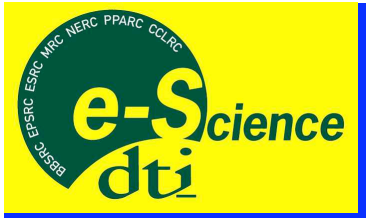
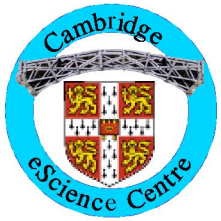
Building Virtual Communities with eScience

Andy Parker
Director, Cambridge eScience Centre



What is e-Science?



"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 the Research Councils, OST

"e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet.

Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation delivered to the individual user scientists."

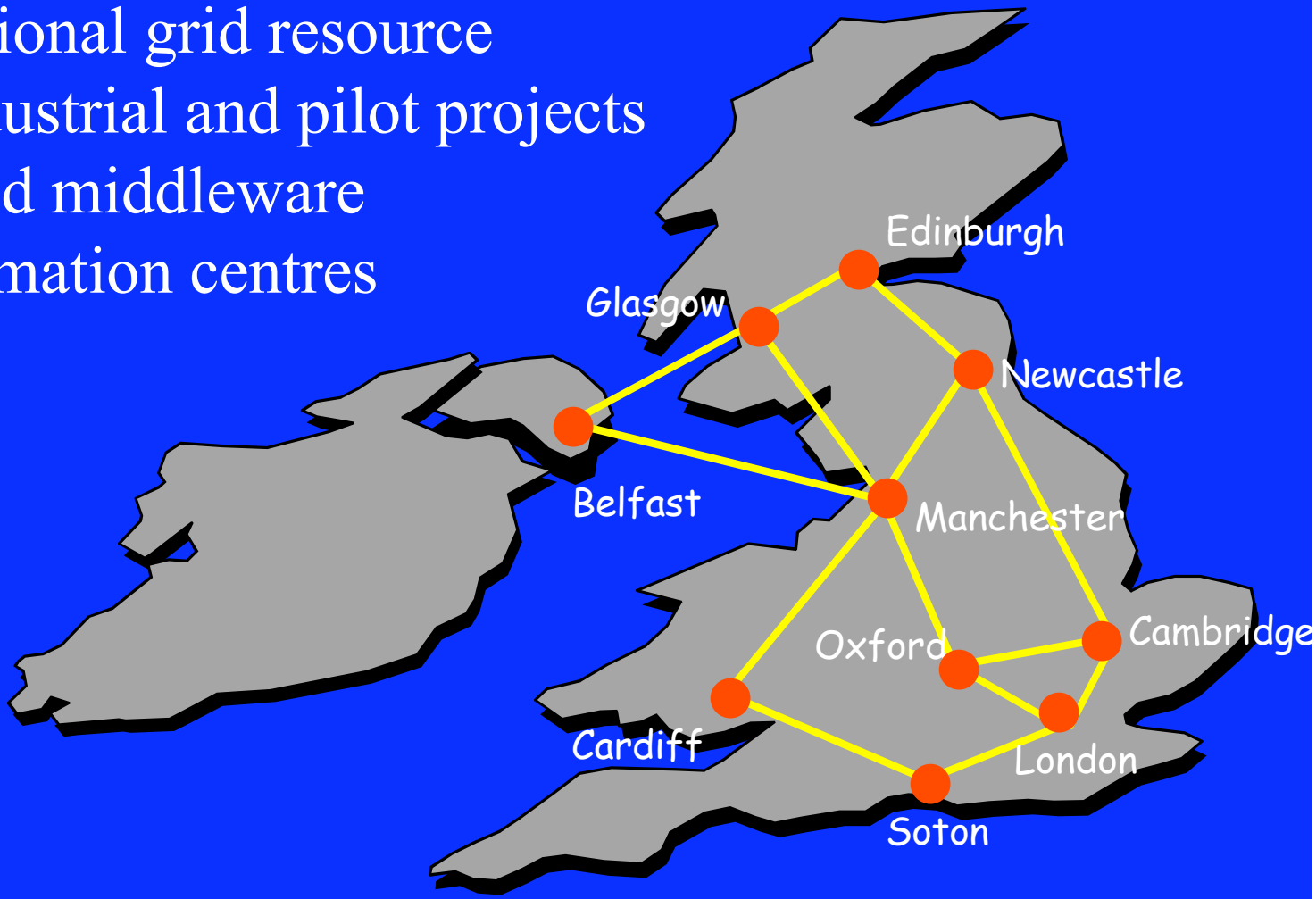
Research Councils Website

What is the Grid?

- The Grid will allow virtual organisations to collaborate in a transparent manner:
  - Remote automatic job submission to all VO resources by intelligent scheduling system
  - Access to distributed data via metadata tagging
  - very high bandwidth connectivity to allow realtime access to large remote data collections
  - High quality video conferencing and remote visualisation
- Requires large computing facilities connected by high quality network.

UK e-Science Grid

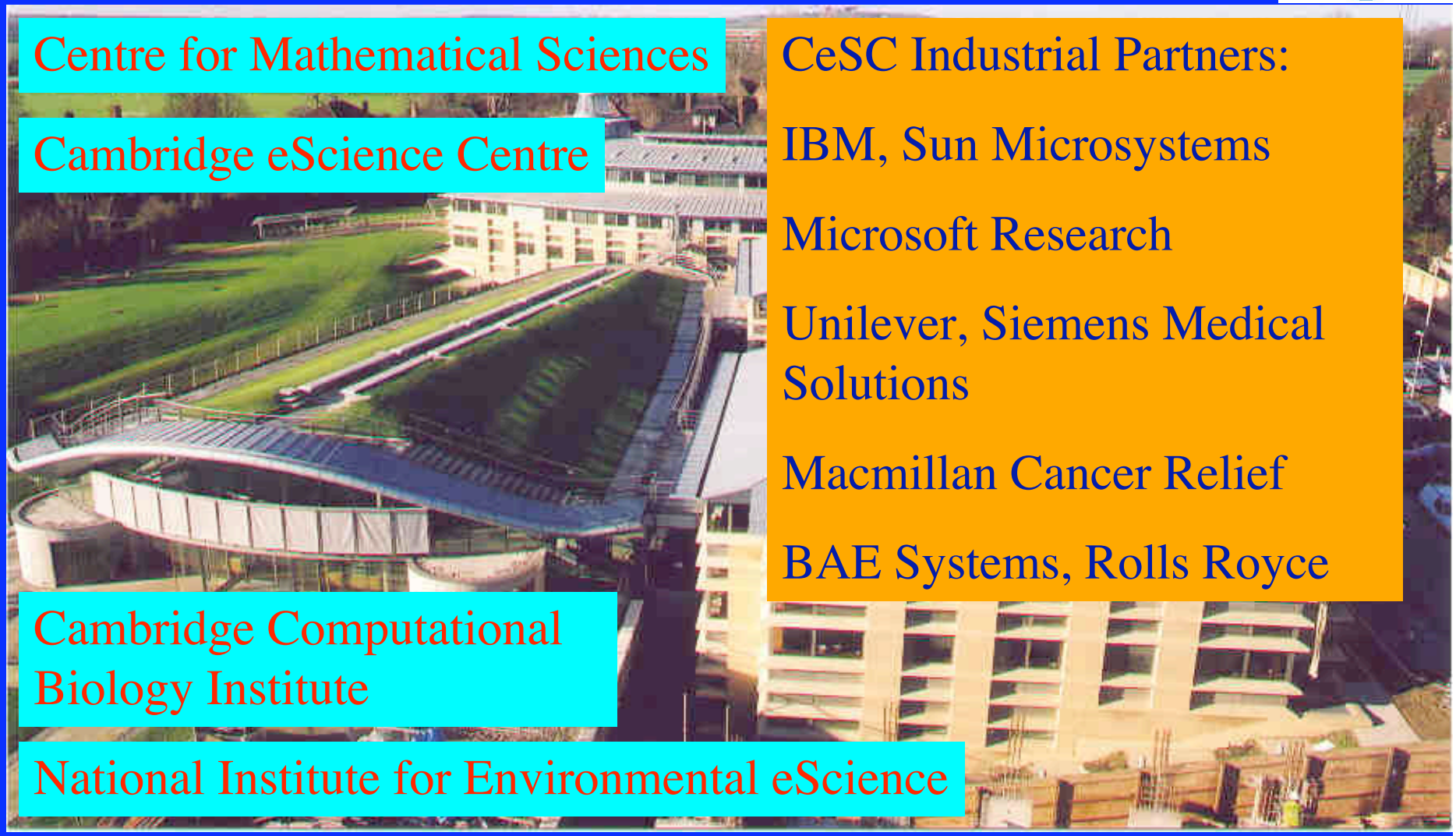
– provide national grid resource
– through industrial and pilot projects advance grid middleware
– act as information centres



