Overview of e-Infrastructure

Mike Mineter
National e-Science Centre
mjm@nesc.ac.uk

14 February 2006

- **The presentations are (or will soon be) on the Agenda page http://agenda.cern.ch/fullAgenda.php?ida=a057811**

- **Please give an evaluation of this event on the form provided. Additional feedback is always welcome: email mjm@nesc.ac.uk**

- **This is the first of several similar events. Talks for each of these events will be available via the NeSC training page http://www.nesc.ac.uk/training/ . To find them, follow the UK link.**
  (Dates: 22 February, 20 March, 28 March)

- **Additional information can be found in some slides that are hidden and in notes on some other slides.**

- **Also please note that:**

  - **Please let us know if you have other requirements for training by email to training-support@nesc.ac.uk**

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

Contents

- **Introduction to**
  - e-Research and e-Science
  - Grids
  - e-Infrastructure
- **More about Grids**
- **Two key themes**

EGEE is a European-funded project that is establishing an international production Grid.

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

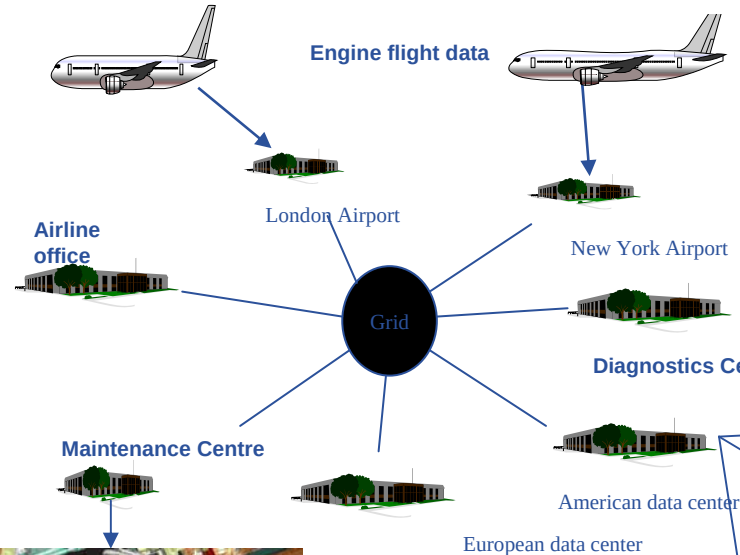
- **Began with focus in the "big sciences" hence initiatives are often badged as "e-science"…. Ever-broadening user communities (arts, humanities…)**

Global Drivers of e-Research

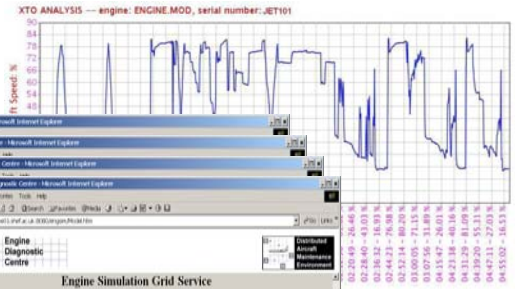
- **Digital technology – exponential growth - e.g. bandwidth**
- **Opportunities for e-Infrastructure to support faster, better, different research**
- **Both Government and Industry investment**
- **Increase in collaborative research**
  - Sharing expertise
    - Support for cooperation and communication
  - 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

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

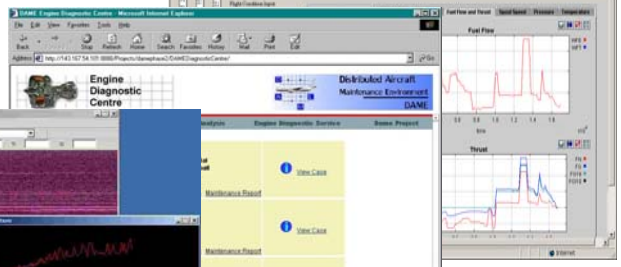
Distributed Aircraft Maintenance Environment
DAME



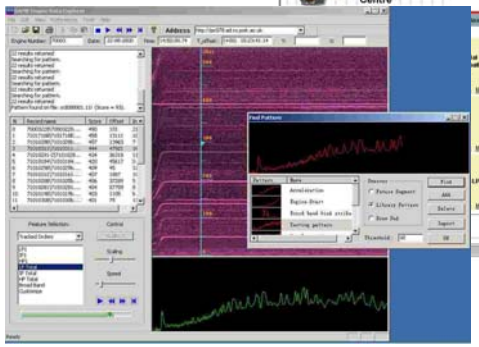
•"A Significant factor in the success of the Rolls-Royce campaign to power the Boeing 7E7 with the Trent 1000 was the emphasis on the new aftermarket support service for the engines provided via DS&S. Boeing personnel were shown DAME as an example of the new ways of gathering and processing the large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

