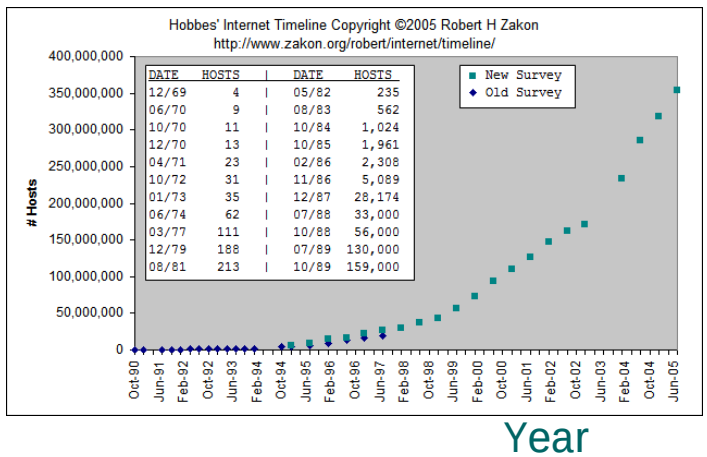
GridPP
UK Computing for Particle Physics

Introduction to Grids and GridPP

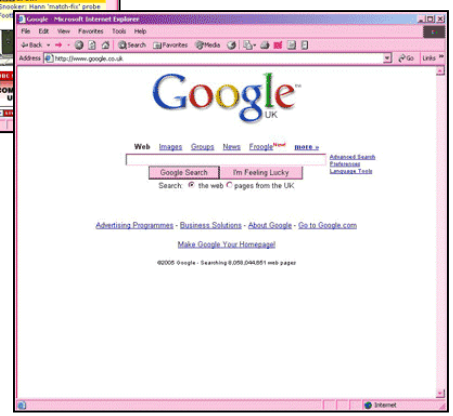
Steve Lloyd
Queen Mary, University of London
London Tier-2 Workshop April 2007
Edited by P Hobson, Brunel March 2008

Web: Information Sharing

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



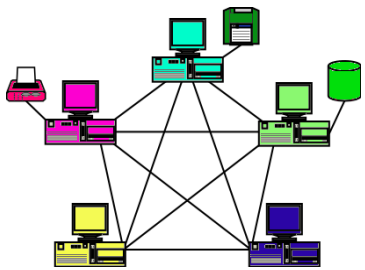
No. of Internet hosts (millions)



- 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



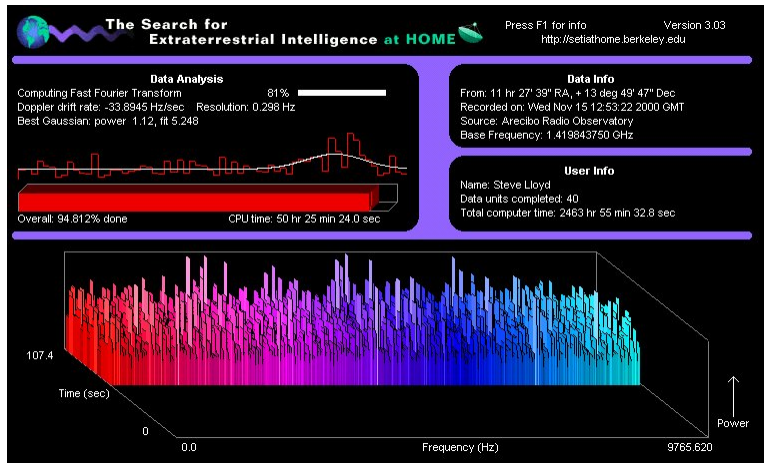
Peer-to-peer network

BitTorrent

Distributed Resource Sharing @Home Projects

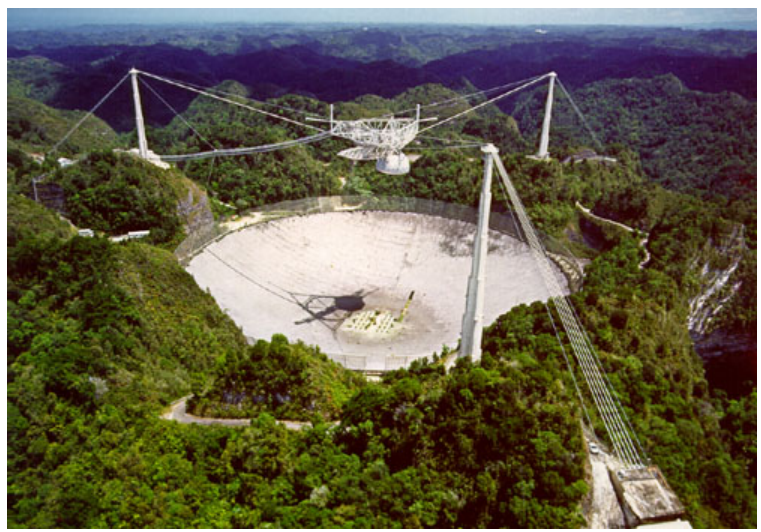
- Uses home PCs to run numerous calculations with dozens of variables.
- Distributed computing project, not a Grid
- Some @home projects
  - BBC Climate Change Experiment
    SETI @ Home
  - FightAIDS@home

SETI@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
Years of CPU Time –	2,121,057
Extraterrestrials found –	0



