

Grids in the UK

Peter Watkins

Head of Particle Physics

University of Birmingham, UK

p.m.watkins@bham.ac.uk

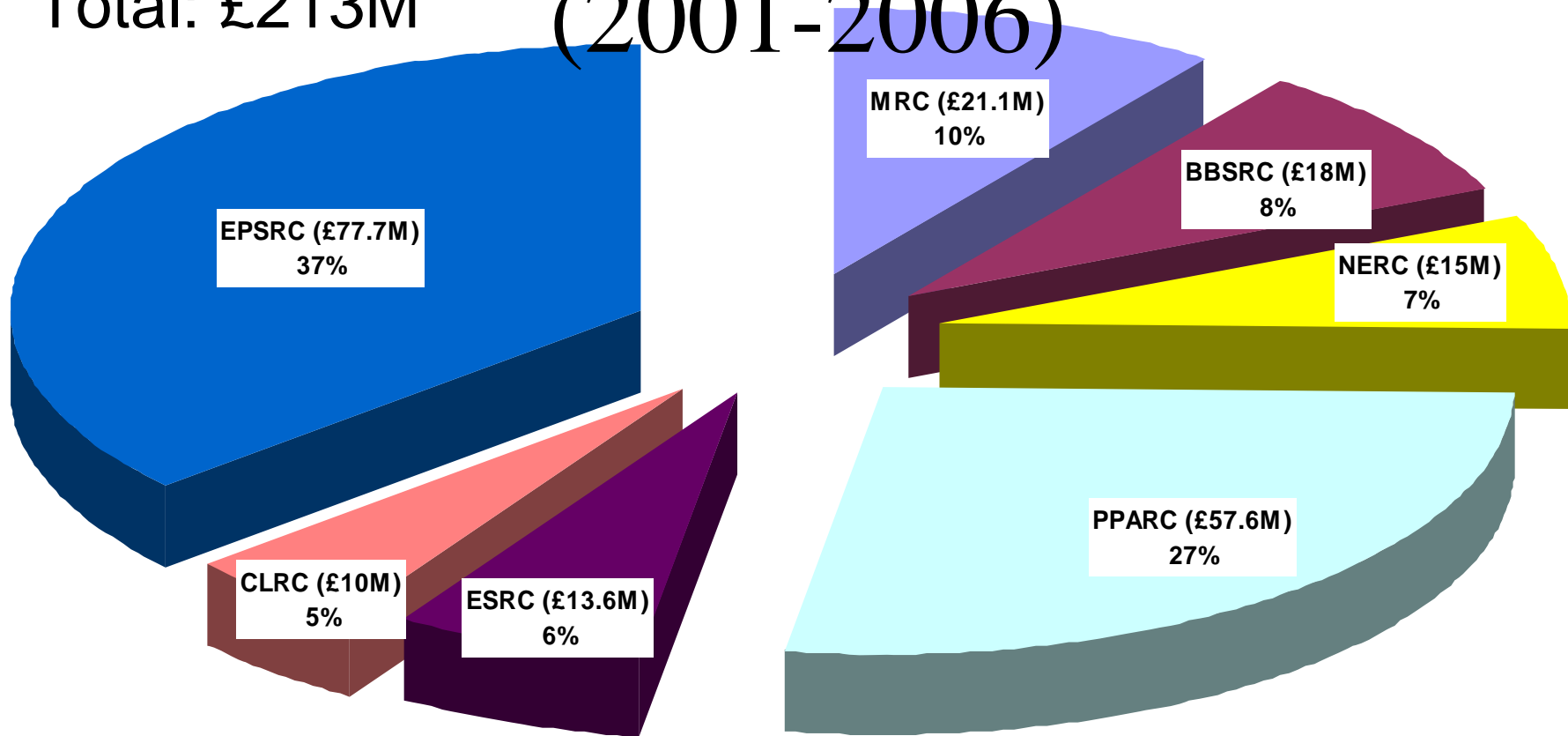
Outline

- Brief Overview
- GridPP
- National Grid Service (NGS)
- Summary

Acknowledgements - Colleagues in GRIDPP and NGS including Stephen Burke, Yves Coppens, Tony Doyle, Neil Geddes and Steve Lloyd

UK e-Science Budget

Total: £213M (2001-2006)

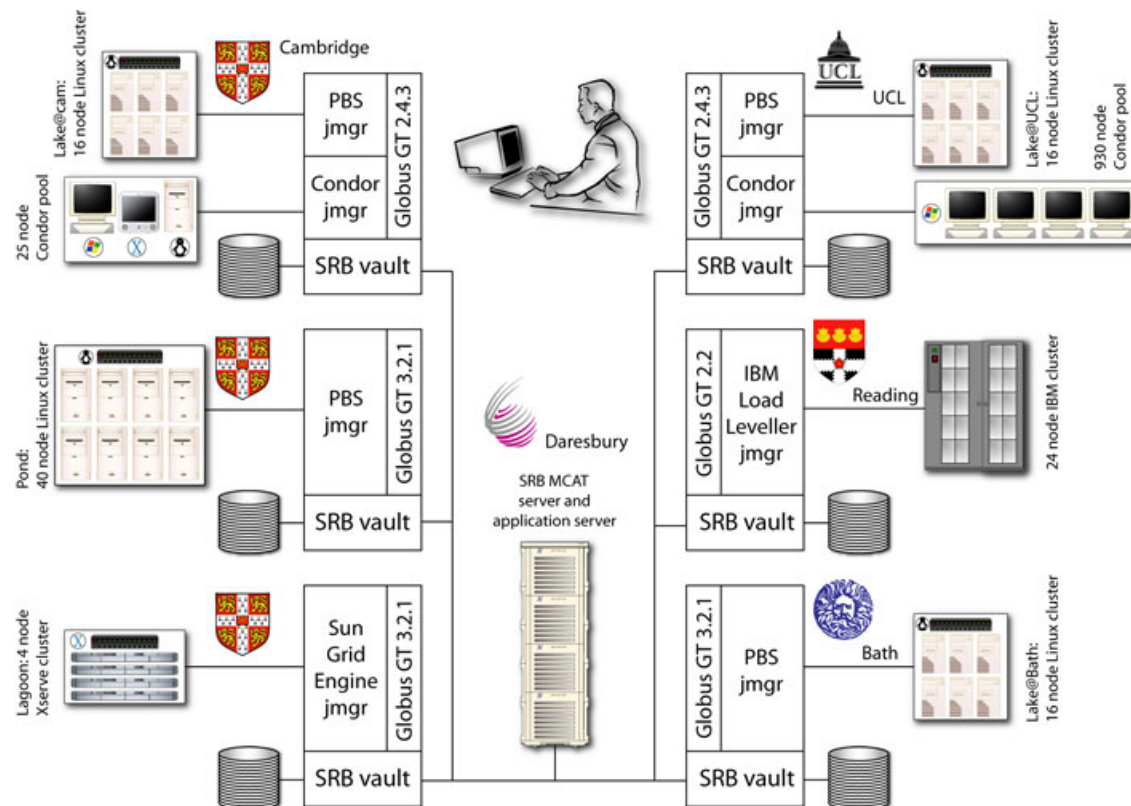
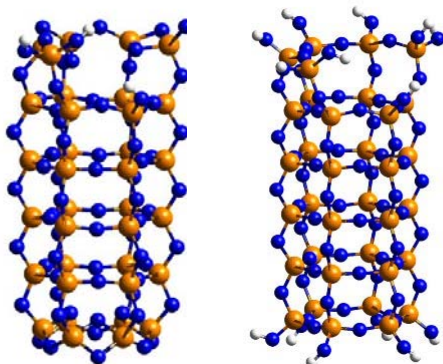
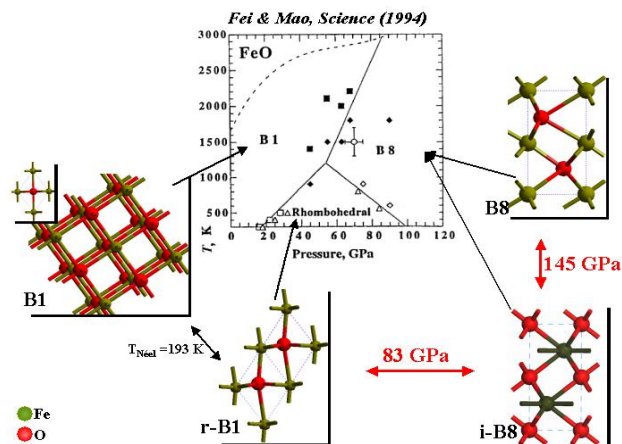


Source: Science Budget 2003/4 – 2005/6, DTI(OST)

e-Minerals

simulations performed on molecular length
and time scales to address environmental issues:

- adsorption of pollutants
- crystal growth and dissolution
- radiation damage
- diffusion of pollutant atoms



CCLRC e-Science

Who are GridPP?