Access Grid

Each Centre has an Access Grid Node
-high specification video conferencing

Cambridge eScience



Centre for Mathematical Sciences

Cambridge eScience Centre

Cambridge Computational Biology Institute

National Institute for Environmental eScience

CeSC Industrial Partners:

IBM, Sun Microsystems

Microsoft Research

Unilever, Siemens Medical Solutions

Macmillan Cancer Relief

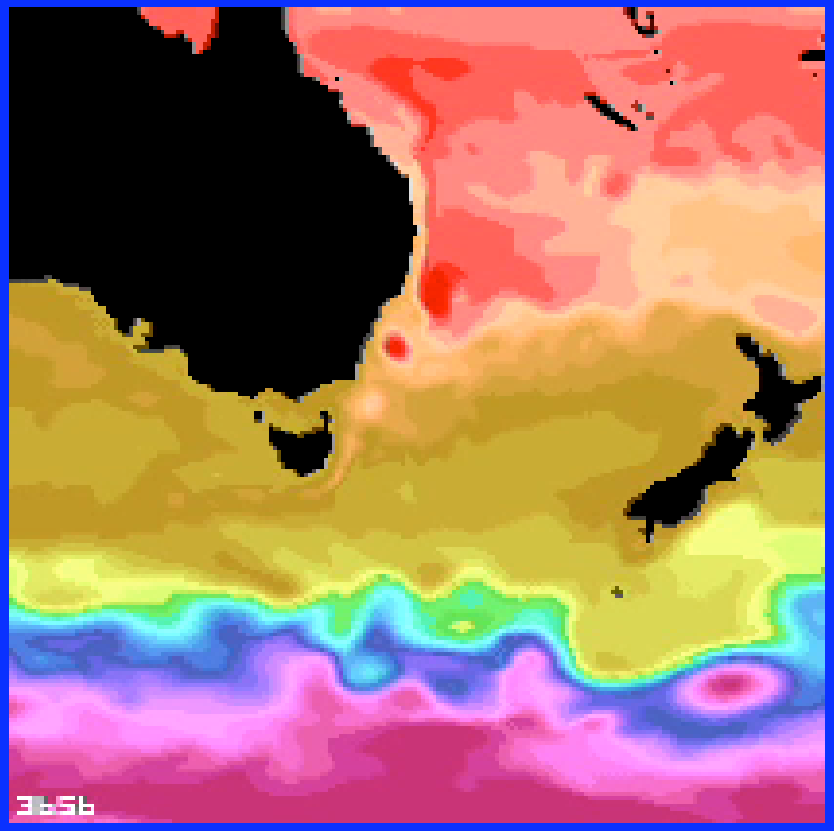
BAE Systems, Rolls Royce

Cambridge Computational Biology Institute

- Link Cambridge expertise in medicine, biology, mathematics and the physical sciences.
- World centre that will develop new knowledge and its application to health, quality of life and wealth creation.
- Research topics: New MPhil Course
  - basic genetics of bacteria
  - developmental biology
  - evolutionary biology
  - complex cell biology of human disease
  - systems biology.
- Multidisciplinary approach using advanced informatics techniques:

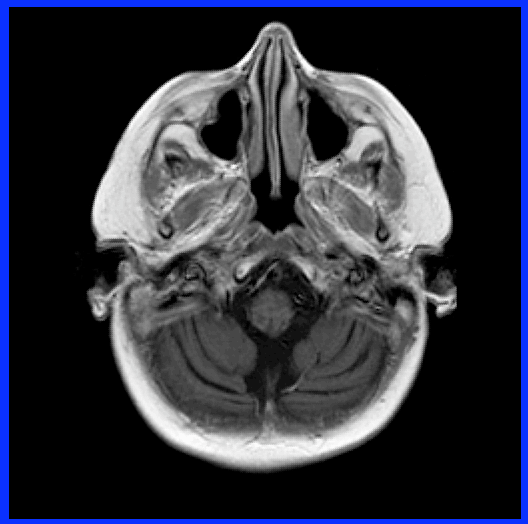
Supported by CeSC and major driver of the Campus Grid

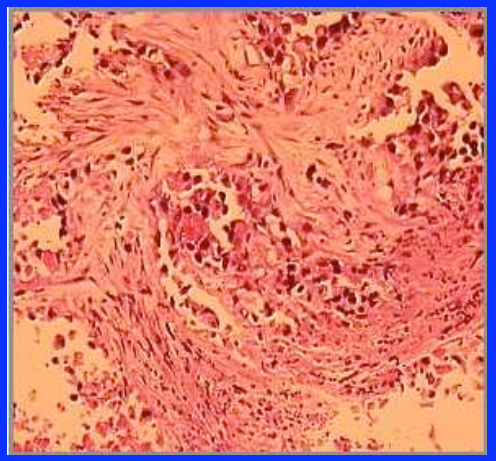
National Institute for Environmental eScience



- The NIES is located with CeSC and shares facilities and staff
- Director: Martin Dove of Dept. of Earth Sciences.
- NIES activities:
  - "Newton Institute" style workshops in Env. Sci. areas
  - Demonstrator projects using Grid technologies

Telemedicine on the Grid





West Anglia NHS Cancer Network

Addenbrooke's NHS Trust

Macmillan cancer relief

UNIVERSITY OF CAMBRIDGE

SCHOOL OF CLINICAL Medicine

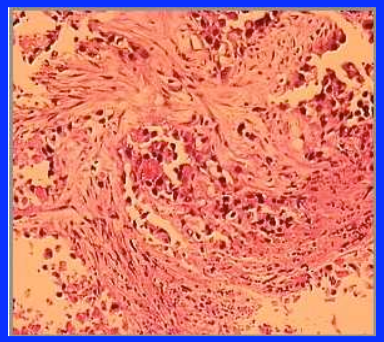
SIEMENS

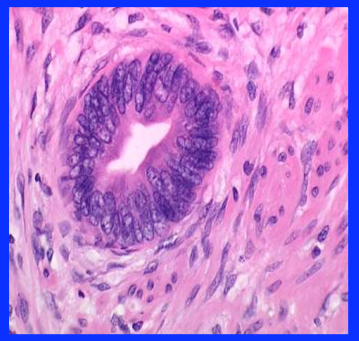
The West Anglia Cancer Network

- Cancer Centre
  - Addenbrooke's/ Papworth
- Cancer Units
  - Bedford
  - Peterborough
  - West Suffolk
  - Harlow
  - Hinchingbrooke
  - King's Lynn

