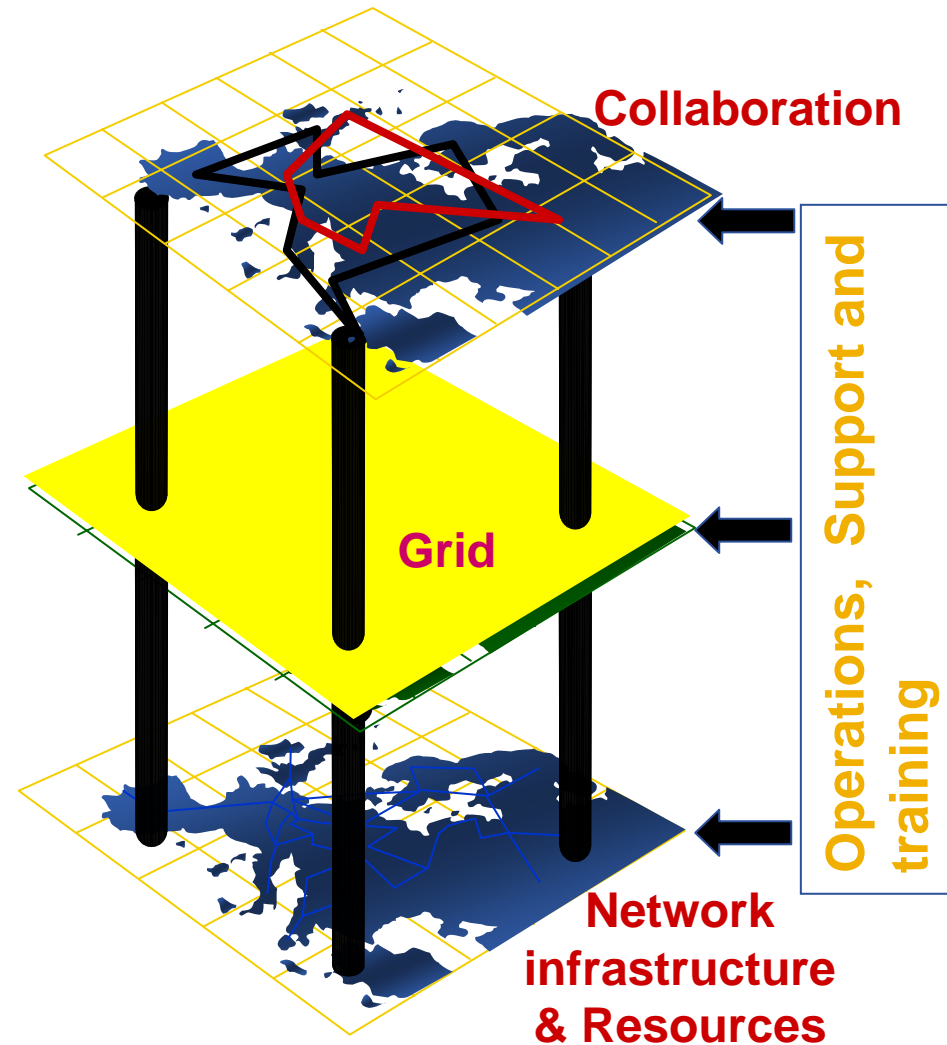
- **Digital technology – exponential growth - e.g. bandwidth**
- **Opportunities for e-Infrastructure to support faster, better, different research**
  - Sharing expertise
    - Support for cooperation and communication
  - Sharing computation services
    - E.g. to serve occasional peaks of high demand 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

Enabling Grids for E-sciencE

- **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 concepts

- **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)**

Enabling Grids for E-sciencE

- **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**

INTERNET

egee

Application

Application toolkits, standards

Middleware: "collective services"