19 UK Universities, CERN and
CCLRC (RAL & Daresbury)

Funded by PPARC:

GridPP1 2001-2004 (£17m)

“From Web to Grid”

GridPP2 2004-2007 (£16m)

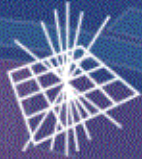
“From Prototype to Production”

GridPP3 2007-2011 (proposed)

“From Production to Exploitation”

Developed a working, highly
functional Grid

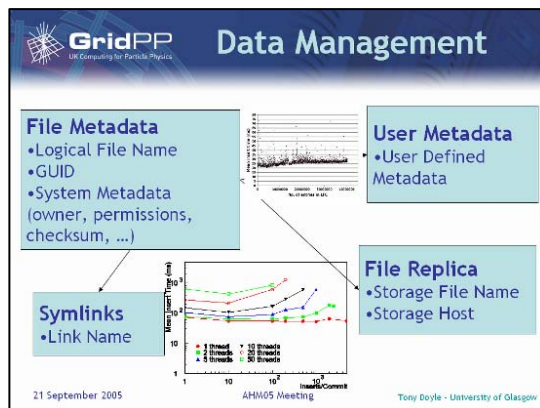




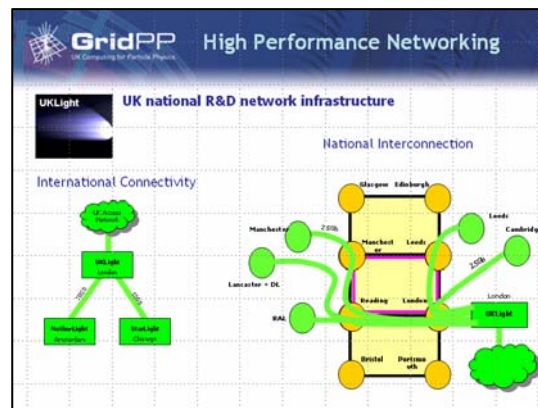
- Grid for UK Particle Physics
- Two phases, 2001-2004-2007
 - Proposal for phase 3 to 2011
- 20 sites, 4354 CPUs, 298 Tb of storage
- Part of EGEE and LCG
 - Working towards interoperability with NGS
- Currently supports 33 VOs, including some non-PP
- For LCG, sites are grouped into four “virtual” Tier 2s
 - Plus RAL as Tier 1
 - Grouping is largely administrative, the Grid sites remain separate
- Runs UK-Ireland ROC (with NGS)
- Grid Operations Centre (GOC) @ RAL (with NGS)
 - Configuration, monitoring and accounting
- Operations and User Support shifts (working hours only)

GridPP Middleware Development

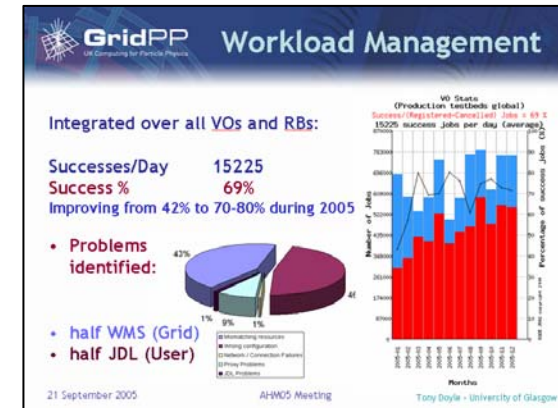
Grid Data Management



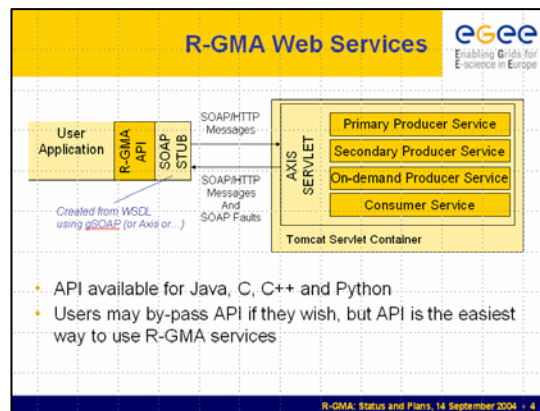
Network Monitoring



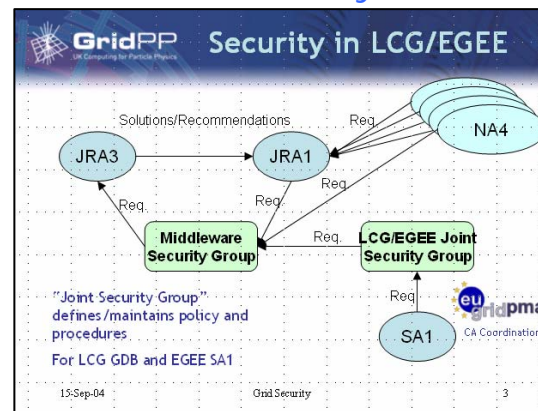
Workload Management



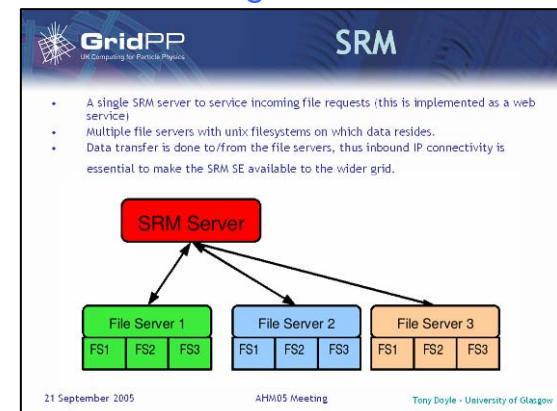
Information Services



Security



Storage Interfaces



International Context

GridPP is part of EGEE and LCG (currently the largest Grid in the world)

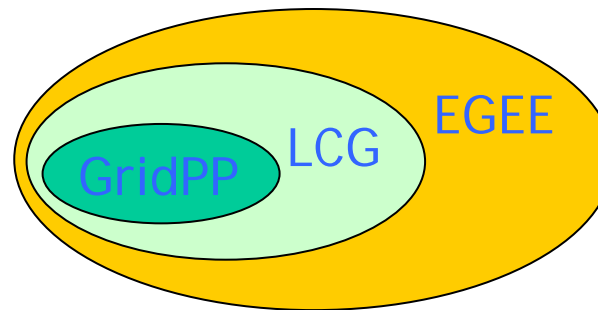
EU Enabling Grids for e-Science (EGEE) 2004-2008

Grid Deployment Project for all disciplines



UK National Grid Service

UK's core production computational and data Grid



LHC Computing Grid (LCG)
Grid Deployment Project for LHC



NorduGrid (Scandinavia)

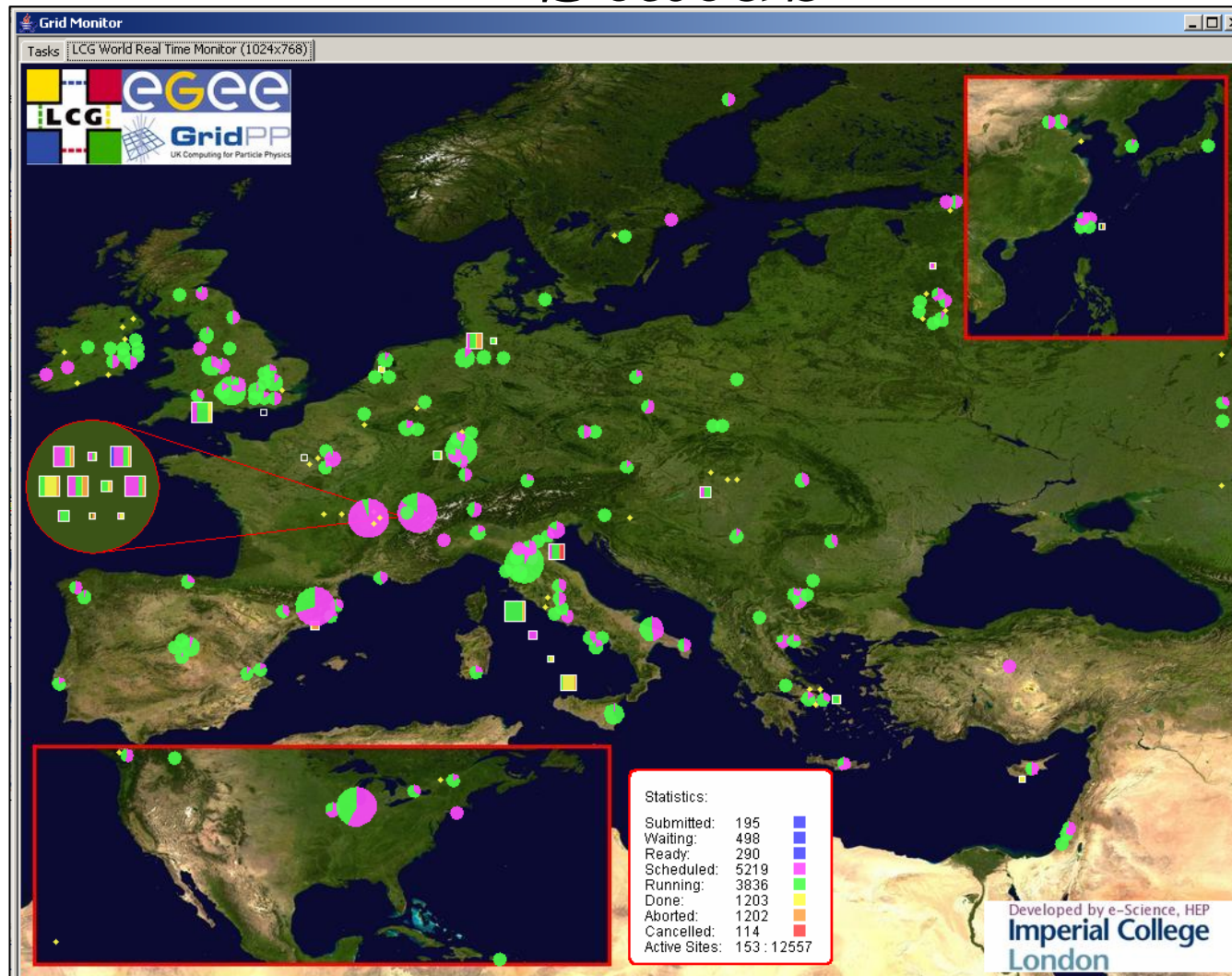
Grid Research and Development collaboration



