



Enabling Grids for E-scienceE

Overview of e-Infrastructure

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

mjm@nesc.ac.uk

www.eu-egee.org



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

- **Introduction to**
 - e-Research and e-Science
 - Grids
 - e-Infrastructure
- **Grid concepts**
- **Grids - Where are we now?**
- **Enabling the research of the future**
 - and for early adopters... the present!

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

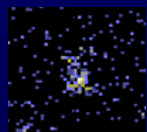
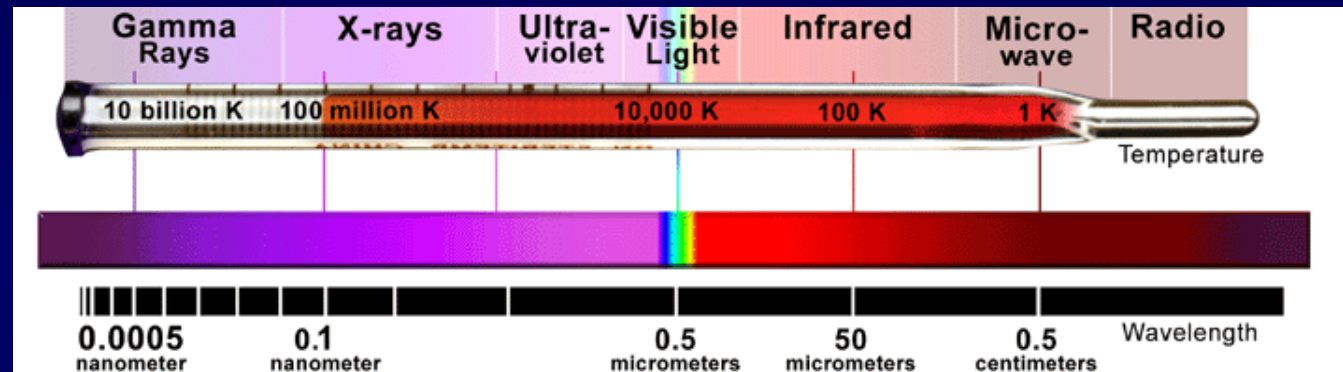
John Taylor

Director General of Research Councils

Office of Science and Technology

Virtual Observatories

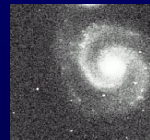
Observations made across entire electromagnetic spectrum



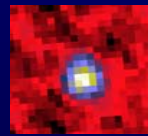
ROSAT ~keV



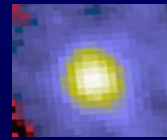
DSS Optical



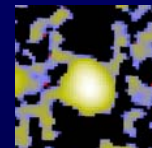
2MASS 2μ



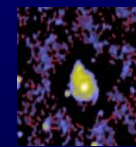
IRAS 25μ



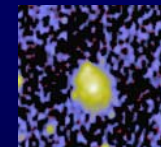
IRAS 100μ



GB 6cm



NVSS 20cm



WENSS 92cm

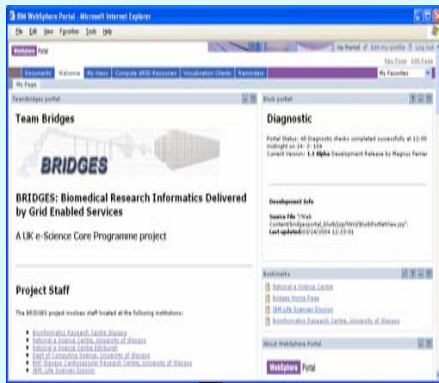
⇒ e.g. different views of a local galaxy

Need all of them to understand physics fully

Databases are located throughout the world

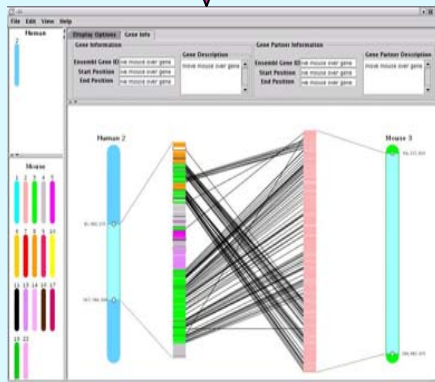
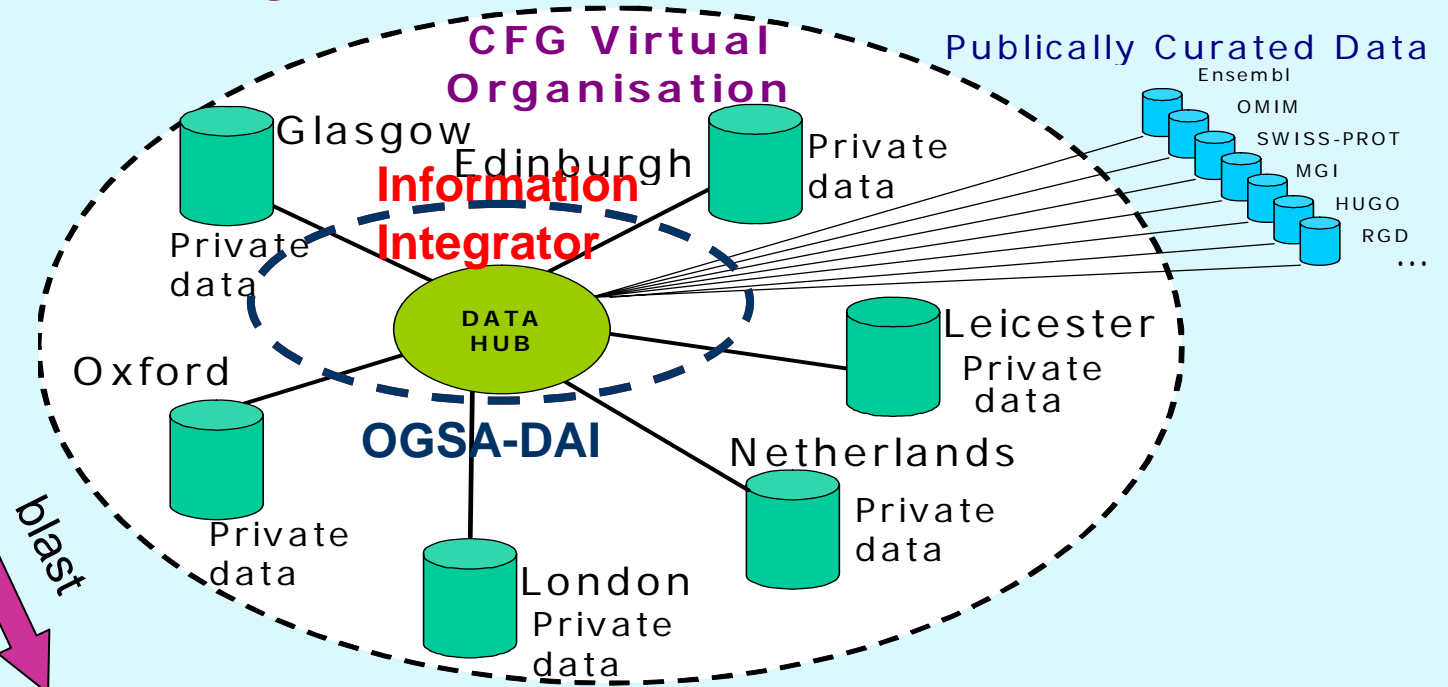
Biomedical Research Informatics Delivered by Grid Enabled Services

