



GridPP

UK Computing for Particle Physics



Introduction to Grids and GridPP

Steve Lloyd

Queen Mary, University of London

London Tier-2 Workshop April 2007

Edited by P Hobson, Brunel March 2008

Week 09

Week 10

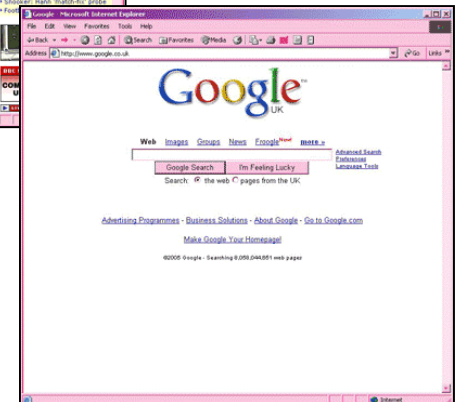
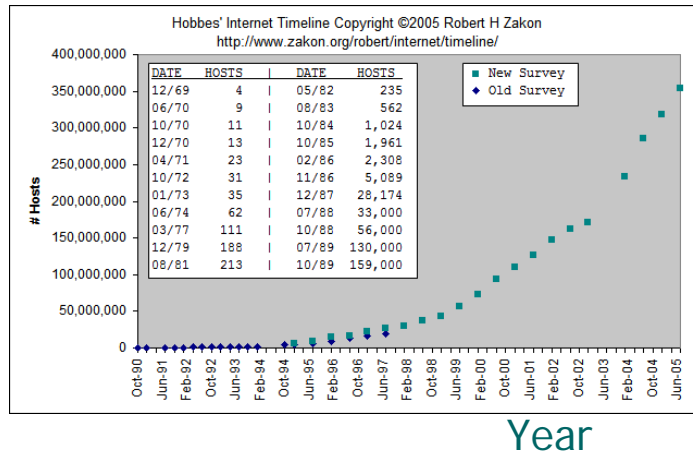
CPUS ■ RUNNING PROCESSES

Memory last month

- Invented at CERN by Tim Berners-Lee
- Quickly crossed over into public use



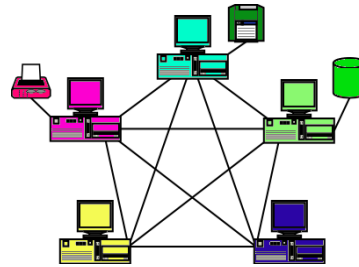
No. of Internet hosts (millions)



- Agreed protocols: HTTP, HTML, URLs
- Anyone can access information and post their own

Distributed File Sharing Peer To Peer Networks

- No centralised database of files
- Legal problems with sharing copyrighted material
- Security problems



Peer-to-peer network



Distributed Resource Sharing @Home Projects

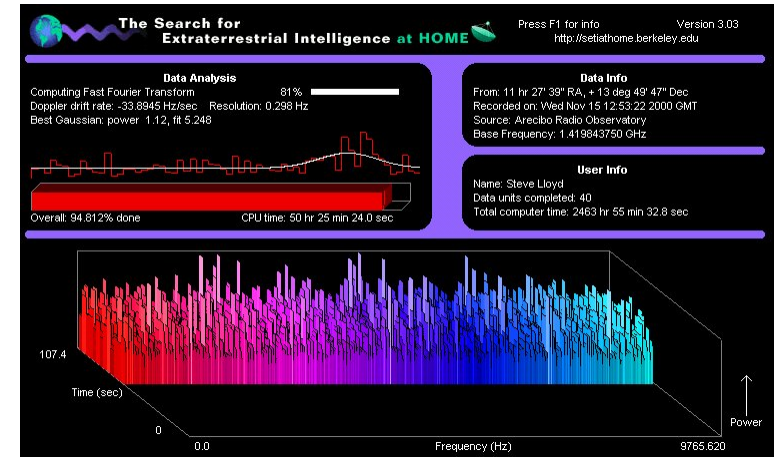
- Uses home PCs to run numerous calculations with dozens of variables.
- Distributed computing project, not a Grid
- Some @home projects
 - BBC Climate Change Experiment
 - SETI @ Home
 - FightAIDS@home





- A distributed computing project - not really a Grid project
- You pull the data from them rather than they submit the job to you

Users -	5,240,038
Results received -	1,632,106,991
<u>Years</u> of CPU Time -	2,121,057
Extraterrestrials found -	0



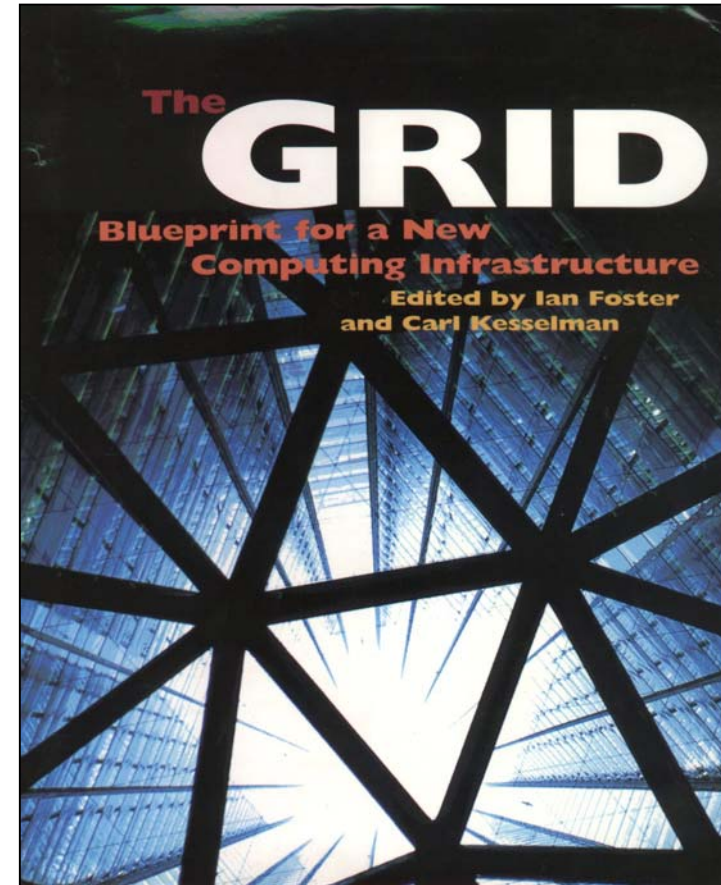
Arecibo telescope in Puerto Rico

1999 - The "Grid"

'Grid' means different things to different people

Ian Foster / Carl Kesselman:

"A computational Grid is a hardware and software infrastructure that provides **dependable, consistent, pervasive** and **inexpensive** access to high-end computational capabilities."



All agree it's a funding opportunity!

2001 - Establishment of UK e-Science Programme

Dr John Taylor - Director General of Research Councils:

"Science increasingly done through distributed global collaborations enabled by the internet using very large data collections, terascale computing resources and high performance visualisation"

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

"e-Science will change the dynamic of the way Science is undertaken"



NGS

National Grid Service

Core Sites

- White Rose (Compute)
- Oxford (Compute)
- RAL (Data)
- Manchester (Data)