Open Science Grid (USA)

Science applications from HEP to biochemistry

The LCG Grid Status



Worldwide

182 Sites

23,438 CPUs

9.2 PB Disk

2,200 Years of
CPU time

UK

21 Sites

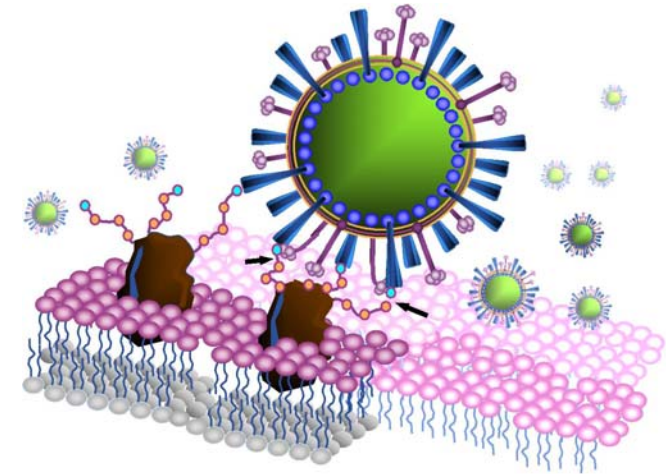
4,482 CPUs

180 TB Disk

593 Years of
CPU time

What GridPP Has Done So Far

- Analysed 300,000 possible drug components in the fight against the Avian Flu virus
- Simulated 46 million molecules for medical research in 5 weeks, which would have taken over 80 years on a single PC



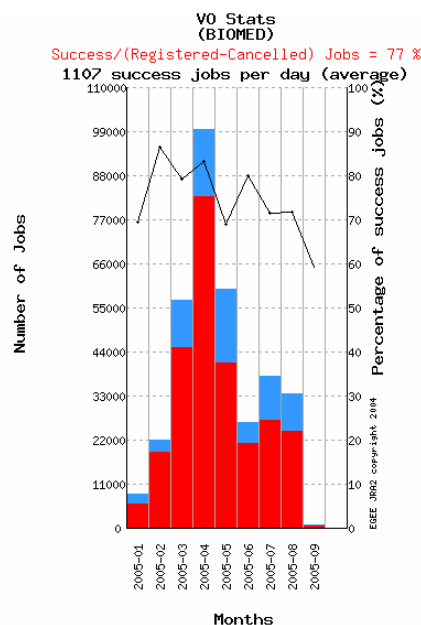
- Reached transfer speeds of 1 Gigabyte per second in high speed networking tests from CERN - a DVD every 5 seconds
- Simulated 500 million particle physics collisions with the BaBar experiment
- *Transformed the way particle physics computing problems are approached*

“UK contributes to EGEE's battle with malaria”

BioMed

Successes/Day **1107**

Success % **77%**

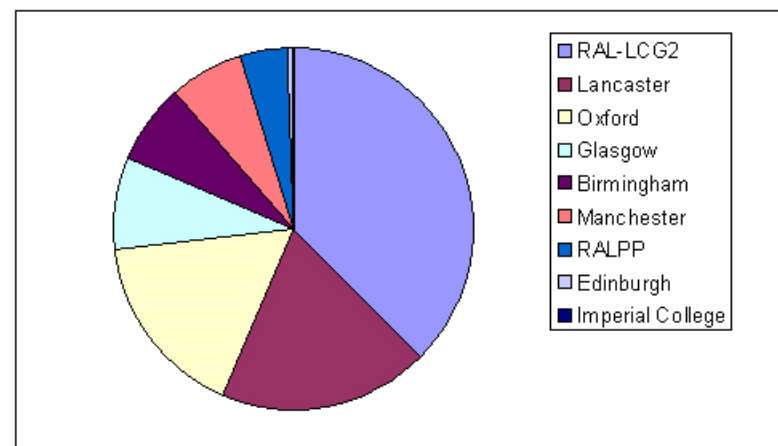
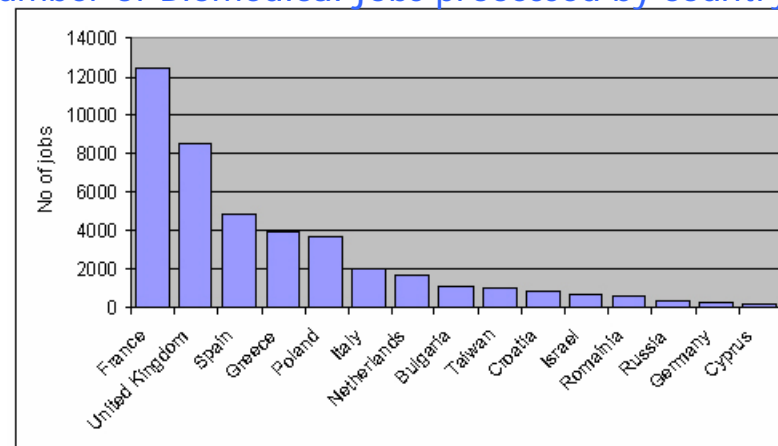


WISDOM (Wide In Silico Docking On Malaria)

The first biomedical data challenge for drug discovery, which ran on the EGEE grid production service from 11 July 2005 until 19 August 2005.

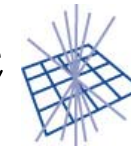
GridPP resources in the UK contributed ~100,000 kSI2k-hours from 9 sites

Number of Biomedical jobs processed by country



Normalised CPU hours contributed to the biomedical VO for UK sites, July-August 2005

