Concepts of Grid Computing

Guy Warner
gcw@nesc.ac.uk

Acknowledgements

- **This talk was originally prepared by Mike Mineter of NeSC and includes slides from previous tutorials and talks delivered by:**
  - Dave Berry, Richard Hopkins (National e-Science Centre)
  - the EDG training team
  - Ian Foster, Argonne National Laboratories
  - Jeffrey Grethe, SDSC
  - EGEE colleagues

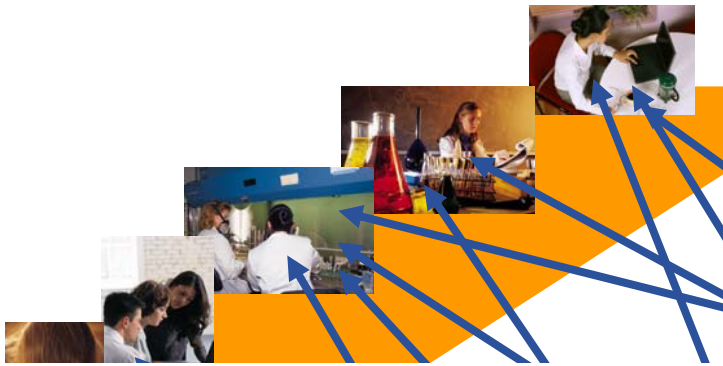
Overview

- **Goal: To introduce the concepts of Grid computing assuming no previous knowledge**
- **Cover the topics of:**
  - Why Grids?
  - What is a grid?
  - Is it secure?
  - Who uses grids? – Some Examples.
  - Is the problem solved? – current status of technology.

Contents

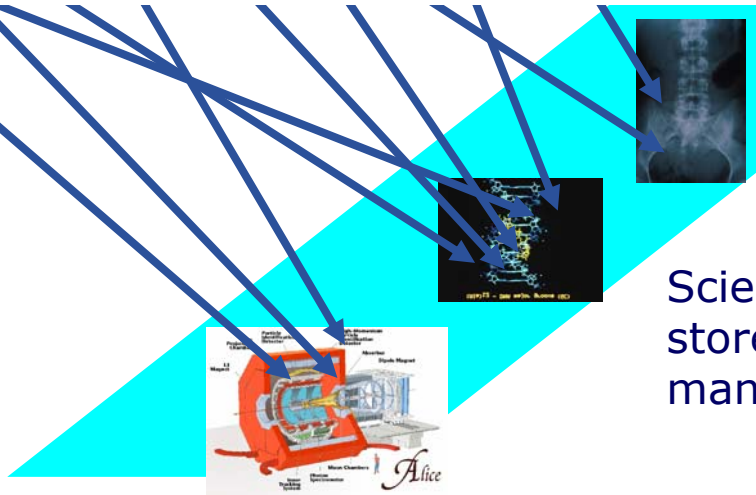
Before Grids

Researchers in many locations need to share resources



FTP, telnet, blood, sweat and tears… and little support for collaboration

There must be a better way of doing this!!!