Basic Grid services: AA, job submission, info, …
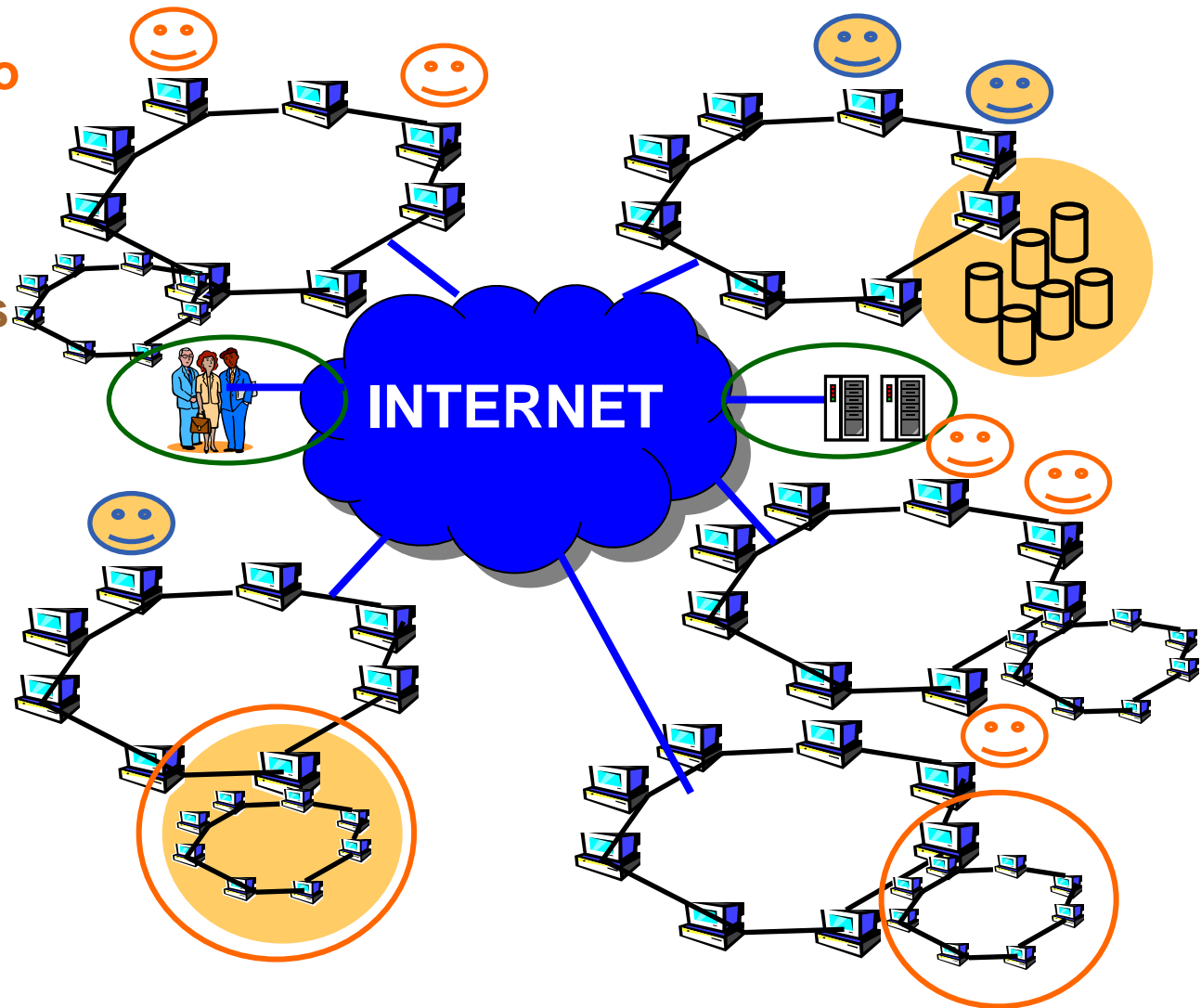