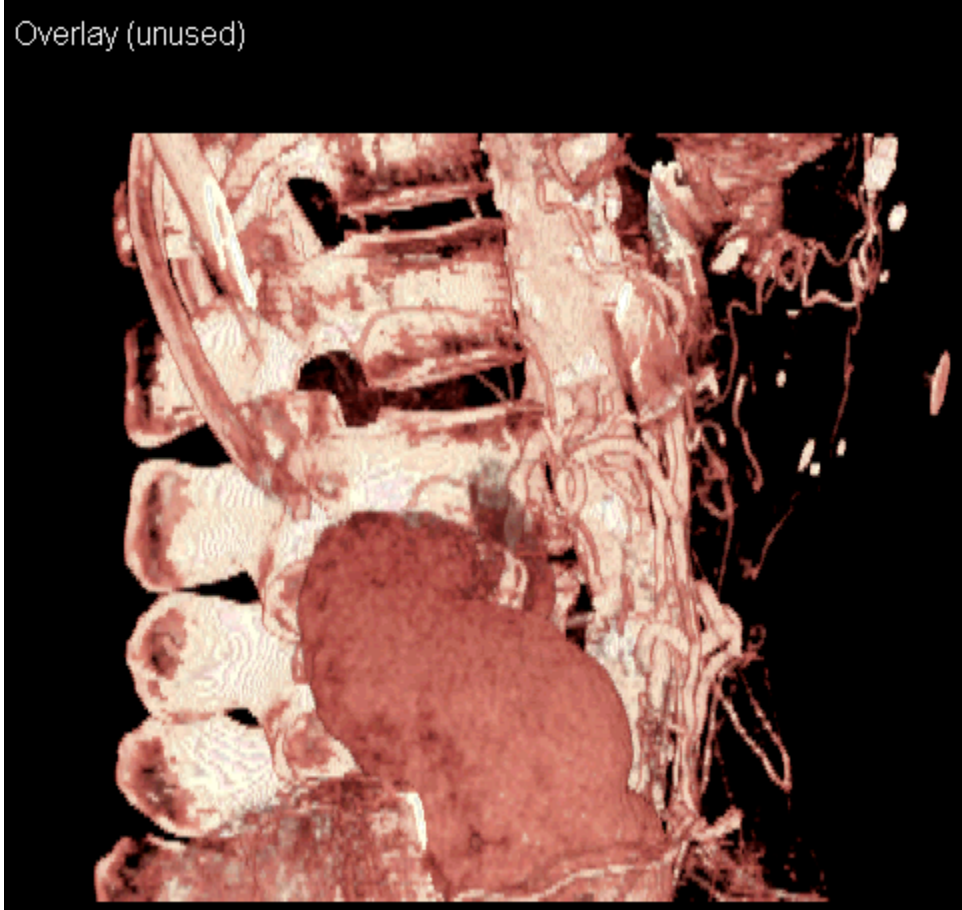
Requirements

- Multi-site videoconferencing
- Access to pathology & radiology images
  - Live microscopy
  - DICOM
- Access to remotely stored patient records through organisational LANs





3D Image Visualization

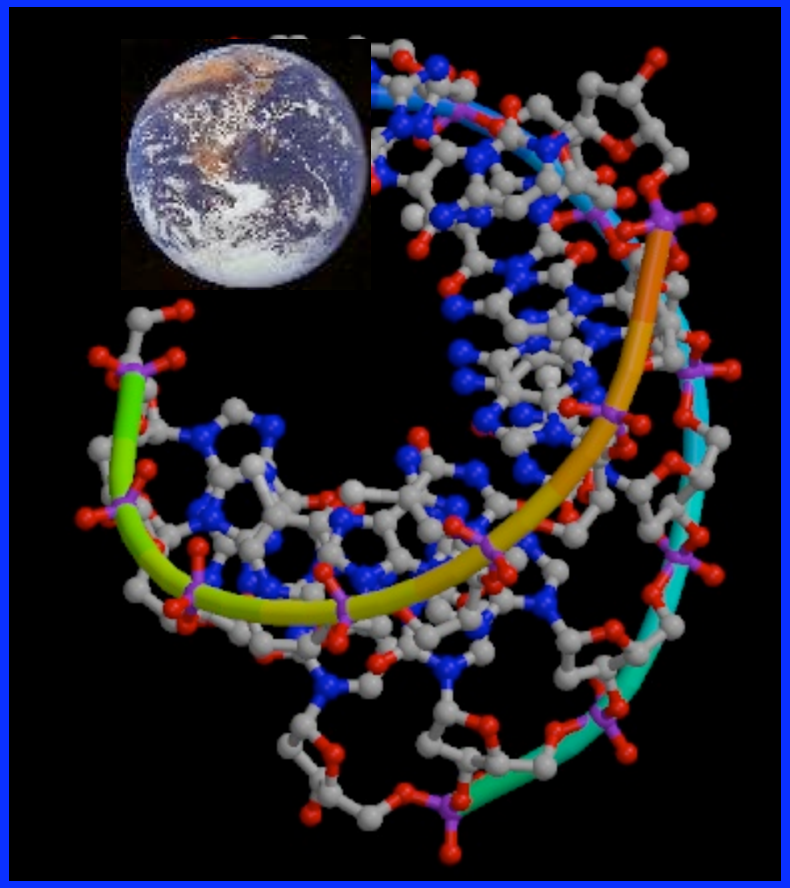


- 3D Volume rendered images
- Access to mass imaging data
- Visualization of complex medical imaging

Progress and future

Collaboration with DEST project - Prof Burrage, Queensland

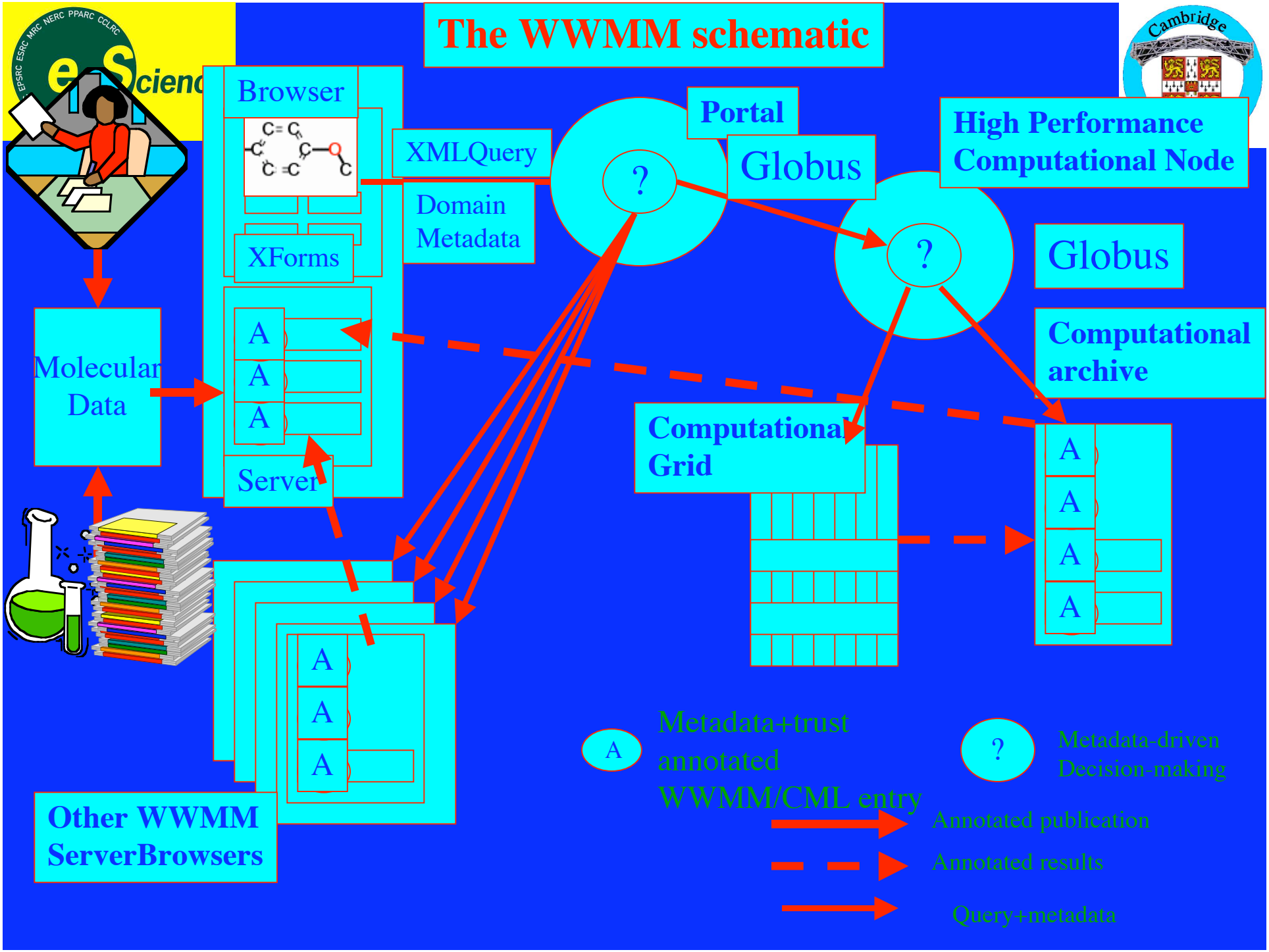
- Telemedicine has been adopted by most cancer MDTs in Cambridge, and is also used for training, management and general communications by the participating trusts.
- Telemedicine has been rolled out to 5 other Cancer Networks, and a National Programme is under discussion.
- Telemedicine is proposed for use in CancerGrid, for running clinical trials, as part as a broader data management project.