Partner Sites

- Belfast
- Bristol
- Cardiff
- Lancaster
- Westminster

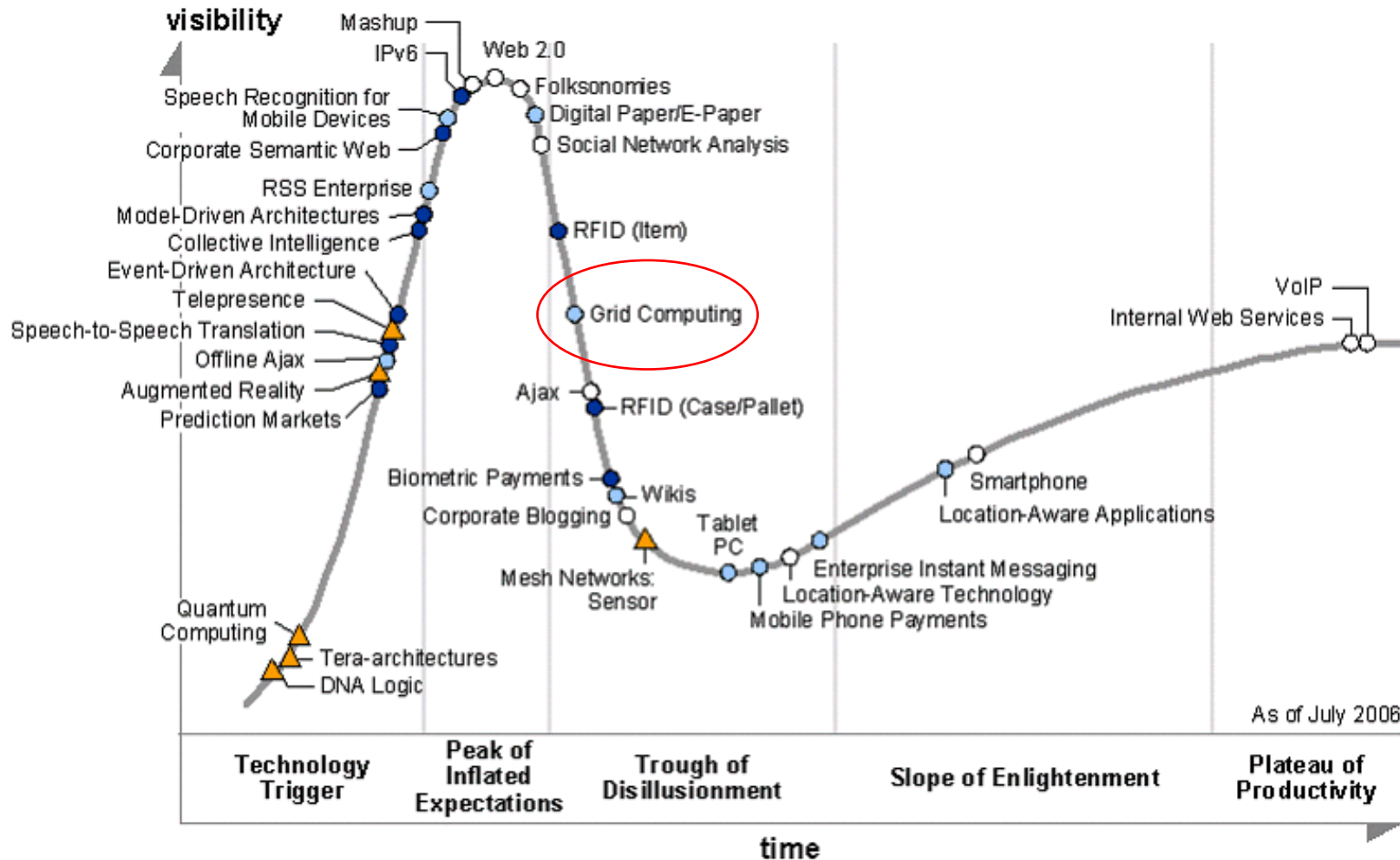
Affiliates

- Edinburgh (NeSC)

National HPC Facilities

- HPCx





Years to mainstream adoption:

○ less than 2 years

● 2 to 5 years

● 5 to 10 years

▲ more than 10 years

○ obsolete

⊗ before plateau

19 UK Universities, CERN,
RAL & Daresbury
Funded by PPARC/STFC:

GridPP1 2001-2004

“From Web to Grid”

GridPP2 2004-2008

“From Prototype to Production”

GridPP3 2008-2011

“From Production to Exploitation”

Developed a working, highly
functional Grid



THE UNIVERSITY
OF BIRMINGHAM



University of Bristol



UNIVERSITY OF
CAMBRIDGE



Imperial College
London



THE UNIVERSITY
of LIVERPOOL



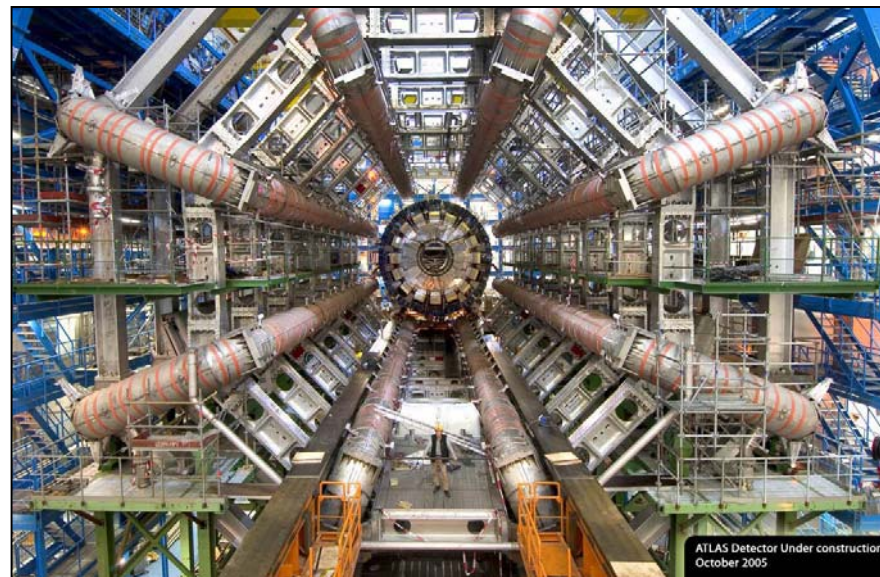
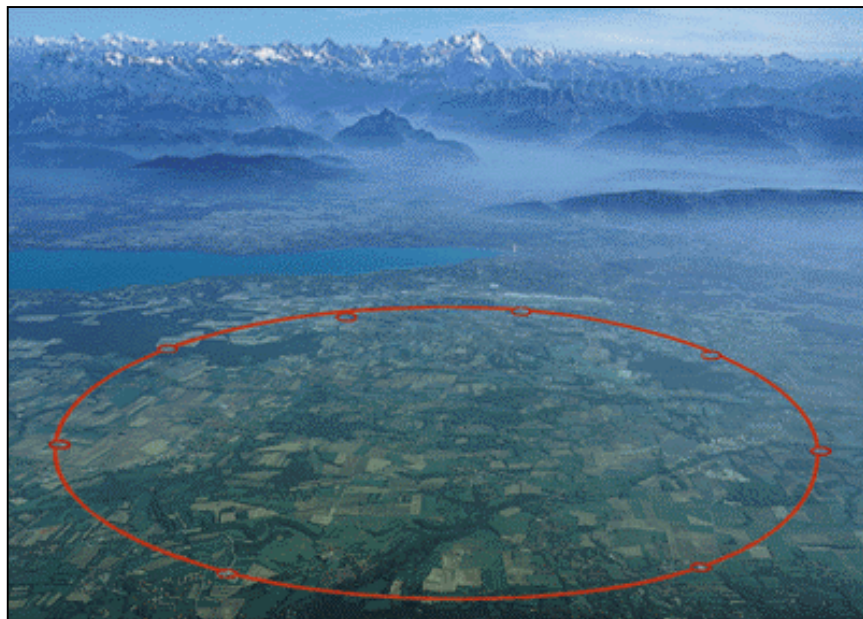
PRIFYSGOL CYMRU ABERTAW
UNIVERSITY OF WALES SWANSEA

THE UNIVERSITY OF SHEFFIELD

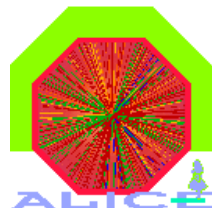
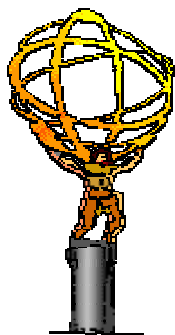


The CERN Large Hadron Collider - LHC

The world's most powerful particle accelerator - Starting 2008



4 Large Experiments



- ~100,000,000 electronic channels
- 800,000,000 proton-proton interactions per second
- 0.0002 Higgs per second
- 10 PBytes of data a year
- (10 Million GBytes = 14 Million CDs)

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



National Grid Service

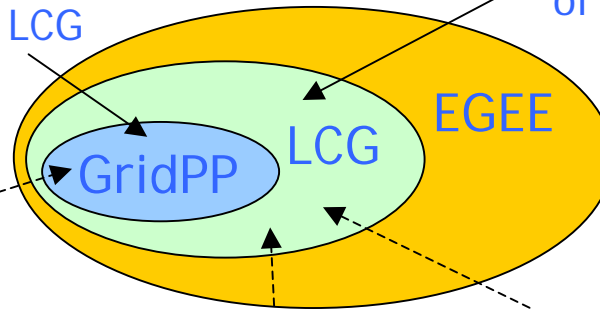
UK part of LCG

PP part of EGEE



LHC Computing Grid (LCG)
Grid Deployment Project for the Large Hadron Collider (LHC)