UK Tier-1/A Centre Rutherford Appleton Lab

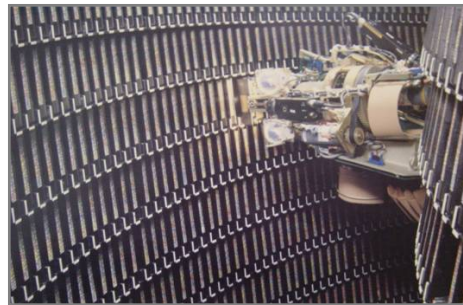


GridPP
UK Computing for Particle Physics

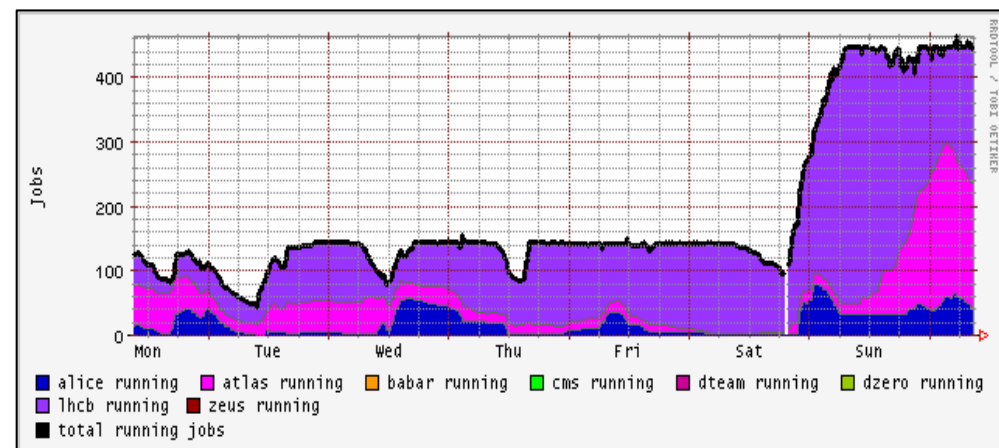
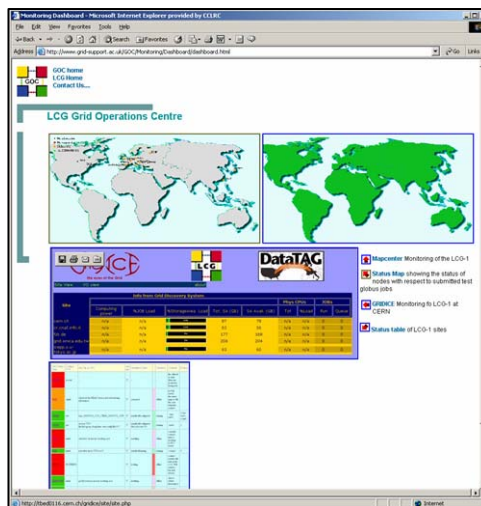
- High quality data services
- National and International Role
- UK focus for International Grid development

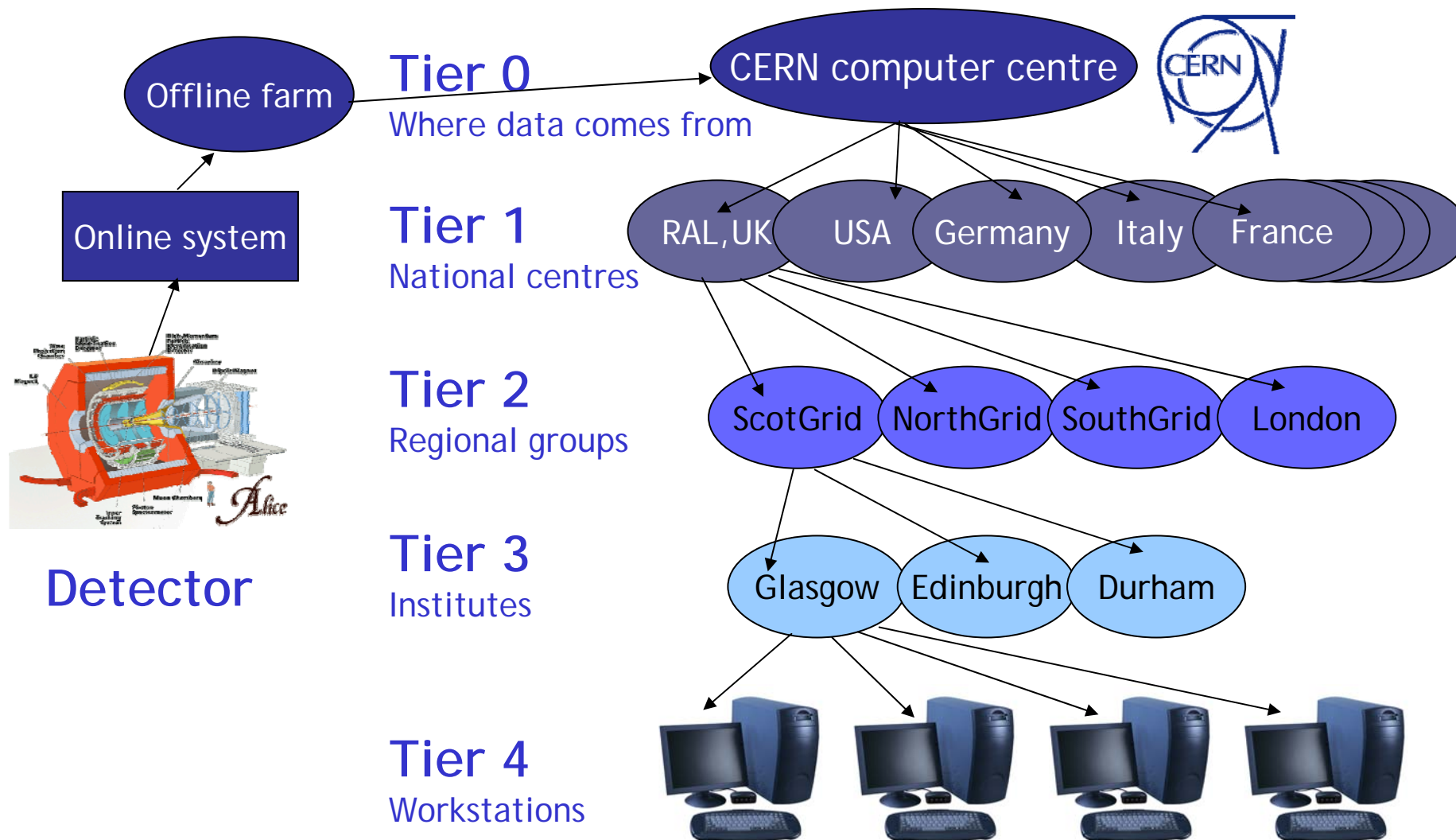


- 700 Dual CPU
- 80 TB Disk
- 60 TB Tape
(Capacity 1PB)



Grid Operations Centre





UK Tier-2 Centres



ScotGrid

Durham, Edinburgh, Glasgow

NorthGrid

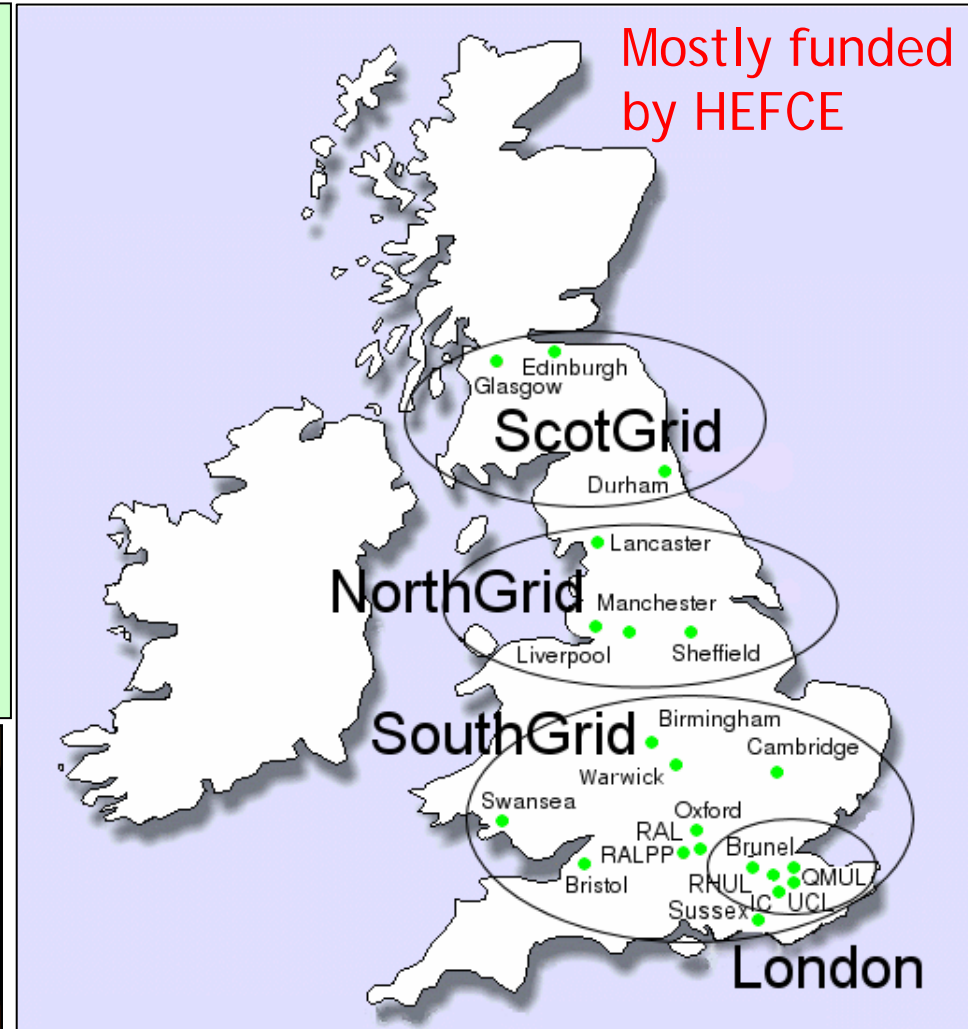
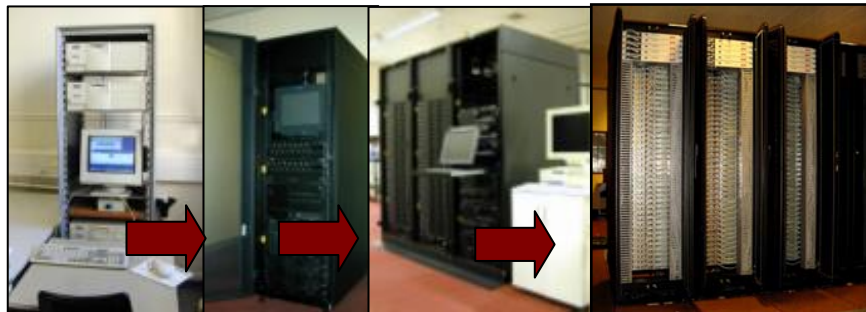
Daresbury, Lancaster, Liverpool, Manchester, Sheffield