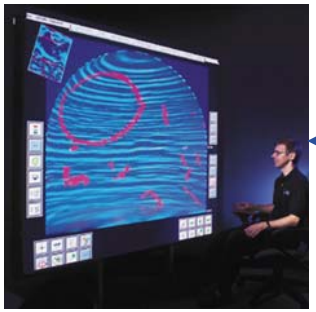


Scientific instruments, data stores and computers in many locations

Middleware

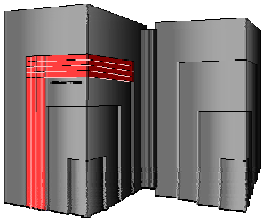
Mobile Access

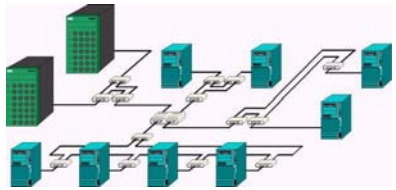
Workstation



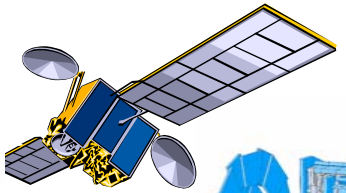
Visualising

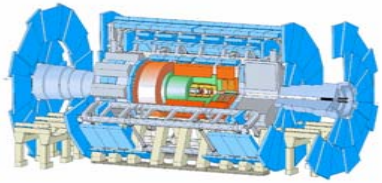
LEWARE



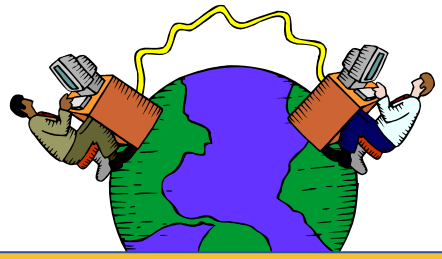


Supercomputer, PC-Cluster





Data-storage, Sensors, Experiments

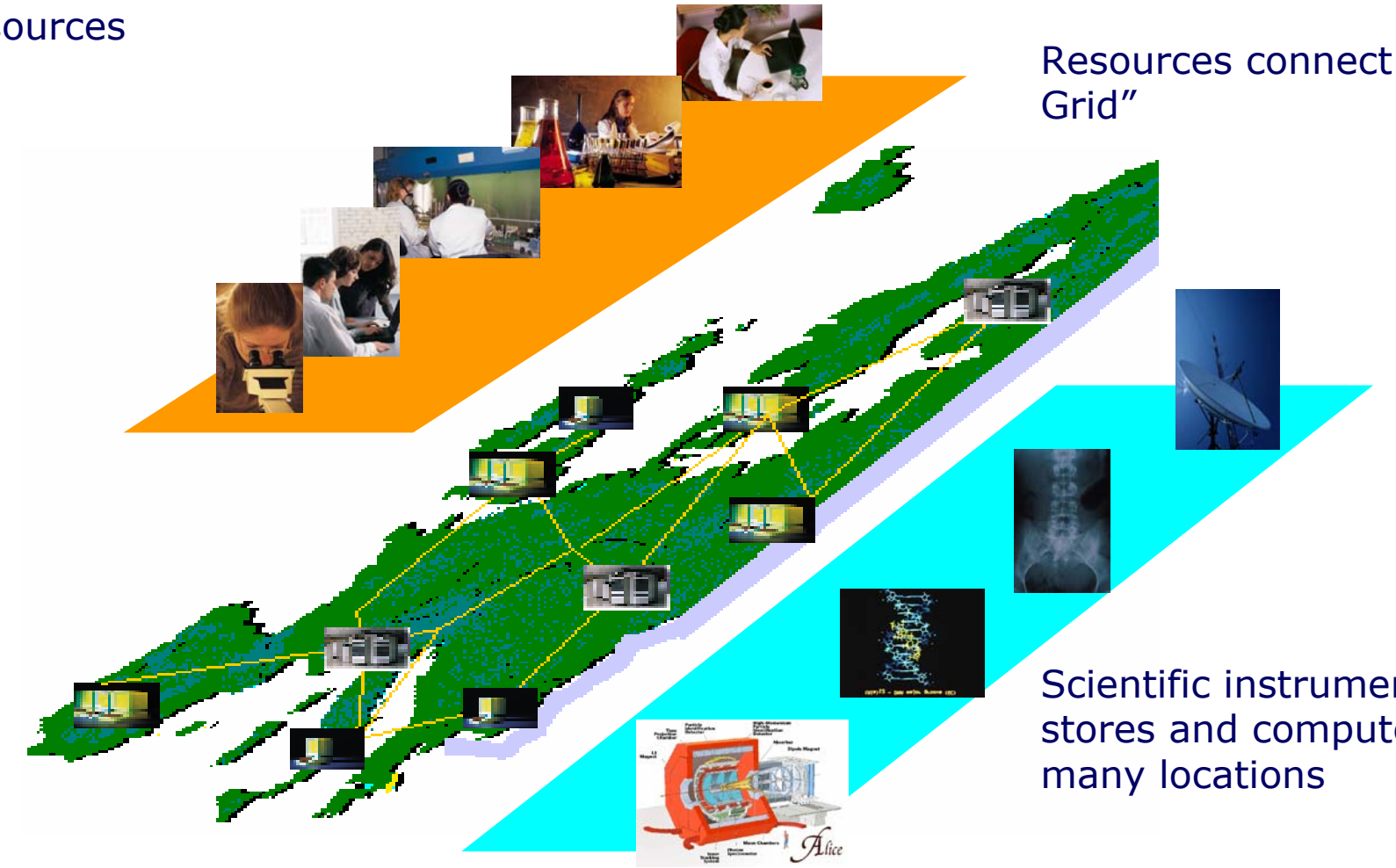


Internet, networks

With Grids

Researchers in many locations need to share resources

Resources connect to "The Grid"



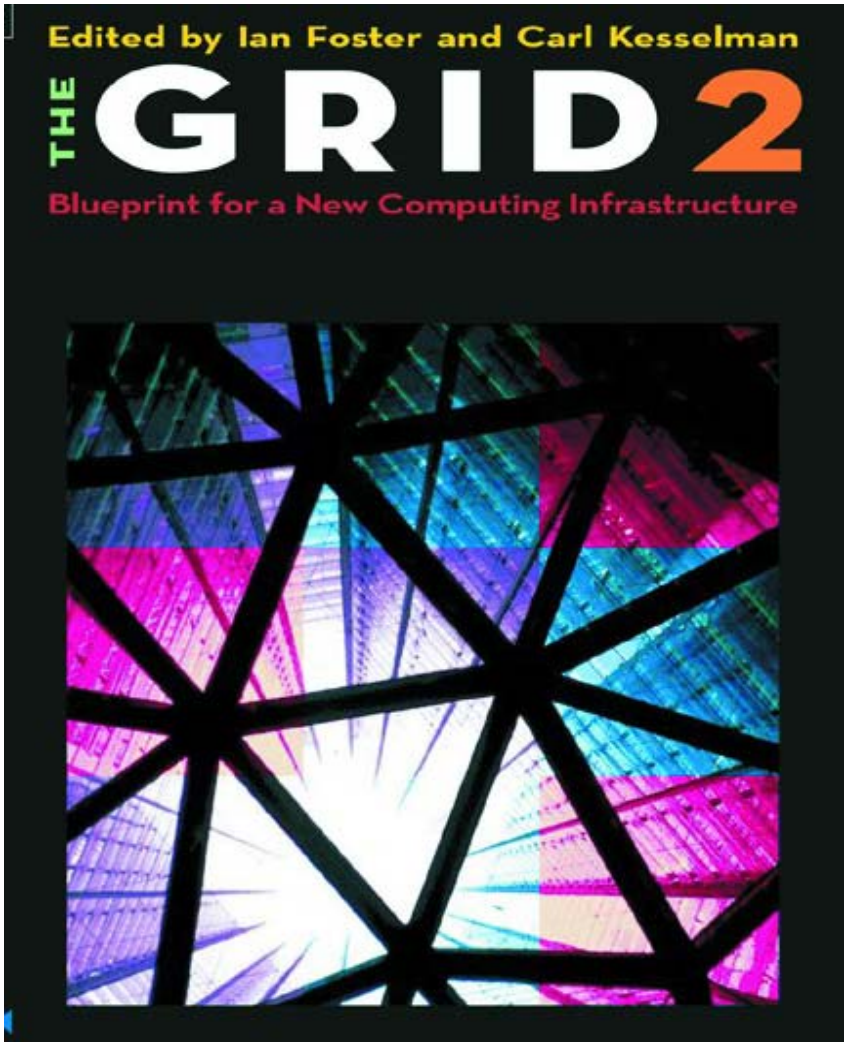
Scientific instruments, data stores and computers in many locations

Slide derived from EDG / LCG tutorials

Why the word "Grid"?

"The word Grid is used by analogy with the electric power grid, which provides pervasive access to electricity … and has had a dramatic impact on human capabilities and society. Many believe that by allowing all components of our information technology infrastructure … the Grid will have a similar transforming effect, allowing new classes of applications to emerge."

Foster and Kesselman – The Grid 2



The grid vision

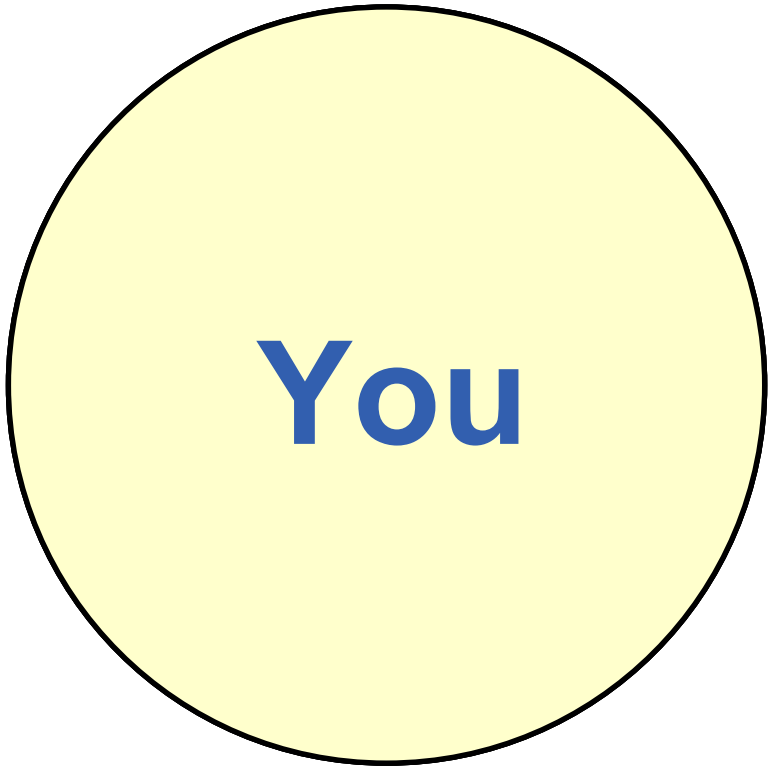
- **The grid vision is of "Virtual computing" (and the information services used to locate computation and storage resources)**
  - Compare this to the web:
    - Web pages are "virtual documents" – different parts come from different sources
    - A search engine is used to locate these resources.