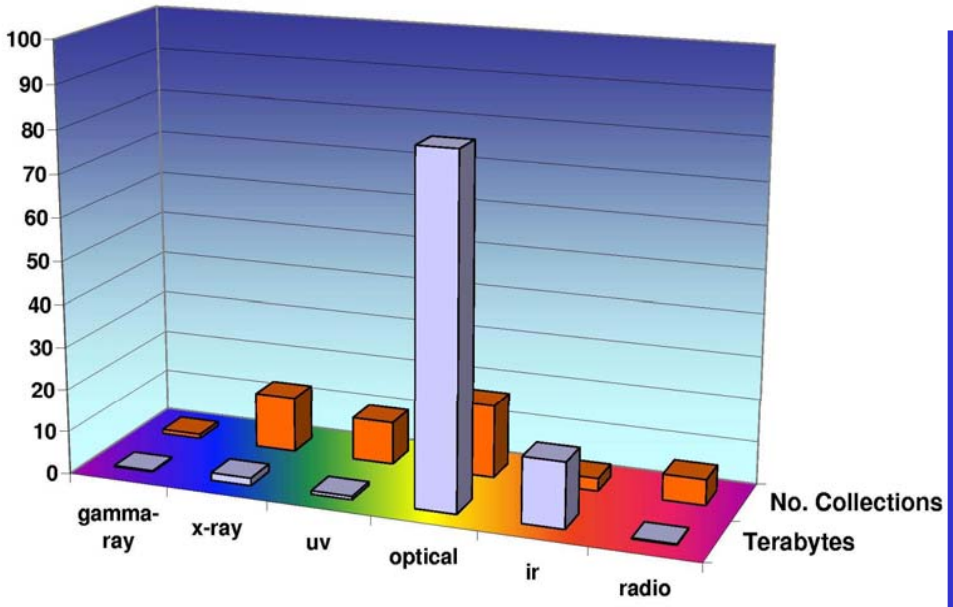


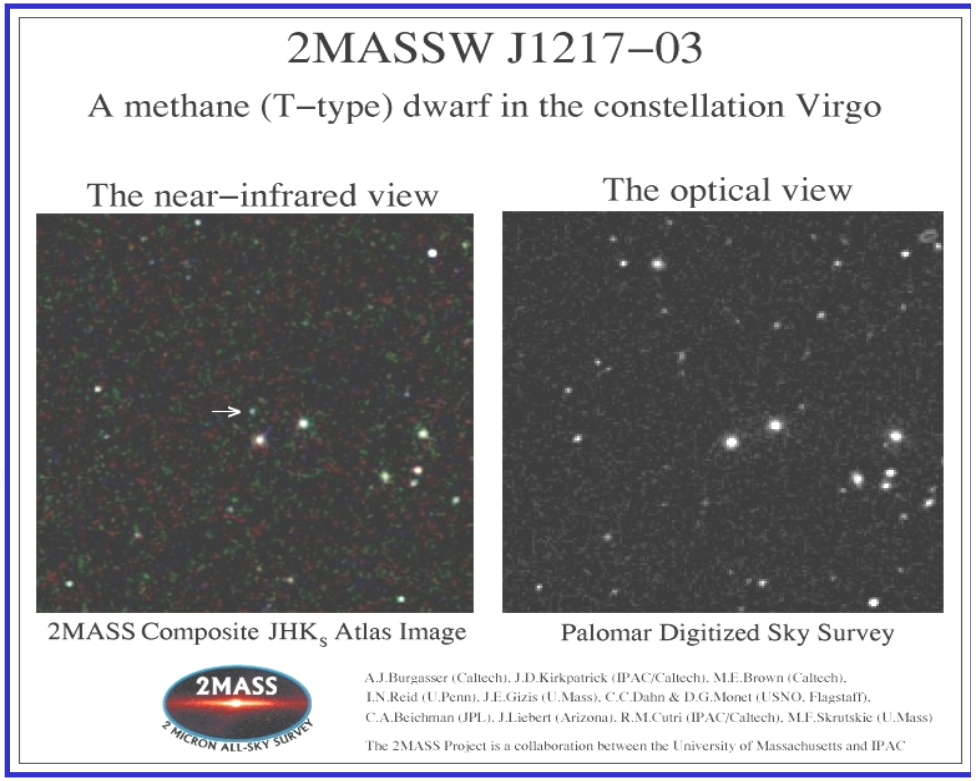
Case Based Reasoning

Companies:
Rolls-Royce
DS&S
Cybula

Universities:
York,
Leeds,
Sheffield, Oxford

Example: Astronomy





No. & sizes of data sets as of mid-2002, grouped by wavelength

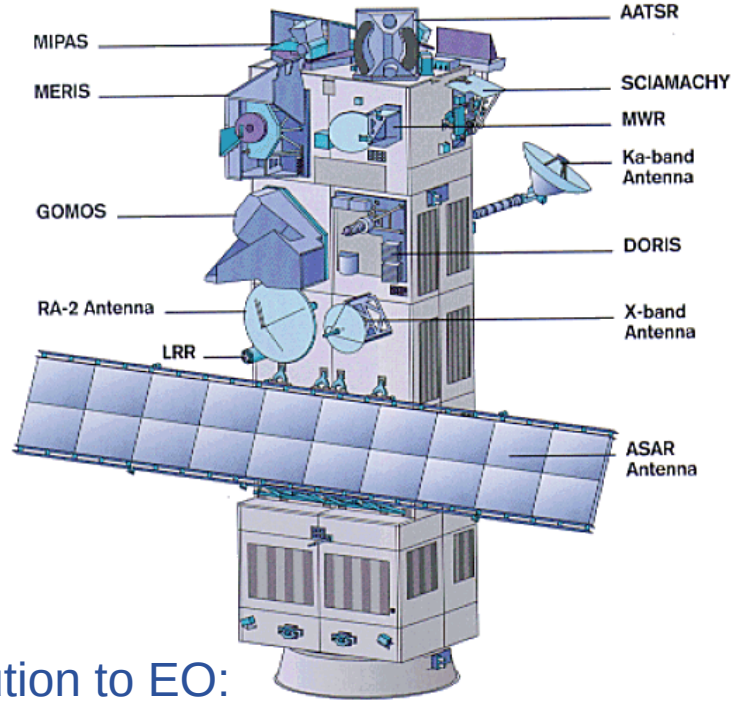
- 12 waveband coverage of large areas of the sky
- Total about 200 TB data
- Doubling every 12 months
- Largest catalogues near 1B objects

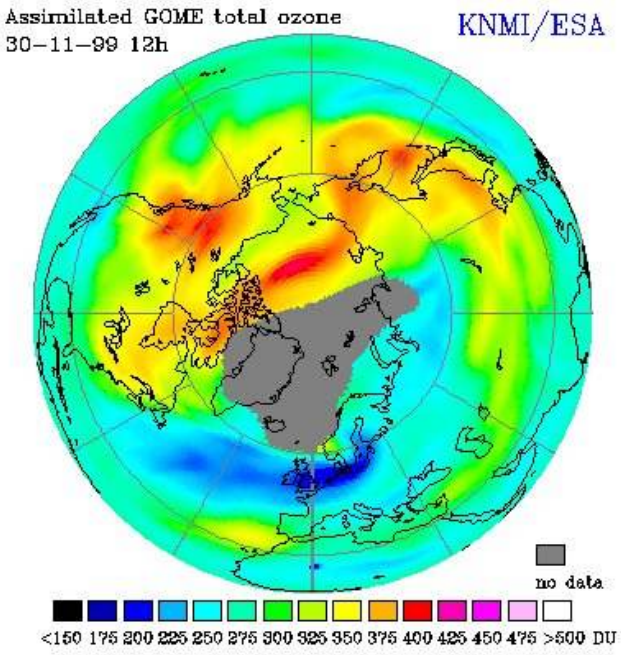
Data and images courtesy Alex Szalay, John Hopkins University

Example: Earth Observation

ESA missions:

- About 100 Gbytes of data per day (ERS 1/2)
- 500 Gbytes, for the next ENVISAT mission (2002).





Grid contribution to EO:

- Enhance the ability to access high level products
- Allow reprocessing of large historical archives
- Improve Earth science complex applications (data fusion, data mining, modelling …)