UK National Grid Service
UK's core production computational and data Grid



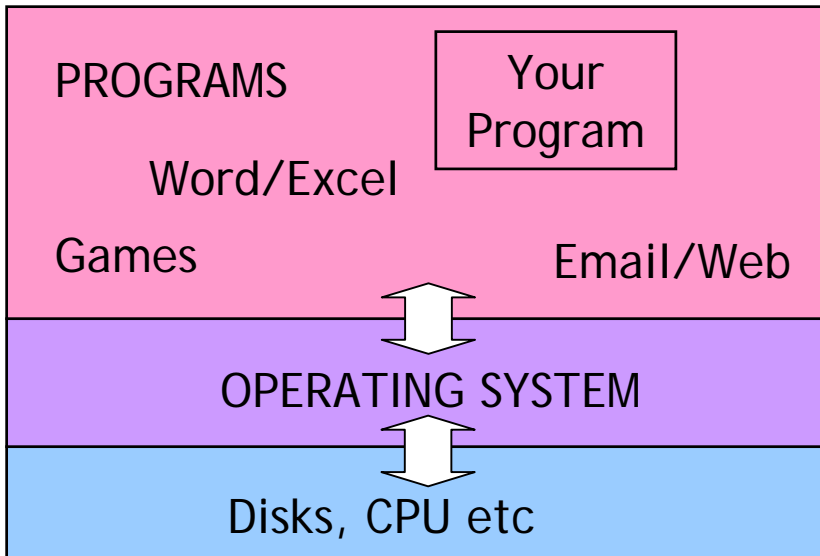
Nordugrid (Scandinavia)
Grid Research and Development collaboration



Open Science Grid (USA)
Science applications from HEP to biochemistry

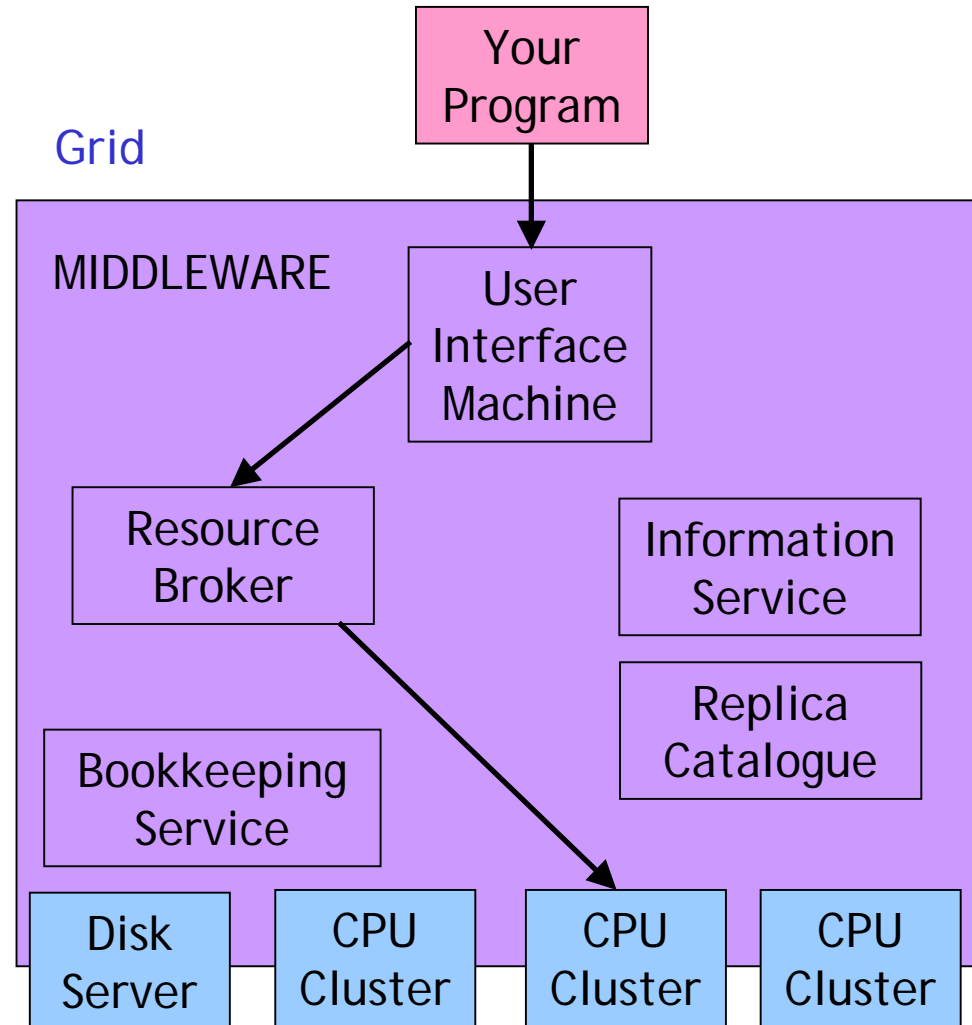
What is (gLite) Middleware?

Single PC



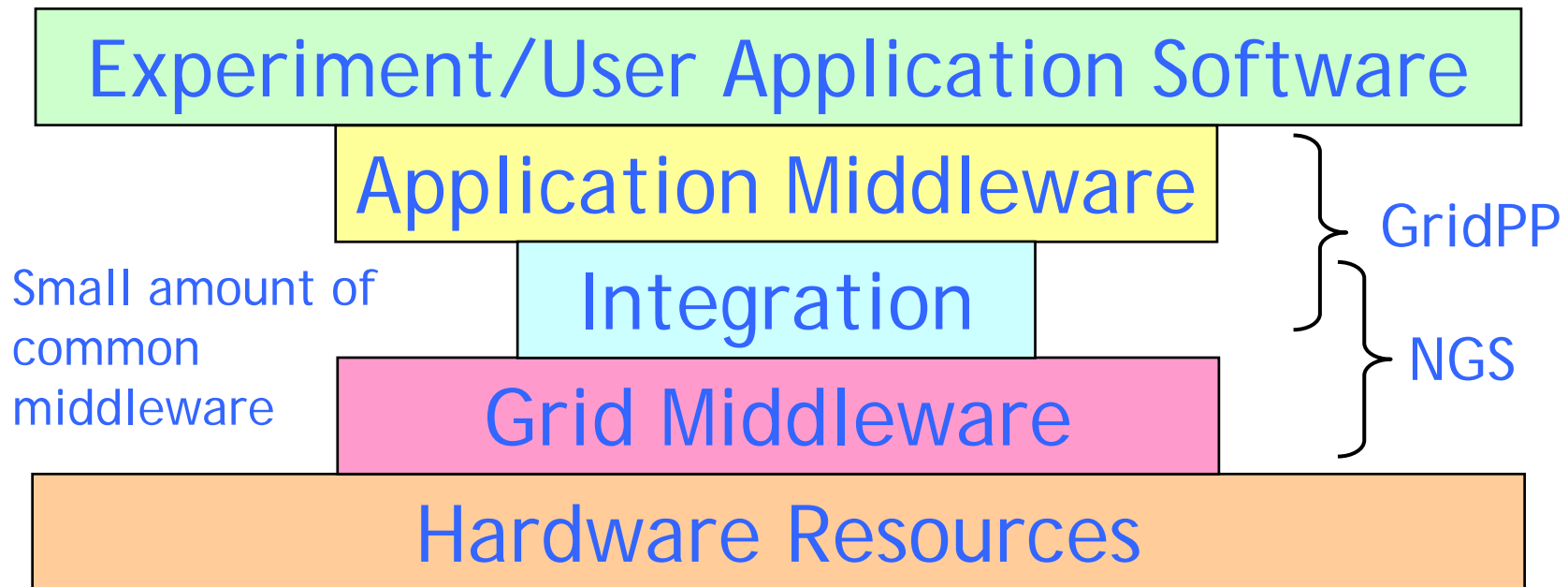
Middleware is the Operating System of a distributed computing system