- **The effect of collaboration through sharing resources (and expertise) is to expand the horizons of**
  - Research
  - Commerce – engineering, … ("the knowledge economy")
  - Public service – health, environment,…

Contents

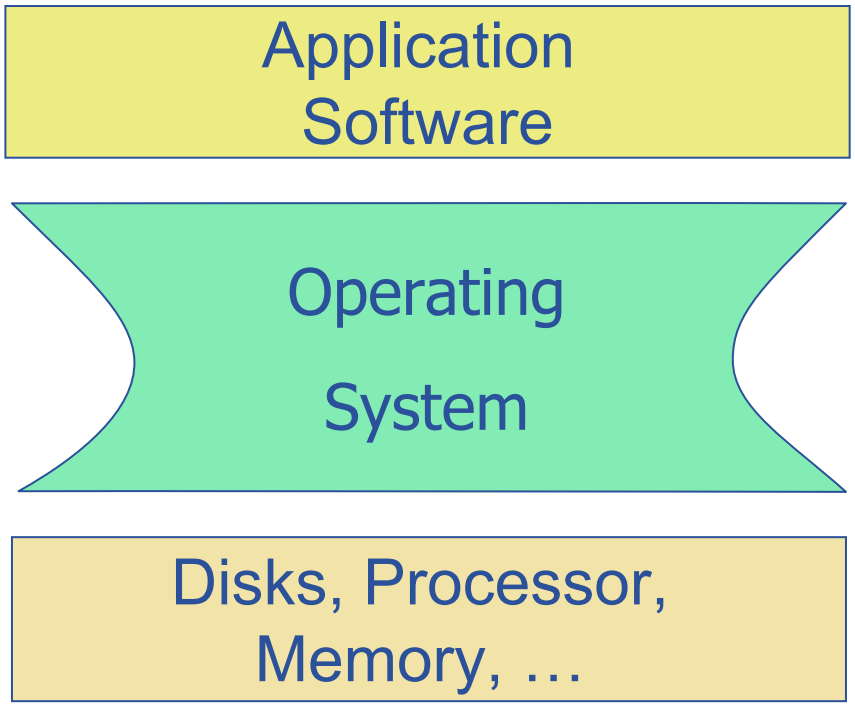
Expanding Horizons

- The initial vision: “The Grid”
- The present reality: Many “grids”
- Each grid is an infrastructure enabling one or more user communities (called “virtual organisations”) to share computing resources
- What makes a user community?
  - People in different organisations seeking to cooperate and share resources across their organisational boundaries



The Single Computer

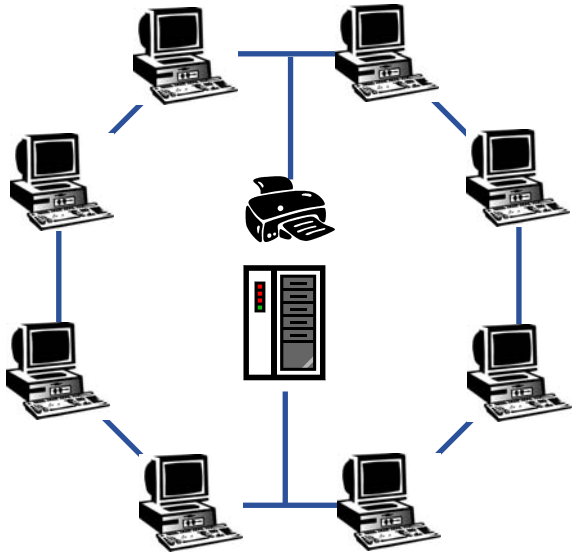
- **The Operating System enables easy use of**
  - Input devices
  - Processor
  - Disks
  - Display
  - Any other attached devices