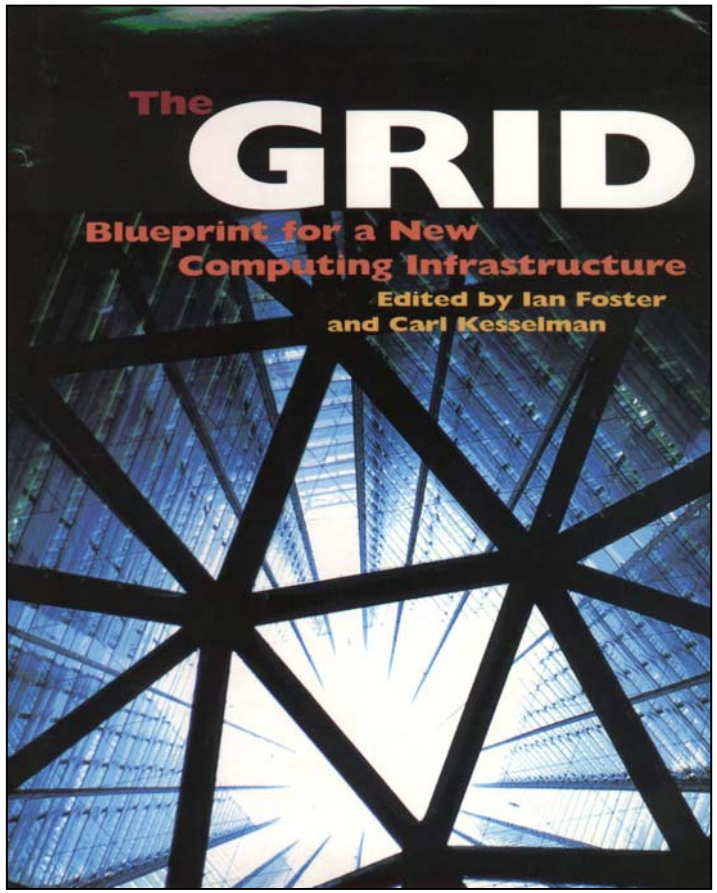
Arecibo telescope in Puerto Rico

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"

e-Science Centres

NGS
National Grid Service

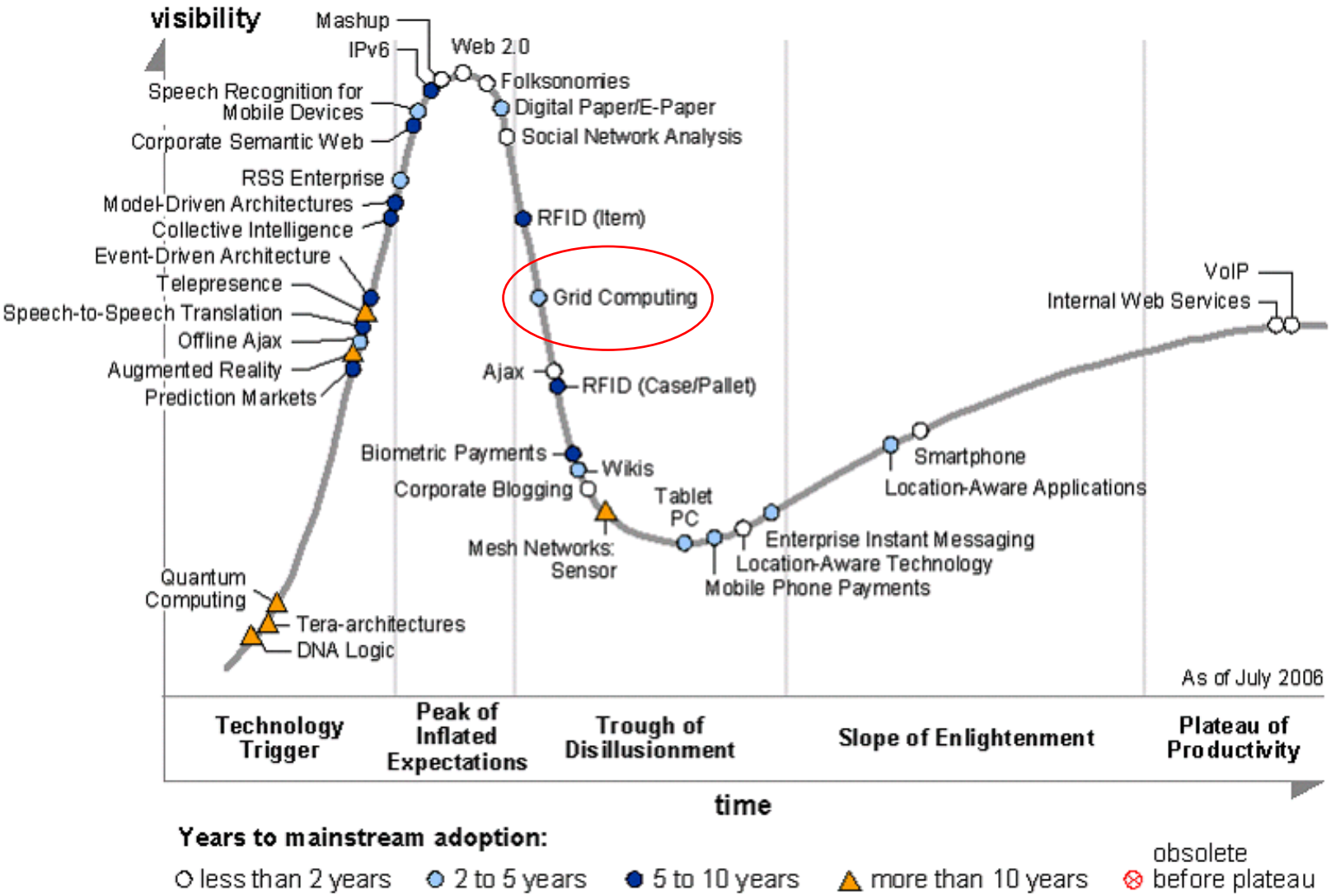
Core Sites
- White Rose (Compute)
- Oxford (Compute)
- RAL (Data)
- Manchester (Data)

Partner Sites
- Belfast
- Bristol
- Cardiff
- Lancaster
- Westminster

Affiliates
- Edinburgh (NeSC)

National HPC Facilities
- HPCx

Gartner Hype Cycle



Who are GridPP?