Grid



A number of different software layers

Many diverse user communities



Large amount of accessible hardware

GridPP Middleware Development

Grid Data Management

GridPP Data Management

- File Metadata**
 - Logical File Name
 - GUID
 - System Metadata (owner, permissions, checksum, ...)
- User Metadata**
 - User Defined Metadata
- File Replica**
 - Storage File Name
 - Storage Host
- Symlinks**
 - Link Name

21 September 2005 AHM05 Meeting Tony Doyle - University of Glasgow

Network Monitoring

GridPP High Performance Networking

UK national R&D network infrastructure

National Interconnection

International Connectivity

21 September 2005 AHM05 Meeting Tony Doyle - University of Glasgow

Workload Management

GridPP Workload Management

Integrated over all VOs and RBs:

Successes/Day: 15225
Success %: 69%
Improving from 42% to 70-80% during 2005

Problems identified:

- 43% Managing resources
- 9% Wrong configuration
- 1% Checksum / Connection failures
- 4% Missing Problems
- 1% JDL Problems

VO Status (Production tests/reqs global): Success/Registered/Cancelled Jobs = 69% / 15225 success jobs per day (Coverage) / 83000

21 September 2005 AHM05 Meeting Tony Doyle - University of Glasgow

Information Services

R-GMA Web Services eGEE Enabling Grids for E-science in Europe

User Application → R-GMA API → SOAP STUB → SOAP/HTTP Messages → AXIS SERVICET → Primary Producer Service, Secondary Producer Service, On-demand Producer Service, Consumer Service → Tomcat Servlet Container

Created from WSDL using gSOAP (or Axis or ...)

- API available for Java, C, C++ and Python
- Users may by-pass API if they wish, but API is the easiest way to use R-GMA services

R-GMA: Status and Plans, 14 September 2004 - 4

Security

GridPP Security in LCG/EGEE

Solutions/Recommendations

JRA3 → JRA1 → NA4

Req. → JRA1 → NA4

Req. → JRA1 → Middleware Security Group

Req. → JRA1 → LCG/EGEE Joint Security Group

Req. → JRA1 → SA1

"Joint Security Group" defines/maintains policy and procedures.

For LCG GDB and EGEE SA1

15-Sep-04 Grid Security 3

Storage Interfaces

GridPP SRM

- A single SRM server to service incoming file requests (this is implemented as a web service)
- Multiple file servers with unix filesystems on which data resides.
- Data transfer is done to/from the file servers, thus inbound IP connectivity is essential to make the SRM SE available to the wider grid.

SRM Server

File Server 1 (FS1, FS2, FS3)

File Server 2 (FS1, FS2, FS3)

File Server 3 (FS1, FS2, FS3)