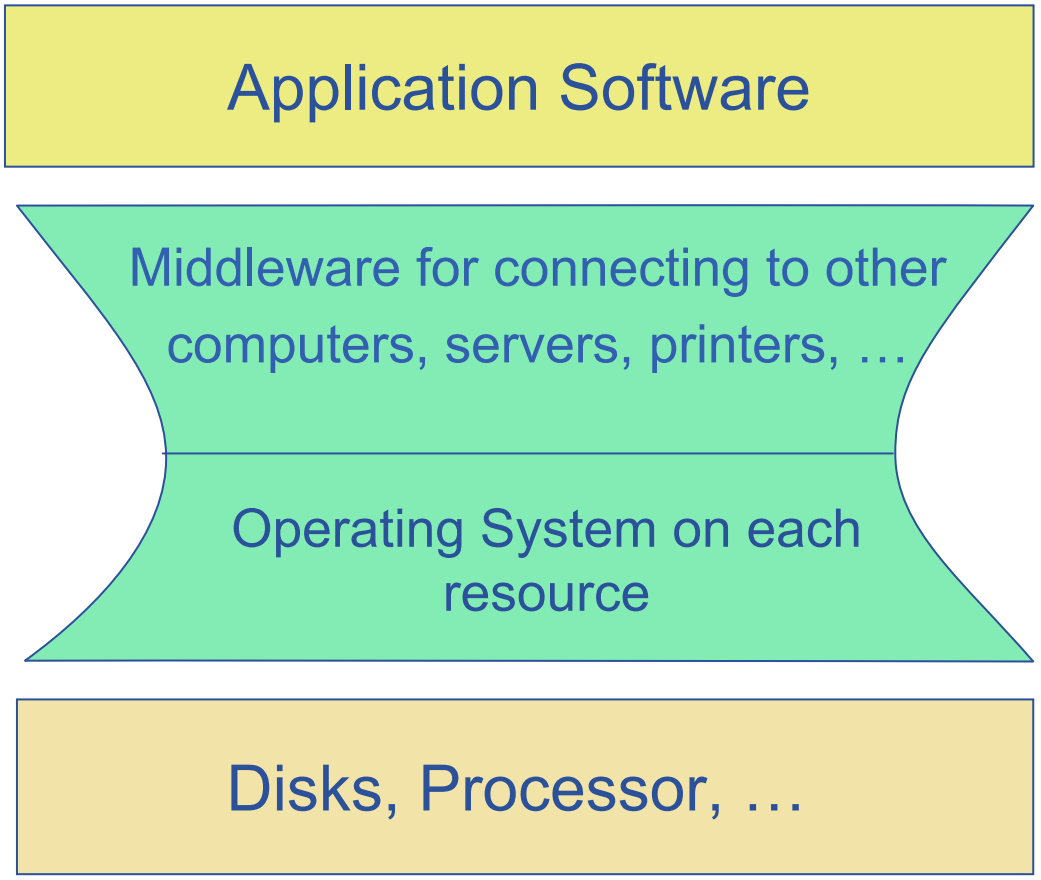


The Local Area Network

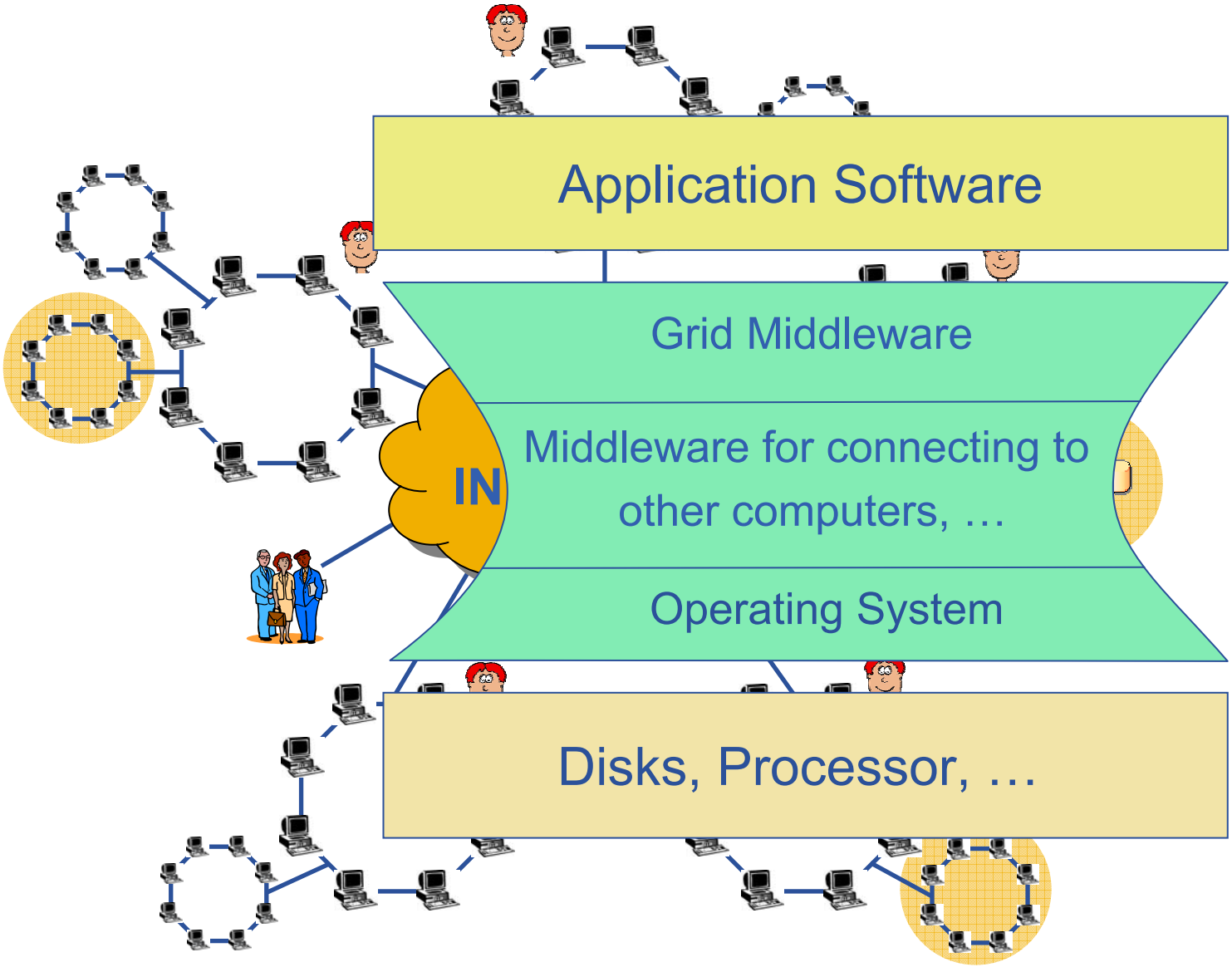
User just perceives “shared resources”, with no regard to location in the building:

- Authenticated by username / password
- Authorised to use own files,…



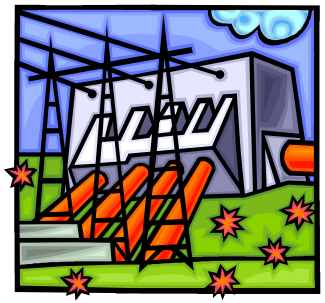
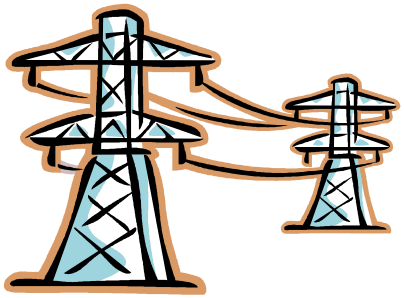


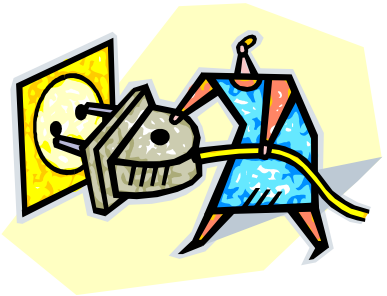
A Grid of Resources



The components of a Grid

- **Infrastructure**
  - networking, computational resources, storage resources, …
- **Middleware**
  - the operating system of the grid, running on all resources.
- **Operations infrastructure**
  - Run enabling services (people + software)
- **Virtual Organization management**
  - Procedures for gaining access to resources