The World Wide Molecular Matrix WWMM

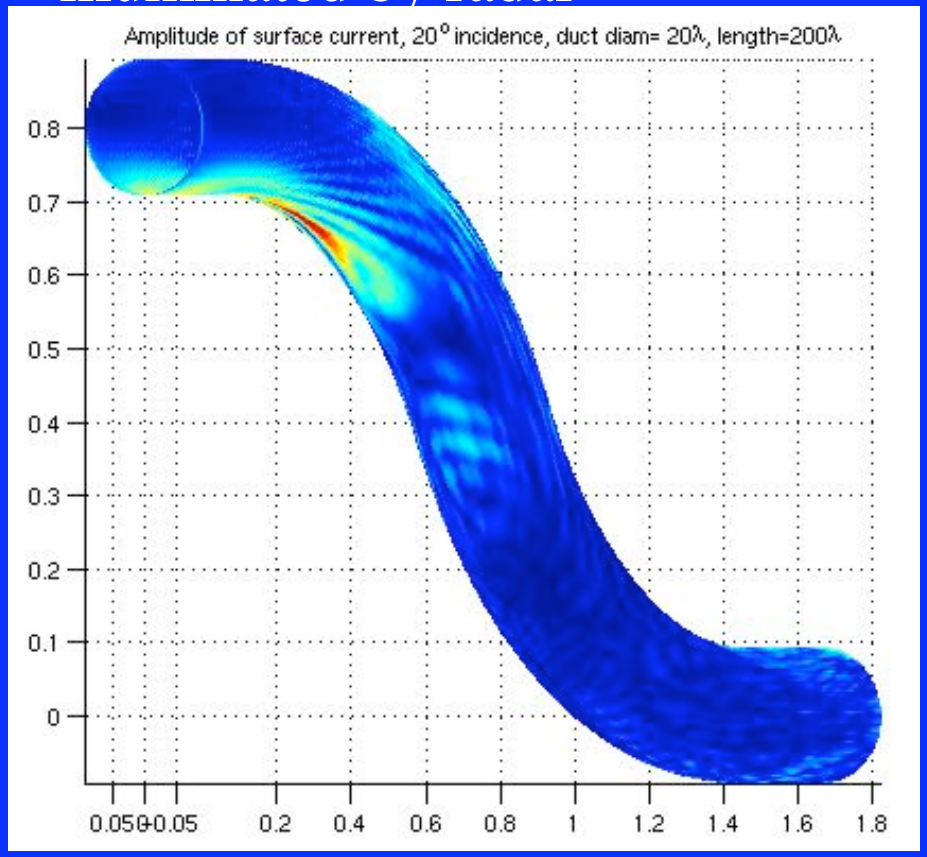
Unilever Centre for Molecular Informatics Cambridge

Grid Technology in Molecular Sciences


The WWMM schematic
Cambridge
Browser
XMLQuery
Portal
Globus
High Performance Computational Node
Domain Metadata
XForms
?
?
Globus
Molecular Data
A
A
A
Server
Computational archive
Computational Grid
A
A
A
A
A
A
A
Other WWMM ServerBrowsers
A
Metadata+trust annotated WWMM/CML entry
?
Metadata-driven Decision-making
Annotated publication
Annotated results
Query+metadata

EM Scattering project

Surface current in a tube illuminated by radar



- Collaborative project with BAE Systems to investigate radar reflection from aircraft
- BAE design aircraft shapes
- Cambridge mathematicians calculate EM scattering from rough surfaces for complex shapes on HPCF

EM Scattering project

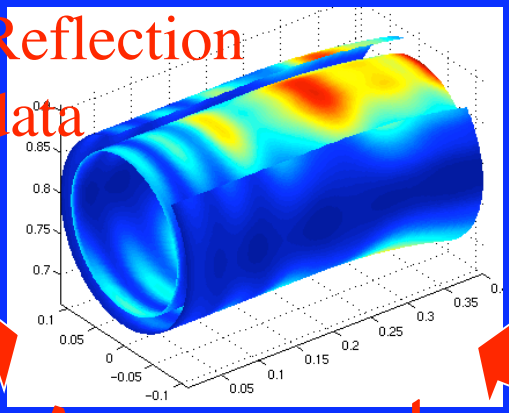
- Link engineering simulations at BAE with EM scattering calculations in Cambridge with Grid based feedback loop