SouthGrid

Birmingham, Bristol, Cambridge, Oxford, RAL PPD, Warwick

LondonGrid

Brunel, Imperial, QMUL, RHUL, UCL

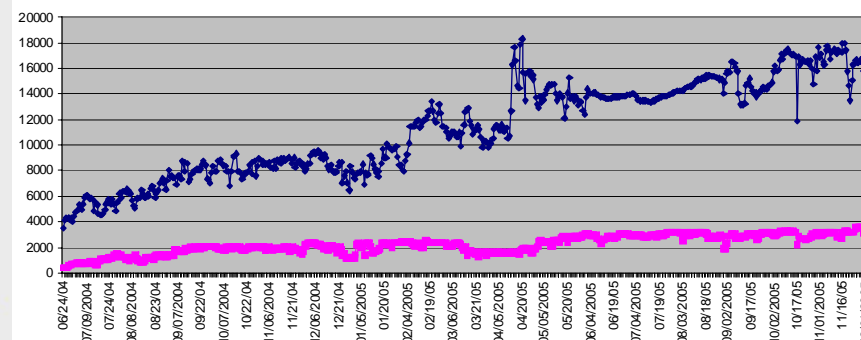
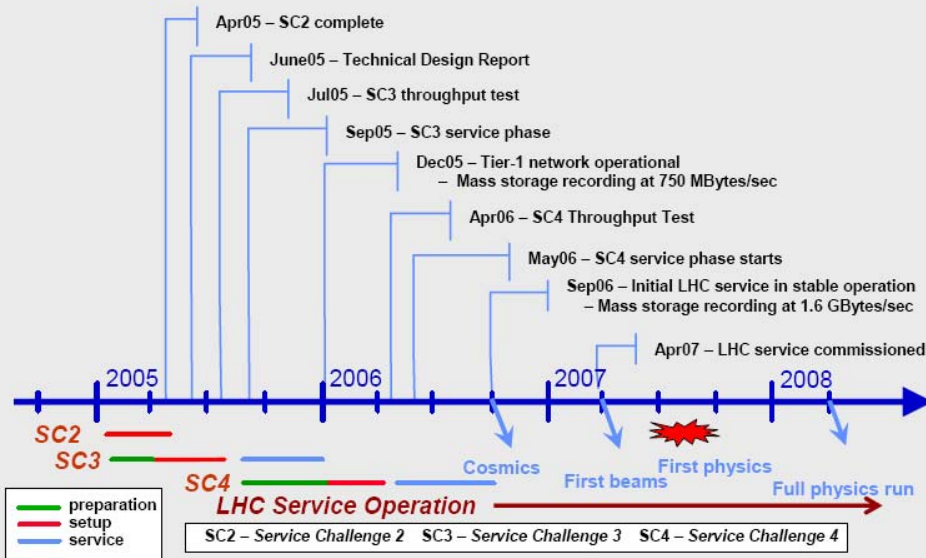
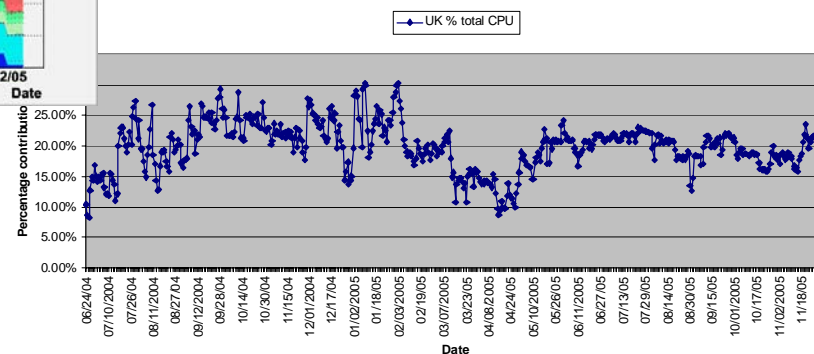
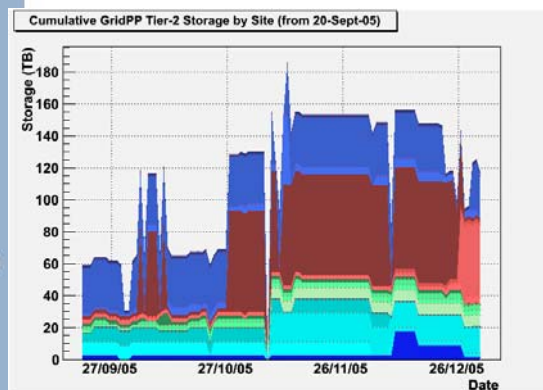


GridPP/LCG/EGEE Continues

High performance services

T0 to T1 streaming at 2Gb

T1 to T2 streaming at 200Mb



Neil Geddes
CCLRC e-Science

Lessons learnt

- “Good enough” is not good enough
 - Grids are good at magnifying problems, so must try to fix everything
- Exceptions are the norm
 - 15,000 nodes * MTBF of 5 years = 8 failures a day
 - Also 15,000 ways to be misconfigured!
 - Something somewhere will **always** be broken
 - But middleware developers tend to assume that everything will work
 - It needs a lot of manpower to keep a big system going
- Bad error reporting can cost a lot of time
 - And reduce people's confidence
- Very few people understand how the whole system works
 - Or even a large subset of it
 - Easy to do things which look reasonable but have a bad side-effect
- Communication between sites and users is an $n*m$ problem
 - Need to collapse to $n+m$

- LHC turns on in 1 year - we must focus on delivering a high QOS
- Grid middleware is still immature, developing rapidly and in many cases a fair way from production quality
- Experience is that new middleware developments take ~ 2 years to reach the production system, so LHC will start with what we have now
- The underlying failure rate is high - this will always be true with so many components, so middleware and operational procedures must allow for it
- We need procedures which can manage the underlying problems, and present users with a system which appears to work smoothly at all times
 - Considerable progress has been made, but there is more to do
- GridPP is running a major part of the EGEE/LCG Grid, which is now a very large system operated as a high-quality service, 24*7*365
- We are living in interesting times!

UK Core e-Science programme -



Goals for NGS

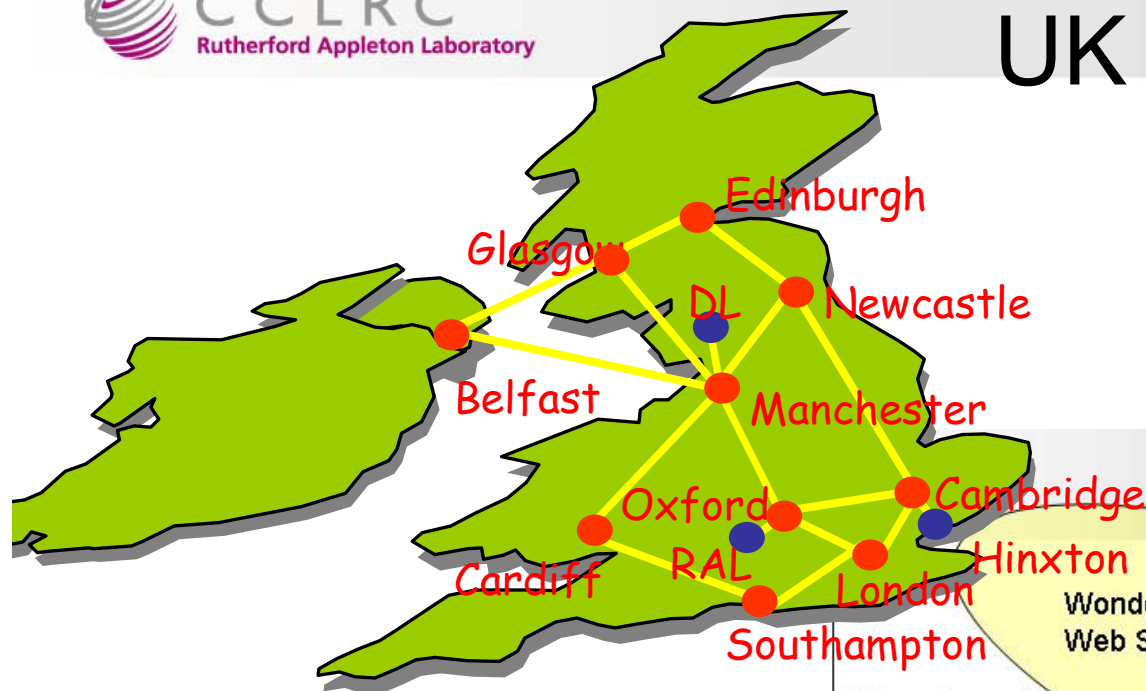
- Provide coherent electronic access for UK researchers to all resources and facilities that they require to carry out their research.
 - independent of location of resource or researcher.
 - Common/standard Interfaces
 - Common/standard procedures
- Services support the interface and applications use it
- Goal can not be achieved by NGS alone
 - NGS will collaborate with
 - Resources and Resource providers
 - Users
 - Researcher identity and authorisation will be established through standard institutional mechanisms,
 - Close integration with international partner infrastructures