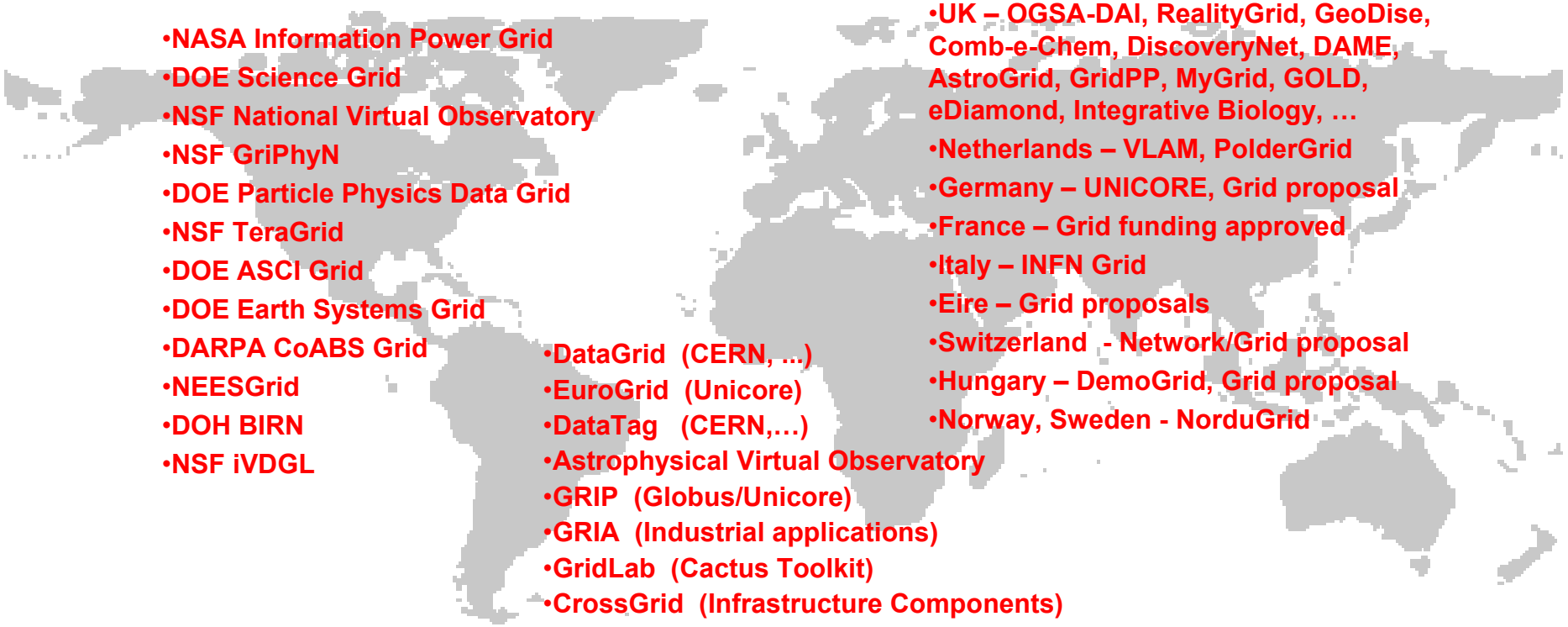


The term “Grid computing”

- **“Grid computing” is a trendy phrase!**
- **It’s therefore also a misused term**
  - Sometimes in Industry : “Grids” = clusters
  - Also used to refer to the harvesting of unused compute cycles, e.g.
    - SETI@home, Climateprediction.net

Grid projects

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

Summary (so far)

- Virtual organisation: people and resources collaborating - crosses admin, organisational boundaries
- Grid middleware runs on each resource
- User just perceives “shared resources” with no concern for location or owning organisation

Contents

Why Grids?

What is a Grid?

Is it Secure?

Some Examples

Current Status

Conclusion

Different Perspectives

- **Users need**
  - single sign-on: the ability to logon to a machine and have the user's identity passed to other resources as required
  - To trust owners of the resources they are using
- **Providers of resources (computers, databases,..) need**
  - risks to be controlled: they are asked to trust users they do not know
  - Minimise impact on security
  - An ability to trace who did what.
- **The solution comes from**
  - Virtual Organisations
  - Digital Certificates

Users and Virtual Organisations

- **Virtual Organisations and trust**
  - User joins a Virtual Organisation
  - Digital certificate is basis of Authentication and Authorisation.
  - Identity passed to other resources you use, where it is mapped to a local account – the mapping is maintained by the Virtual Organisation.
  - The user trusts the Virtual Organisation to only use resources that are safe and secure

- **User just perceives "shared resources" with no concern for location or owning organisation**

Resource-Provider's and VO's

- **Virtual Organisations and trust**
  - A Resource-Provider trusts a Virtual Organisation
  - The Virtual Organisation trusts its users
- **Common *agreed policies* establish rights for a Virtual Organization to use resources**
  - Each resource provider has different usage and security considerations that must be accounted for.

Digital Certificates

- **A digital certificate is the basis for:**
  - Single Sign on:
  - *Authentication*: How do I identify myself to a resource without username/password for each resource I use?
  - *Authorisation*: what can I do? Determined by
    - User's membership of a Virtual Organisation
    - Virtual Organisation negotiations with resource providers
  - *Non-repudiation*: the ability to prove who did what
- **Certificate Authorities issue digital certificates**
  - Certificate Authorities provide digital certificates after certifying user's identity – for example by showing a passport
  - Digital certificates issued by national Certification Authorities are recognized internationally – a pre-requisite of an international grid.

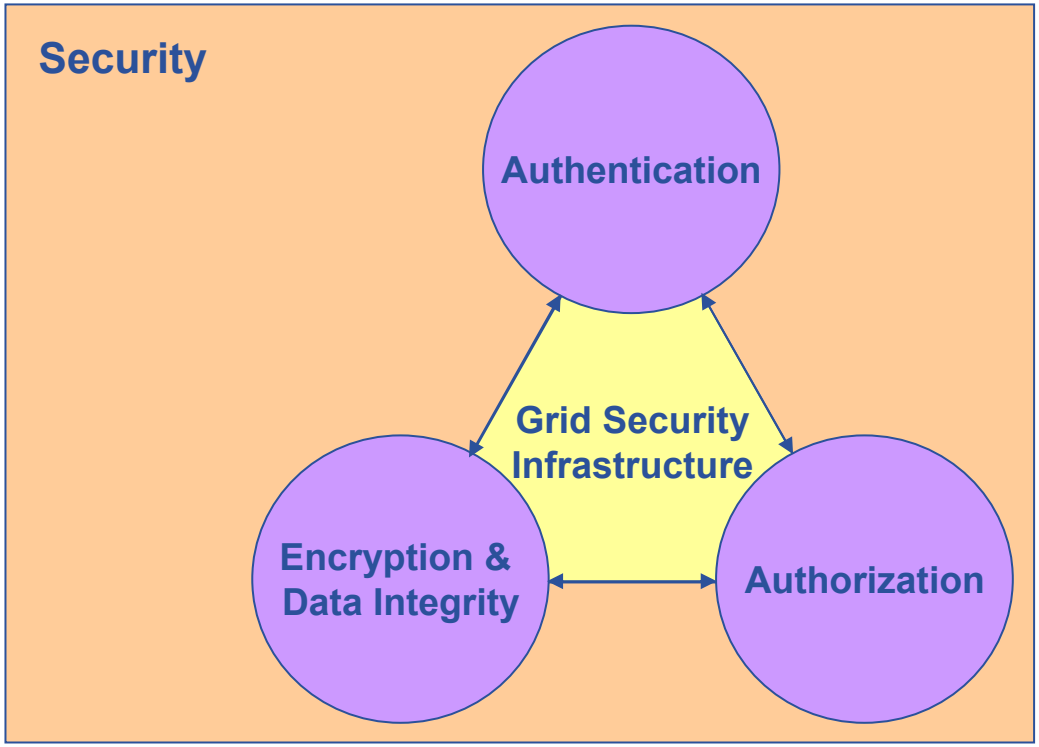
Certificate Authority Websites

- **A list of Certificate Authorities that mutually recognize each other:**
  http://www.gridpma.org/.