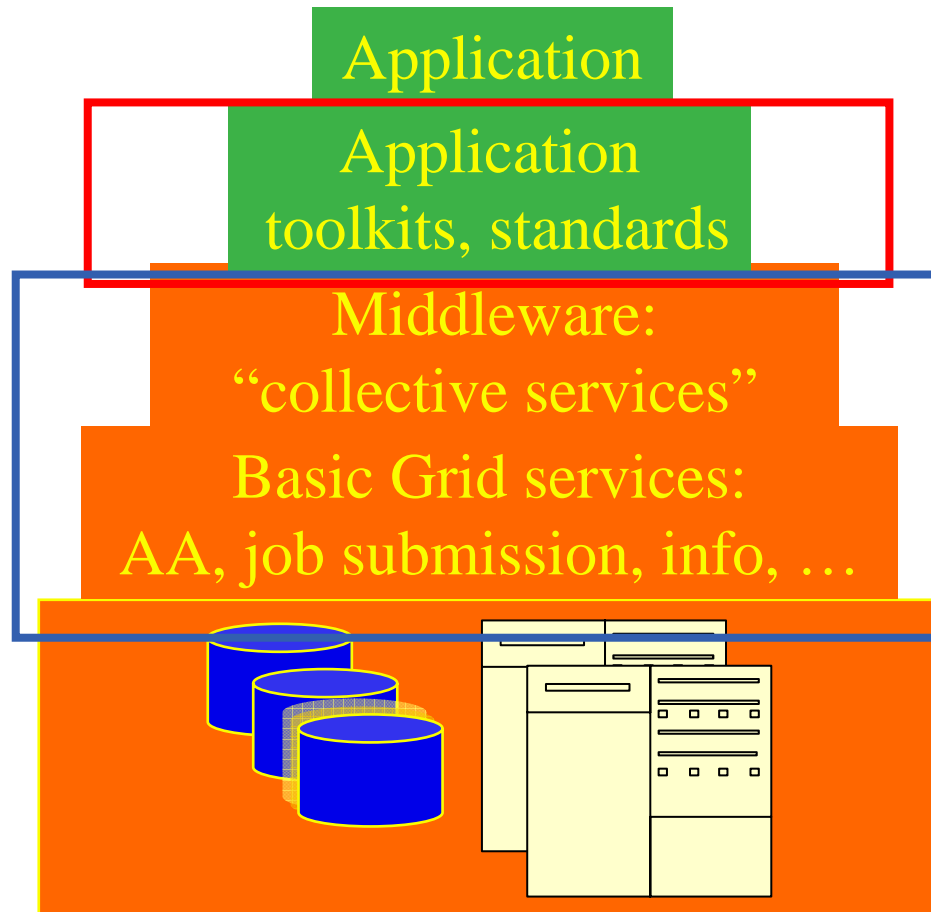
Where computer science meets the application communities!