Distributed simulation

Visualisation

Security

Reflection data

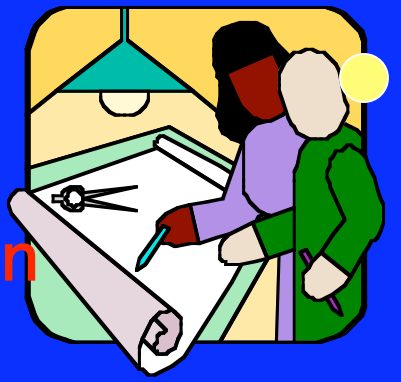


Globus

AccessGrid

CAD

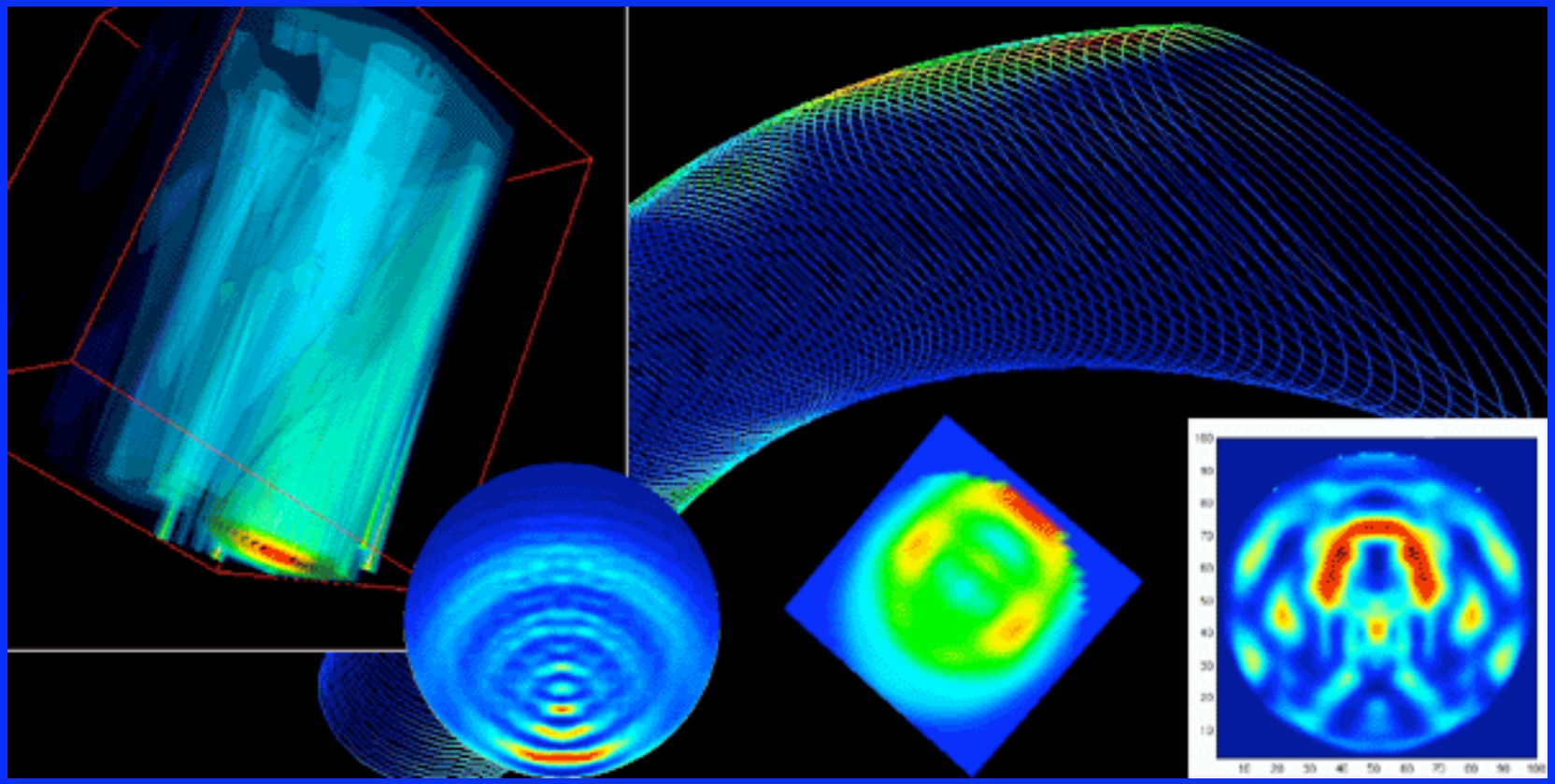
Design



EM Grid visualization

Use portal to execute scattering code or launch the visualisation software.

View isosurfaces, ie surfaces of equal intensity in 3D

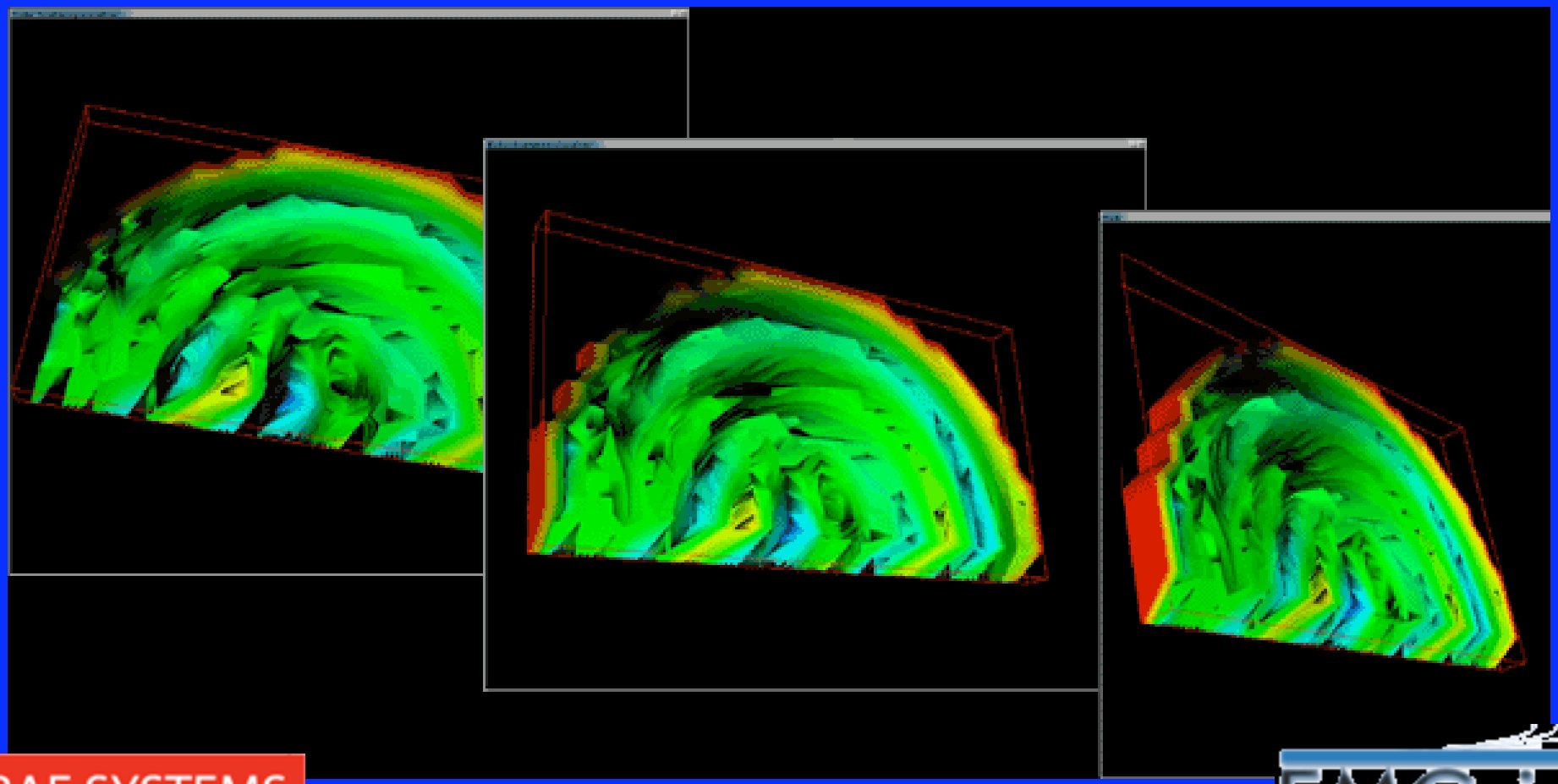


BAE SYSTEMS

EMGrid

EM Grid visualization

Alternatively view colour contour plots
Reveal high-intensity areas by steering a cutting plane interactively along the structure, in a virtual 'fly-through'