21 September 2005 AHM05 Meeting Tony Doyle - University of Glasgow



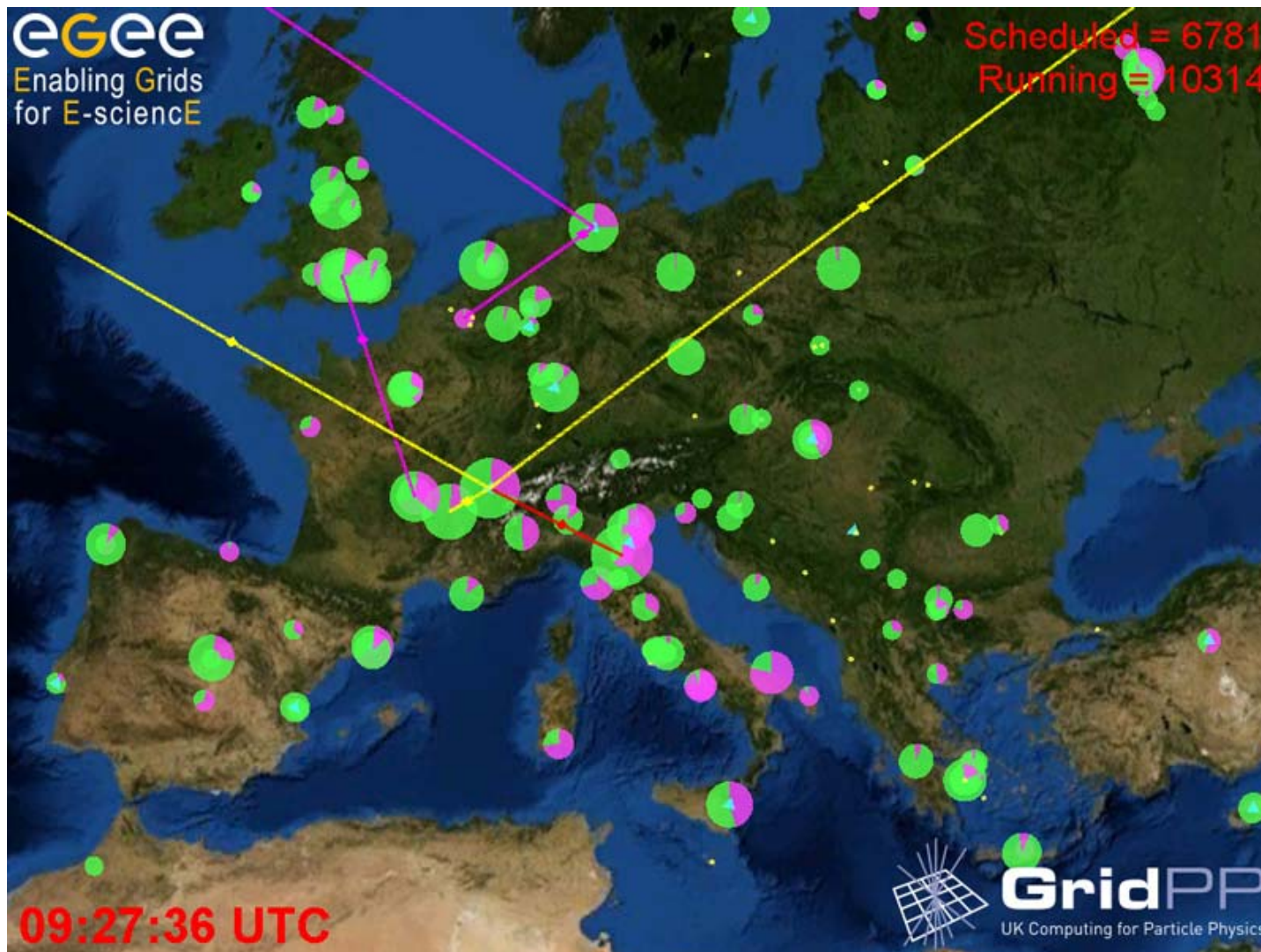
Steve Lloyd



London Tier-2 Workshop - 16 Apr 2007



Slide 14

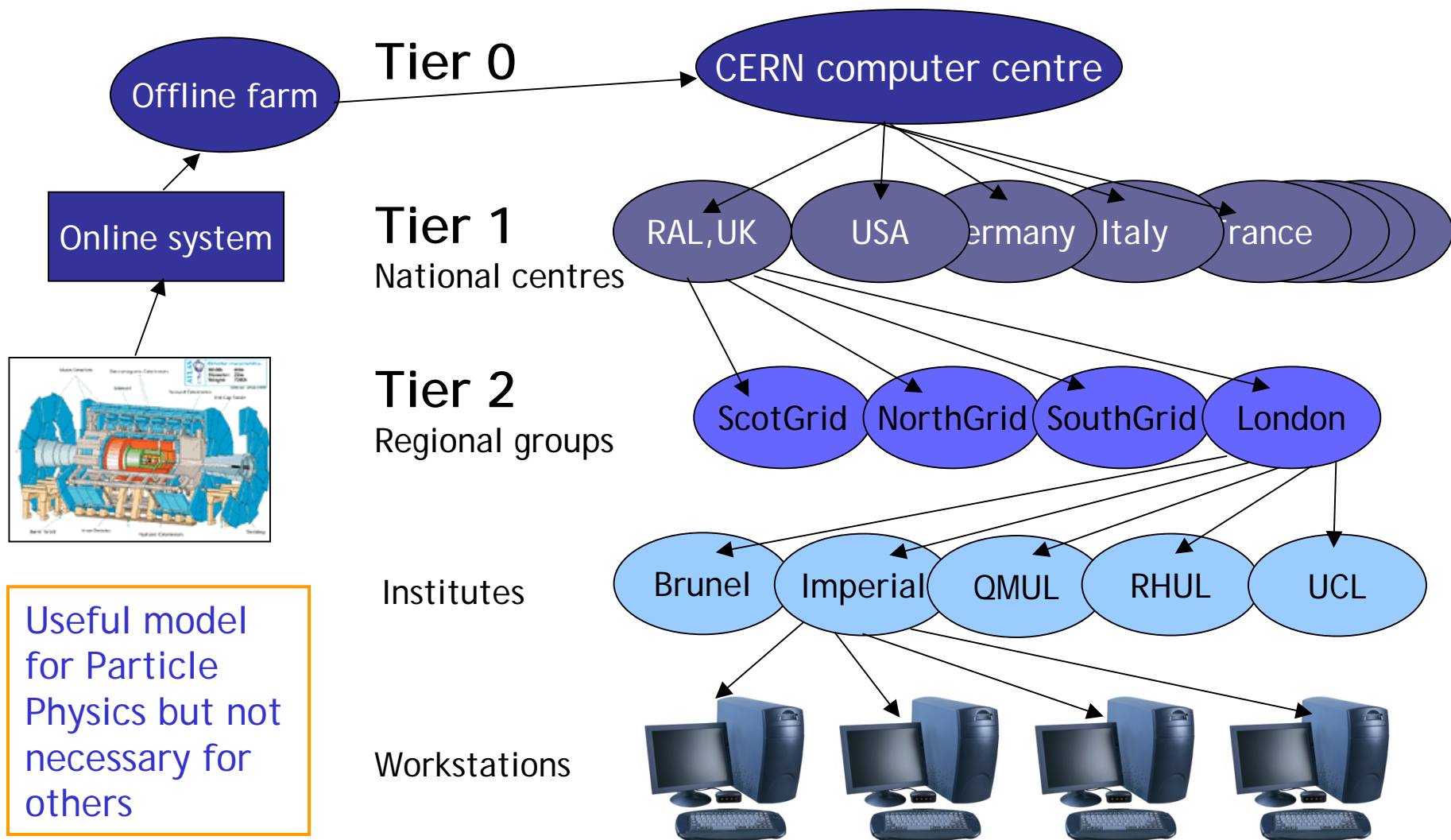


Worldwide

237 Sites
50 Countries
35,716 CPUs
21.3 PB Disk
10,579 Years
of CPU time

UK

21 Sites
8089 CPUs
876 TB Disk
3,361 Years of
CPU time



ScotGrid

Durham, Edinburgh, Glasgow

NorthGrid

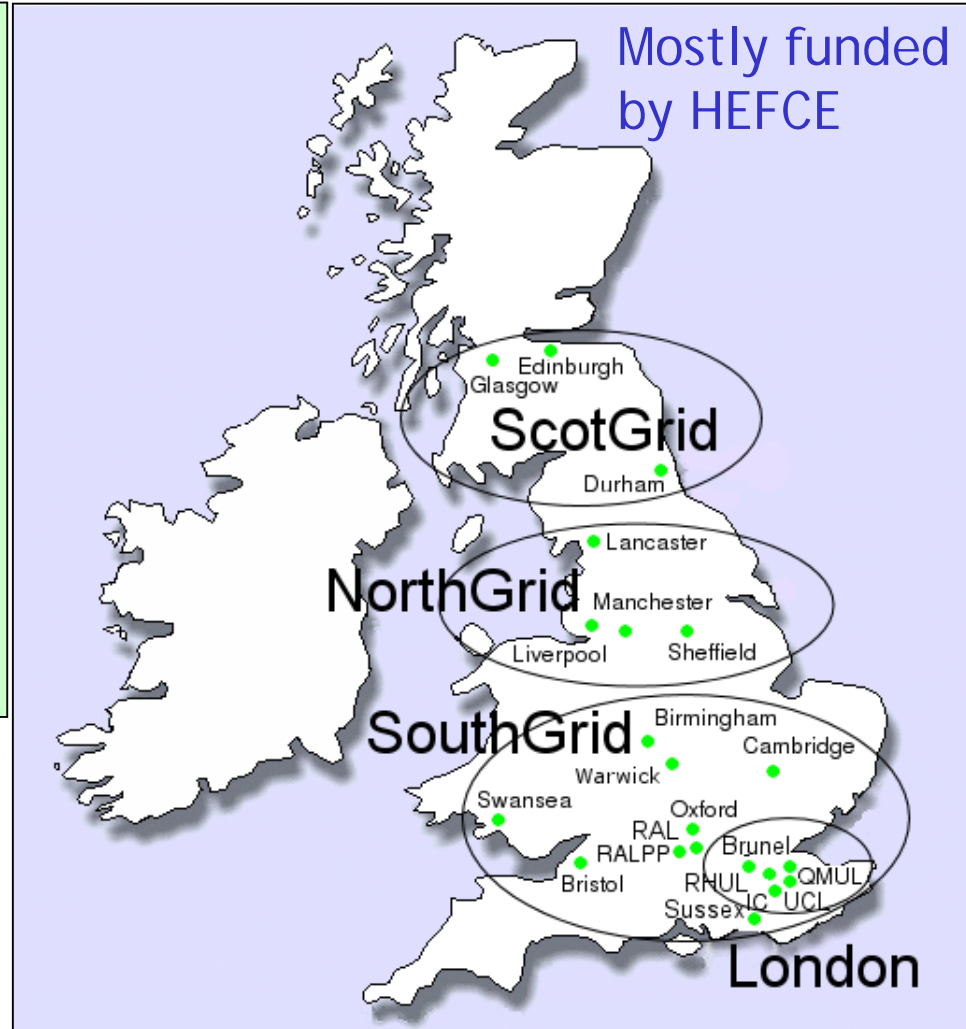
Daresbury, Lancaster, Liverpool,
Manchester, Sheffield