VO-specific developments:

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

Production grids provide these services.

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**
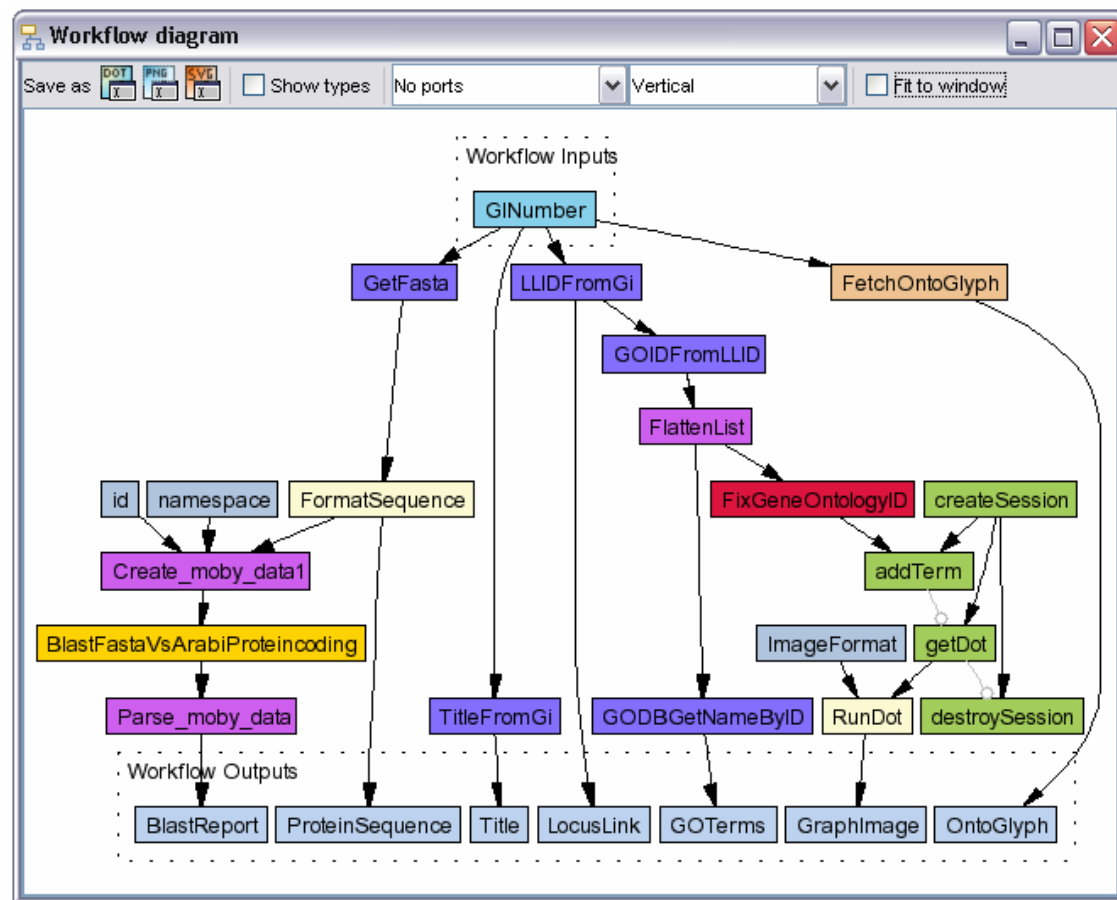
International instruments,..

International grid (EGEE)

National datacentres, HPC, instruments

National grids (e.g. National Grid Service)

Wider collaboration greater resources

Regional grids (e.g. White Rose Grid)

Institutes' data;

Condor pools

Campus grids

Desktop

Enabling Grids for E-sciencE

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

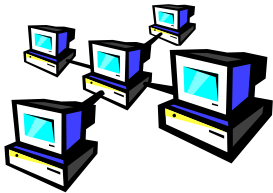**Information System**: Characteristics and status of resources

Computing Element (CE): A batch queue on a site's computers where the user's job is executed
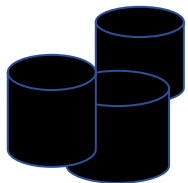
Storage Element (SE): provides (large-scale) storage for files