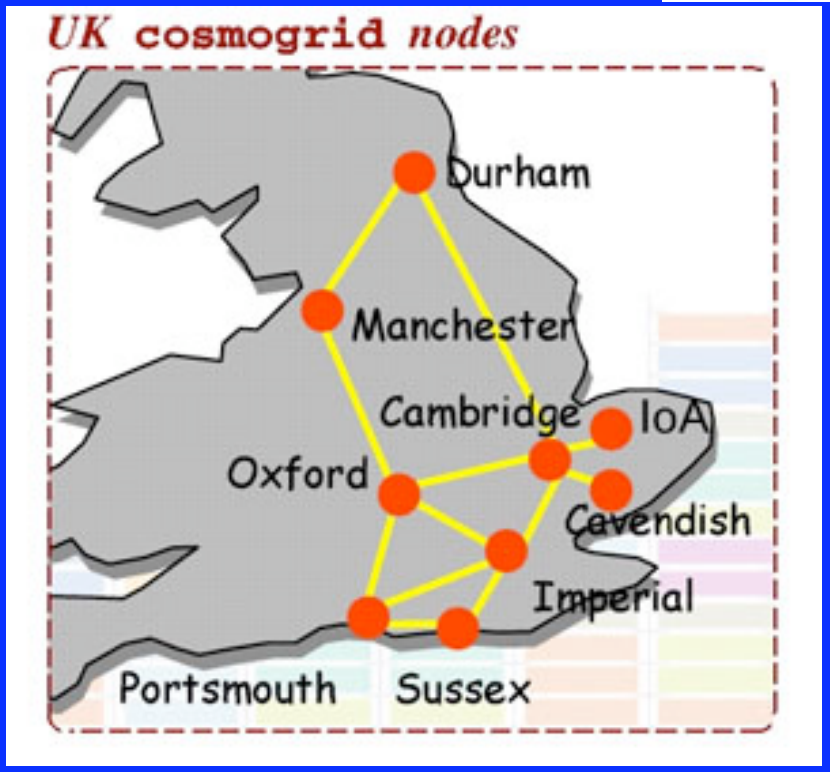


BAE SYSTEMS

EMGrid

CosmoGrid

The COSMOS consortium, led by Prof Stephen Hawking, employs large-scale supercomputer resources to advance our knowledge of the origin and structure of our universe



COSMOS, the National Cosmology Supercomputer, is an SGI Altix 3800 (128 IA64 cpus, 128Gb memory, 10Tb storage) housed in Cambridge.

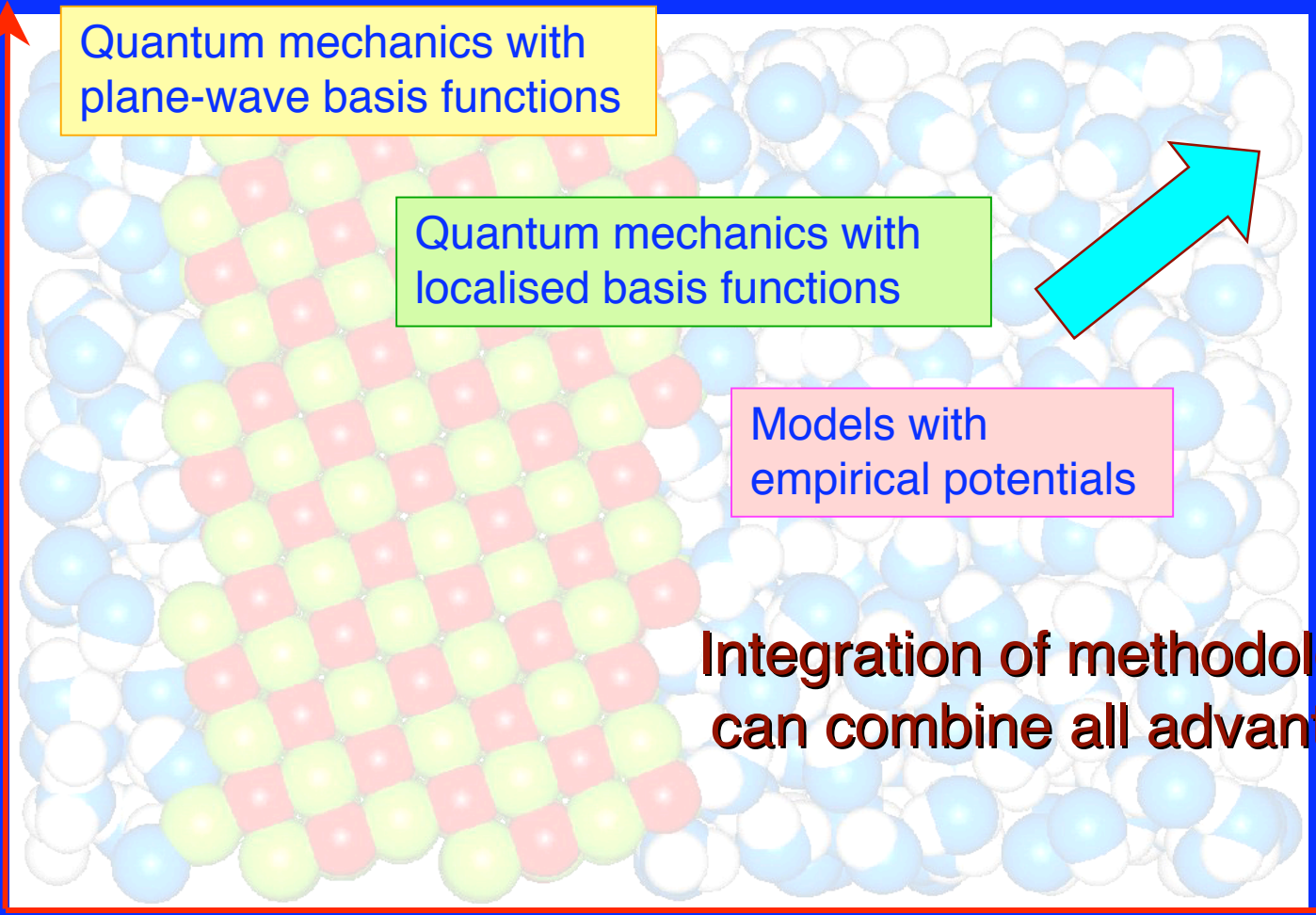
Cosmos consortium

- Formation of a galaxy cluster

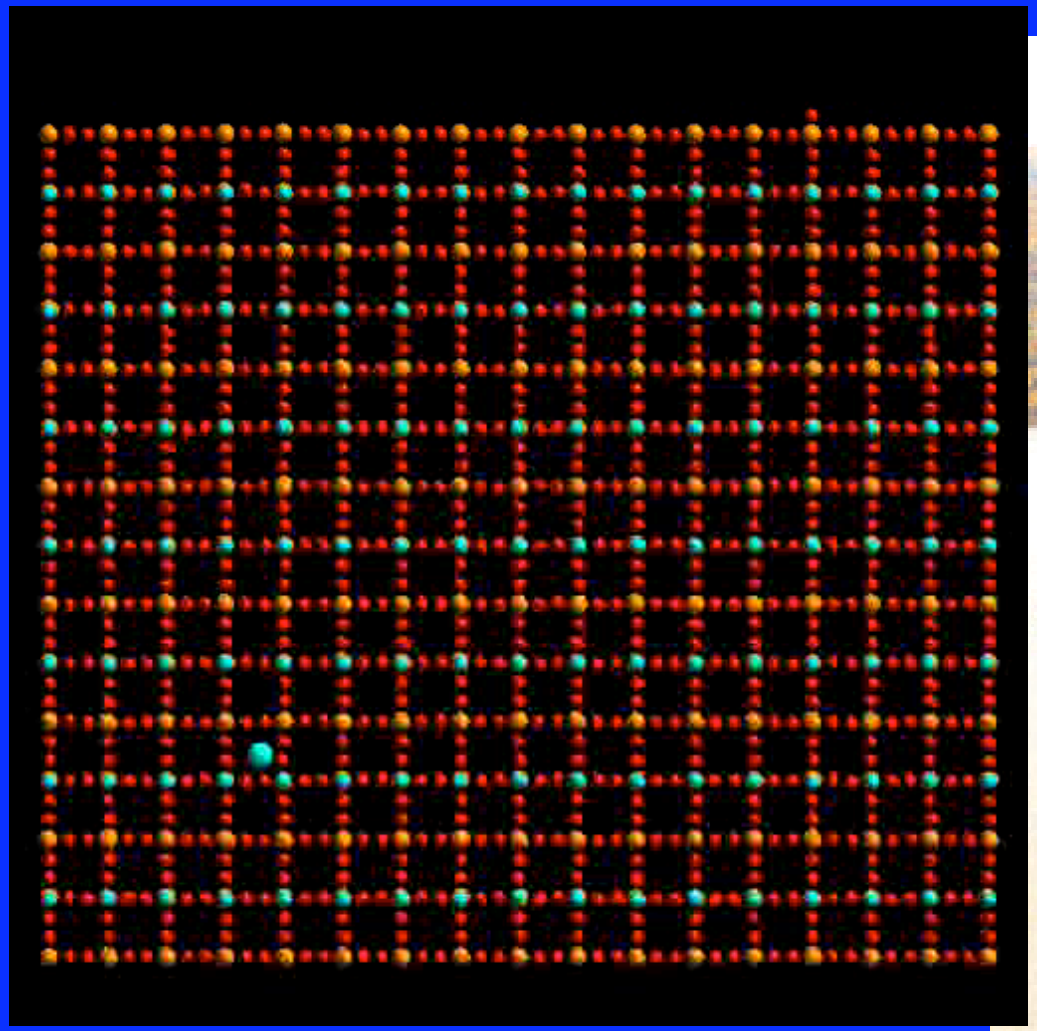
Environment from the molecular level:

An e-science proposal for modelling the atomistic processes involved in environmental issues

Molecular models



Example of radioactive waste containment

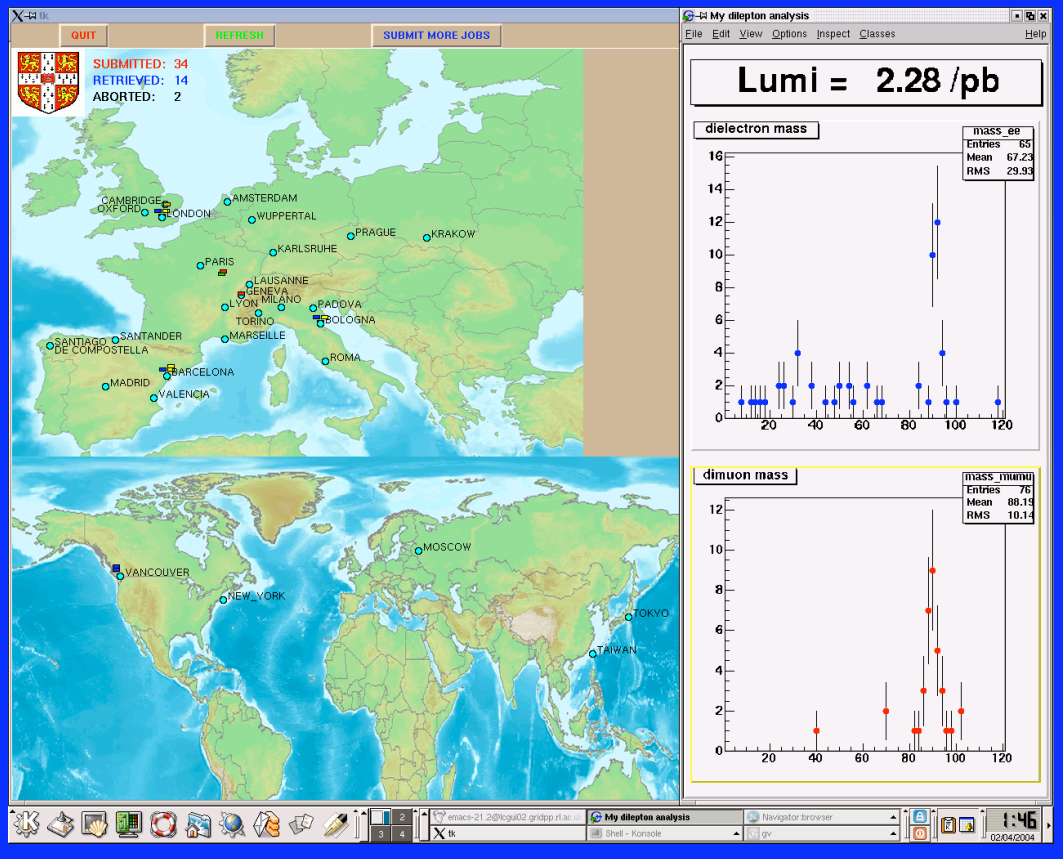


- Issues:
- Scale up in space and time
- Access to simulated data
- Visualisation of results
- Commercial security

LHC Computing Grid

- CeSC supports the GridPP project to handle Petabytes of data per year from the Large Hadron Collider.

Cambridge is a Core node on the new LHC Computing Grid.

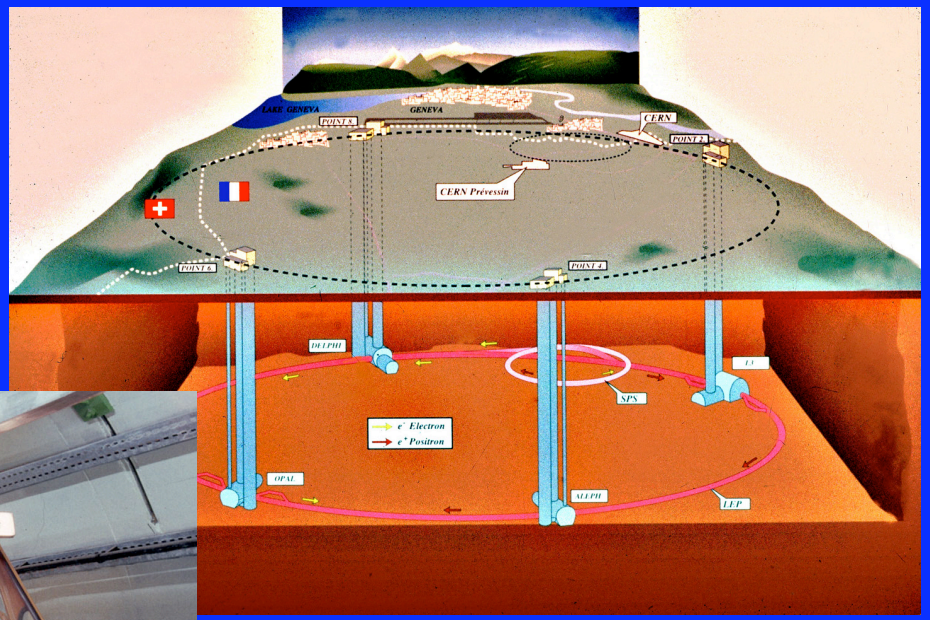


Interactive analysis in Cambridge of ATLAS data worldwide.

GridPP

1Pb=1000Gb=1km stack of DVDs

- PPARC funded project (£17M) to enable data analysis for the Large Hadron Collider experiments.
- Links with EU DataGrid and Grid projects in the USA