SouthGrid

Birmingham, Bristol, Cambridge,
Oxford, RAL PPD

London

Brunel, Imperial, QMUL, RHUL, UCL

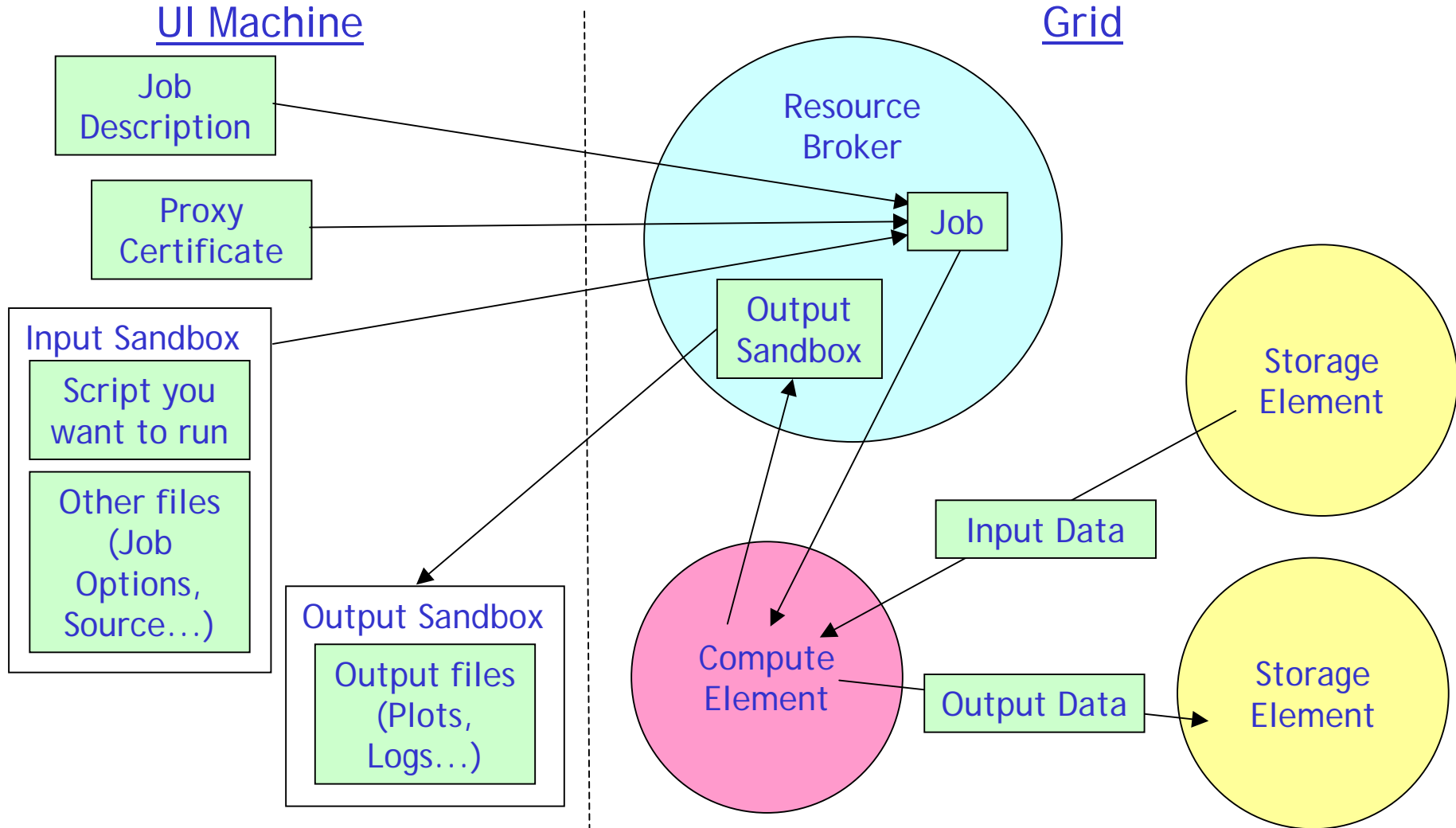


What you need to use the Grid

1. Get a digital certificate (UK Certificate Authority)
Authentication - who you are
2. Join a Virtual Organisation (VO)
Authorisation - what you are allowed to do
3. Get access to a local User
Interface Machine (UI) and copy
your files and certificate there
4. Write some Job Description Language (JDL) and scripts
to wrap your programs

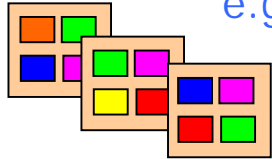


```
##### HelloWorld.jdl #####  
Executable = "/bin/echo";  
Arguments = "Hello welcome to the Grid ";  
StdOutput = "hello.out";  
StdError = "hello.err";  
OutputSandbox = {"hello.out", "hello.err"};  
#####
```

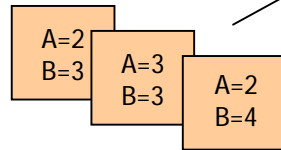


Problems that are highly parallelizable

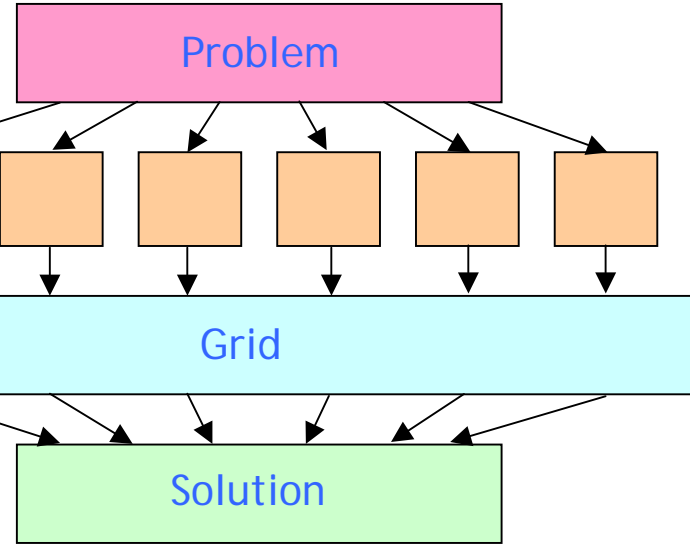
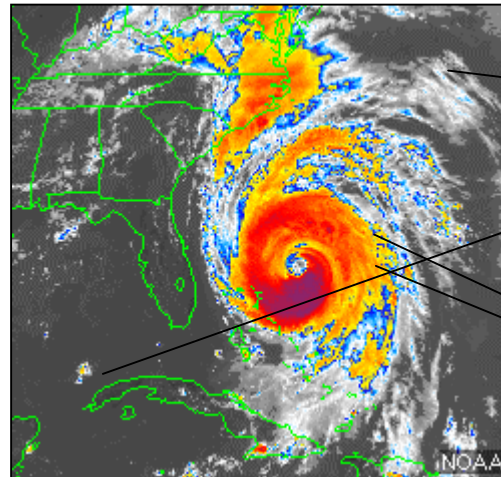
Input data is independent
e.g. Images:



Simulation using
different parameters:



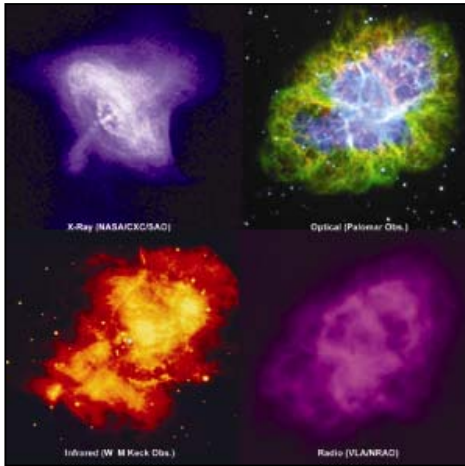
Not so good for closely
coupled problems



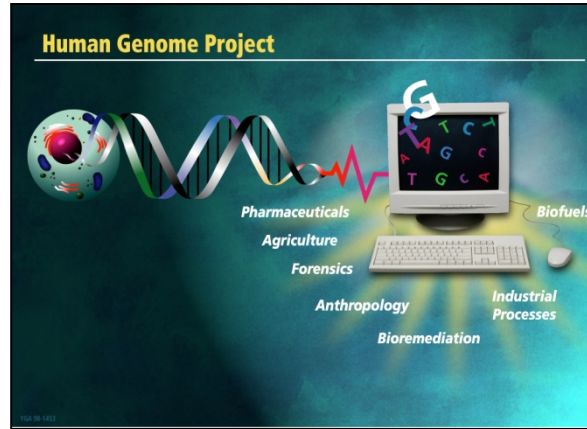
These pieces may be
independent

These pieces will have
to interact

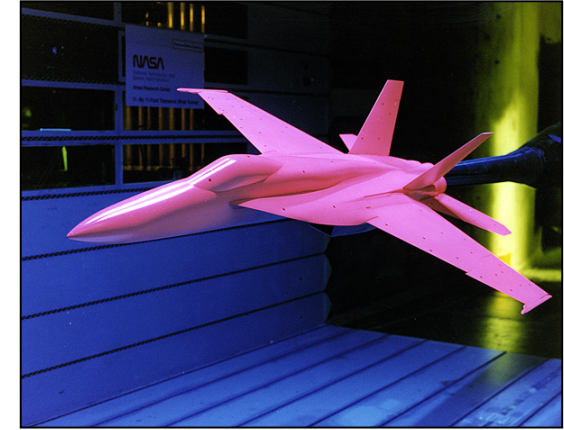
- Astronomy



- Bioinformatics



- Engineering



- Healthcare



- Commerce



- Gaming



BBC NEWS | Technology - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://news.bbc.co.uk/1/hi/technology/default.stm

Google Groups GridPP BBC News The Register Steve Lloyd's Logbook

BBC NEWS | Technology

bbc.co.uk Home TV Radio Talk Where I Live A-Z Index

BBC NEWS

UK version International version About the versions Low graphics Help

News Front Page World UK England Northern Ireland Scotland Wales Business Politics Health Education Science/Nature Technology Entertainment

Have Your Say Magazine In Pictures Country Profiles In Depth Programmes

What is RSS?

RELATED BBC SITES SPORT WEATHER CBBC NEWSROUND ON THIS DAY NEWSWATCH

Done

BBC NEWS | Technology | Grid searches for avian flu cure - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://news.bbc.co.uk/1/hi/technology/4977150.stm

Google Groups GridPP BBC News The Register Steve Lloyd's Logbook

BBC NEWS | Technology | Grid search...

bbc.co.uk Home TV Radio Talk Where I Live A-Z Index Search

BBC NEWS

UK version International version About the versions Low graphics Help Contact us

Last Updated: Friday, 5 May 2006, 13:26 GMT 14:26 UK

E-mail this to a friend Printable version

Grid searches for avian flu cure

A cure for bird flu is being sought by computers that usually search for the fundamental elements of matter.