NGS Overview: User view

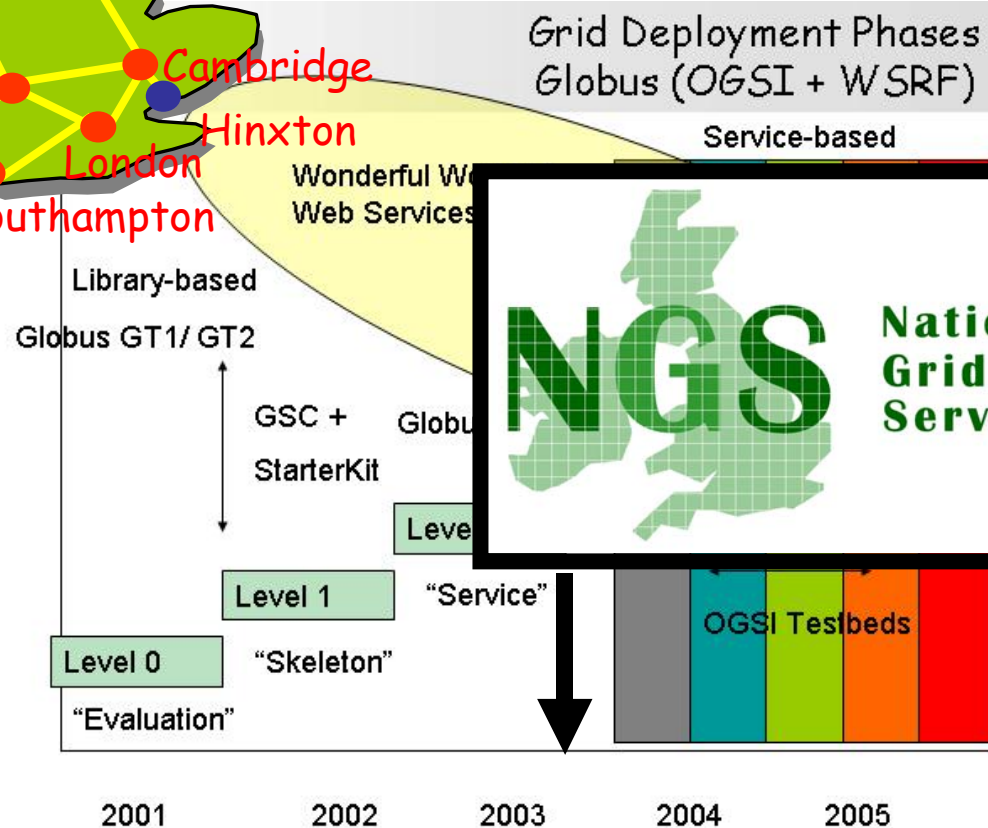


- Resources
 - 4 Core clusters
 - UK's National HPC services
 - A range of partner contributions
 - clusters, shared mem. portals, data ...
 - Partners and affiliates
- Access
 - Support UK academic researchers
 - All partners support common user base + whoever they want
 - Free at the point of use
 - Light weight peer review for limited “free” resources
 - Partners can provide larger commitments as required
- Central help desk
 - www.grid-support.ac.uk

UK e-Science Grid



**Application
independent**



NGS partnership programme

Goals:

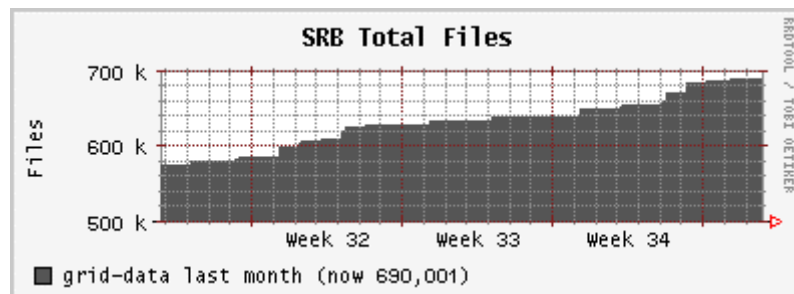
- Increase the range and depth of services and resources that NGS can offer to its users
 - by enlisting partners
- Provide leadership and guidance to sites wanting to put their resources “on the Grid”
 - by defining NGS compatibility and certifying compliant sites
- Two levels of partnership
 - An **NGS Affiliate** is a site certified to be NGS-compatible.
 - An **NGS Partner** is a site that offers significant resources or services to NGS users.



NGS & Partners Today

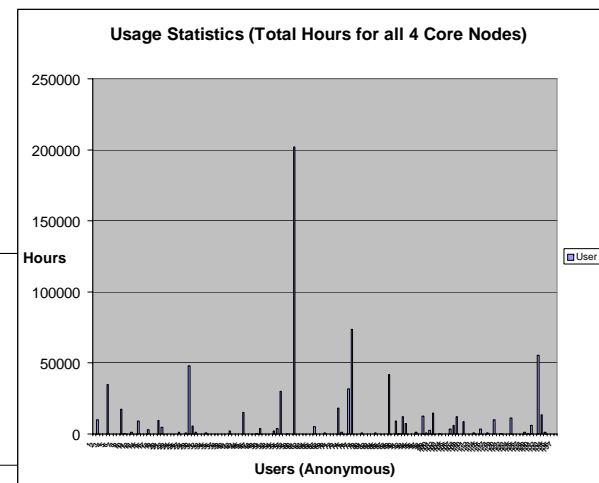
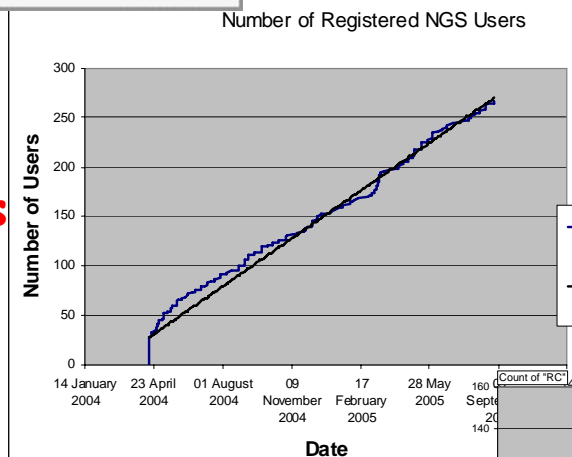