19 UK Universities, CERN, RAL & Daresbury

Funded by PPARC/STFC:

GridPP1 2001-2004

"From Web to Grid"

GridPP2 2004-2008

"From Prototype to Production"

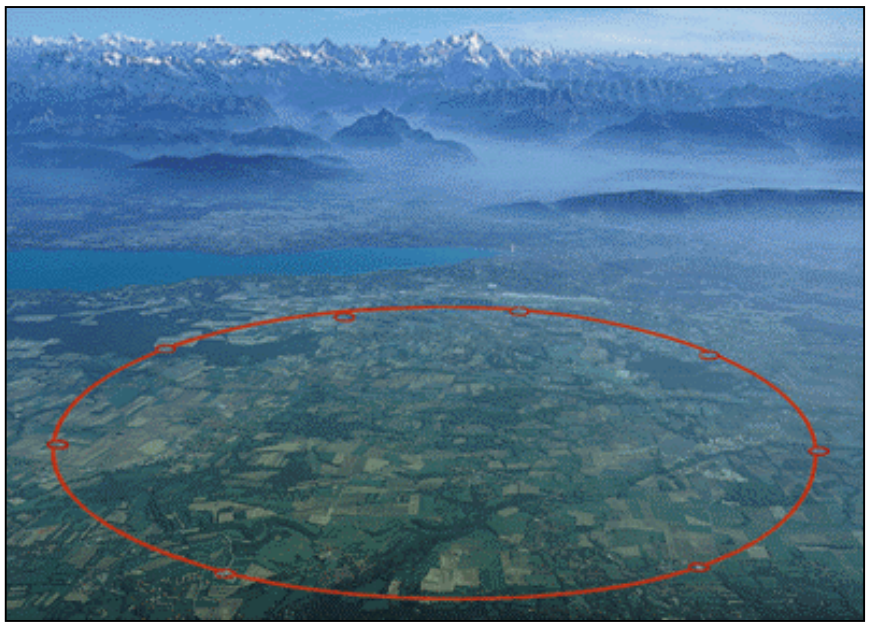
GridPP3 2008-2011

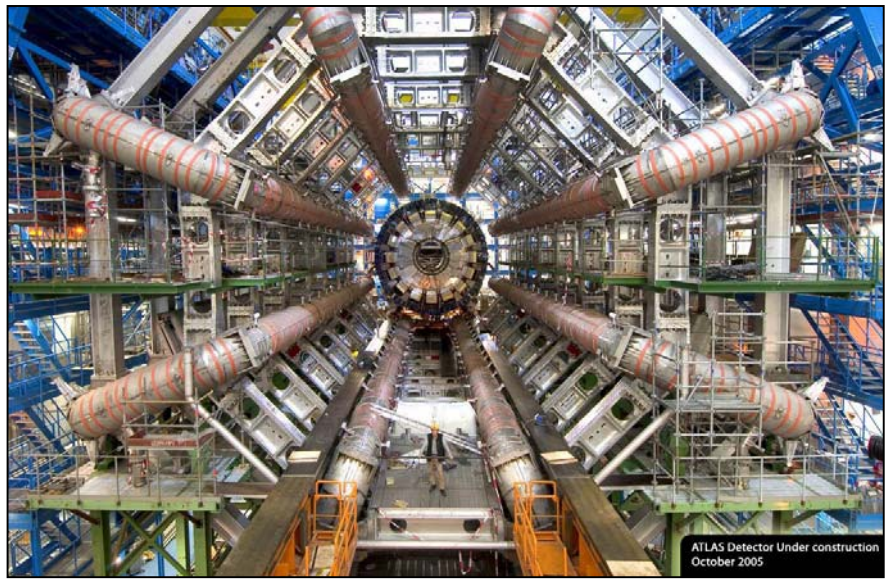
"From Production to Exploitation"

Developed a working, highly functional Grid

The CERN Large Hadron Collider - LHC

The world's most powerful particle accelerator – Starting 2008





4 Large Experiments

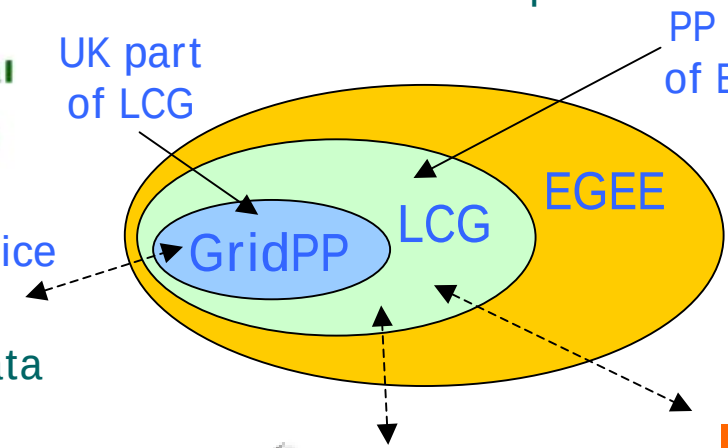
- ~100,000,000 electronic channels
- 800,000,000 proton-proton interactions per second
- 0.0002 Higgs per second
- 10 PBytes of data a year
- (10 Million GBytes = 14 Million CDs)