VO Authorisation



Synteny
Grid
Service

blast



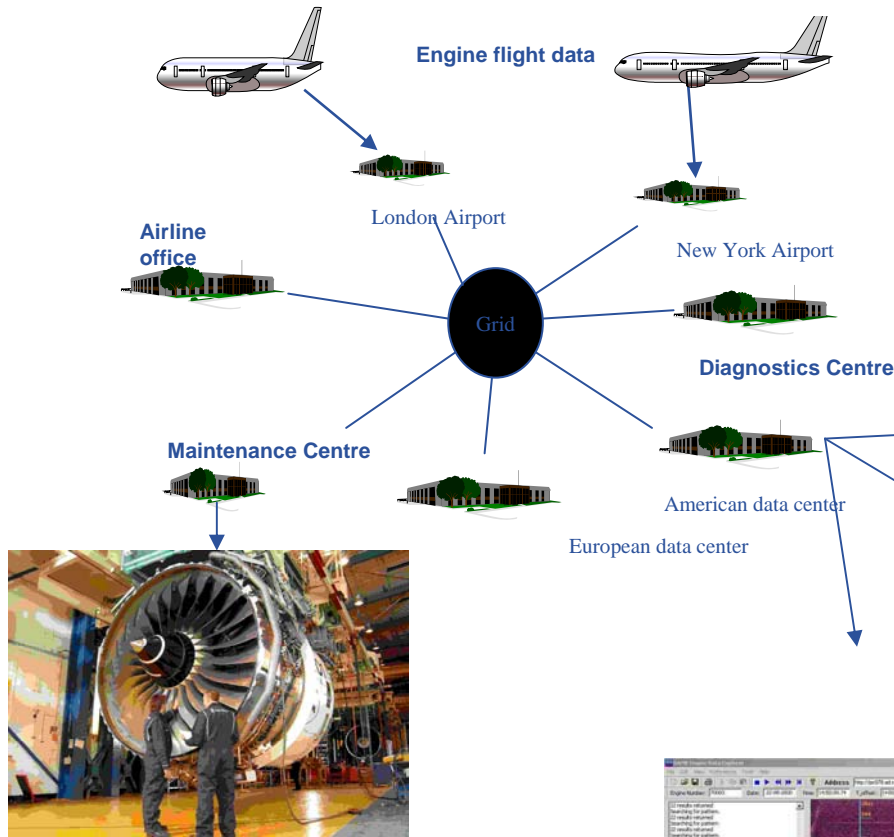
+



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



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

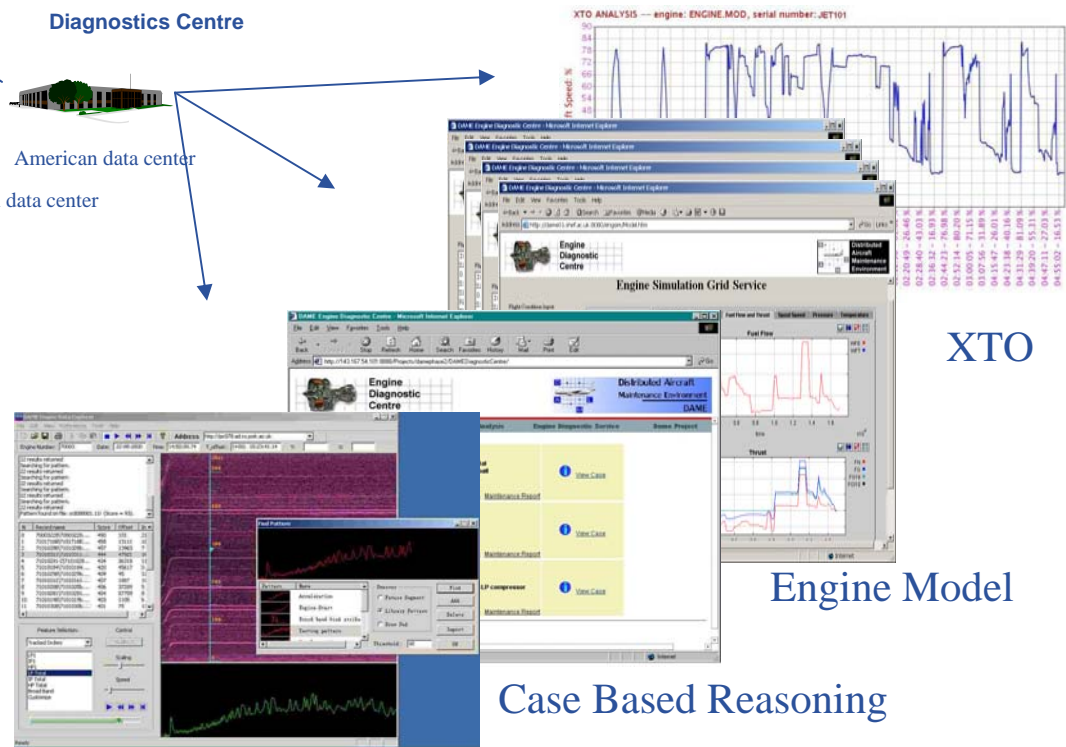


“A Significant factor in the success of the Rolls-Royce campaign to power the Boeing 7E7 with the Trent 1000 was the emphasis on the new aftermarket support service for the engines provided via DS&S. Boeing personnel were shown DAME as an example of the new ways of gathering and processing the large 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

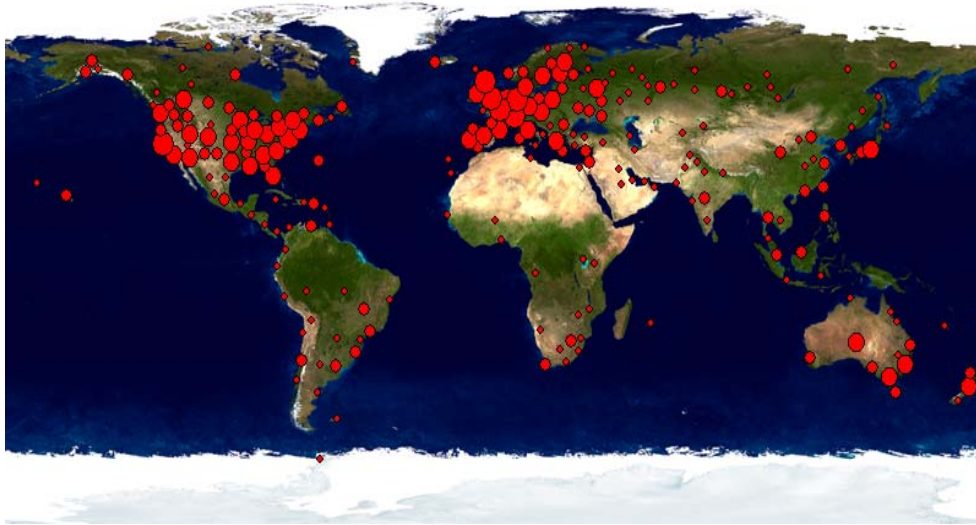


XTO

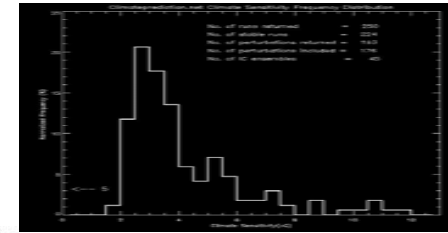
Engine Model

Case Based Reasoning

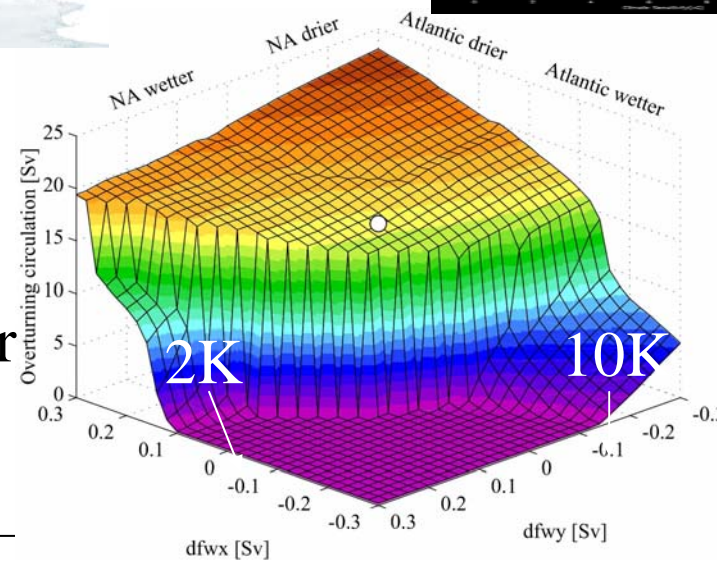
climateprediction.net and GENIE



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

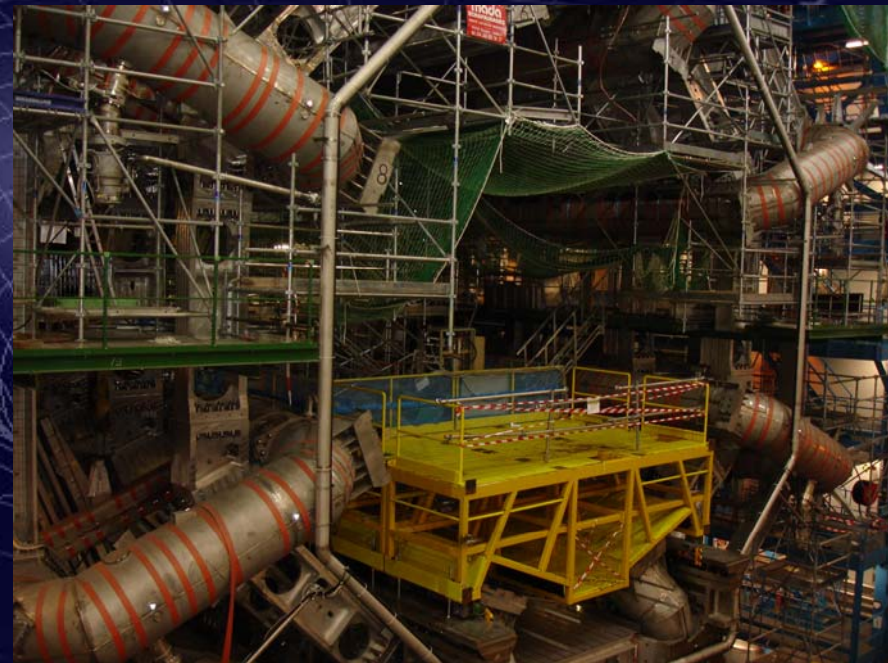


Response of Atlantic
circulation to freshwater
forcing



PPARC

UK Grid for Particle Physics



GridPP www.gridpp.ac.uk

ATLAS detectors, 2/3/06

<http://www.accessgrid.org/>

Cameras



Microphones

- **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 ($\times 1000/\text{decade}$)
 - ▶ enabling a whole-system approach