NGS Use

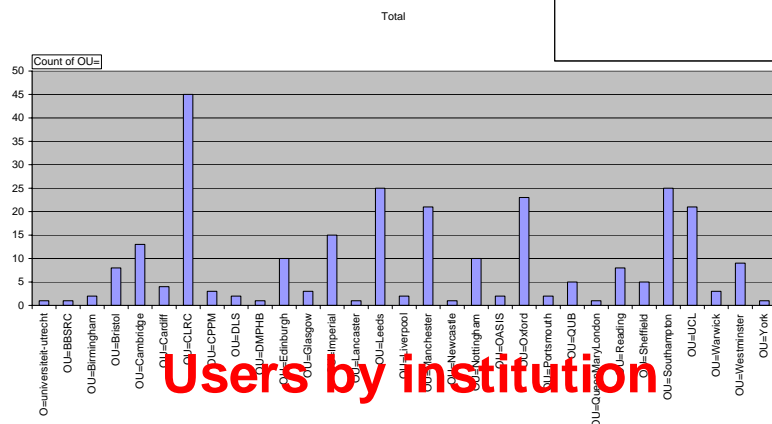


Files stored

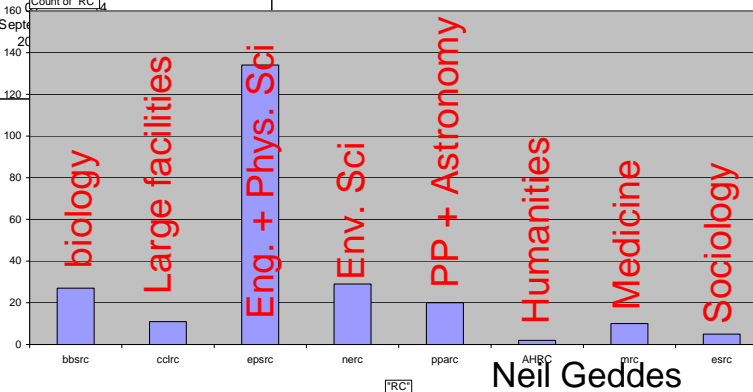
Over 320 users



CPU time by user



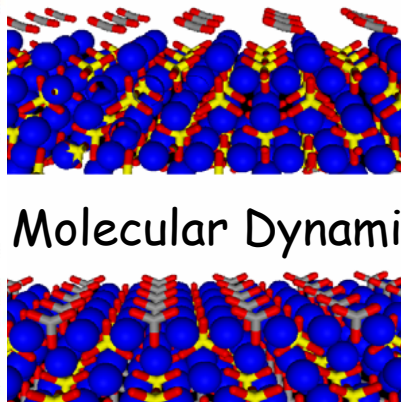
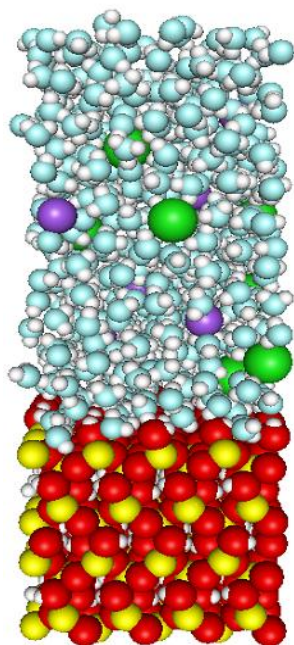
Users by institution



Users by discipline

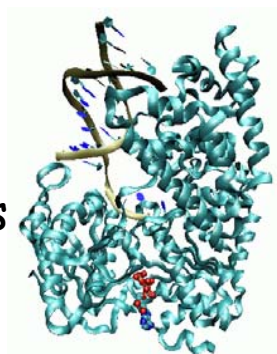
Neil Geddes
CUK Science

Applications: 1

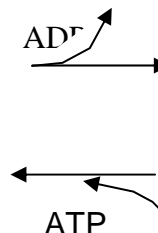
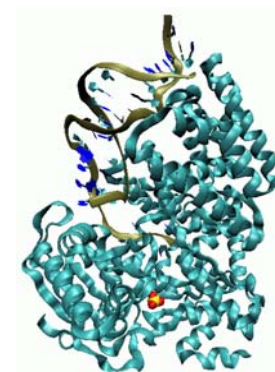


Molecular Dynamics

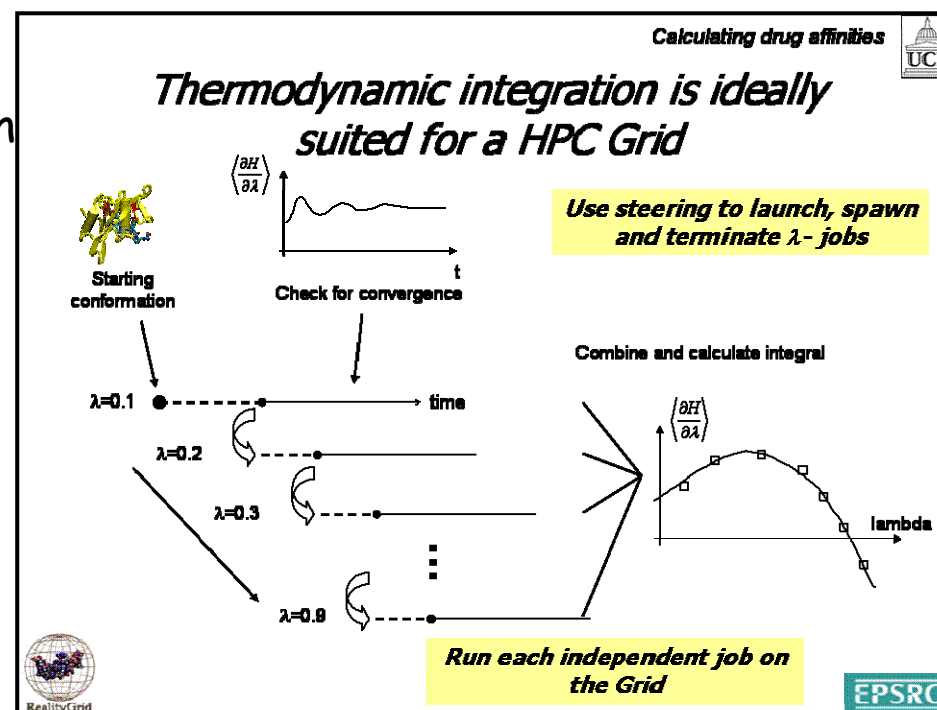
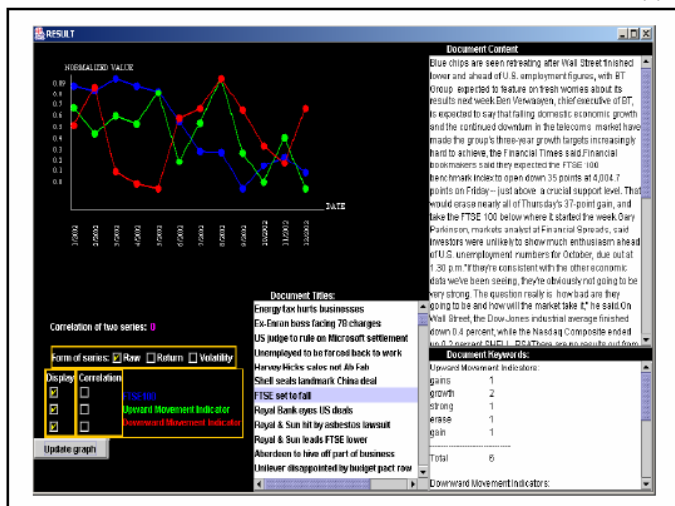
substrate complex



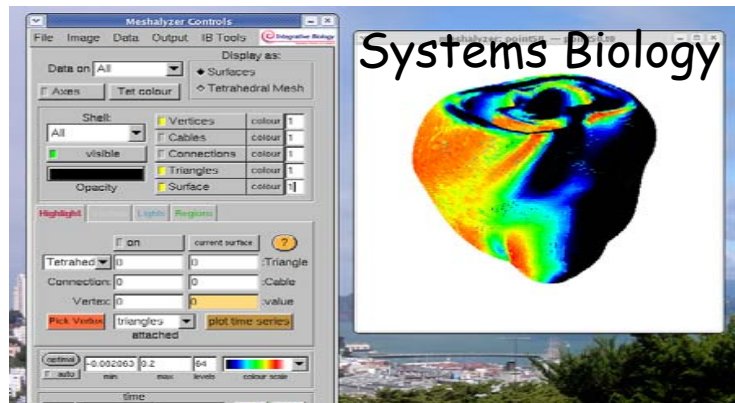
product complex



Lattice Boltzmann
Text mining



Applications: 2



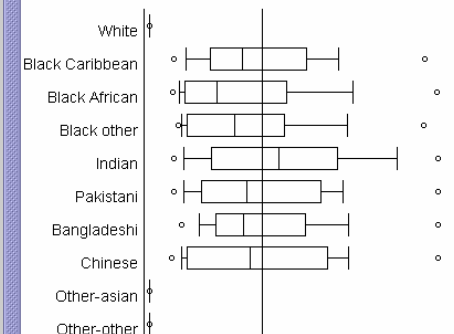
☐ [Met2_pov = 0]
☐ [Met2_pov > 0.0] AND [Met2_pov <= 10.0]
☐ [Met2_pov > 10.0] AND [Met2_pov <= 20.0]
☐ [Met2_pov > 20.0] AND [Met2_pov <= 25.0]
☐ [Met2_pov > 25.0] AND [Met2_pov <= 33.3]
☐ [Met2_pov > 33.3] AND [Met2_pov <= 50.0]
☐ [Met2_pov > 50.0] AND [Met2_pov <= 66.6]
☐ [Met2_pov > 66.6] AND [Met2_pov <= 100.0]