- **A list of the Certificate Authorities in the EU:**
  http://marianne.in2p3.fr/datagrid/ca/ca-table-ca.html
  - E.g. In UK go to http://www.grid-support.ac.uk/ca/ralist.htm

The Grid Security Infrastructure

- The “Grid Security Infrastructure” middleware is the basis of (most) production grids



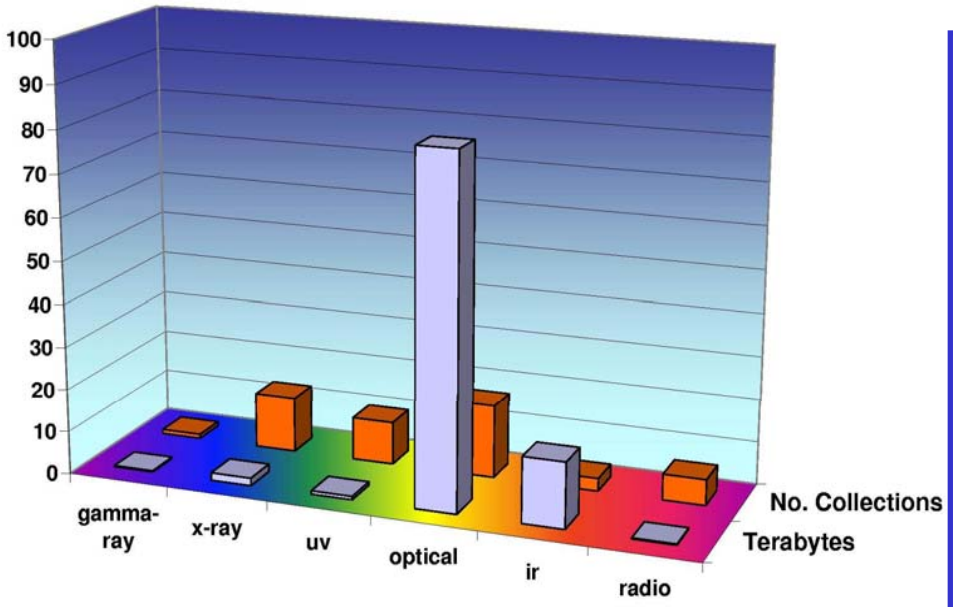
For all of this to work you must keep your digital certificate secure

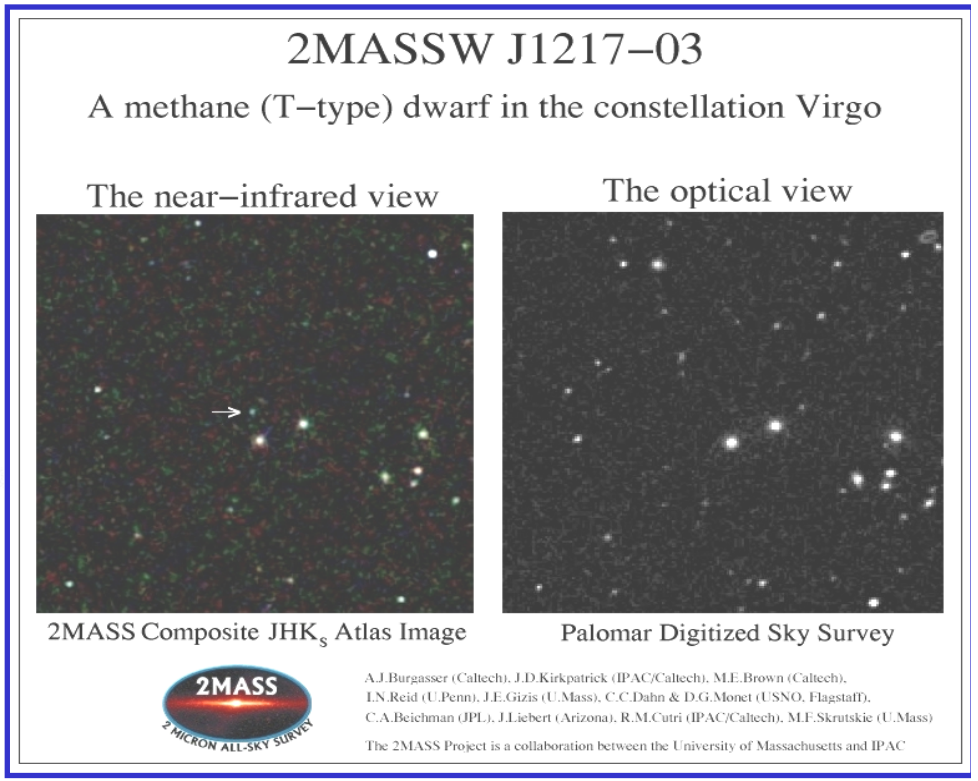
Contents

An early driver of Grids: e-Science

- **What is e-Science?**
  Collaborative science that is made possible by the sharing across the Internet of resources (data, instruments, computation, people's expertise...)
  - Often very compute intensive
  - Often very data intensive (both creating new data and accessing very large data collections) – data deluges from new technologies
  - **Crosses organisational boundaries**

Astronomy





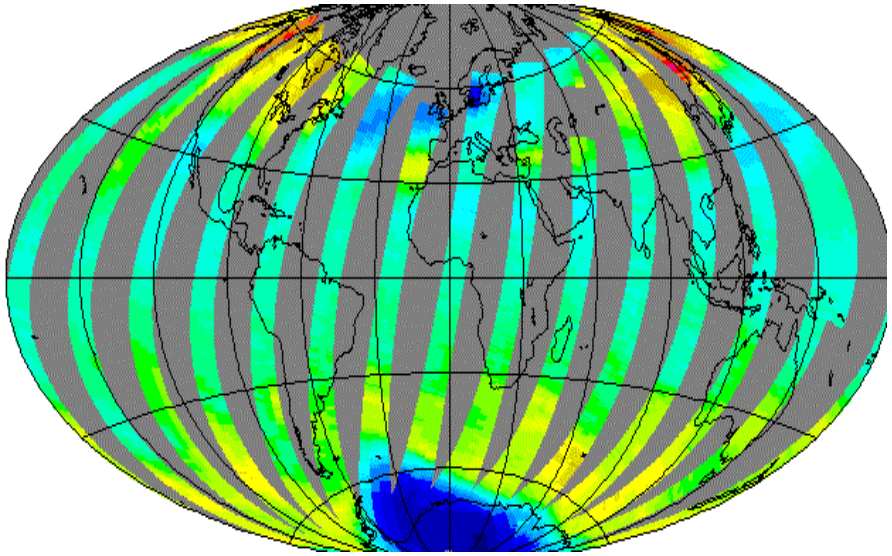
No. & sizes of data sets as of mid-2002, grouped by wavelength