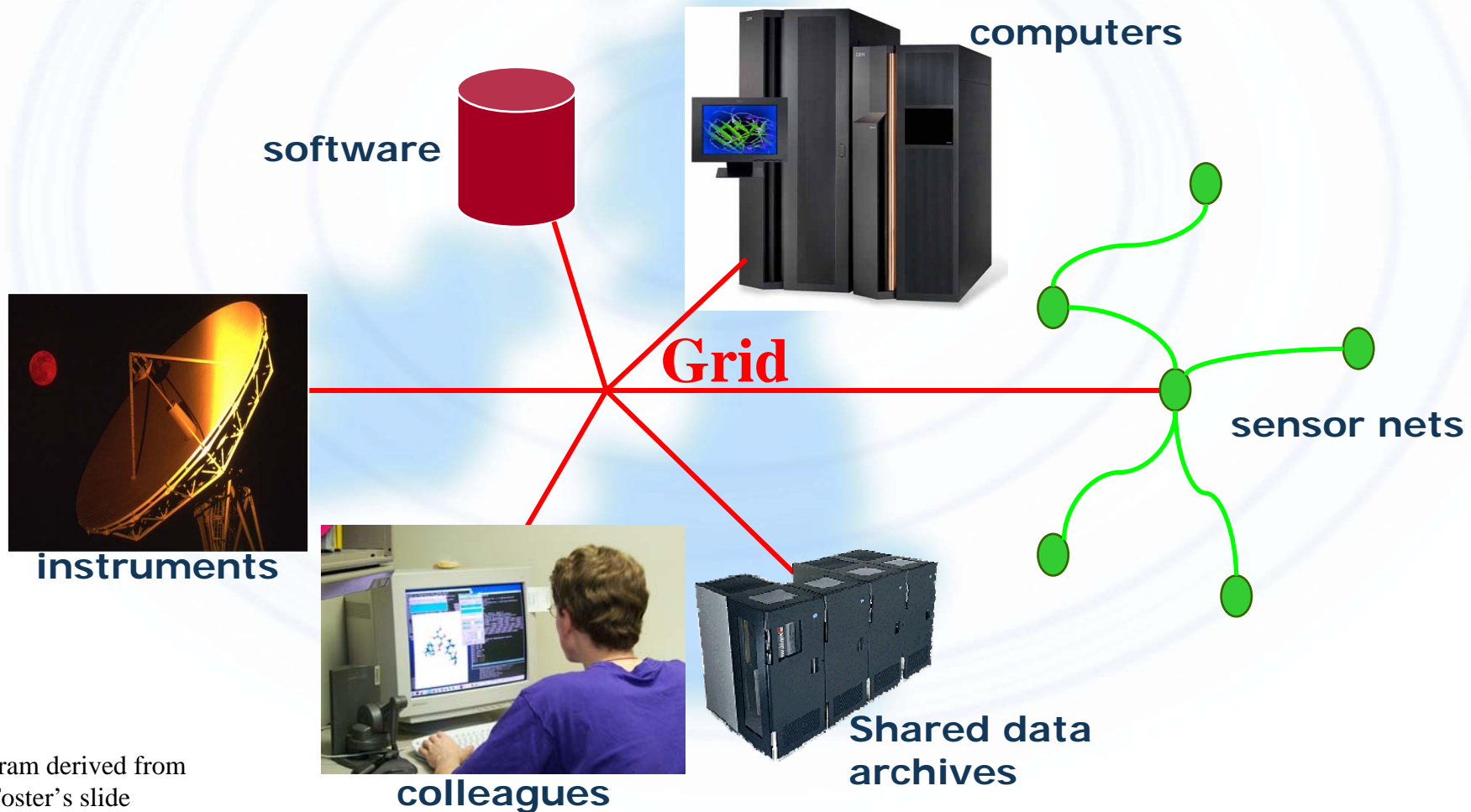
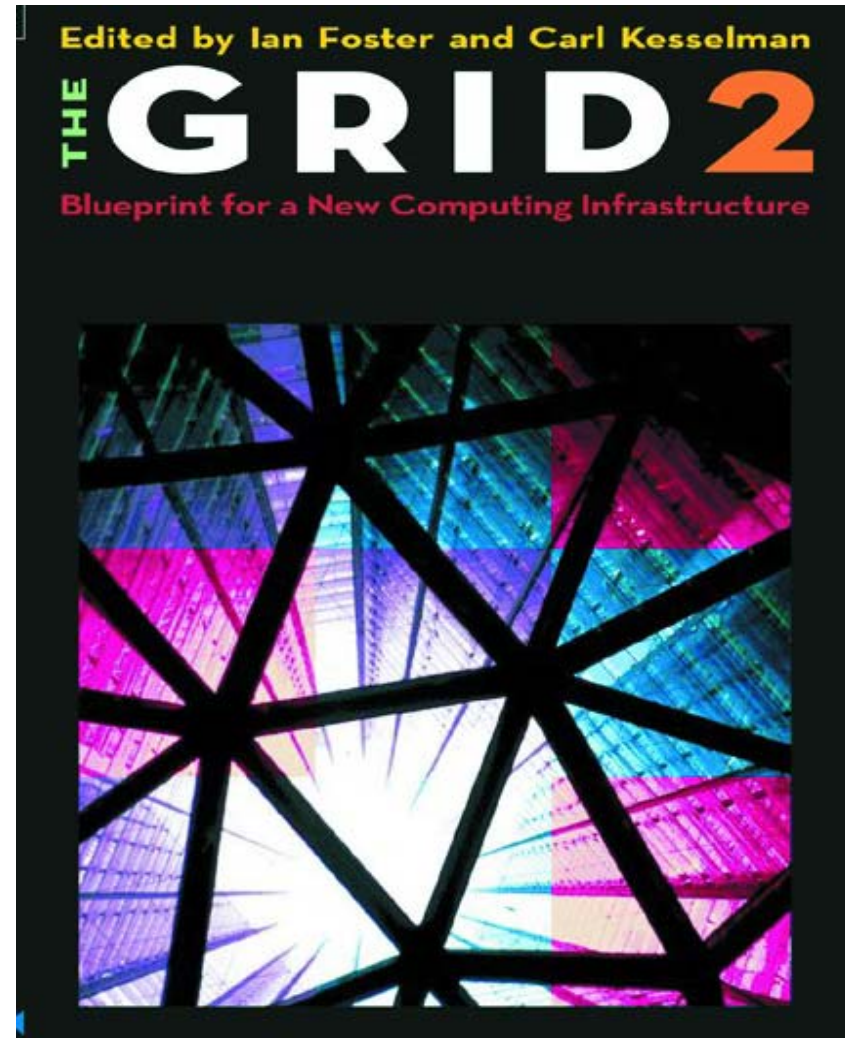
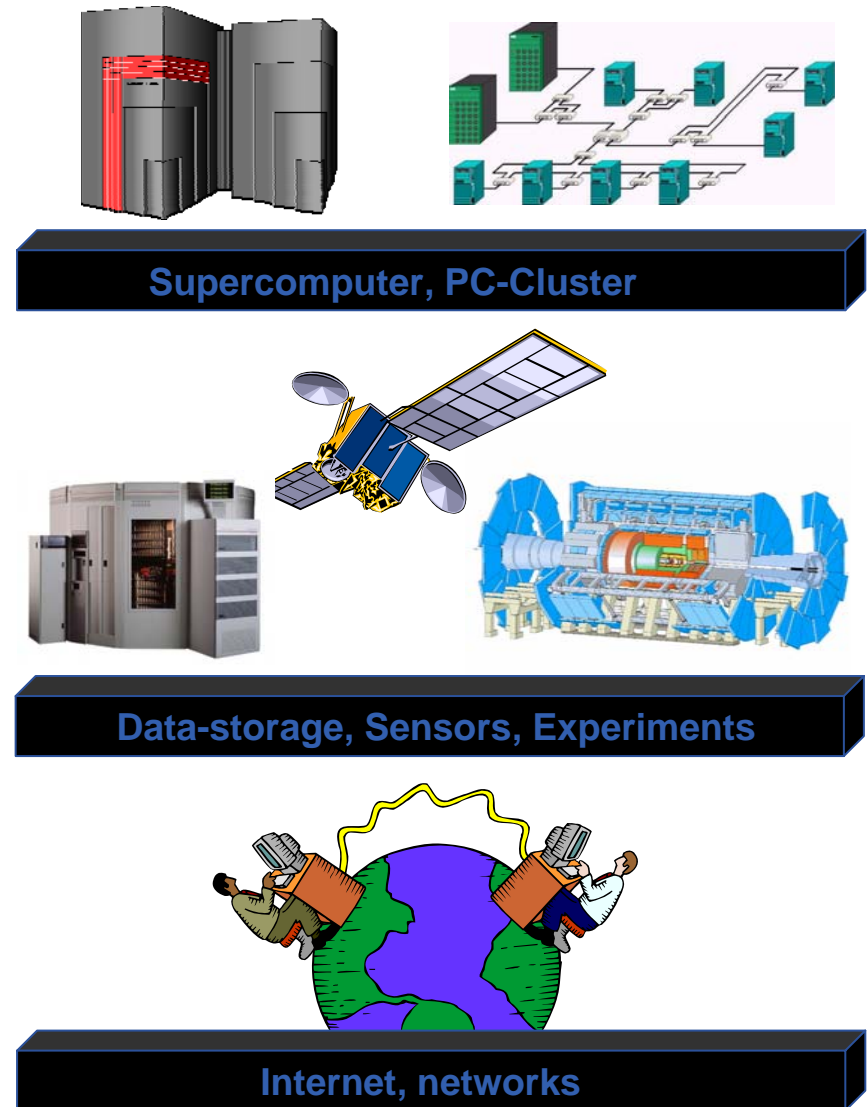
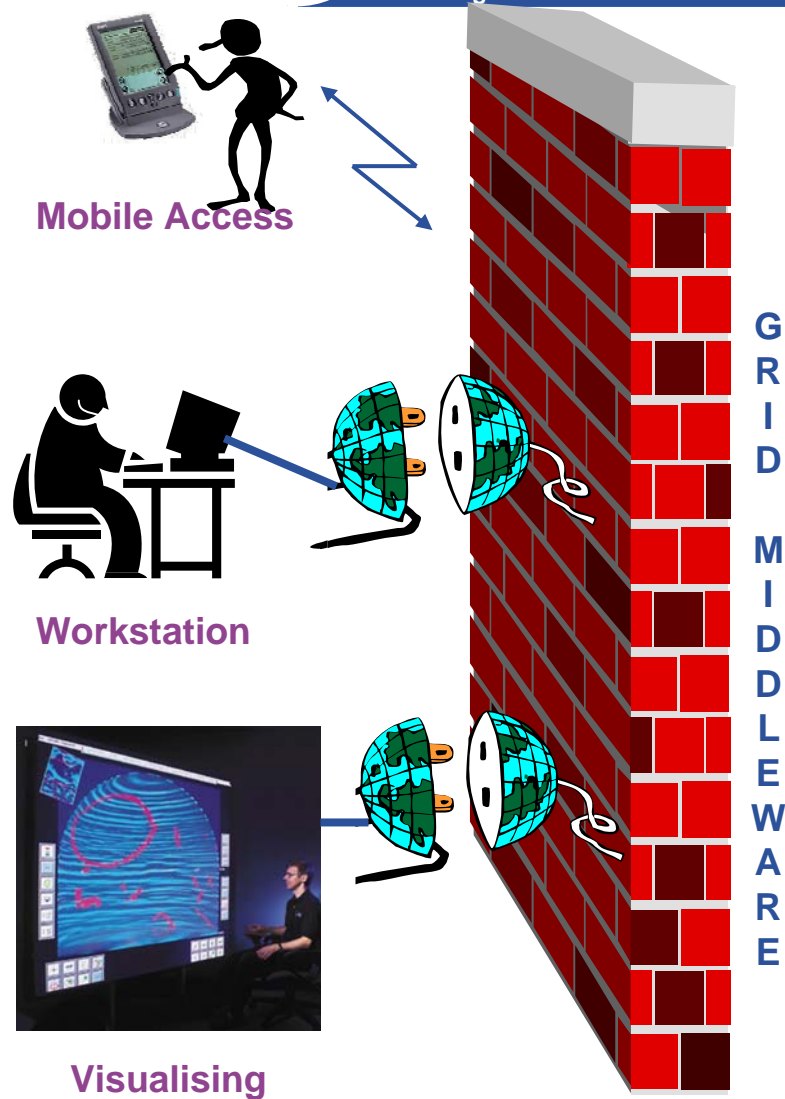


Diagram derived from
Ian Foster's slide

- 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,...