In April, the UK grid of computers that crunches data from particle physics experiments was used by scientists to study the avian H5N1 virus.

The grid simulated the reactions of 300,000 chemical compounds to find which ones might combat the bug.

The project aims to make grids available to scientists that need to call on lots of computer power.

Drug trials

Computer grids tie together ranks of low-powered machines to create a virtual supercomputer.

For this project, computers at 11 British universities and research labs were used to search through candidate compounds that might help combat H5N1.

Before now, the grid of computers was organised as the GridPP project, which harnessed machines to comb through the data generated by physics experiments.

OTHER: File-sh, Piracy, Net ce, Toxic, Sky s

ALSO I

RELATED BBC SITES SPORT WEATHER CBBC NEWSROUND ON THIS DAY NEWSWATCH

UK ON BIRD FLU ALERT

KEY STORIES

- Slow-death fear for poultry culls
- Two farms in bird flu infection
- Bird flu confirmed in farm worker
- Chicken cull after bird flu find
- Bird flu swan was from outside UK
- Risk of human flu outbreak 'low'

FEATURES AND BACKGROUND

Wildfowl watch

The 10 species most likely to bring bird flu into the UK

- Experts answer your questions
- Bird flu 'not hitting business'
- Is it so hard to identify a swan?
- Are garden birds under threat?
- Bird flu plans
- Q&A: Bird flu in the UK
- Timeline: Bird flu in the UK

GLOBAL PICTURE

Bird flu journey

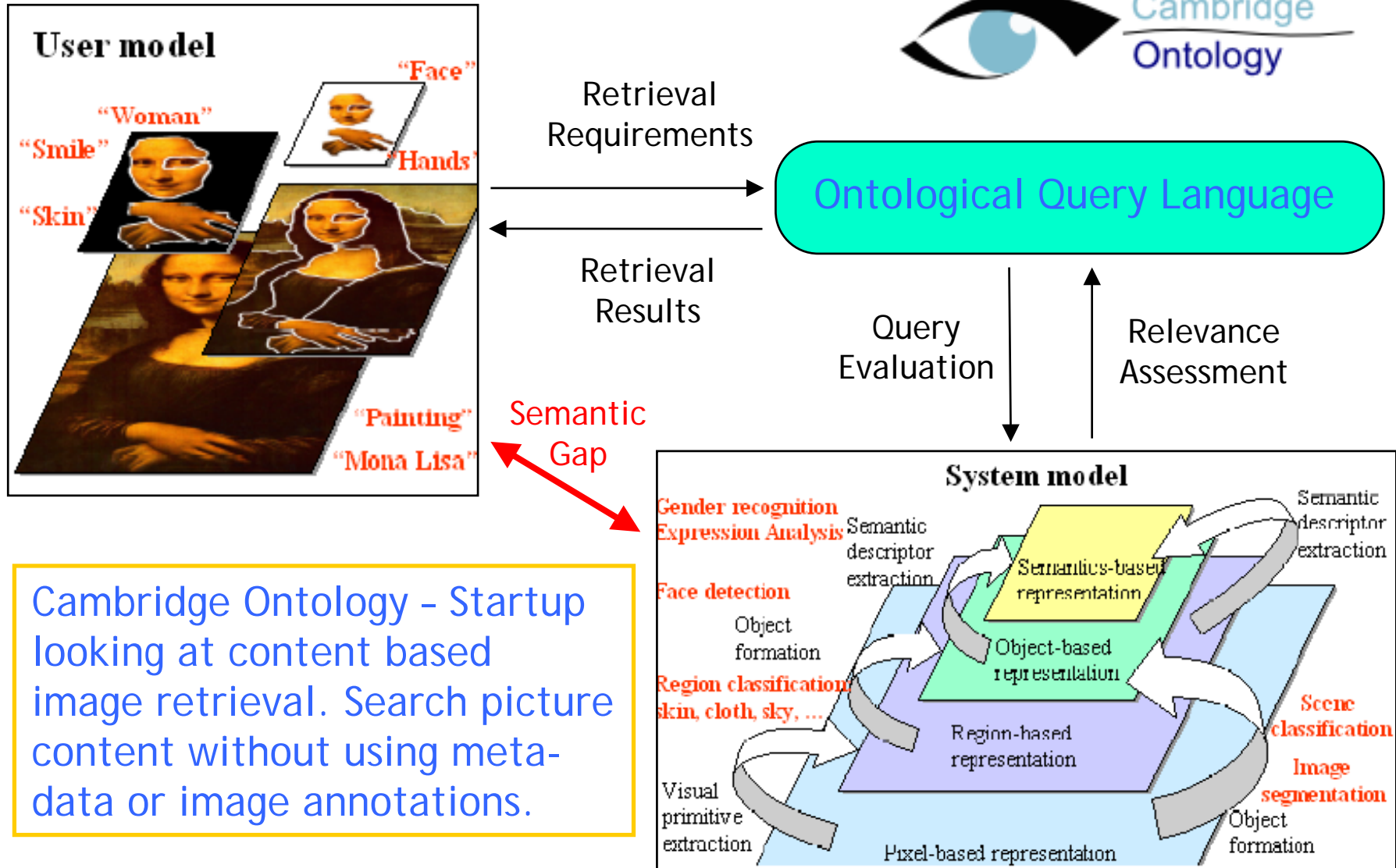
Watch how the lethal virus has spread

- Map: Global impact
- Deadly bird flu in Czech swans
- Egypt reports new bird flu death
- Case confirmed in West Bank

VIDEO AND AUDIO

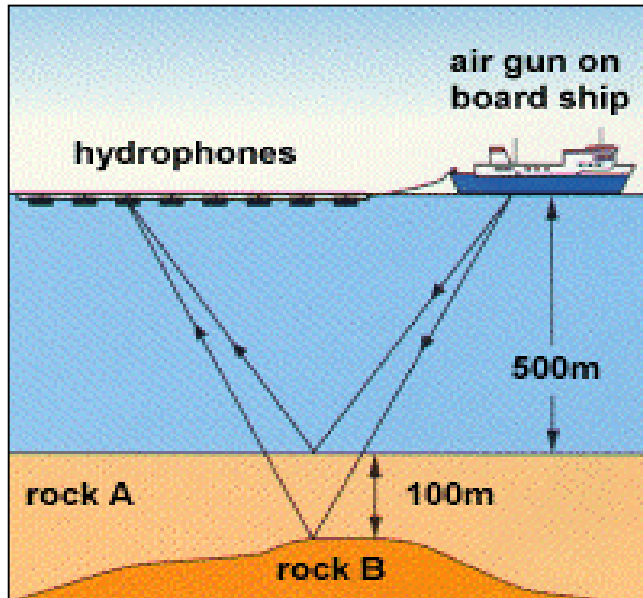
VIDEO TV reports

"GridPP has been developed to help answer questions about the conditions in the Universe just after the Big Bang," said Professor Keith Mason, head of the Particle Physics and Astronomy Research Council (PPARC).
 "But the same resources and techniques can be exploited by other sciences with a more direct benefit to society."

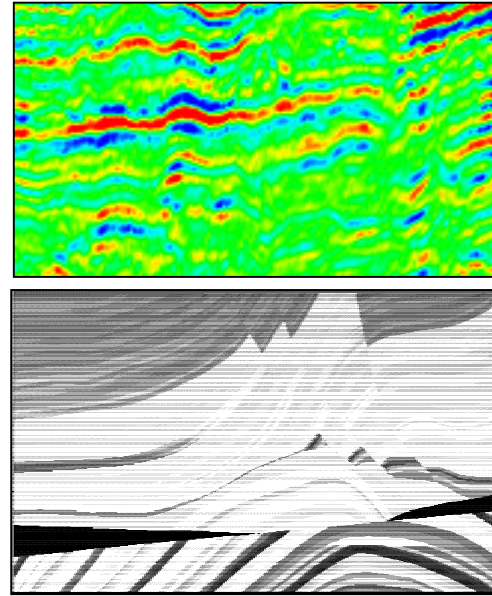


Cambridge Ontology - Startup looking at content based image retrieval. Search picture content without using meta-data or image annotations.