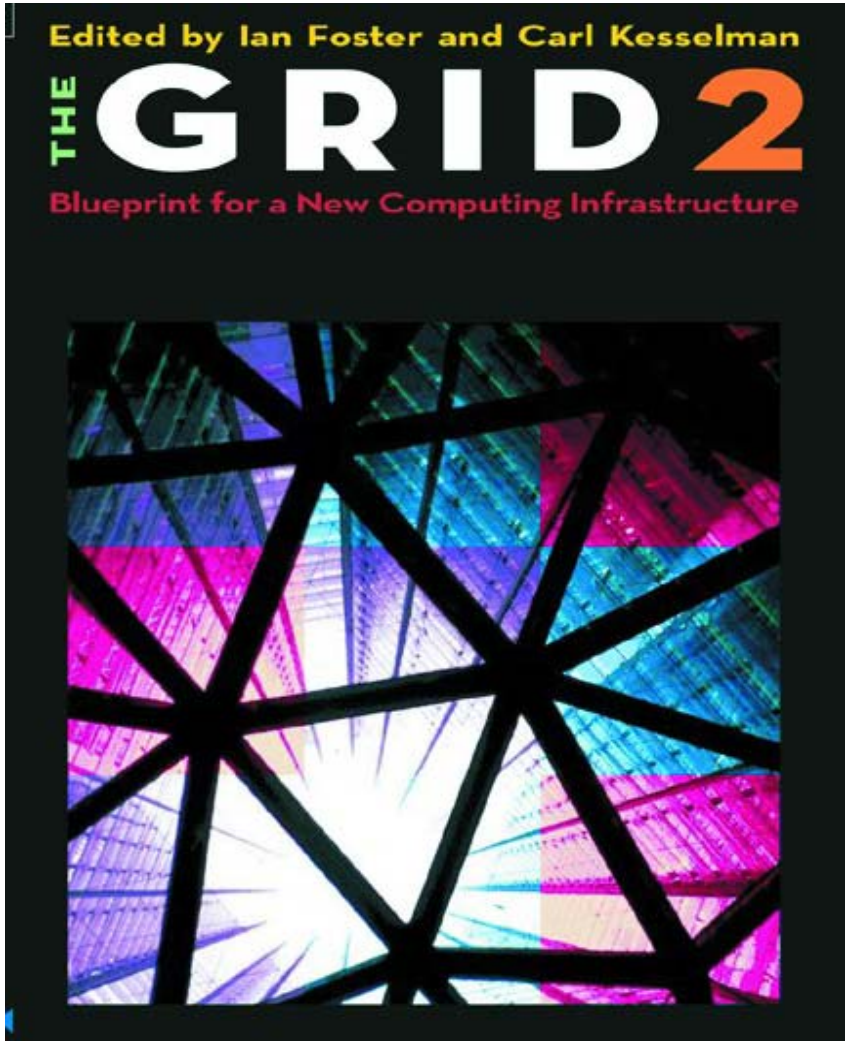
Source: L. Fusco, June 2001

Connecting *people*: Access Grid

http://www.accessgrid.org/

What is Grid Computing?

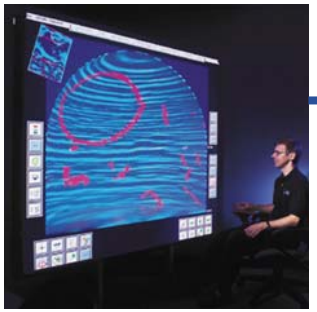
- **The grid vision is of "Virtual computing" (+ information services to locate computation, storage resources)**
  - Compare: The web: "virtual documents" (+ search engine to locate them)
- **MOTIVATION: collaboration through sharing resources (and expertise) to expand horizons of**
  - Research
  -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



The Grid Metaphor

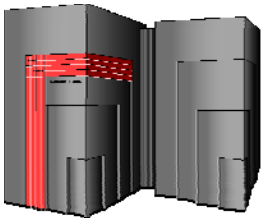
Mobile Access

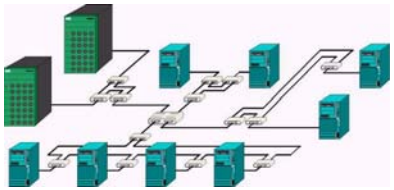
Workstation



Visualising

GRID MIDDLEWARE





Supercomputer, PC-Cluster

Data-storage, Sensors, Experiments

Internet, networks

What is 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)

- **In the e-Research world it means:**
  - “An infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources”
    Ian Foster and Carl Kesselman

  - 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)

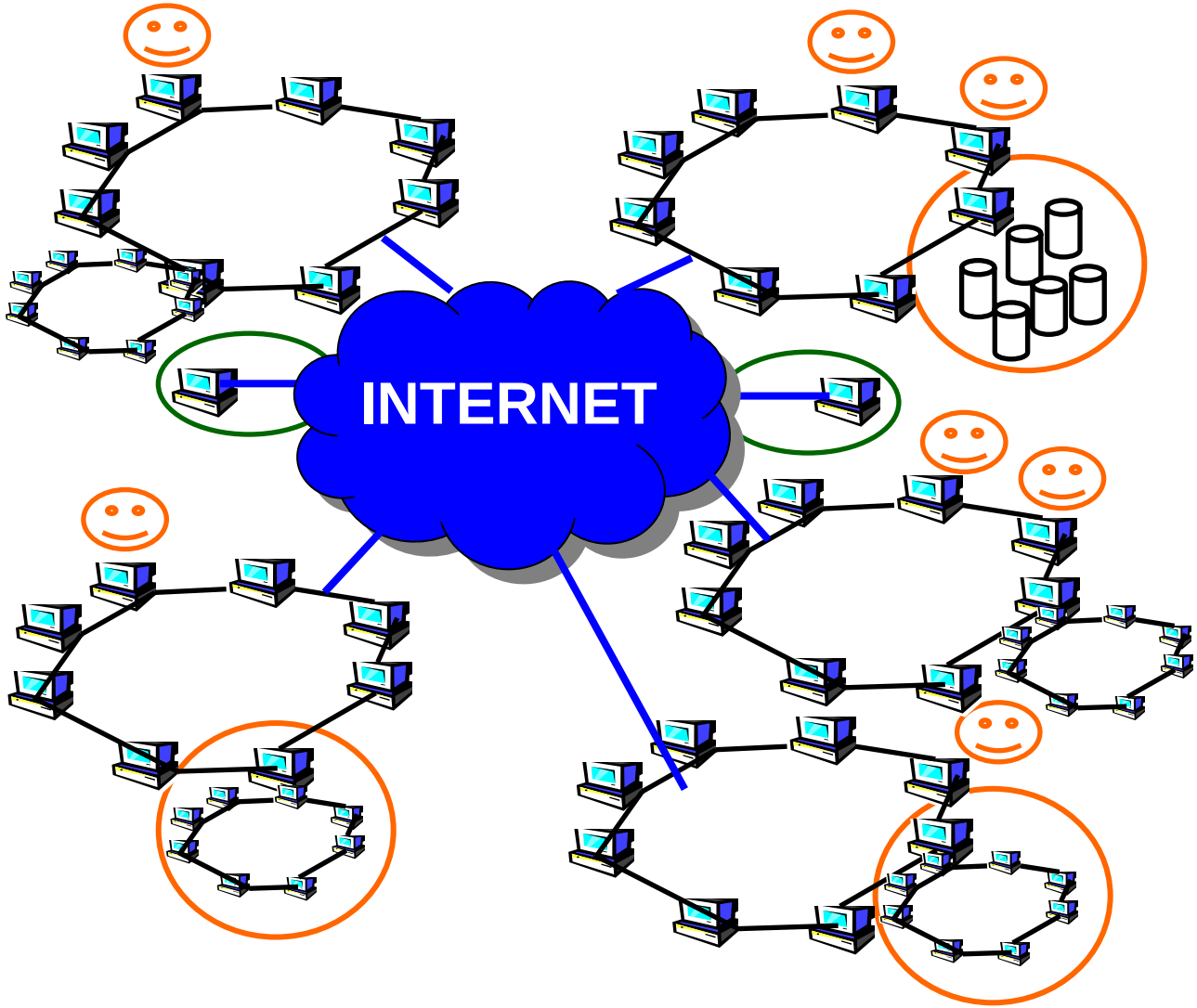
"A grid"

- **The initial vision: "The Grid"**
- **The present reality: Many "grids"**
- **What's a VO?**
  - 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**
  - FOR **COLLABORATION** – not a new architecture for *individuals* to achieve high performance computation – go to HPC centres for that!
  - FOR RESOURCE SHARING / ACCESS across administrative boundaries