- 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

Earth Observation: Ozone

- **Building on European Datagrid experience**
- **To produce and store the Ozone profiles or columns**
  - Enhance availability
- **To extend the processing capabilities**
  - Validation against other data
  - Mid-latitude ozone studies
  - ...
- **To facilitate collaboration**
  - Including with emerging large scale European projects

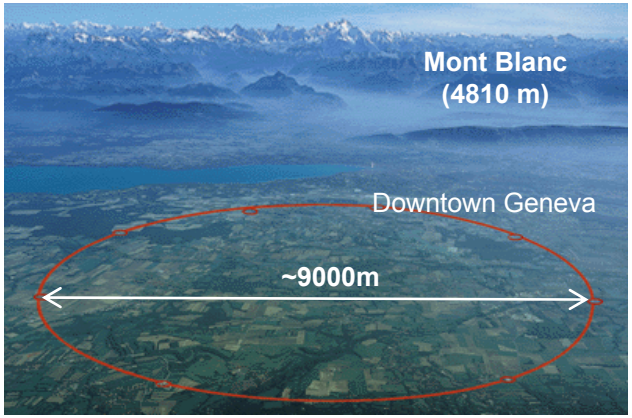


GOME instrument
(~75 GB - ~5000 orbits/y)

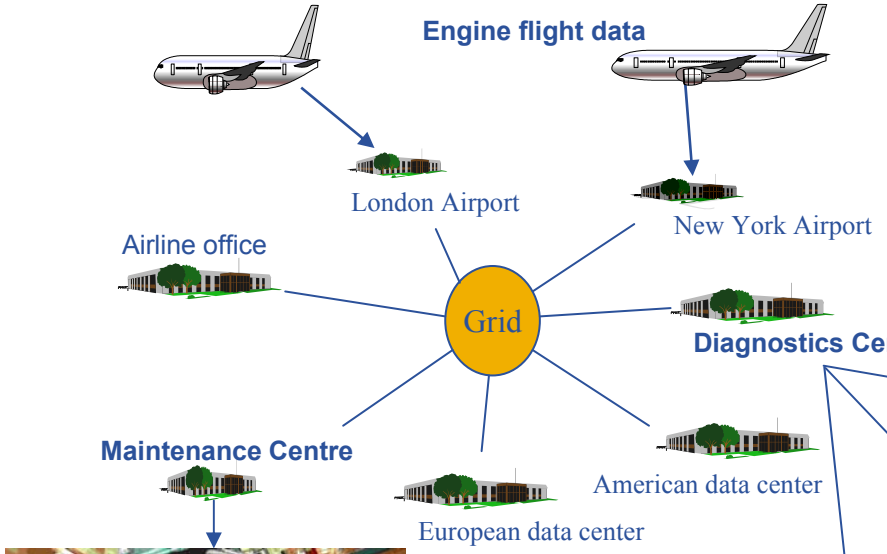
~28000 profiles/day

Large Hadron Collider at CERN

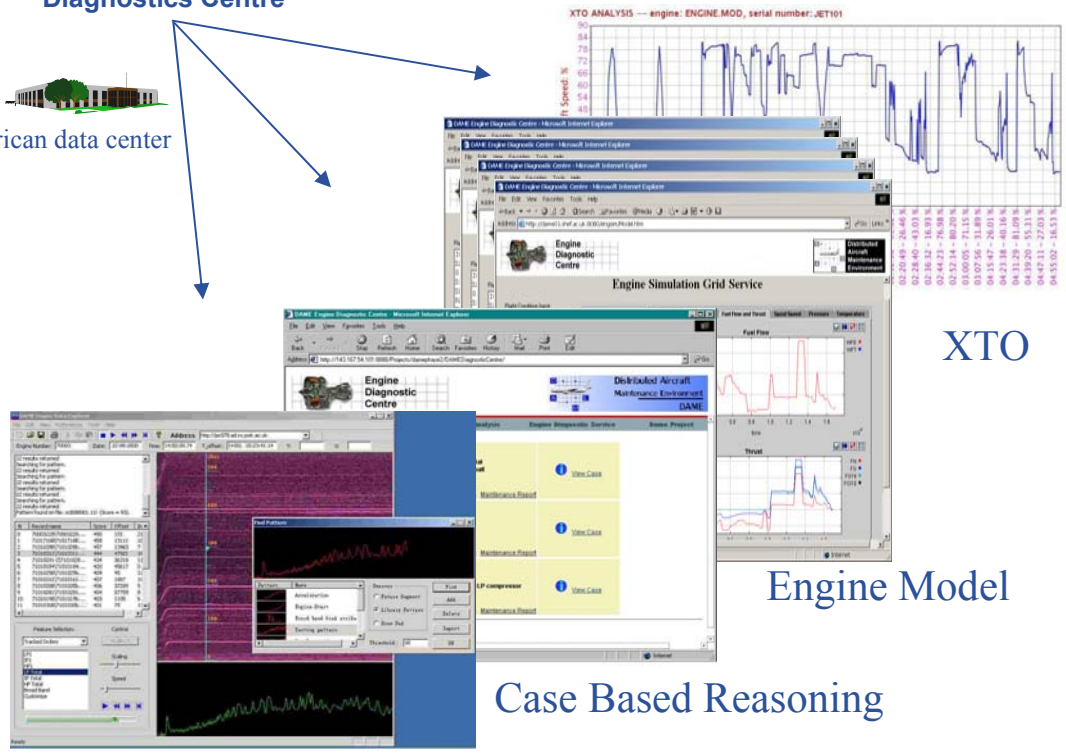
- **Data Challenge:**
  - 10 Petabytes/year of data !!!
  - 20 million CDs each year!
- **Simulation, reconstruction, analysis:**
  - LHC data handling requires computing power equivalent to ~100,000 of today's fastest PC processors!
- **Operational challenges**
  - Reliable and scalable through project lifetime of decades