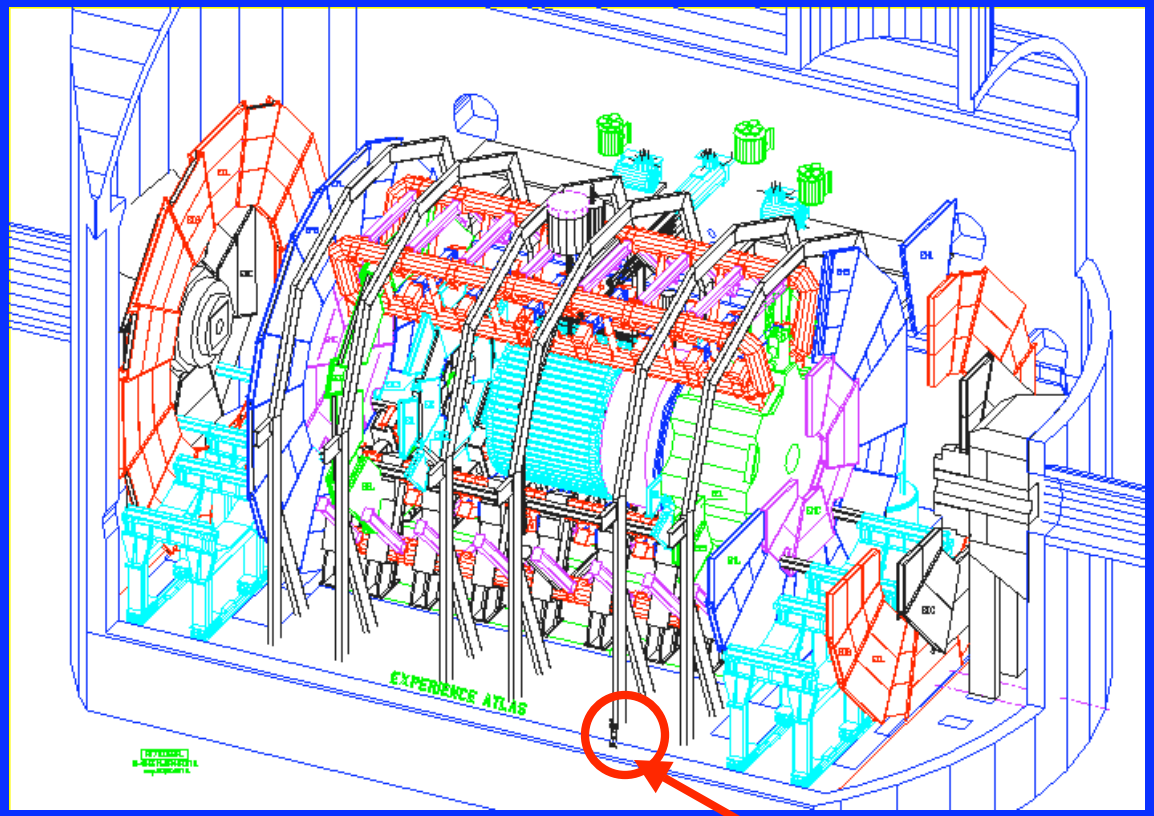


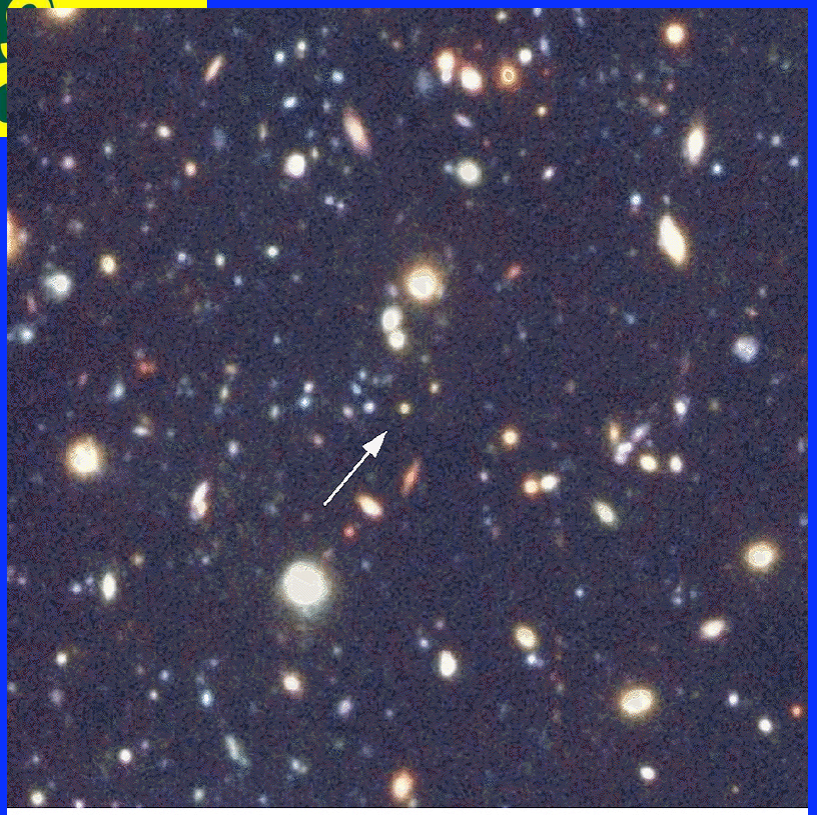
LHC under construction at CERN: will generate a few Petabytes of data every year from 2007

1 TeV proton-proton collider

The Atlas Experiment

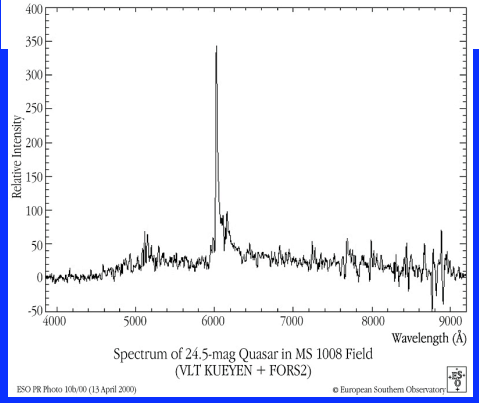
- 150 participating institutes worldwide
- 1700 scientists and engineers involved
- Observe 40 million collisions per second
- 1000 tracks per collision
- >1 Petabyte of data/year





Quasar (24.5mag ; z~4) in MS 1008 Field (VLT ANTU + FORS1)

© European Southern Observatory



Spectrum of 24.5-mag Quasar in MS 1008 Field
(VLT KUEYEN + FORS2)

ESO PR Photo 10b/00 (13 April 2000) © European Southern Observatory

Astro Grid

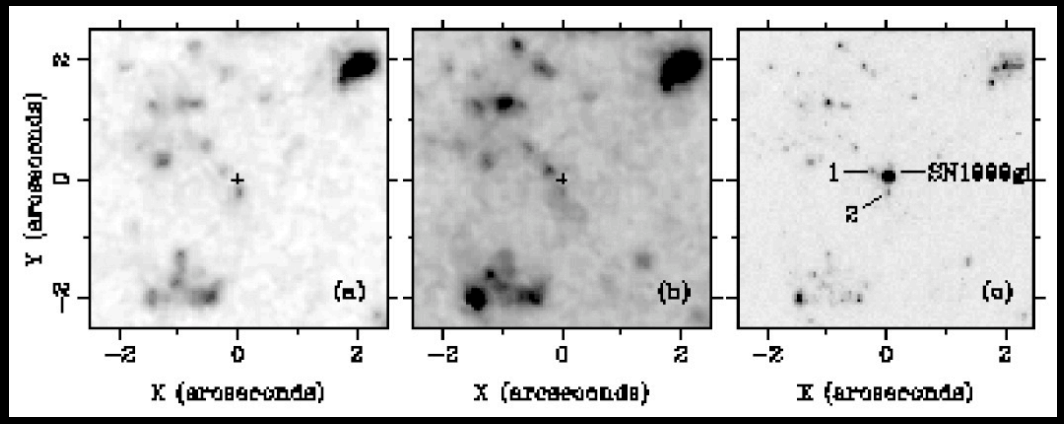
(http://www.astrogrid.ac.uk)

Grid for Astrophysics:
Federated databases
Real-time telescope operations
Virtual Observatory

Astronomical Drivers: Pre-Discovery Mining

- Investigating the progenitors of sources that show variability
  - Dark matter revealed by microlensing events
  - Planets revealed by stellar variability
  - Formation of neutron stars revealed by GRB's
  - Death of massive stars revealed by Type II SN

The progenitor of SN1999gi is <9 M$_\odot$: found from mining pre-discovery HST images. (Smartt et al, 2001, ApJ, 556, L29)



Conclusions

- The Grid started in academic supercomputing
- The key features:
  - Instant access to worldwide collections of reliable data
  - Effective use of large distributed computing systems
  - Collaborative environments

- Grids are now rolling out in industry and the public sector - anywhere with distributed teams needing to share large amounts of data of any sort.