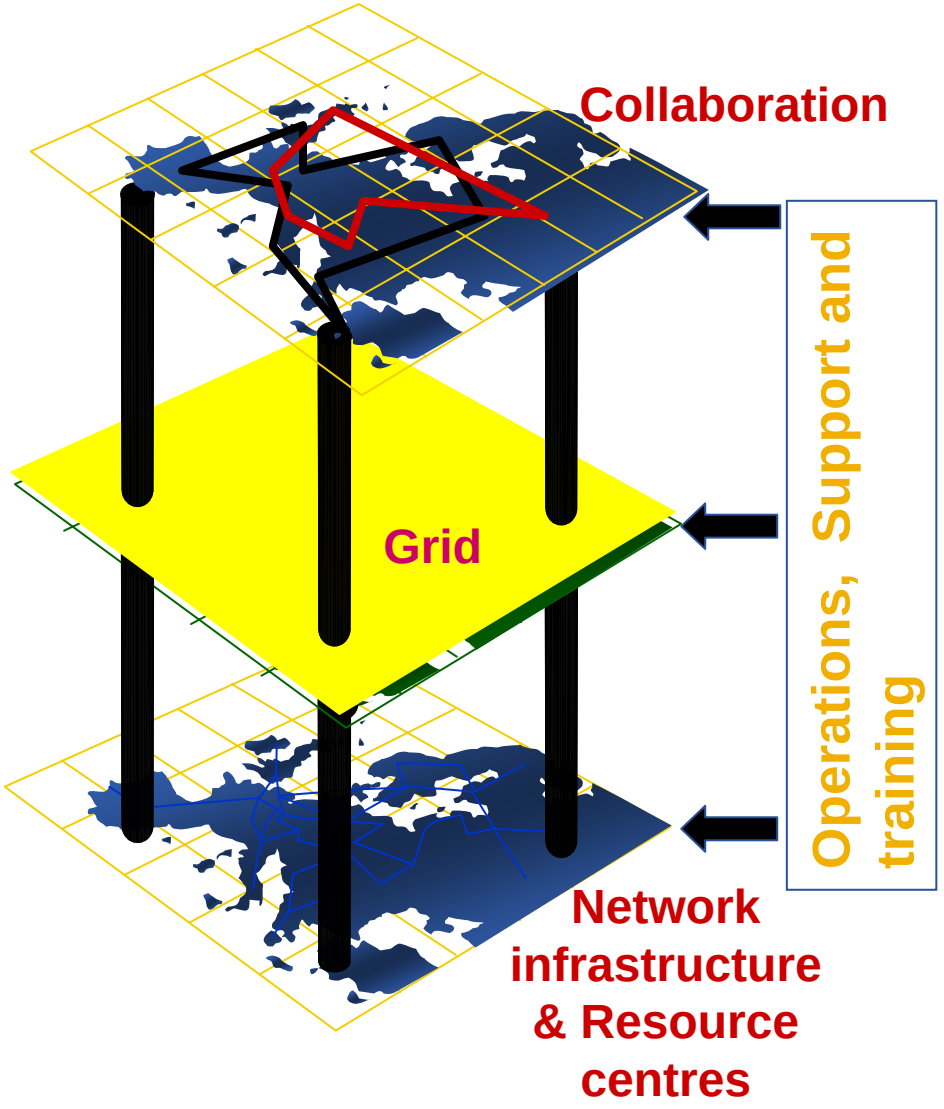
Typical current grid

- **Grid middleware runs on each shared resource**
  - Data storage
  - (Usually) batch jobs on pools of processors
- **Users join VO's**
- **Virtual organisation negotiates with sites to agree access to resources**
- **Distributed services (both people and software) enable the grid**



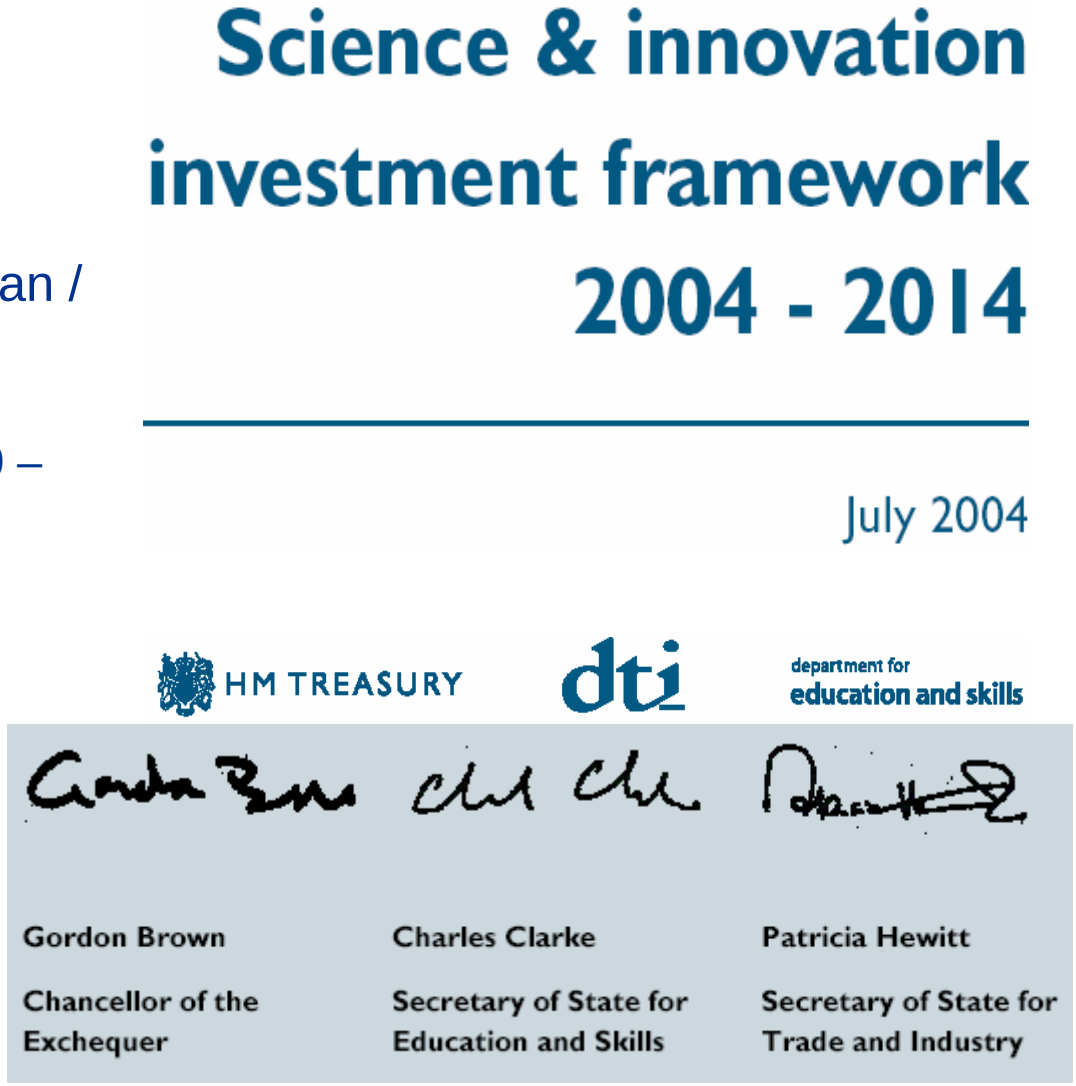
What is e-Infrastructure?