- **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



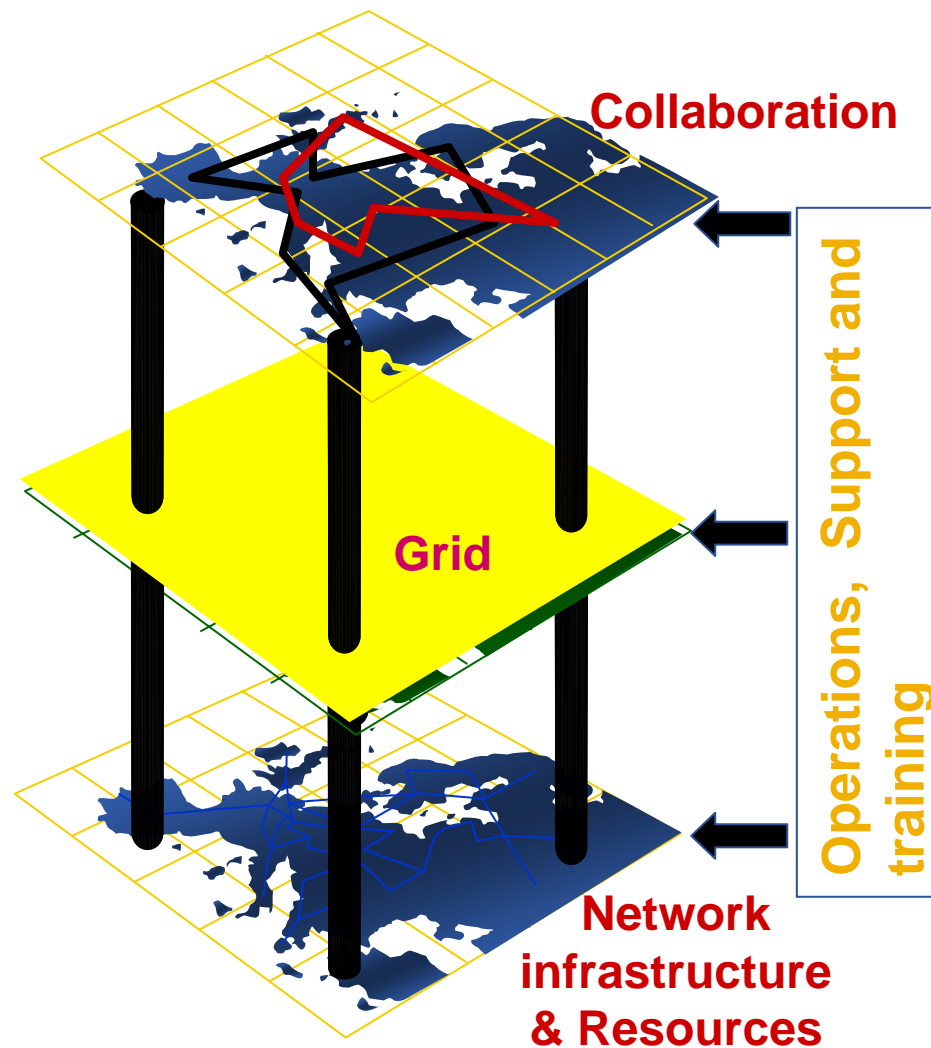

Gordon Brown
Chancellor of the
Exchequer

Charles Clarke
Secretary of State for
Education and Skills

Patricia Hewitt
Secretary of State for
Trade and Industry

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



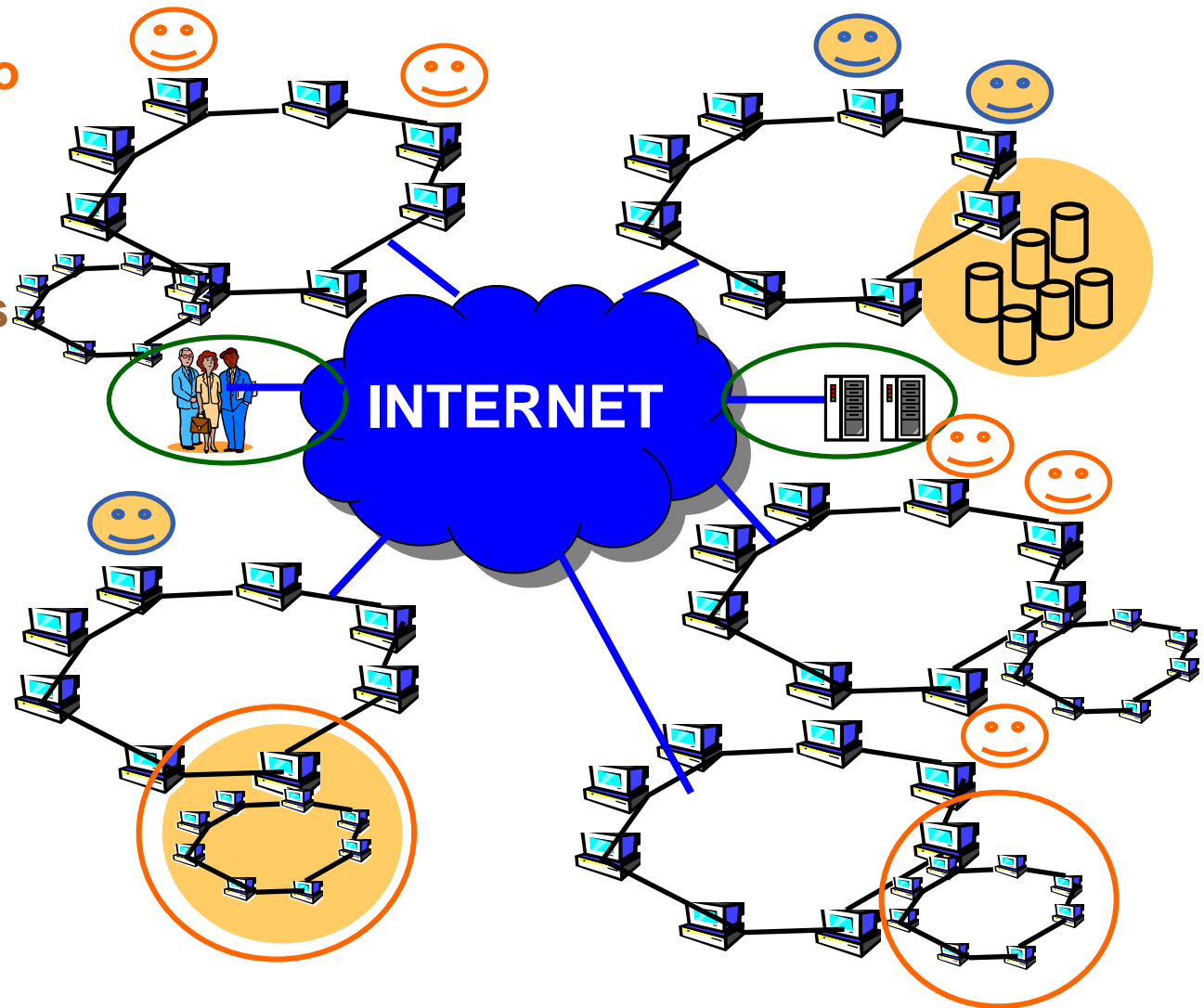
- **Digital technology – exponential growth - e.g. bandwidth**
- **Opportunities for e-Infrastructure to support faster, better, different research**
 - Sharing expertise
 - Support for cooperation and communication
 - Sharing computation services
 - E.g. to serve occasional peaks of high demand for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

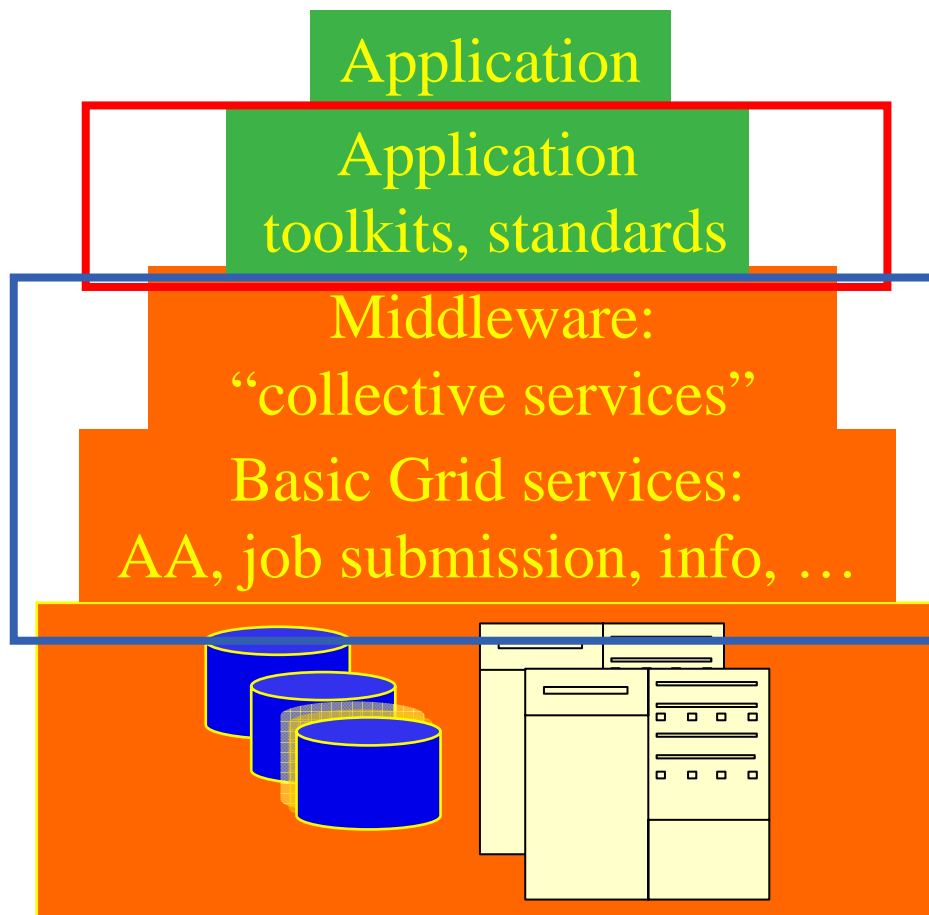
- **The term “Grid” has become popular!**
 - Sometimes in Industry : “Grids” = clusters
 - Motivations: better use of resources; scope for commercial services
 - Also used to refer to the harvesting of donated, unused compute cycles
 - (SETI@home, Climateprediction.net)
 - These are e-Infrastructure 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)**