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 the Large Hadron Collider (LHC)

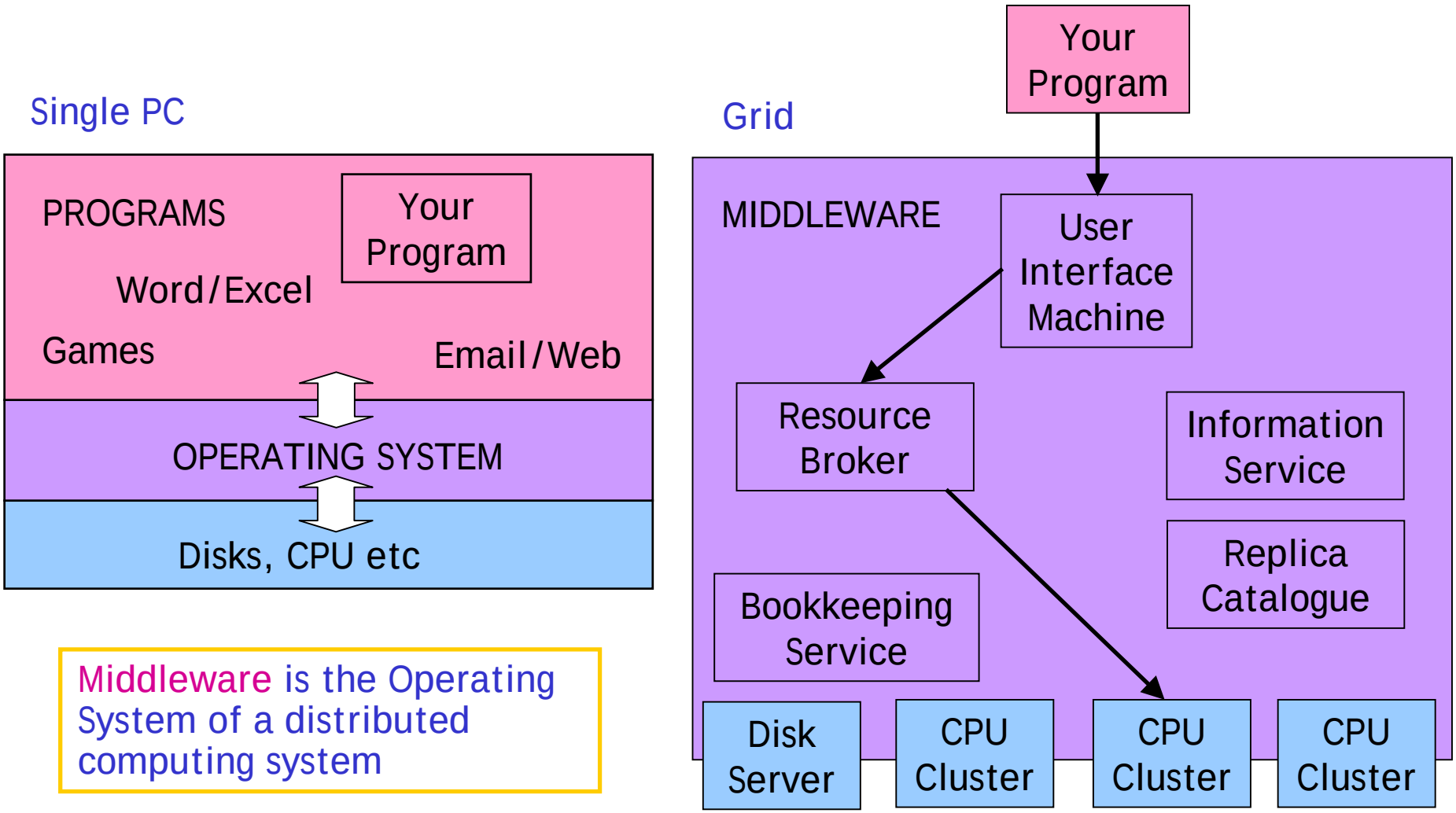
Nordugrid (Scandinavia)

Grid Research and Development collaboration

Open Science Grid (USA)