- **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
- **Middleware**
  - Authentication, Authorisation
  - Registries, search engines
  - Toolkits, environments
    - E.g. for collaboration



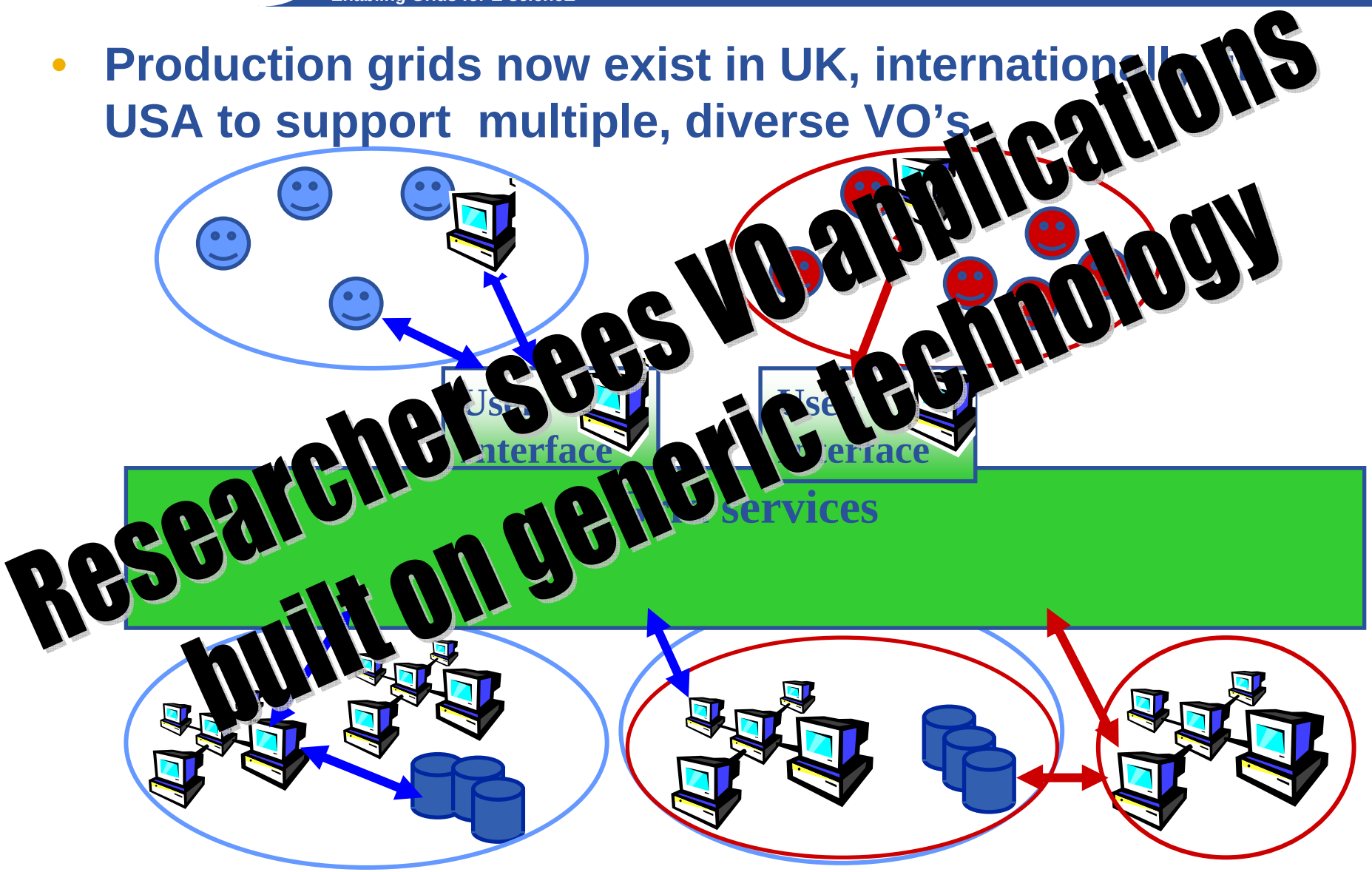
What is e-Infrastructure? – Political view

- **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



A multi-VO Grid

- Production grids now exist in UK, internationally, USA to support multiple, diverse VO's



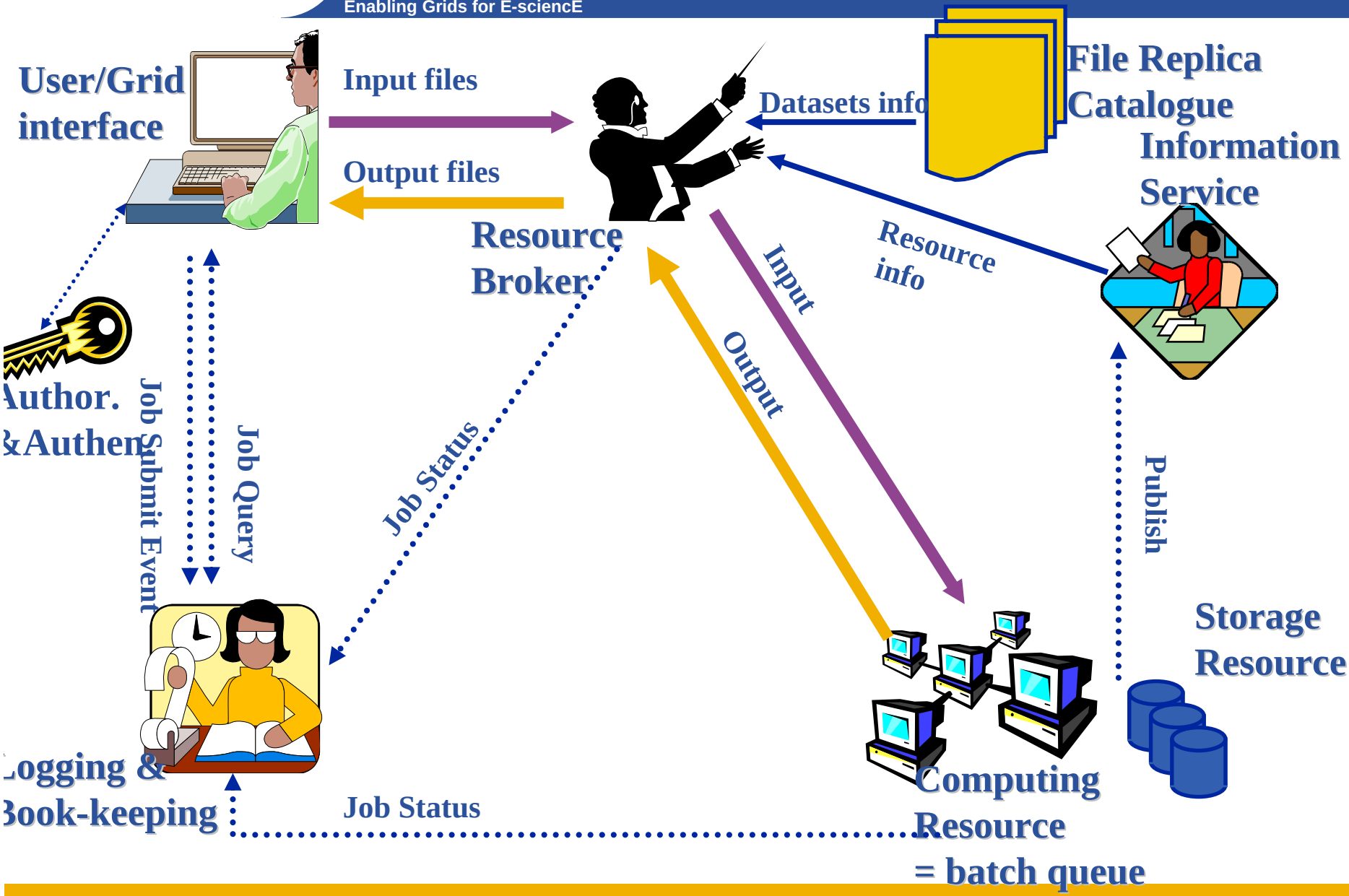
Grid security and trust -1

- **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 users
- **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**