- **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**





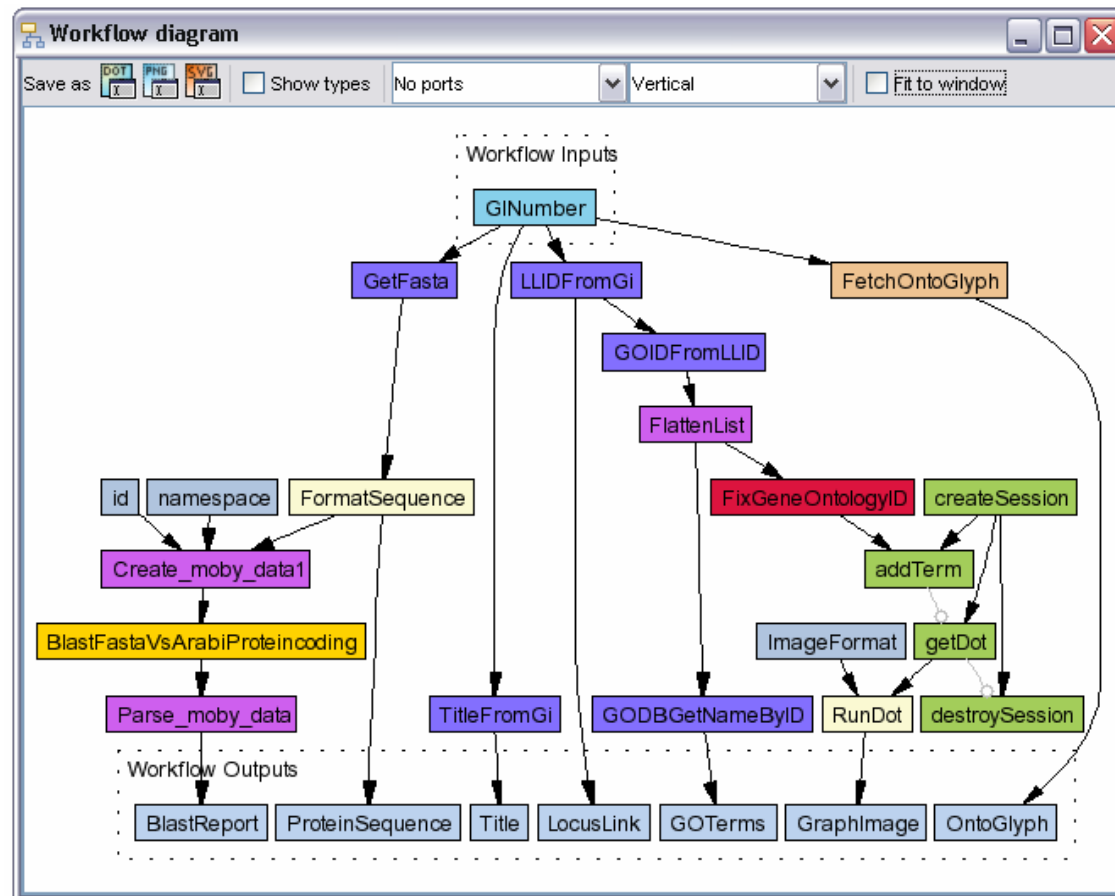
Where computer science meets the application communities!

VO-specific developments:

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

Production grids provide these services.

- 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,..

National datacentres,
HPC, instruments

Institutes' data;
Condor pools

Wider collaboration
greater resources



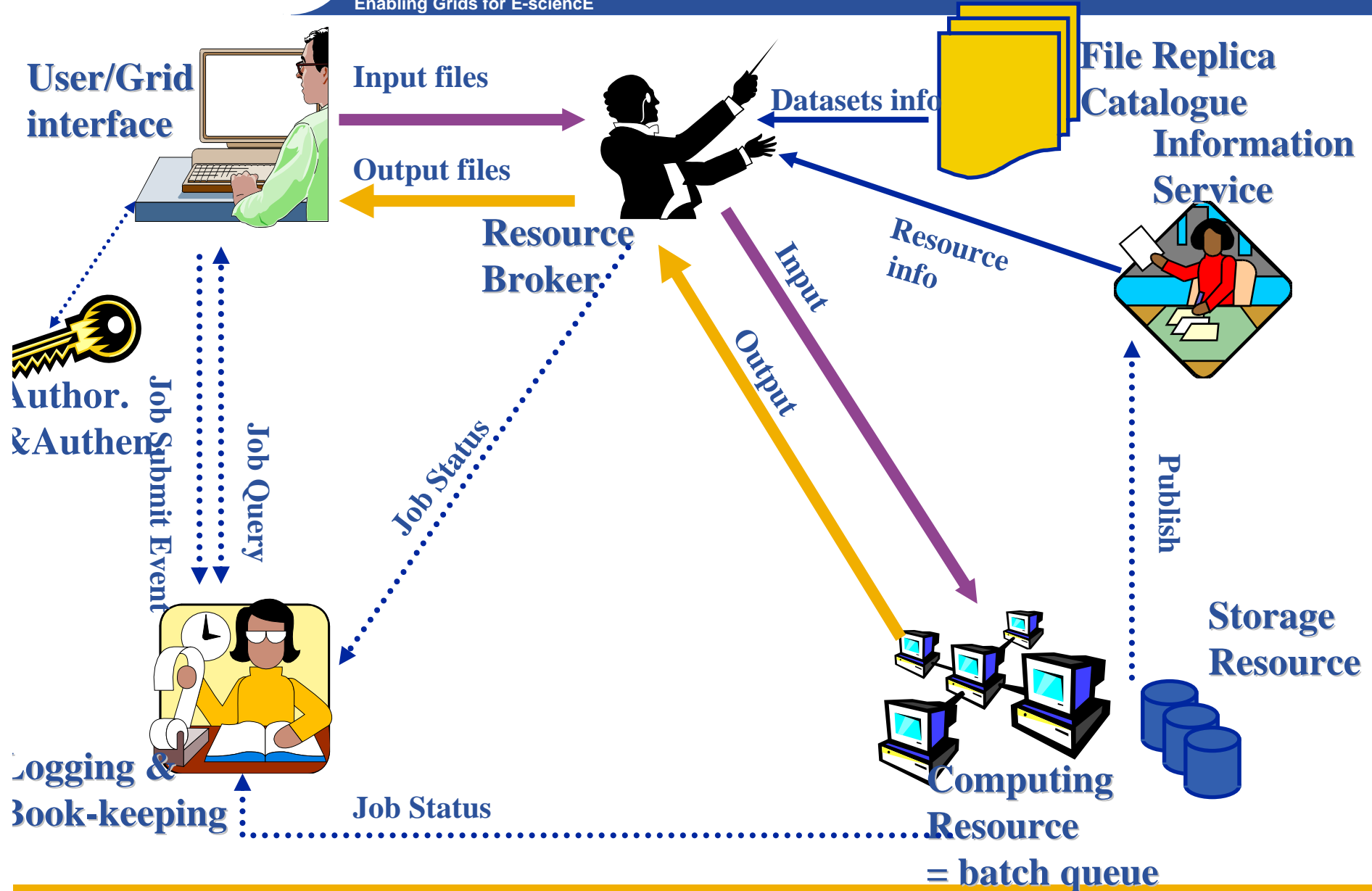
International grid (EGEE)

National grids (e.g.
National Grid Service)

Regional grids (e.g.
White Rose Grid)

Campus grids

Desktop





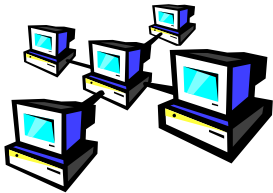
Access service How users logon to a Grid



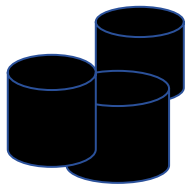
Resource Broker (RB): Matches the user requirements with the available resources on a Grid



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?!

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

“VO”: virtual organisation

“Grid operations”: funded effort

- **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**

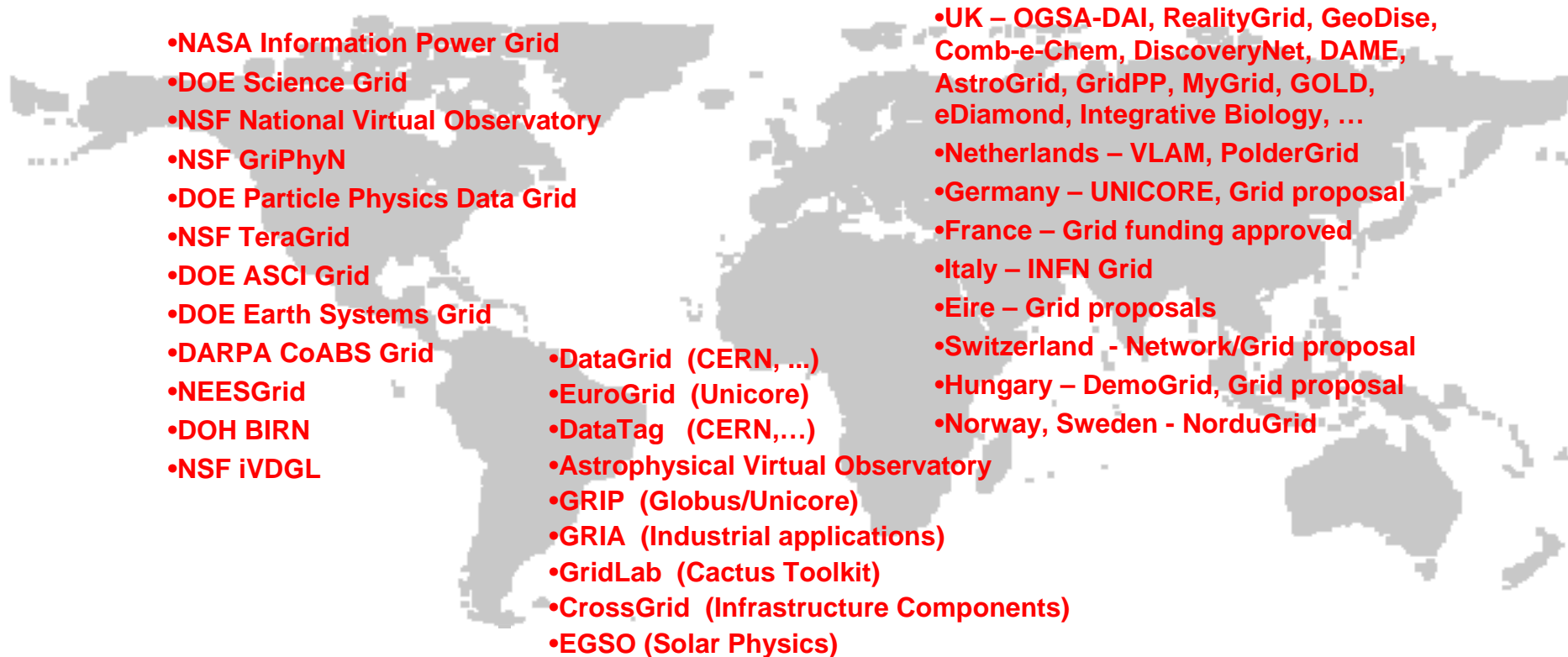
- **Achieved by Certification:**
 - User's identity has to be certified by one of the national *Certification Authorities (CAs)*
 - mutually recognized <http://www.gridpma.org/>, for EU go via here to <http://marianne.in2p3.fr/datagrid/ca/ca-table-ca.html> to find your CA
 - *E.g. In UK go to <http://www.grid-support.ac.uk/ca/ralist.htm>*
 - 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**