Enabling Grids for E-sciencE

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

User/Grid interface

Input files

Output files

File Replica Catalogue

Datasets info

Information Service

Resource Broker

Resource info

Input

Output

Author. &Authen

Job Submit Event

Job Query

Job Status

Publish

Storage Resource

Logging & Book-keeping

Job Status

Computing Resource = batch queue
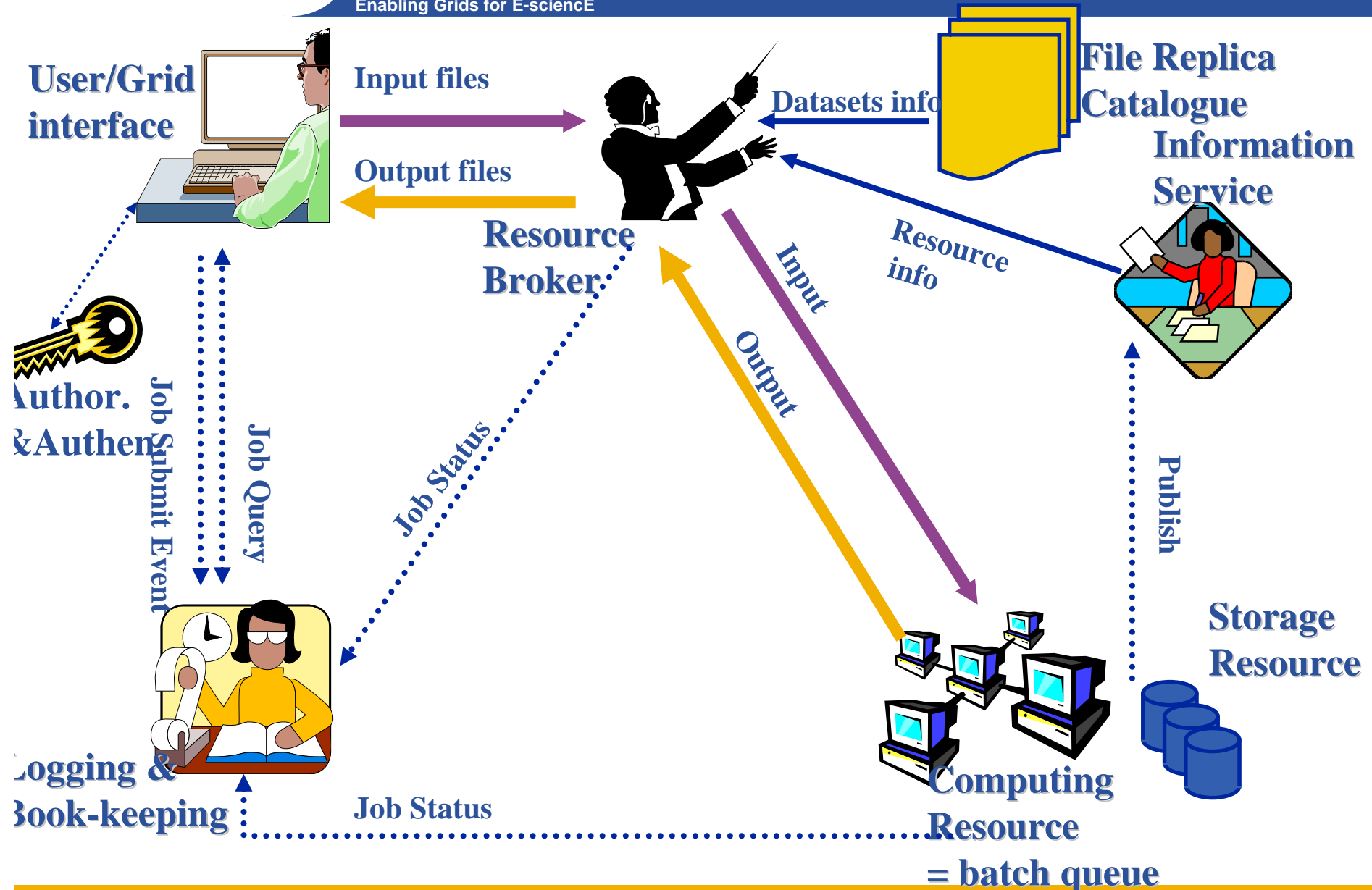
- **EU-funded project that has established the largest multi-VO production grid in the world!**

- **What's happening now?**
  http://gridportal.hep.ph.ic.ac.uk/rtm/

- **What resources are connected?**
  http://goc.grid-support.ac.uk/gridsite/monitoring/