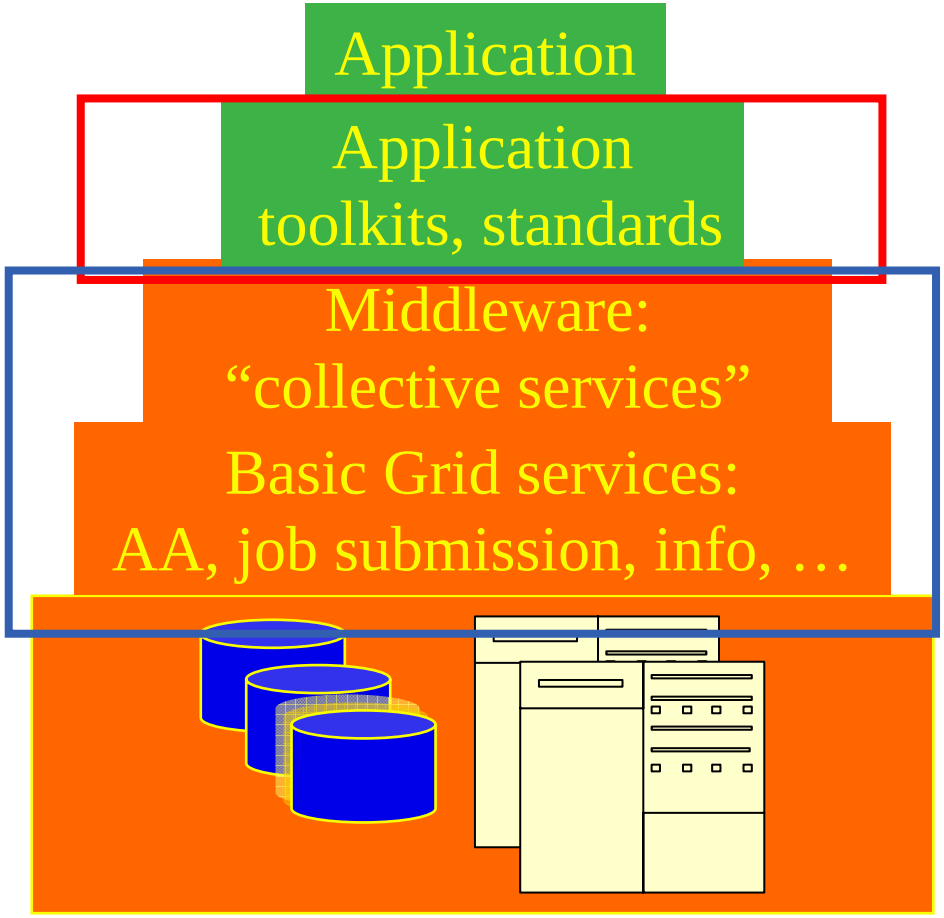
Grid security and trust -2

- **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 other resources you use, where it is mapped to a local account – the mapping is maintained by the VO
- **Common *agreed policies* establish rights for a Virtual Organization to use resources**

Typical current grid



Empowering VO's



VO-specific developments:

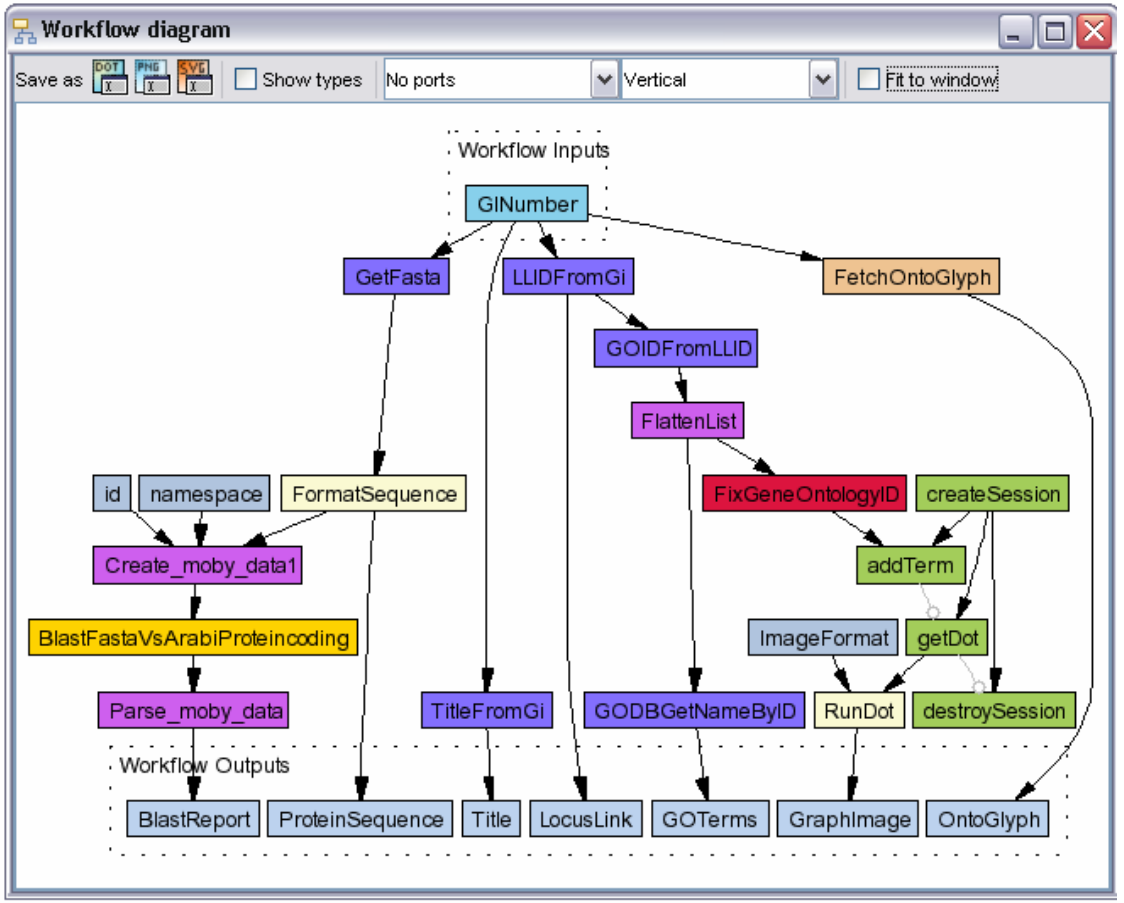
- Portals
- Virtual Research Environments
- Semantics, ontologies
- Workflow
- Registries of VO services

Production grids provide these services.