Econometric analysis

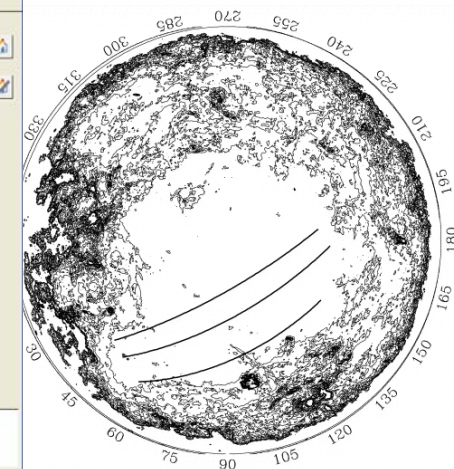
Region/SARs Area

- ☐ White
- ☒ Black Caribbean
- ☐ Black African
- ☐ Black other
- ☐ Indian
- ☐ Pakistani
- ☐ Bangladeshi
- ☐ Chinese
- ☐ Other-asian
- ☐ Other-other
- ☒ Male
- ☐ Female
- ☐ All

UK Male Imputed Income

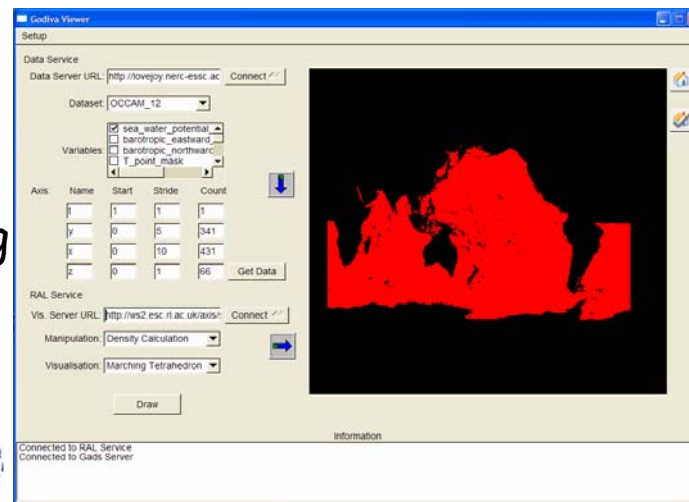
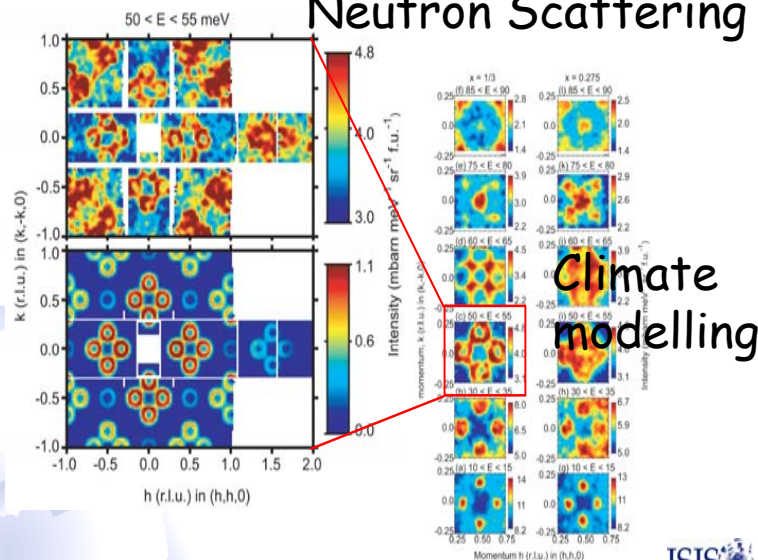


Astronomy



Example: $\text{La}_{2-x}\text{Sr}_x\text{NiO}_4$

Neutron Scattering



H. Woo et al, Phys Rev B 72 064437 (2005)



Applications: 3

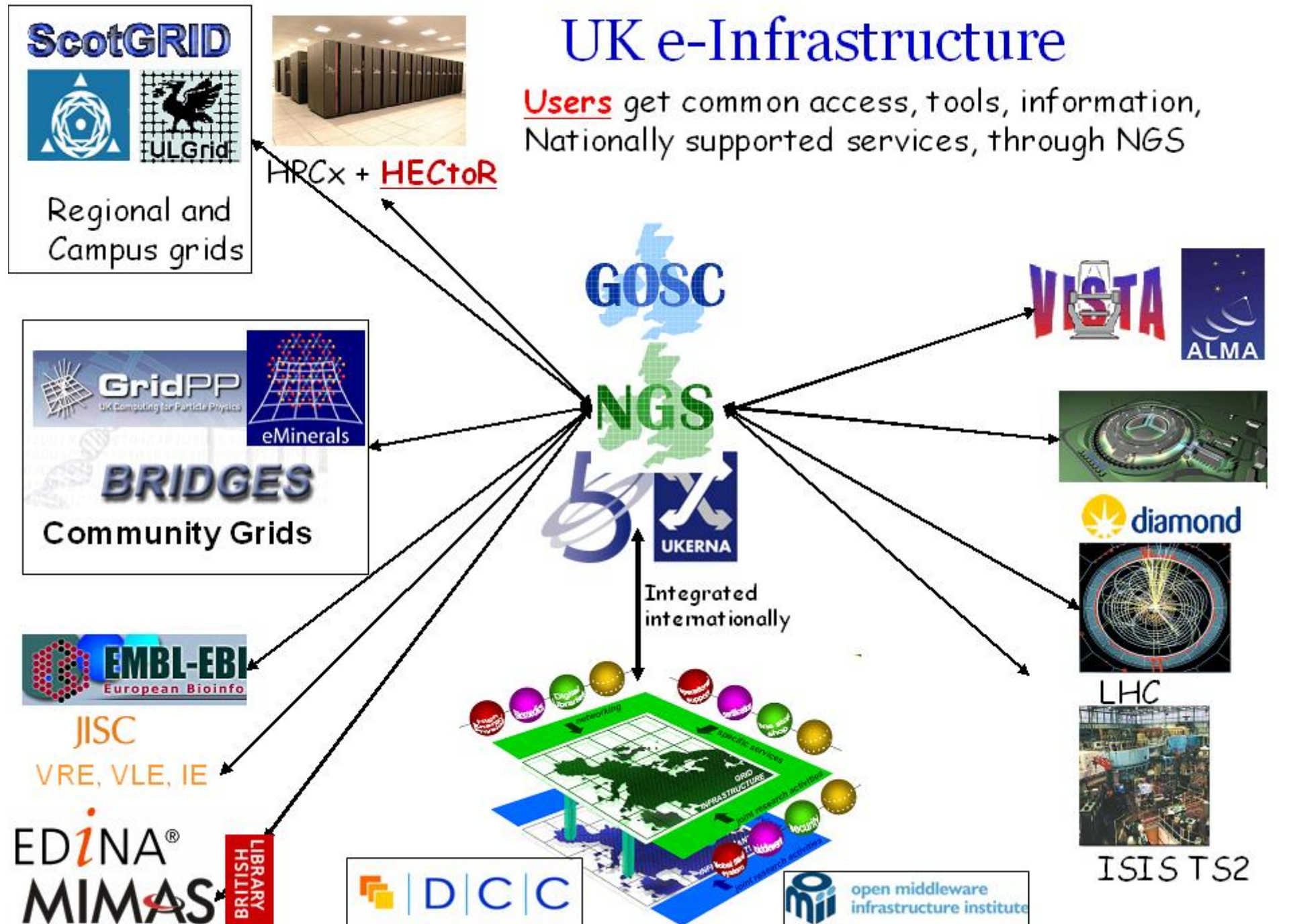
Other Applications:

- nano-particles
 - protein folding
 - ab-initio protein structure prediction
 - radiation transport (radiotherapy)
 - IXI (medical imaging)
 - Biological membranes
 - Micromagnetics
 - Archeology
 - Text mining
 - Lattice QCD (analysis)
 - Astronomy (VO services)
- Many, but not all, applications cover traditional computational sciences
 - Both user and pre-installed software
 - Several data focused activities
 - Common features are
 - Distributed data and/or collaborators
 - Not just pre-existing large collaborations
 - Explicitly encourage new users
 - Common infrastructure/interfaces

- What generic infrastructure/services exist ?
 - Moving tools form project -> infrastructure
- Difference community views of what is critical
- Authorisation
- User/VO management across shared infrastructures
- Bringing infrastructures together
 - Now
 - Common NGS/GridPP/EGEE helpdesk, security, authentication
 - Overlapping local teams
 - Working on VO management and brokering
 - Future
 - Depends on standardisation + commitment from projects
 - Shibboleth
 - ...

UK e-Infrastructure

Users get common access, tools, information,
Nationally supported services, through NGS



NGS Summary

- National Grid Service has now been operational for 18 months
 - Funded through to 2009 and recognised as strategically important
- Focused on Services not Technology
- Ultimate success will be determined by how many resource and users support the interfaces

Conclusions

Major UK Investment in Grid Technology

Many successes but key tests still to come - LHC

Success in broadening the user and application base
but more work needed

Work ongoing to bring the GridPP and NGS efforts
closer together