Science applications from HEP to biochemistry

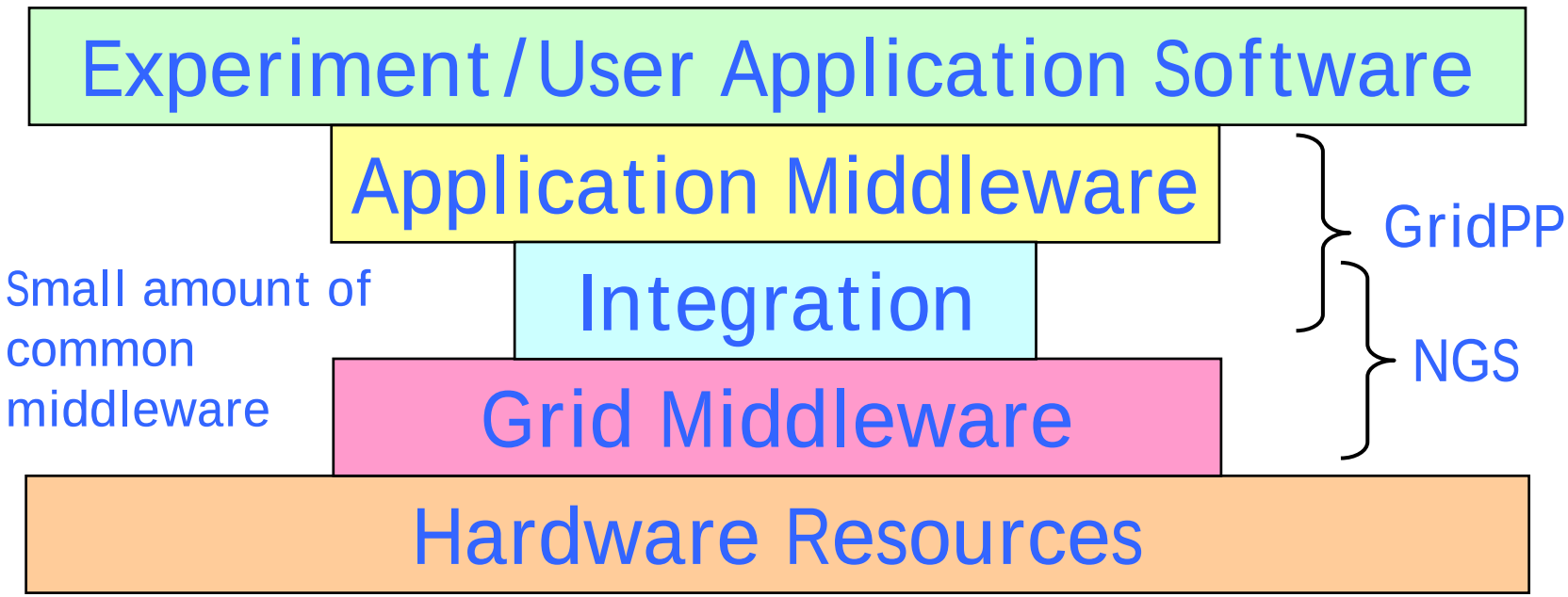
What is (gLite) Middleware?