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

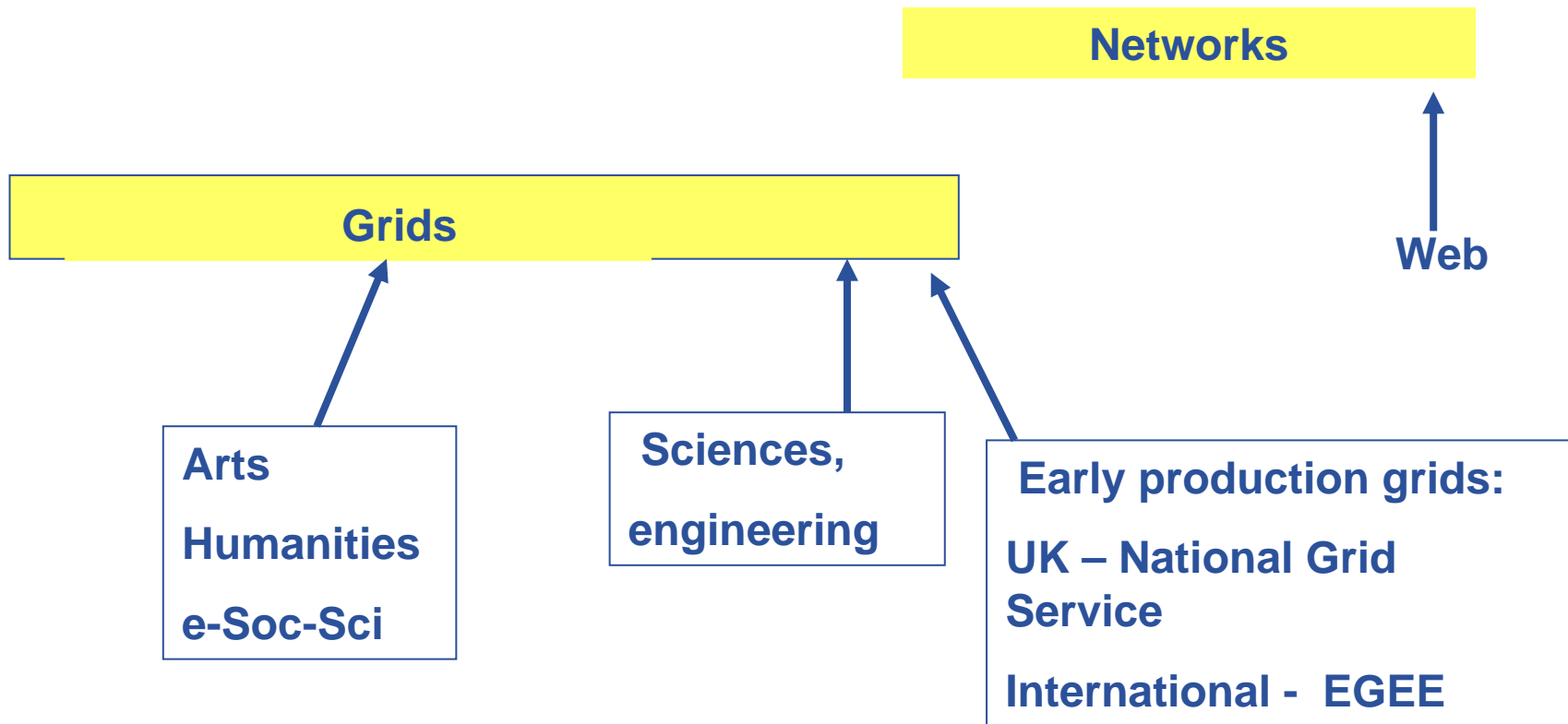
... then where are
we now?

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:
 - **Standards** (a slow process)
(e.g. Global Grid Forum, <http://www.gridforum.org/>)
 - **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
- **“Service orientation” - “the way to build grids”**

Where are we now? –user's view



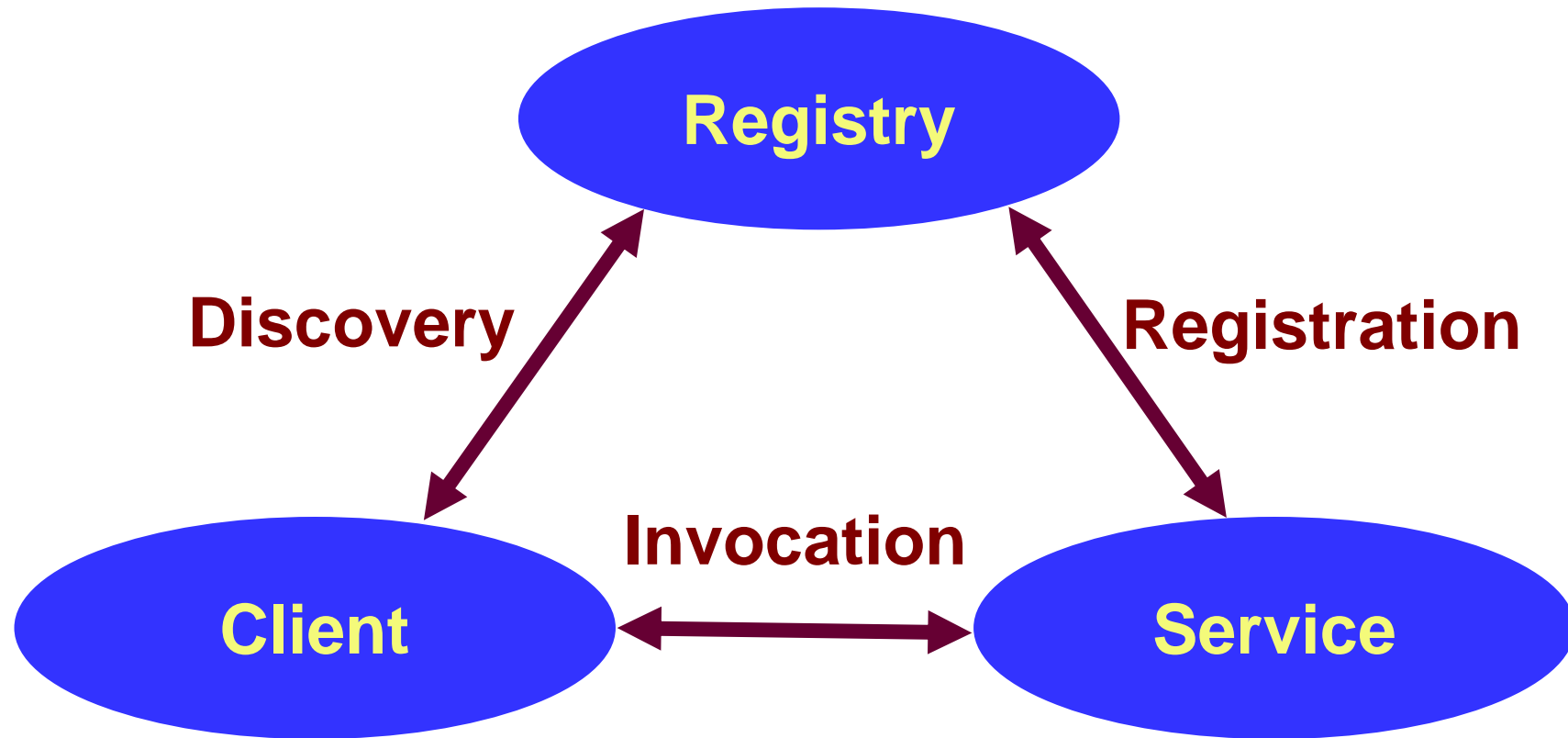
- **Standards are emerging... some near acceptance and some being discarded**
 - Standards bodies:
 - W3C <http://www.w3c.org/>
 - GGF <http://www.ggf.org/>
 - OASIS <http://www.oasis-open.org/home/index.php>
 - IETF <http://www.ietf.org/>
 - For a summary see <http://www.innoq.com/soa/ws-standards/poster/>

- **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

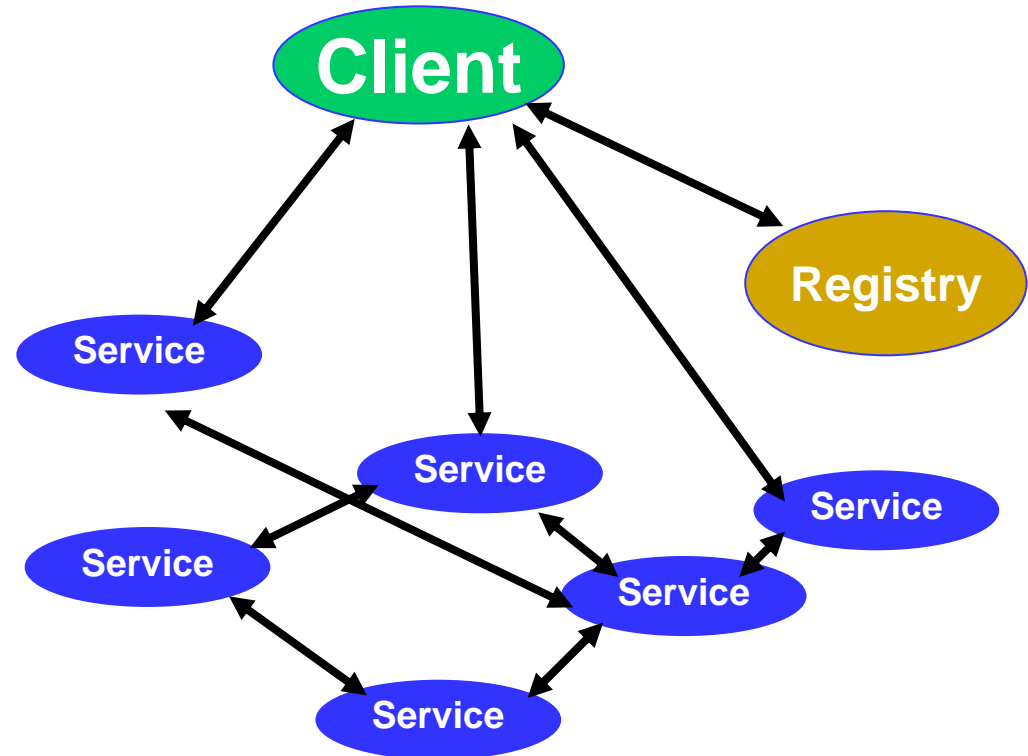
- **Globus Toolkit 4 has been released**