Distributed Aircraft Maintenance Environment



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



XTO

Engine Model

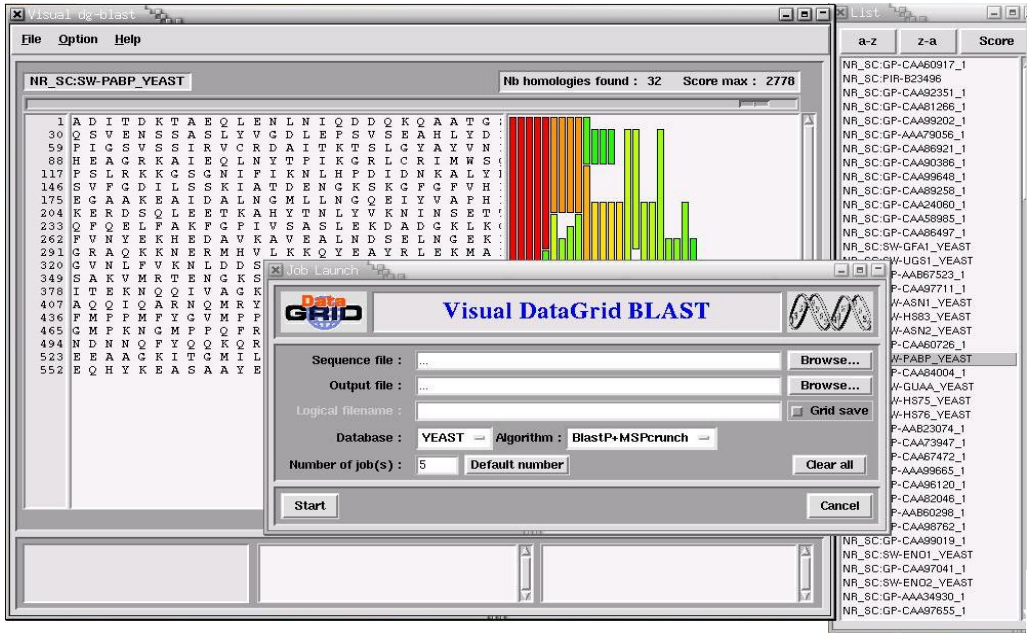
Case Based Reasoning

Companies:
Rolls-Royce
DS&S
Cybula

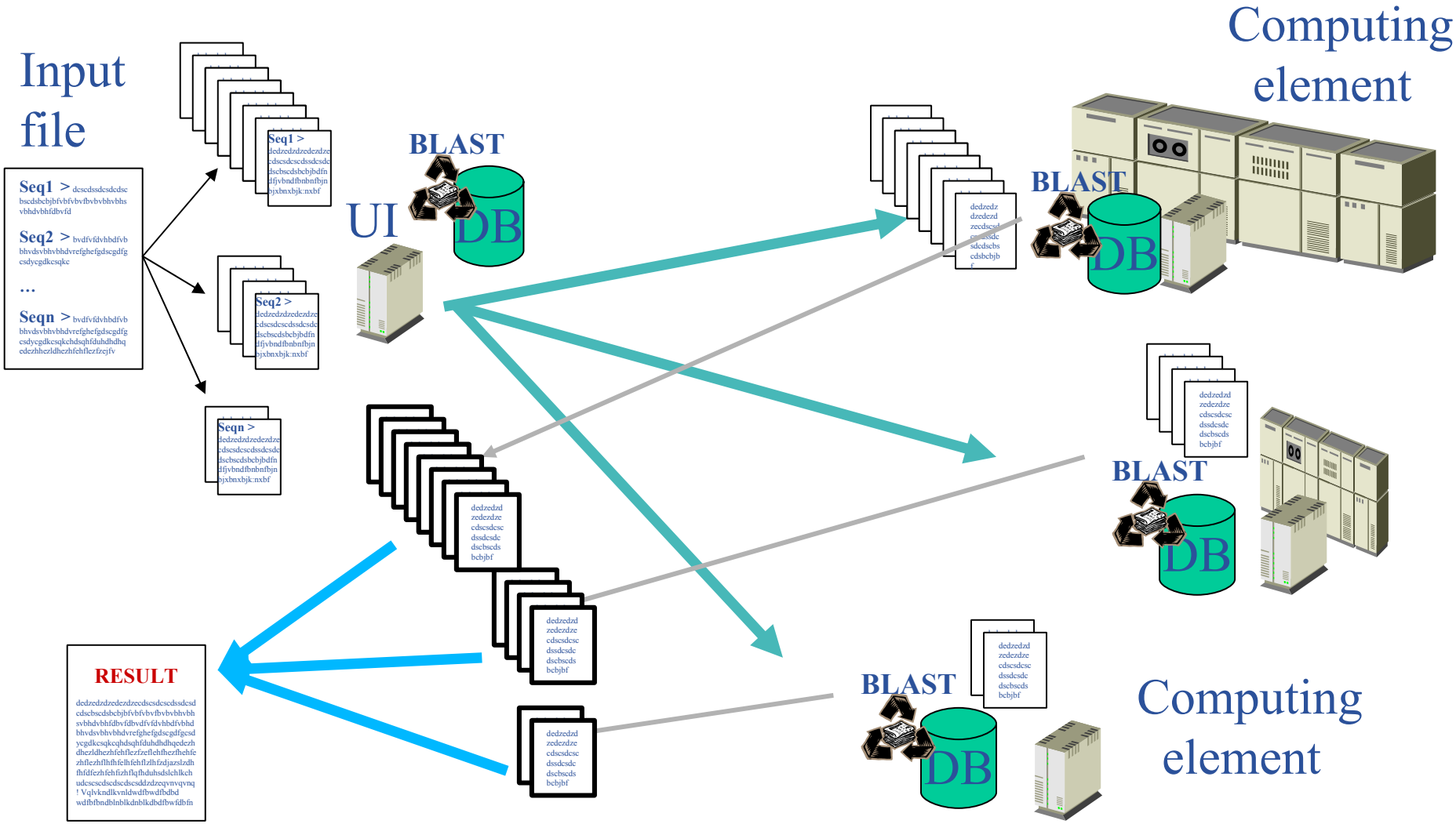
Universities:
York,
Leeds,
Sheffield, Oxford

BLAST – comparing DNA or protein sequences

- **BLAST is the first step for analysing new sequences: to compare DNA or protein sequences to other ones stored in personal or public databases.**
- **Ideal as a grid application – trivial to parallelise as independent concurrent jobs on one or more CEs.**
  - Requires resources to store databases and run algorithms
  - Large user community



BLAST gridification



Contents

Why Grids?

What is a Grid?

Is it Secure?

Some Examples

Current Status

Conclusion


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

Grids: still a work in progress!!!

- **Many key concepts identified and known**
- **Many grid projects have tested these**
- **Major efforts now on establishing:**
  - Standards (a slow process)
    (Global Grid Forum, http://www.gridforum.org/)
  - Production Grids *for multiple Virtual Organisation's*
    - "Production" = Reliable, sustainable, with commitments to quality of service
      - ***In Europe, EGEE***
      - *In UK, National Grid Service*
      - *In US, Teragrid*
    - One stack of middleware that serves many research (and other!!!) communities
    - Operational procedures and services (people!, policy,..)
  - New user communities
- **… whilst research & development continues**

Contents

Summary

- **Collaboration across multiple organisations – sharing data, computers, instruments, application software,..**
- **Single sign-on to resources in multiple organisations**
- **Need for people-services as well as middleware services**
- **Drives are towards**
  - Production services (reliable, sustainable,… – against which research projects, etc… can plan with confidence)
    - In Europe, EGEE
  - Standards
  - New user communities