- **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**

- **Achieved by Certification:**
  - User's identity has to be certified by one of the national *Certification Authorities* (CAs)
    - mutually recognized http://www.gridpma.org/
  - In UK go to http://www.grid-support.ac.uk/ca/ralist.htm to find CA's local "Registration Authorities"
  - Resources are also certified by CAs

- **User**
  - User joins a VO
  - Digital certificate is basis of AA
  - Identity passed to resources you use, where it is mapped to a local account

- ***Policies* express the rights for a Virtual Organization to use resources**

eGee

Enabling Grids for E-sciencE

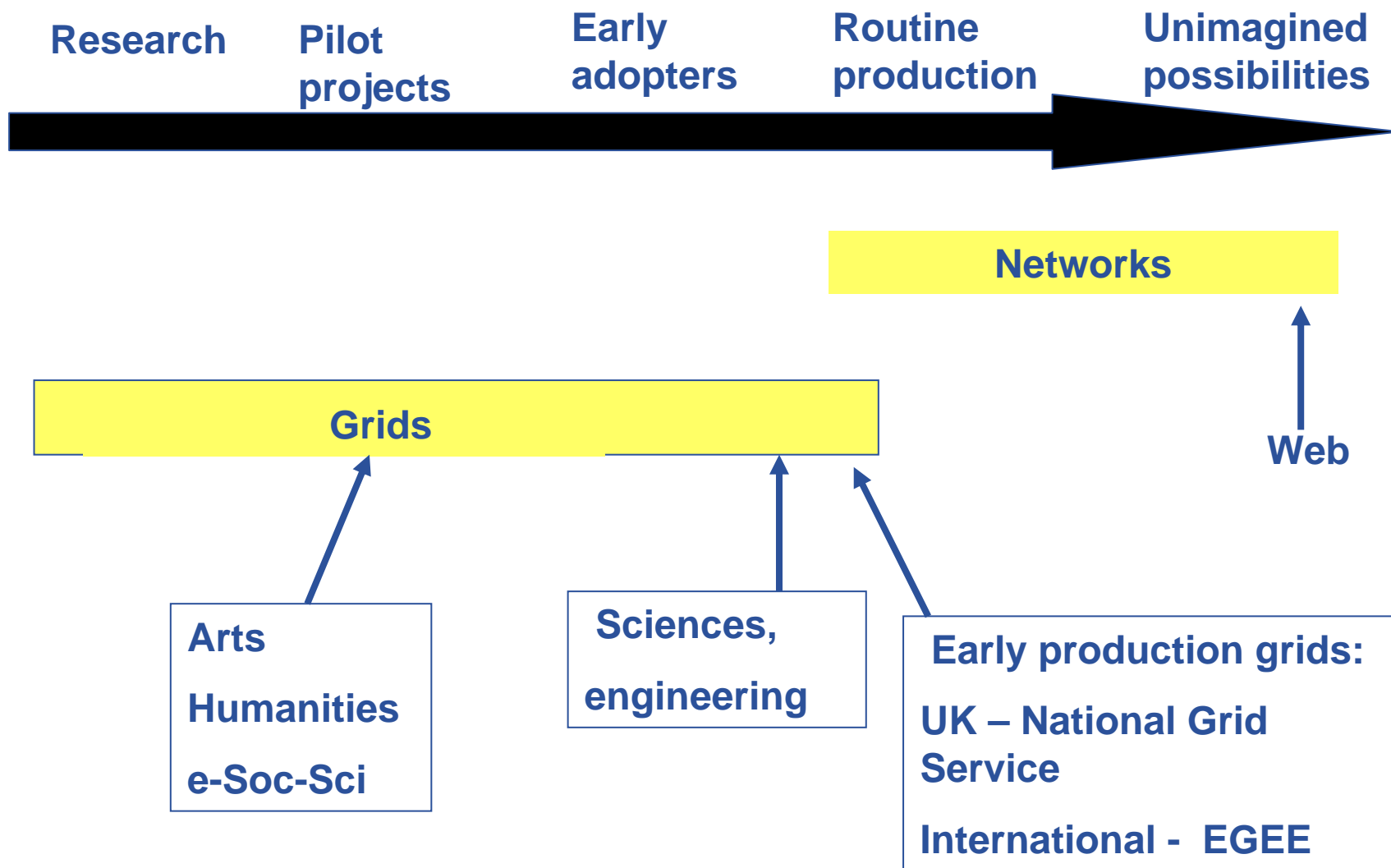## Many Grid development efforts — all over the world