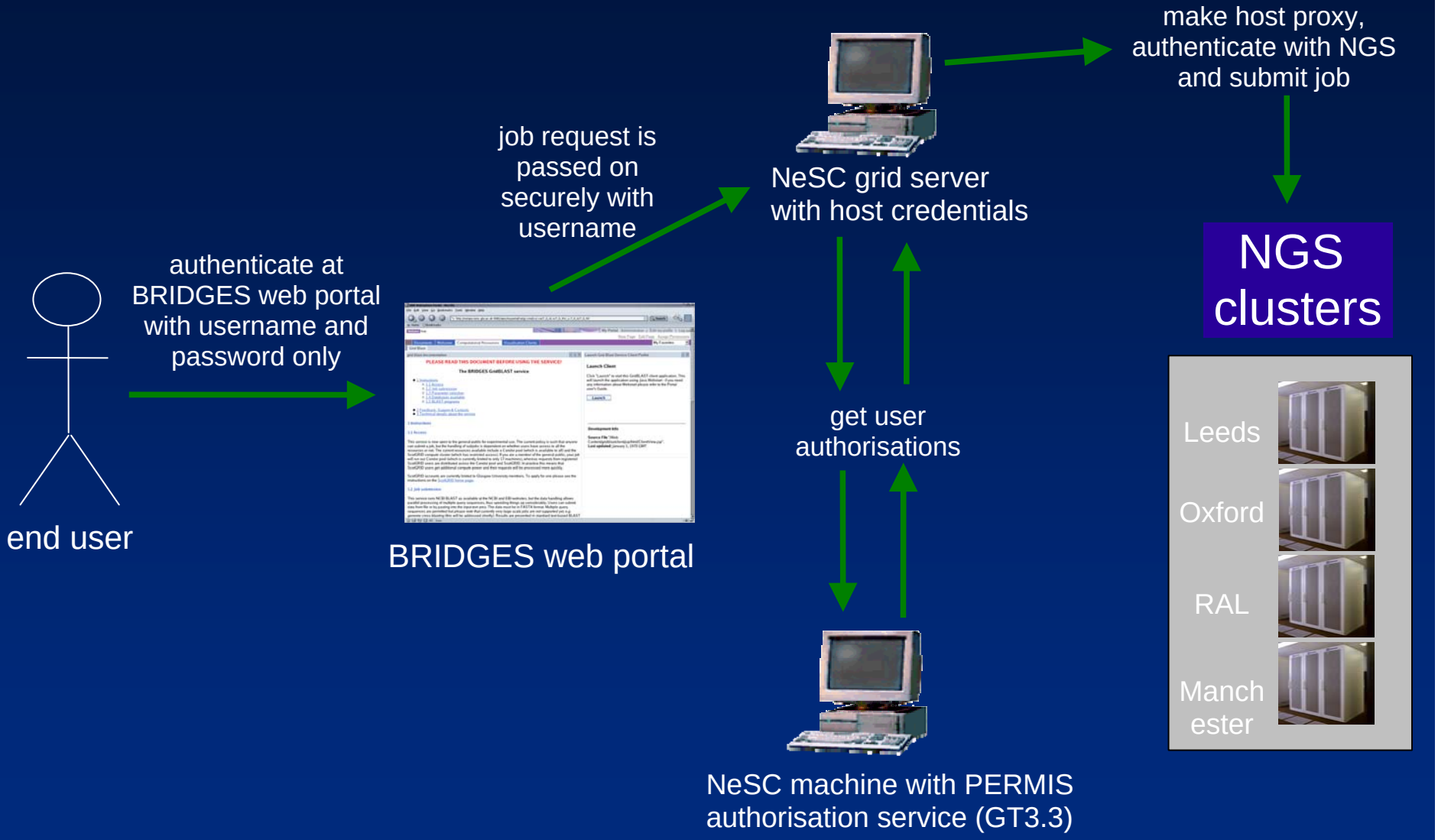
Essential to develop above these to empower non-UNIX specialists!

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



Security in BRIDGES – summary

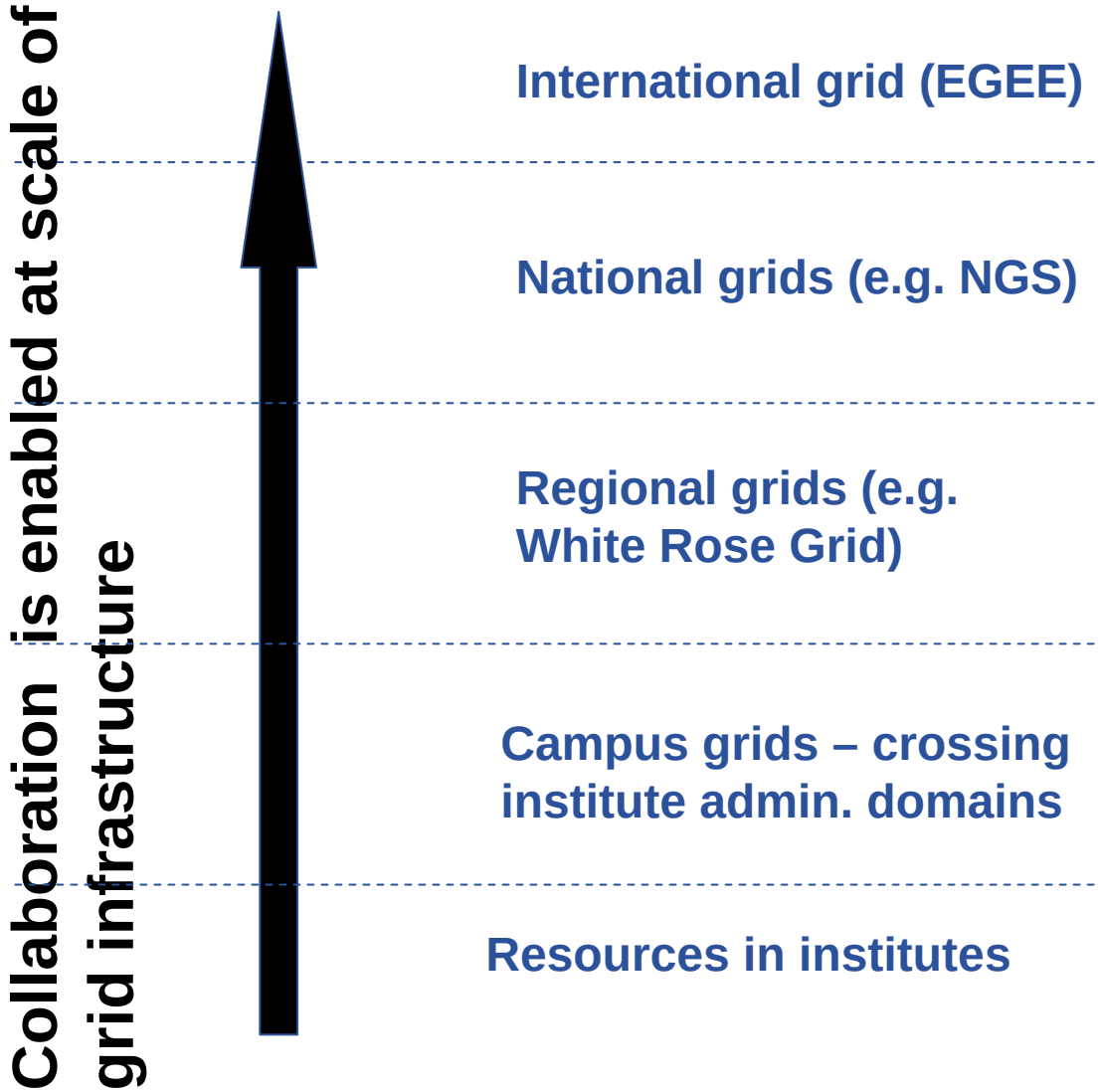


Slide by Micha Bayer, NeSC

The many scales of grids

Two classes of Grids

- **Serving many VOs These can be at many scales**
- **"community grids" for one project**




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

Grid projects - ~ 2003

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

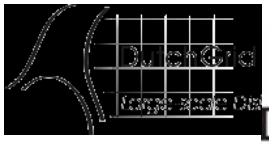
Grids: where are we now?