Middleware is the Operating System of a distributed computing system

A number of different software layers

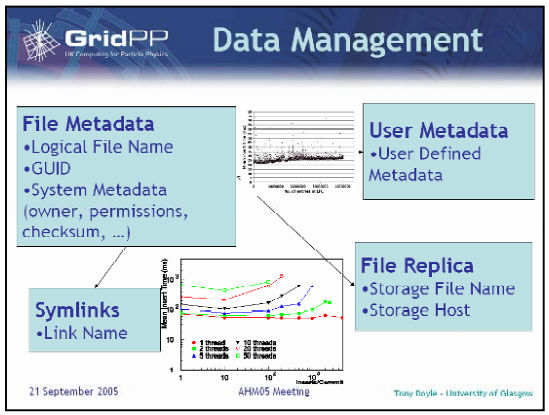
Many diverse user communities



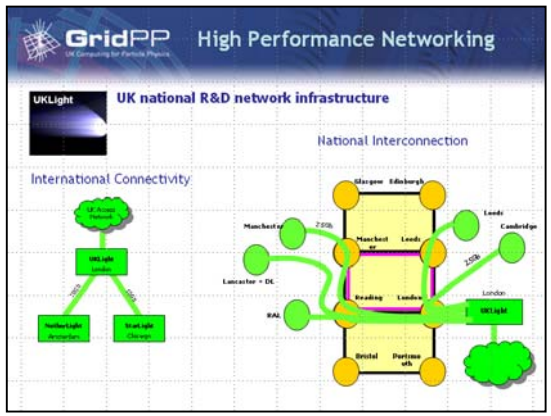
Large amount of accessible hardware

GridPP Middleware Development

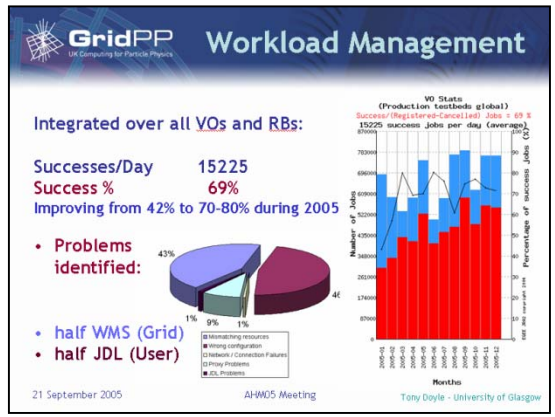
Grid Data Management



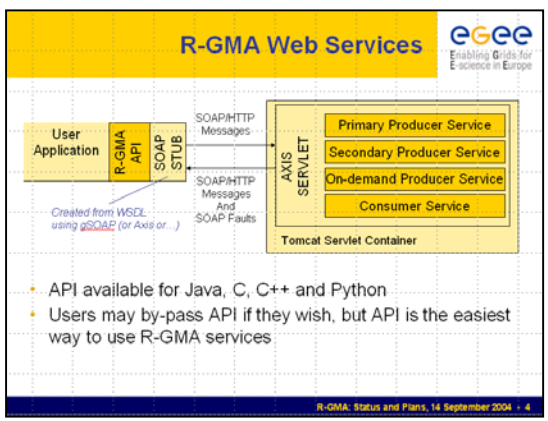
Network Monitoring



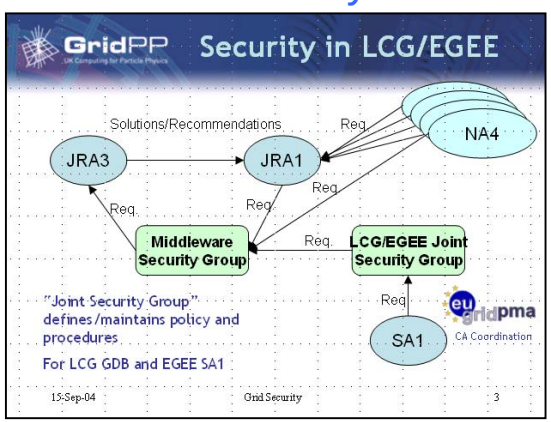
Workload Management



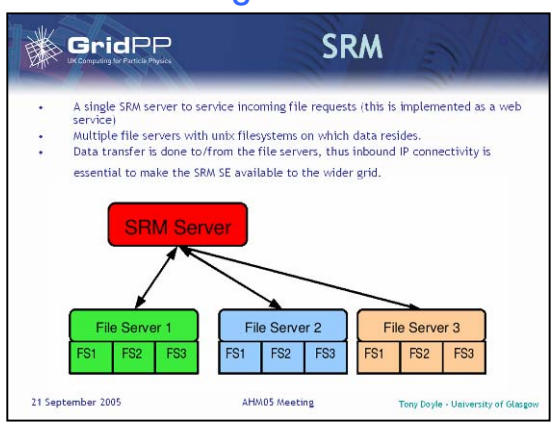
Information Services



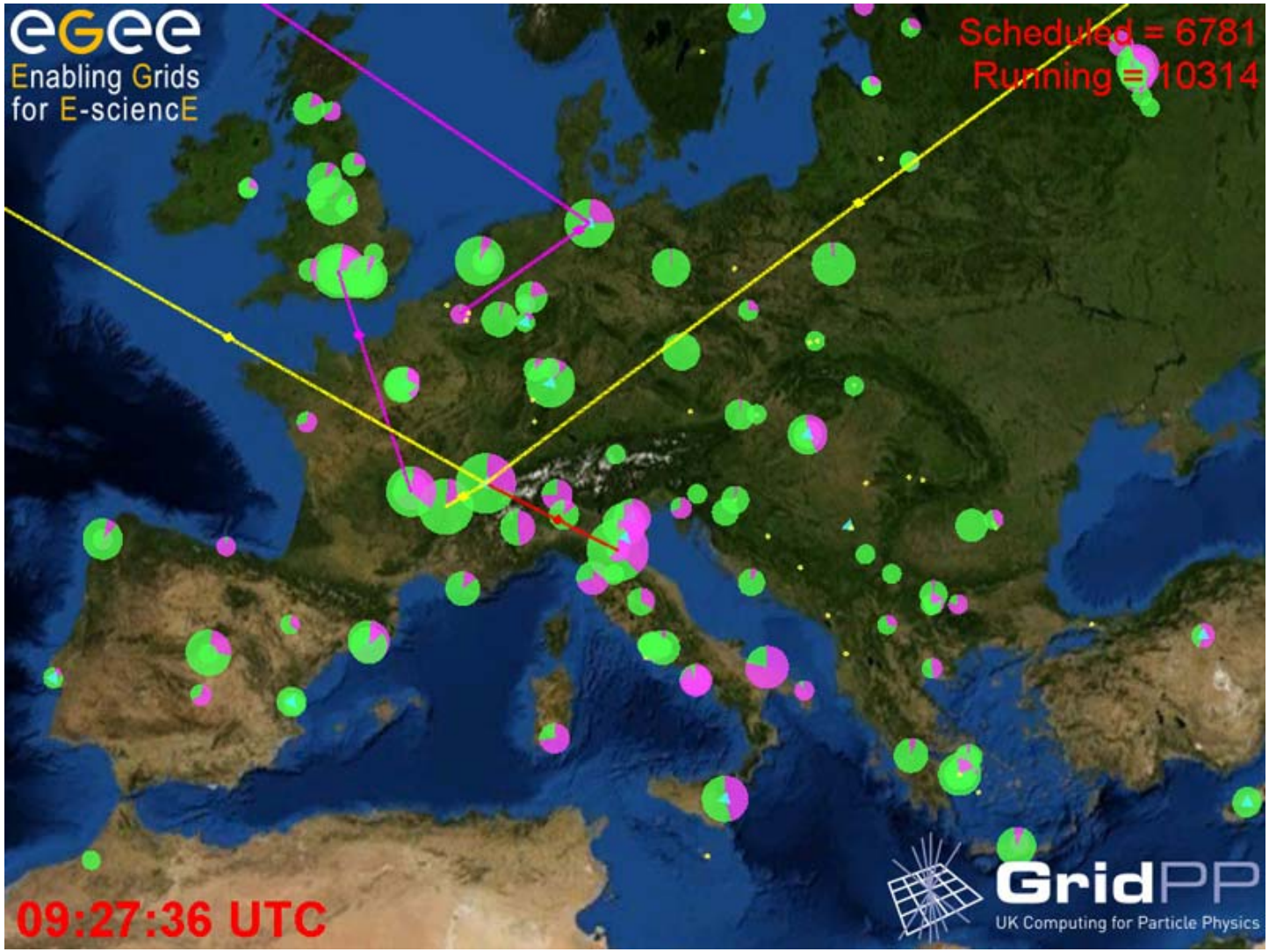
Security



Storage Interfaces



The EGEE Grid Status



Worldwide

237 Sites

50 Countries

35,716 CPUs