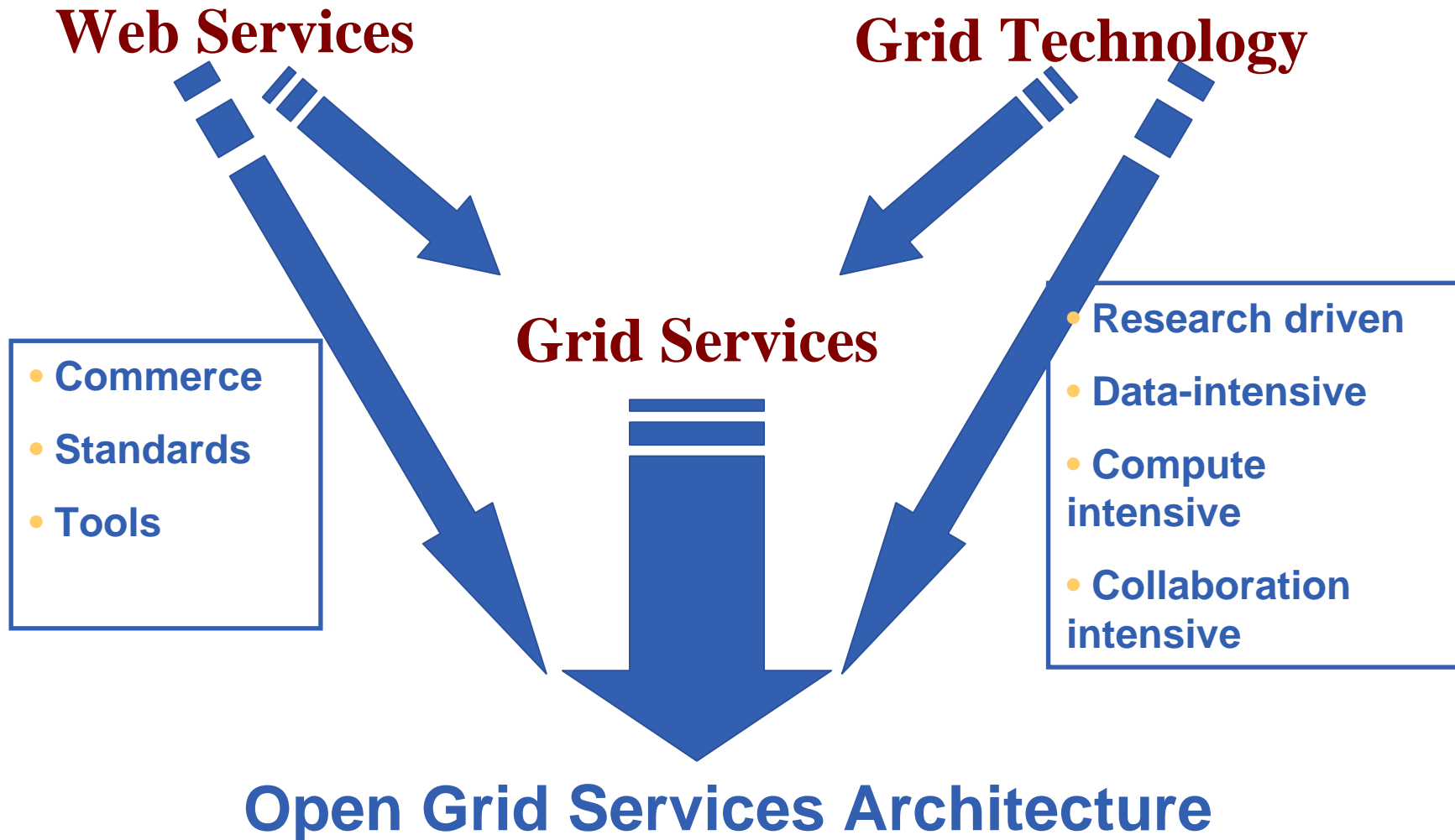
National grid initiatives now include...





- Accessible across a network
- Loosely coupled, defined by the messages they receive / send
- Interoperable: each service has a description that is accessible and can be used to create software to invoke that service
- Based on standards (for which tools do / could exist)
- Developed in anticipation of new uses





- “Open grid services architecture” OGSA– proposed in 2001
- Open Grid Services Infrastructure
 - Globus Toolkit 3 resulted
- Then in January 2004
 - OGSI to be replaced by emerging **WS-RF (Web Services Resource Framework)**: manage “state” without major rewrite of WS standards
- **WS-I used meanwhile: <http://www.ws-i.org/>**
Open standards:
 - SOAP: protocol for message passing
 - Web Service Description Language: to describe services
 - UDDI: Universal Description, Discovery and Integration
 - WS-Security: incorporates security

Web Services

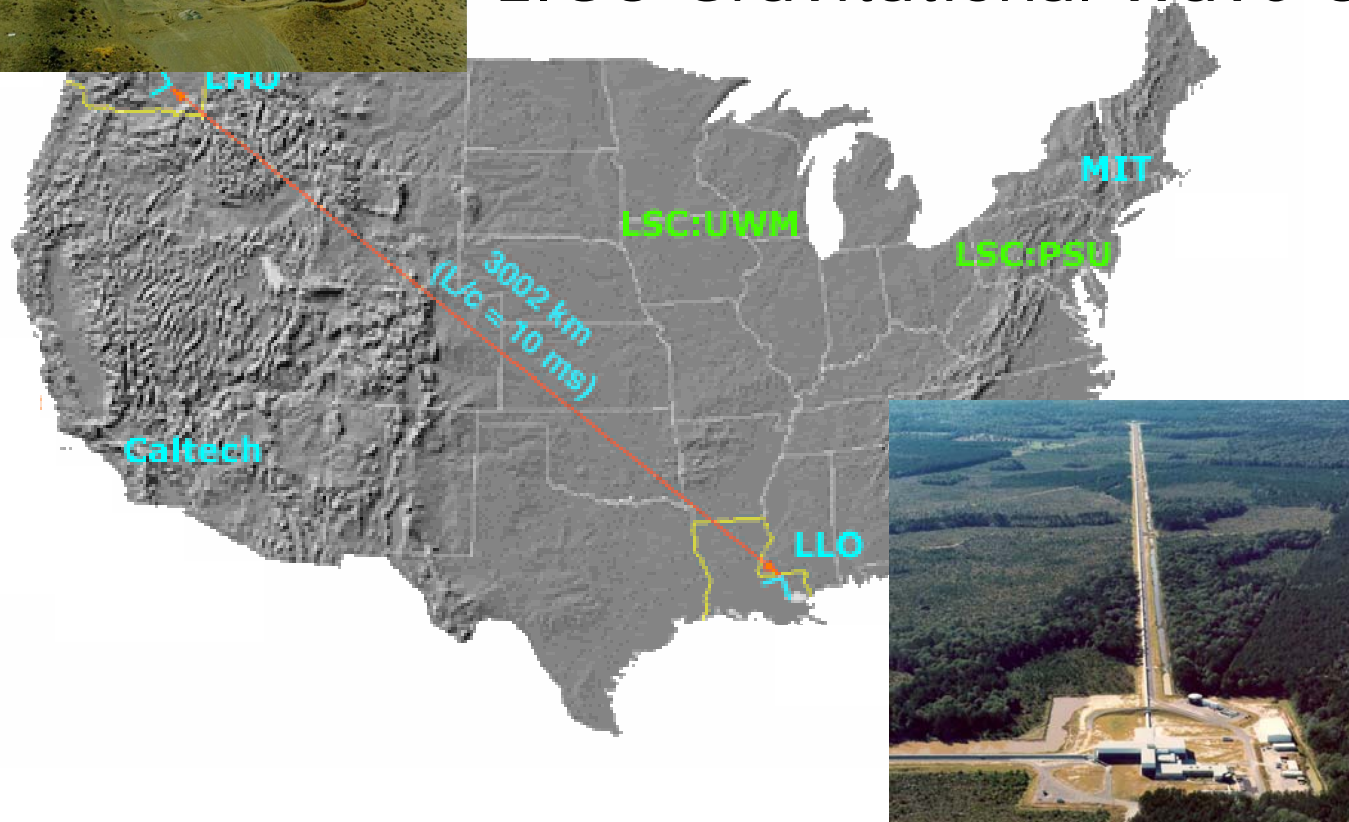
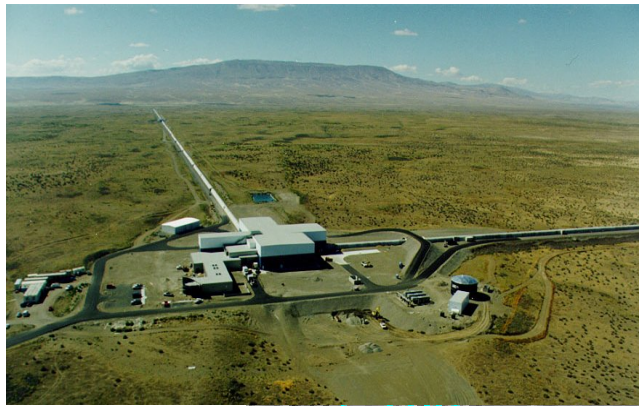
- **Goals**
 - Computational presentation & access of Enterprise services
 - Marketing integrated large scale software and systems
 - Model for independent development
 - Model for independent operation

Grids

- **Goals**
 - Inter-organisational collaboration
 - Sharing information and resources
 - Framework for collaborative development
 - Framework for collaborative operation