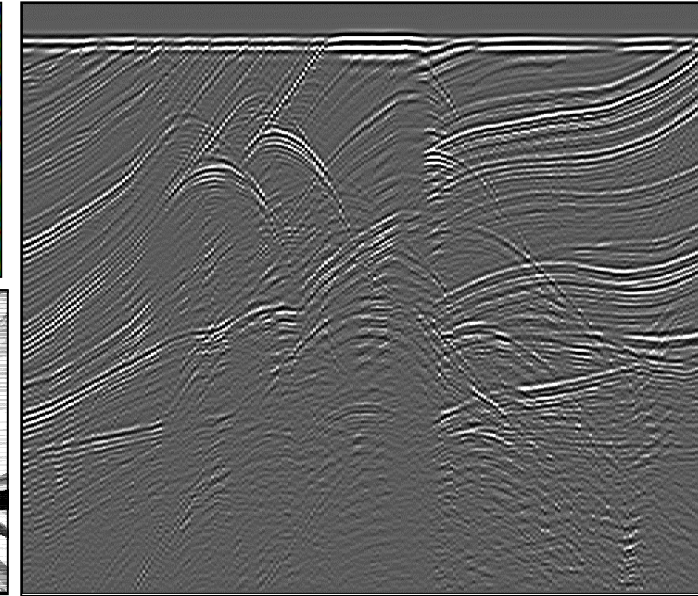
Total Exploration & Production



Marine Experiment



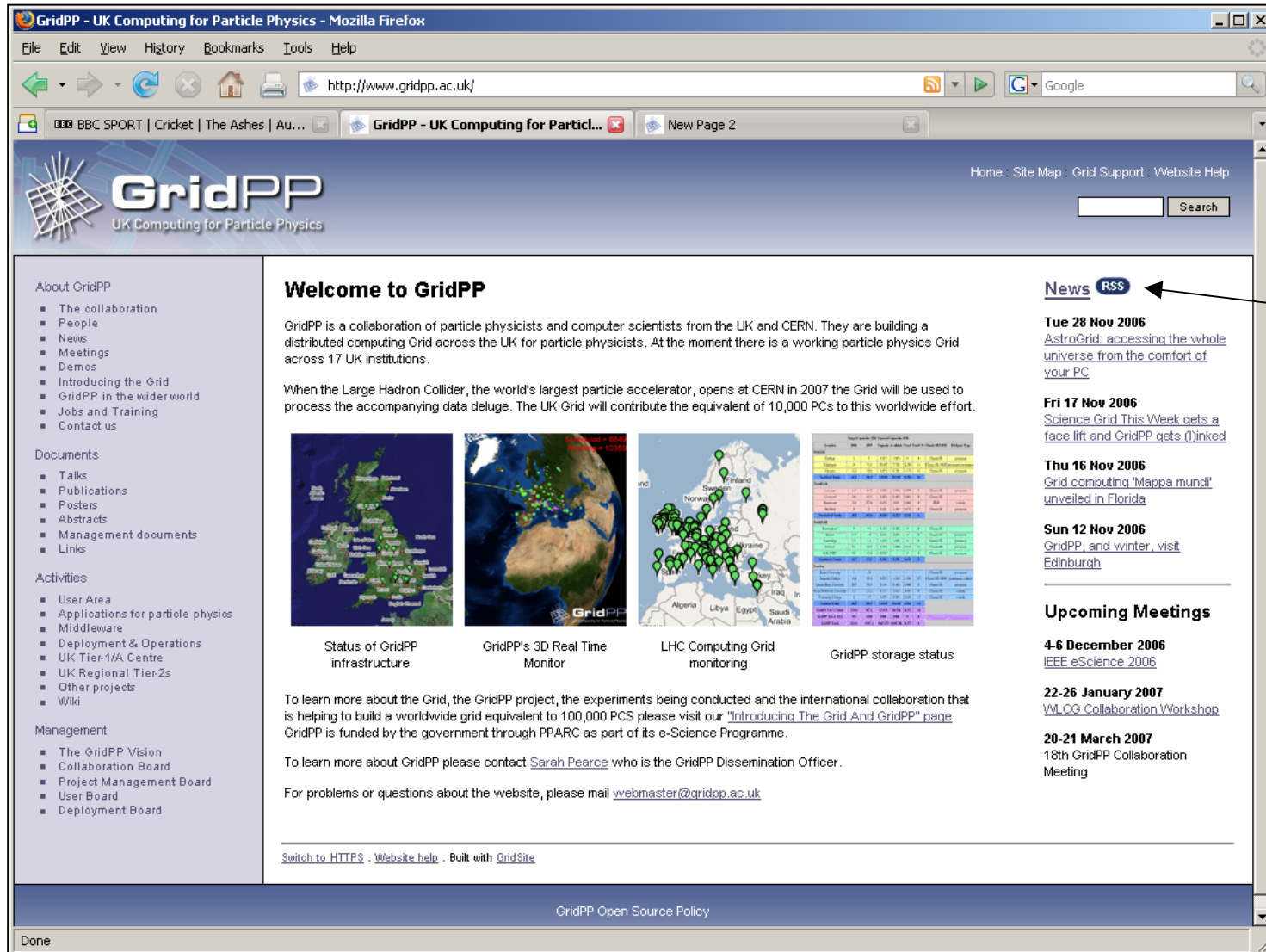
Results of marine experiments



Modelled results based on bore-hole data and wave equations

Use the Grid with modelled data to validate results from marine experiments. Other areas potential areas to port: Seismic Processing. Interpretation of subsurface structures. Reservoir / Field modelling

<http://www.gridpp.ac.uk>



GridPP - UK Computing for Particle Physics - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.gridpp.ac.uk/

BBC SPORT | Cricket | The Ashes | Au... GridPP - UK Computing for Partic... New Page 2

GridPP
UK Computing for Particle Physics

Home : Site Map : Grid Support : Website Help

Search

About GridPP

- The collaboration
- People
- News
- Meetings
- Demos
- Introducing the Grid
- GridPP in the wider world
- Jobs and Training
- Contact us

Documents

- Talks
- Publications
- Posters
- Abstracts
- Management documents
- Links

Activities

- User Area
- Applications for particle physics
- Middleware
- Deployment & Operations
- UK Tier-1/A Centre
- UK Regional Tier-2s
- Other projects
- Wiki

Management

- The GridPP Vision
- Collaboration Board
- Project Management Board
- User Board
- Deployment Board

Welcome to GridPP

GridPP is a collaboration of particle physicists and computer scientists from the UK and CERN. They are building a distributed computing Grid across the UK for particle physicists. At the moment there is a working particle physics Grid across 17 UK institutions.

When the Large Hadron Collider, the world's largest particle accelerator, opens at CERN in 2007 the Grid will be used to process the accompanying data deluge. The UK Grid will contribute the equivalent of 10,000 PCs to this worldwide effort.

Status of GridPP infrastructure

GridPP's 3D Real Time Monitor

LHC Computing Grid monitoring

GridPP storage status

To learn more about the Grid, the GridPP project, the experiments being conducted and the international collaboration that is helping to build a worldwide grid equivalent to 100,000 PCS please visit our ["Introducing The Grid And GridPP" page](#). GridPP is funded by the government through PPARC as part of its e-Science Programme.

To learn more about GridPP please contact [Sarah Pearce](#) who is the GridPP Dissemination Officer.

For problems or questions about the website, please mail webmaster@gridpp.ac.uk

Switch to HTTPS . Website help . Built with [GridSite](#)

GridPP Open Source Policy

News [RSS](#)

Tue 28 Nov 2006
[AstroGrid: accessing the whole universe from the comfort of your PC](#)

Fri 17 Nov 2006
[Science Grid This Week gets a face lift and GridPP gets Dinked](#)

Thu 16 Nov 2006
[Grid computing 'Mappa mundi' unveiled in Florida](#)

Sun 12 Nov 2006
[GridPP, and winter, visit Edinburgh](#)

Upcoming Meetings

4-6 December 2006
[IEEE eScience 2006](#)

22-26 January 2007
[WLCG Collaboration Workshop](#)

20-21 March 2007
18th GridPP Collaboration Meeting

RSS
News
feed