21.3 PB Disk

10,579 Years of CPU time

UK

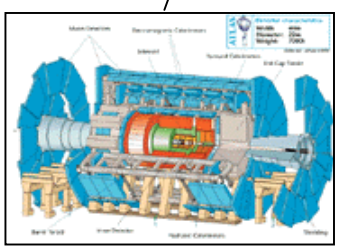
21 Sites

8089 CPUs

876 TB Disk

3,361 Years of CPU time

Tier Structure



Offline farm

Online system

Tier 0 — CERN computer centre

Tier 1 — National centres: RAL,UK; USA; ermany; Italy; rance

Tier 2 — Regional groups: ScotGrid; NorthGrid; SouthGrid; London

Institutes: Brunel; Imperial; QMUL; RHUL; UCL

Workstations

Useful model for Particle Physics but not necessary for others

UK Tier-2 Centres

ScotGrid

Durham, Edinburgh, Glasgow

NorthGrid

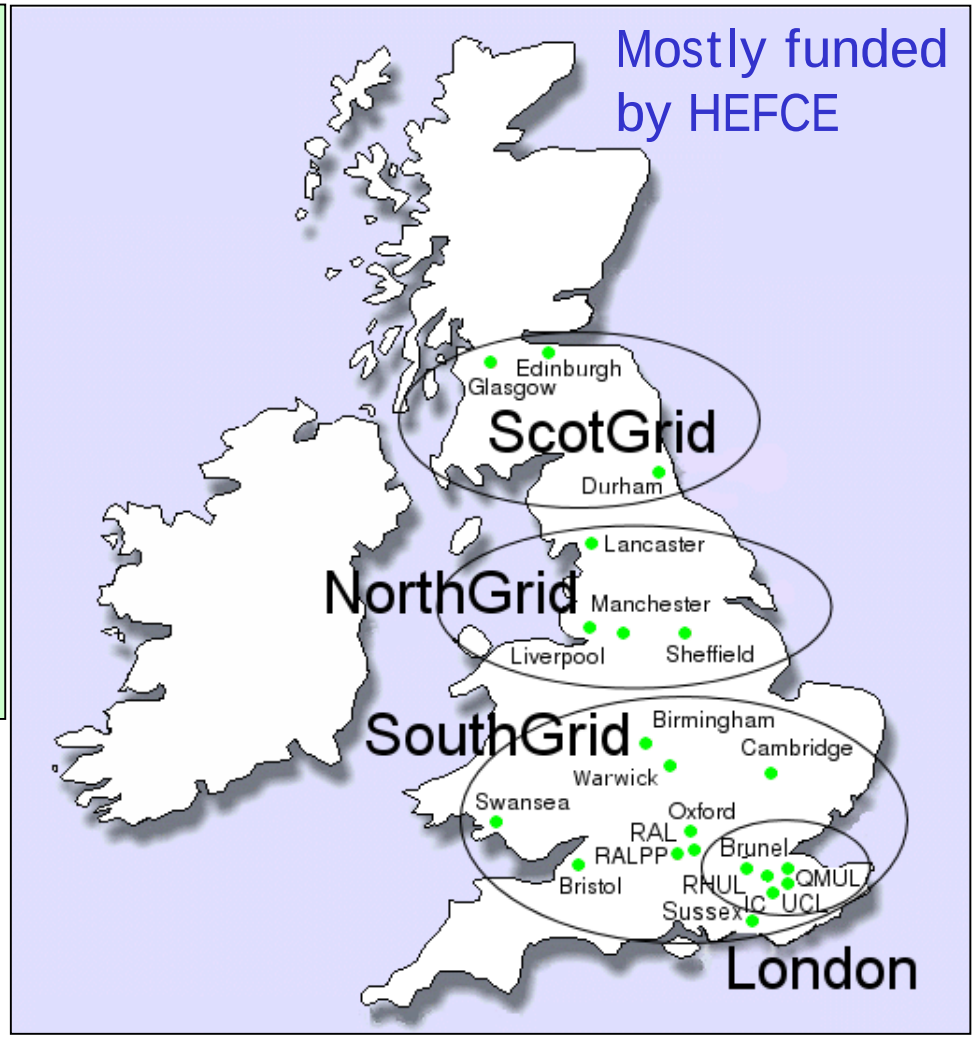
Daresbury, Lancaster, Liverpool, Manchester, Sheffield

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)