- NASA Information Power Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF TeraGrid
- DOE ASCI Grid
- DOE Earth Systems Grid
- DARPA CoABS Grid
- NEESGrid
- DOH BIRN
- NSF iVDGL

- DataGrid (CERN, ...)
- EuroGrid (Unicore)
- DataTag   (CERN,…)
- Astrophysical Virtual Observatory
- GRIP (Globus/Unicore)
- GRIA (Industrial applications)
- GridLab (Cactus Toolkit)
- CrossGrid (Infrastructure Components)
- EGSO (Solar Physics)

- UK – OGSA-DAI, RealityGrid, GeoDise, Comb-e-Chem, DiscoveryNet, DAME, AstroGrid, GridPP, MyGrid, GOLD, eDiamond, Integrative Biology, …
- Netherlands – VLAM, PolderGrid
- Germany – UNICORE, Grid proposal
- France – Grid funding approved
- Italy – INFN Grid
- Eire – Grid proposals
- Switzerland - Network/Grid proposal
- Hungary – DemoGrid, Grid proposal
- Norway, Sweden - NorduGrid

- **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"**

Research

Pilot projects

Early adopters

Routine production

Unimagined possibilities

Networks

Grids

Web

Arts

Humanities

e-Soc-Sci

Sciences, engineering

Early production grids:

UK – National Grid Service

International - EGEE

- **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**

CroGrid

- **Grids: collaboration across administrative domains**

- **Networks: collaboration across geographical distance**

- **Semantics, ontologies: collaboration across disciplines / groups**

- **Storage, ("curation"): collaboration across time**

Collaboration

Grid

Operations, Support and training

Network infrastructure & Resource centres

Enabling Grids for E-sciencE
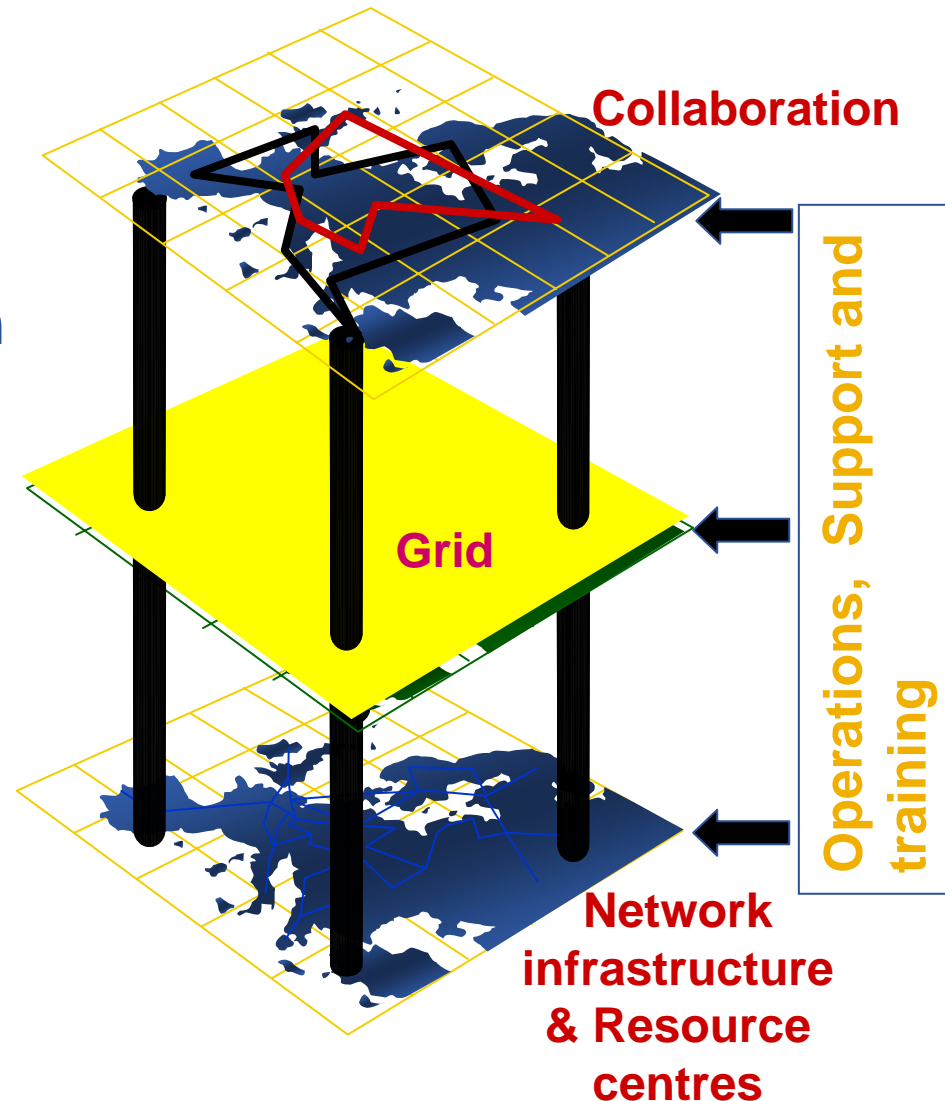
- **Ask not what "the Grid" can do for *you***

- **BUT**

- **With whom do you collaborate?**

- **What resources / services can you provide?**

- **What resources would empower your research?**

People

Data

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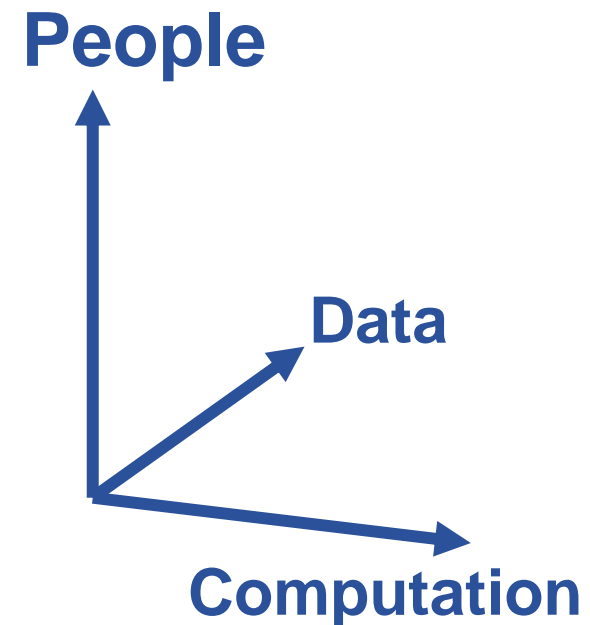
- **The Grid Core Technologies, Maozhen Li and Mark Baker, Wiley, 2005**

- **The Globus Toolkit 4 Programmer's Tutorial Borja Sotomayor, Globus Alliance, http://gdp.globus.org/gt4-tutorial/multiplehtml/index.html**

- **The Web Services Grid Architecture (WSGA) www.nesc.ac.uk/technical_papers/UKeS-2004-05.pdf**

- **http://java.sun.com/xml/webservices.pdf**

- **Global Grid Forum http://www.ggf.org/ (see GGF16)**