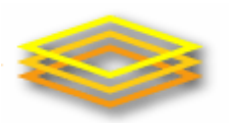
- **Many key concepts identified and known**
- **Many grid projects have tested, and benefit from, these**
- **Major efforts now on establishing:**
  - Standards (a slow process)
    (e.g. Global Grid Forum, http://www.gridforum.org/)
  - Production Grids *for multiple VO's*
    - "Production" = Reliable, 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 (and other!!!) communities
    - Establishing operational procedures and services (people!, policy,..)
- **… whilst research & development continues**

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
- **Some confusion remains after the OGSI era**
  - Many projects sidestepped this by using “pure” WS
- **Globus Toolkit 4 has been released**

National grid initiatives now include…





CroGrid

Contents

- **Introduction to**
  - e-Research and e-Science
  - Grids
  - e-Infrastructure
- **More about Grids**
- **Two key themes**

EGEE is a European-funded project that is establishing an international production Grid.

In progress!

- **In many of its aspects e-Infrastructure is a range of works in progress**
  - Still learning about the operations needed, how to organise at each scale
  - Integrating/developing approaches to security, Authentication, Authorisation, Policy
  - ……

Where are we now? -2

Networks

Grids

Web

Arts

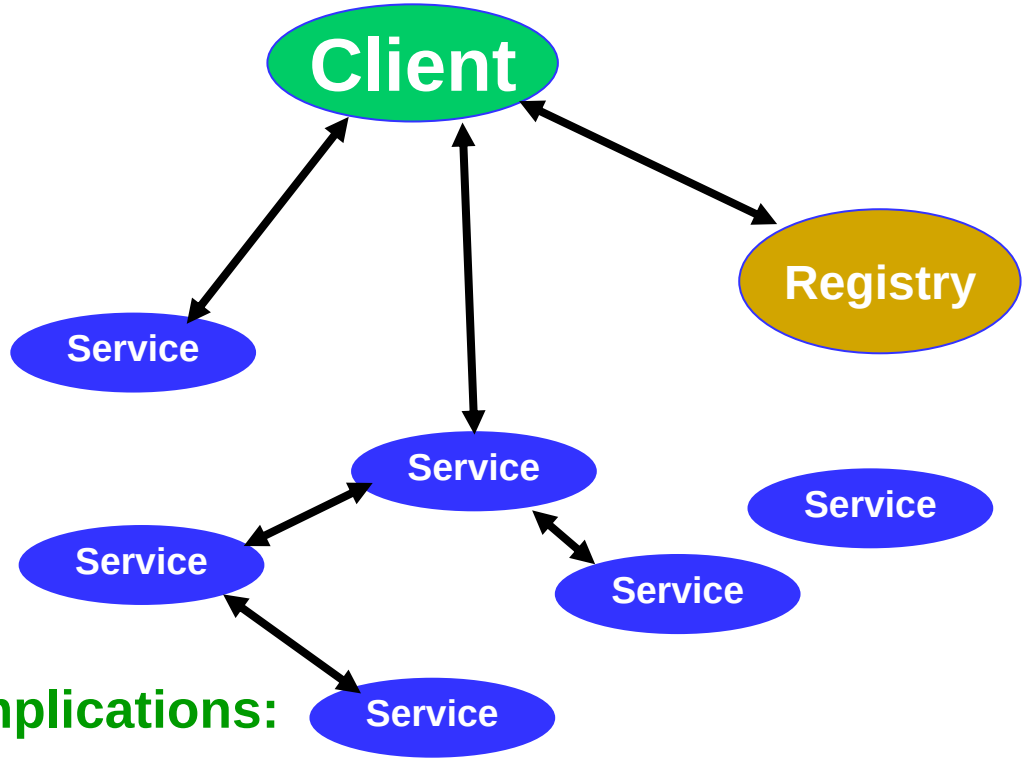
Humanities

E-Soc-Sci

- **Still need skilled developers to "grid-enable" VO's applications & data**
- **Operations effort is very significant**

Service orientation – software components that are…

- 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

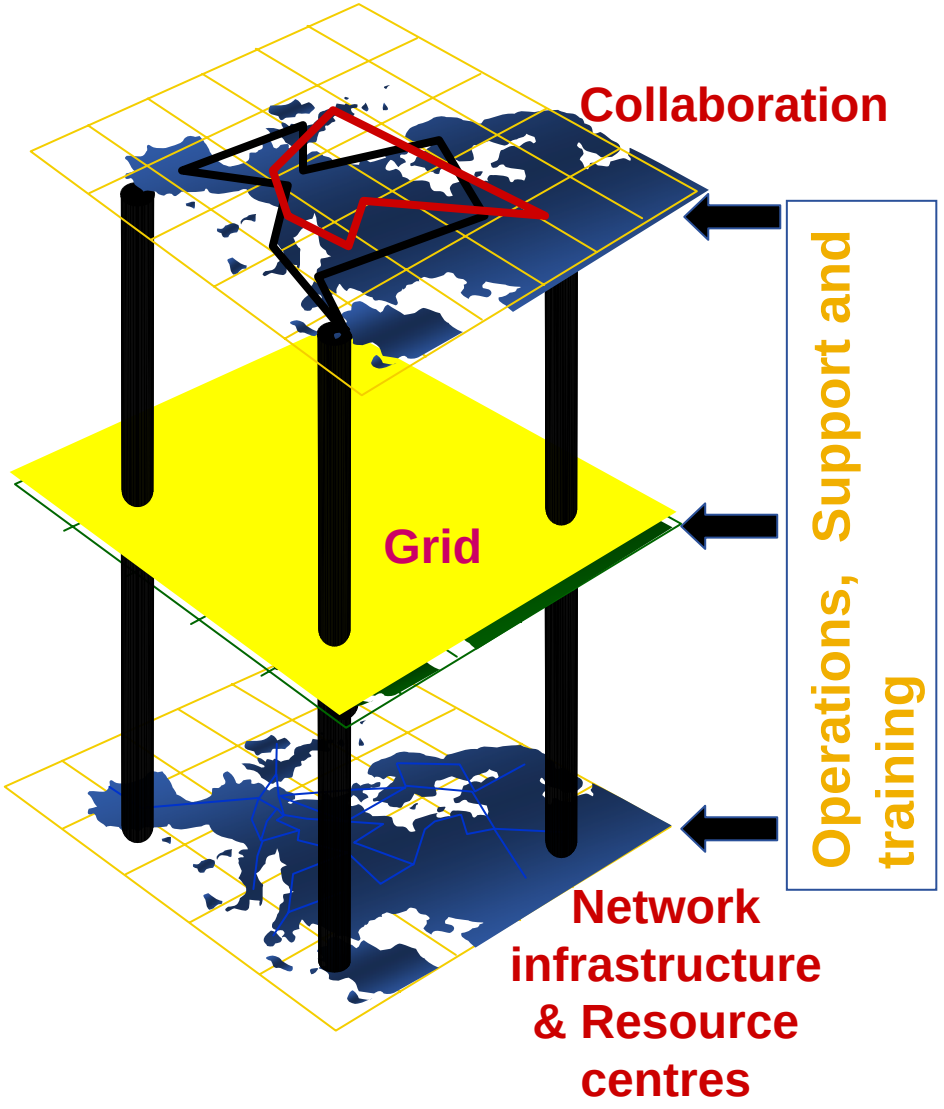


Implications:

- Distinct roles of creating, hosting, using services
- Can compose workflow
- Basis for use of semantics, ontologies

Summary -1: its about collaboration!!

- Grids: collaboration across administrative domains
- Networks: collaboration across geographical distance
- Semantics, ontologies: collaboration across disciplines
- 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 collaboration?