The Globus-Based LIGO Data Grid

LIGO Gravitational Wave Observatory



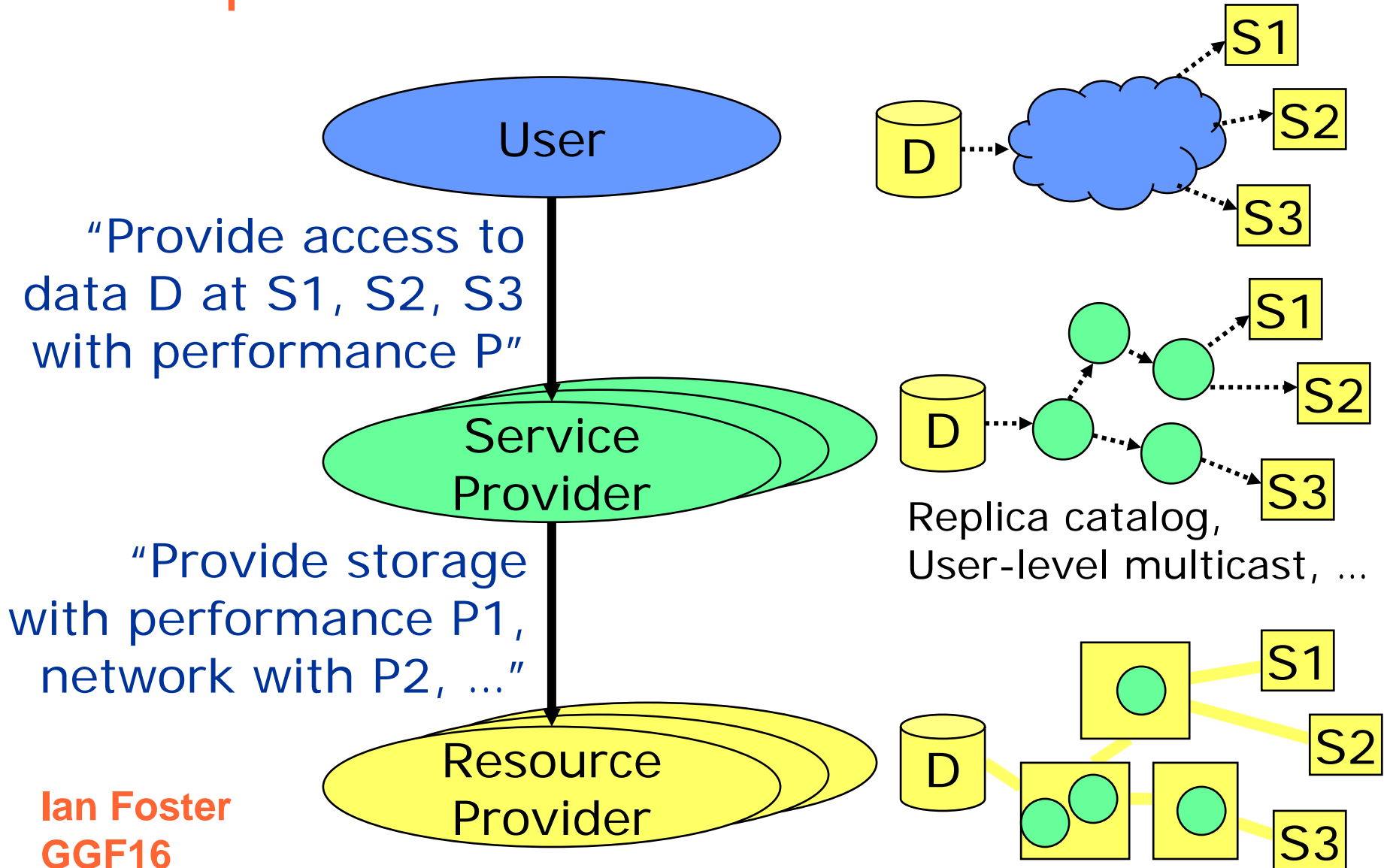
Replicating >1 Terabyte/day to 8 sites

>40 million replicas so far

MTBF = 1 month www.globus.org/solutions



Decomposition Enables Separation of Concerns & Roles



- **Introduction to**
 - e-Research and e-Science
 - Grids
 - e-Infrastructure
- **Grid concepts**
- **Grids - Where are we now?**
- **Enabling the research of the future**
 - Grids already empower a widening spectrum of research but.
 - **What happens if research becomes service oriented??**



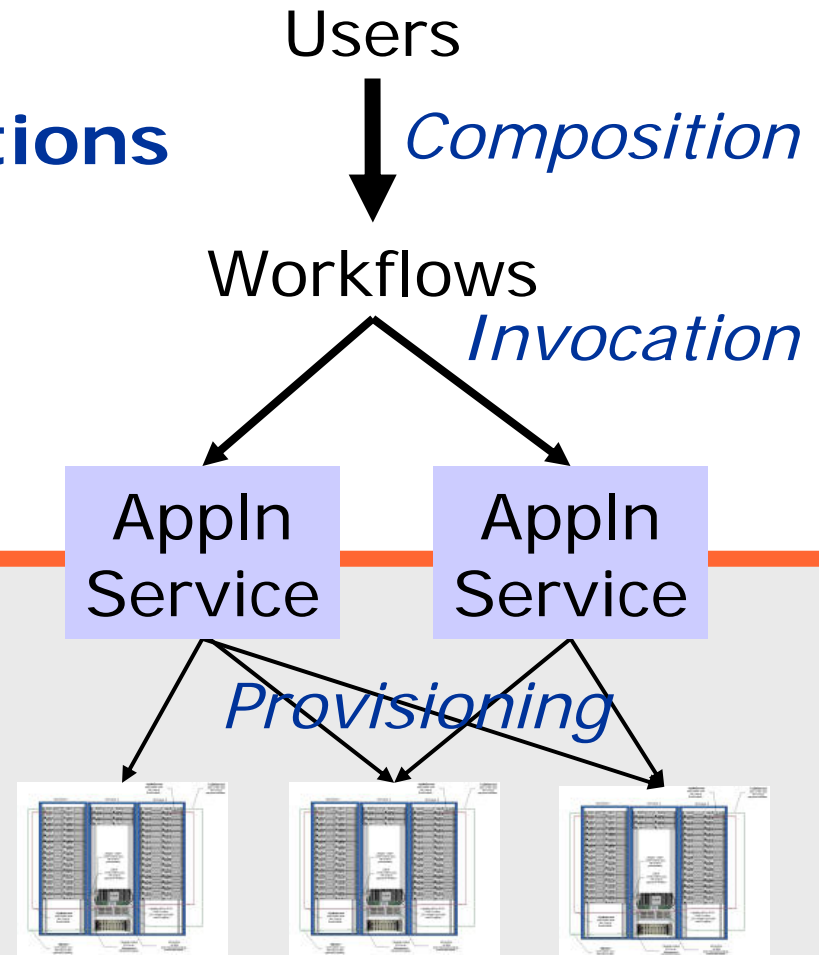
Service-Oriented Systems: The Role of Grid Infrastructure

- Service-oriented **applications**

- ◆ Wrap applications as services
- ◆ Compose applications into workflows

- Service-oriented **Grid infrastructure**

- ◆ Provision physical resources to support application workloads



- “potential to increase individual and collective scientific productivity by making powerful information tools available to all”
- “Ultimately, we can imagine a future in which a community's shared understanding ... is documented also in the various databases and programs that represent—and automatically maintain and evolve—a collective knowledge base. ”

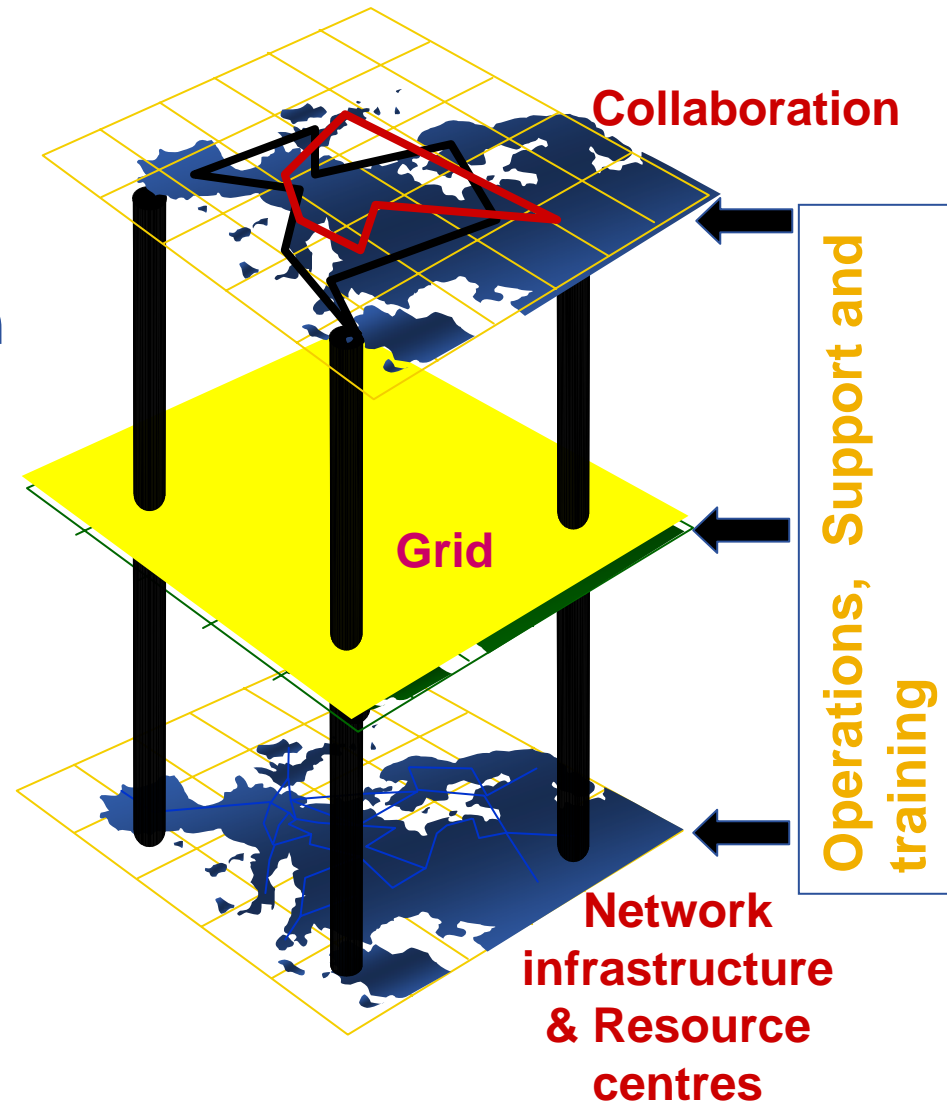
Ian Foster,

<http://www.sciencemag.org/cgi/content/full/308/5723/814?ijkey=aqCCmCFix8LI.&keytype=ref&siteid=sci>

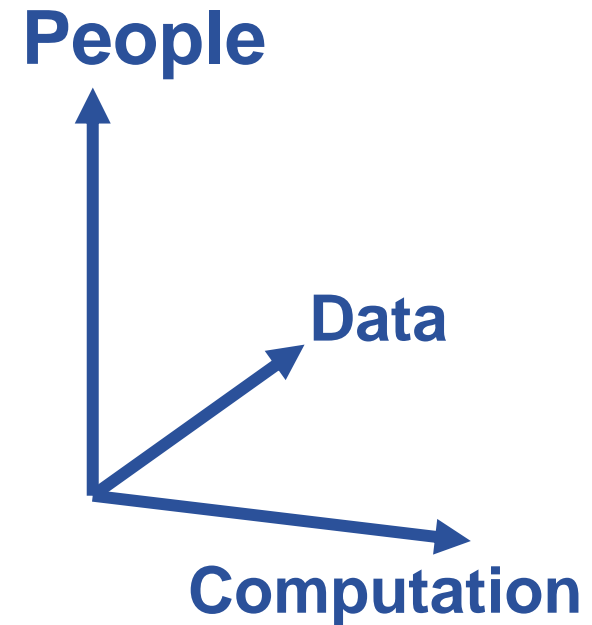
Science 6 May 2005

- **Early grids**
 - Resource utilisation
 - A few big-science VOs
 - Trivial parallelism – many concurrent independent jobs
 - Data management – files only
- **Grid-enabling databases**
 - Pre-existing databases accessible from grids
 - Data integration
- **Service-oriented grid: possibilities for**
 - any collaborative research
 - International / national / university resources become accessible
 - With control and AA (authorisation and authentication)

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



- 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?



- **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>**
- **Globus Grid Forum <http://www.ggf.org/> (see GGF16)**