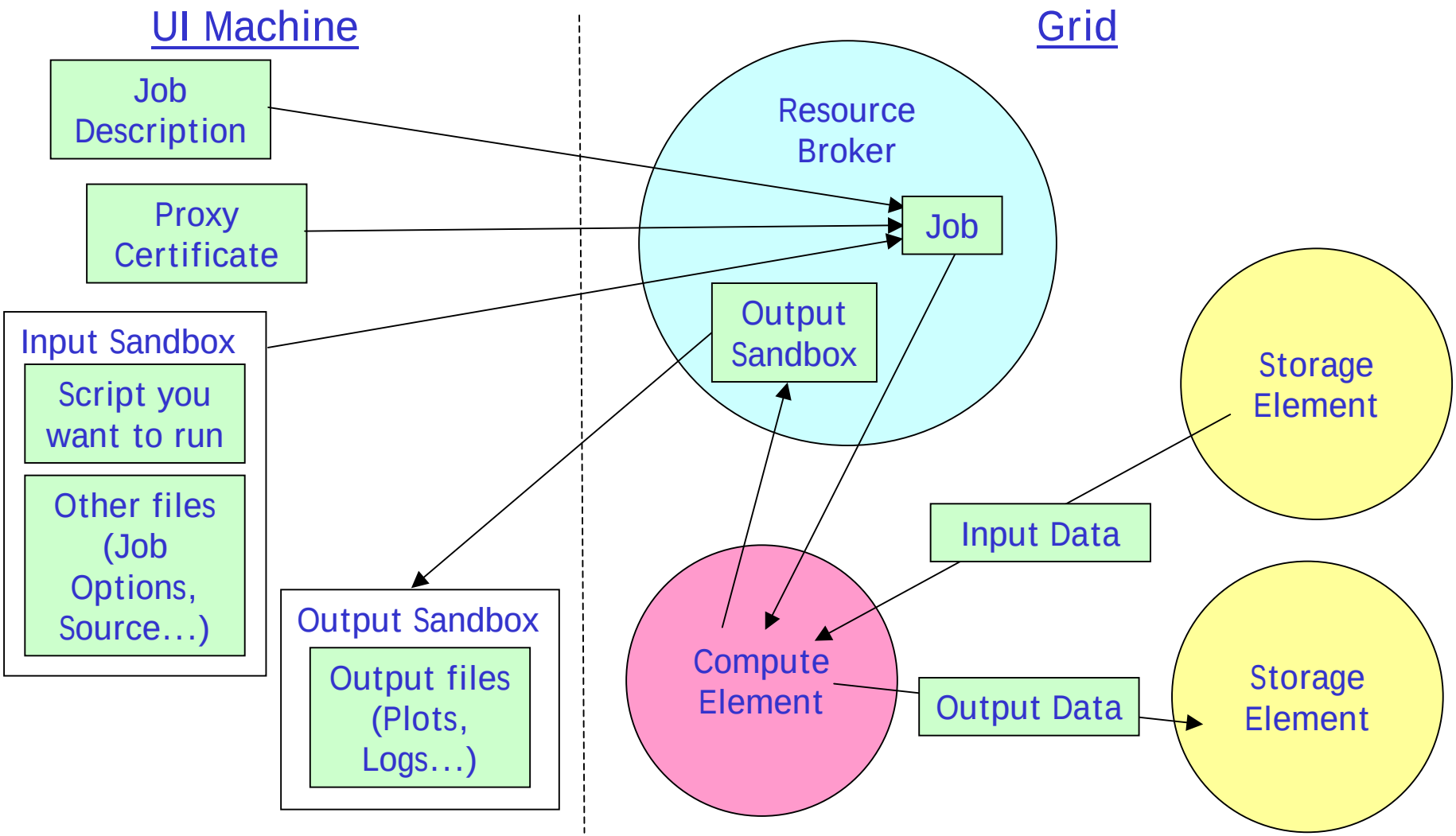
   Authorisation – what you are allowed to do

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"};
##########################################
```

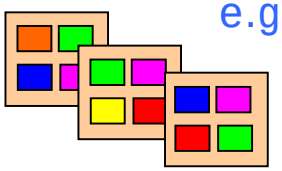
How it works


UI Machine
Grid
Job Description
Proxy Certificate
Input Sandbox
Script you want to run
Other files (Job Options, Source…)
Output Sandbox
Output files (Plots, Logs…)
Resource Broker
Job
Output Sandbox
Compute Element
Input Data
Output Data
Storage Element
Storage Element

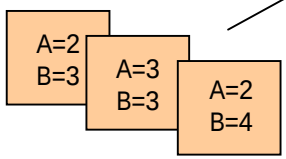
What is it good for?

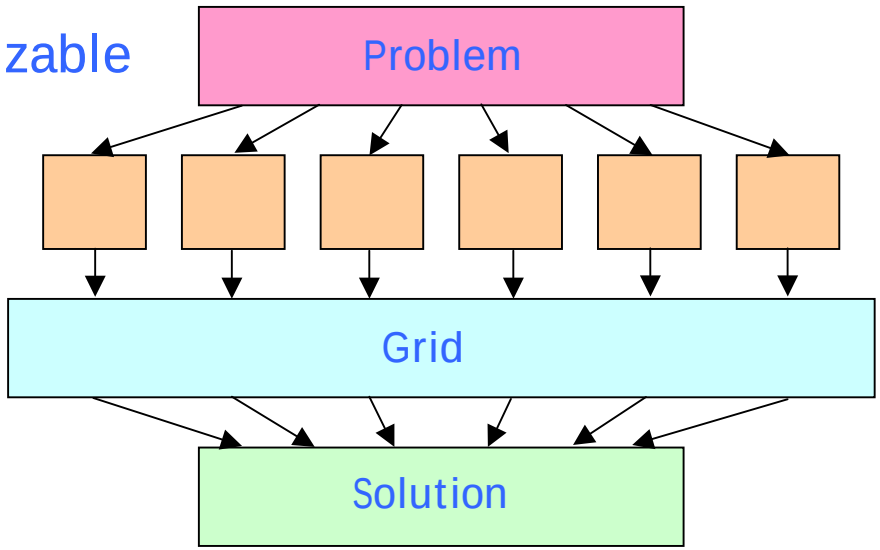
Problems that are highly parallelizable

Input data is independent
e.g. Images:

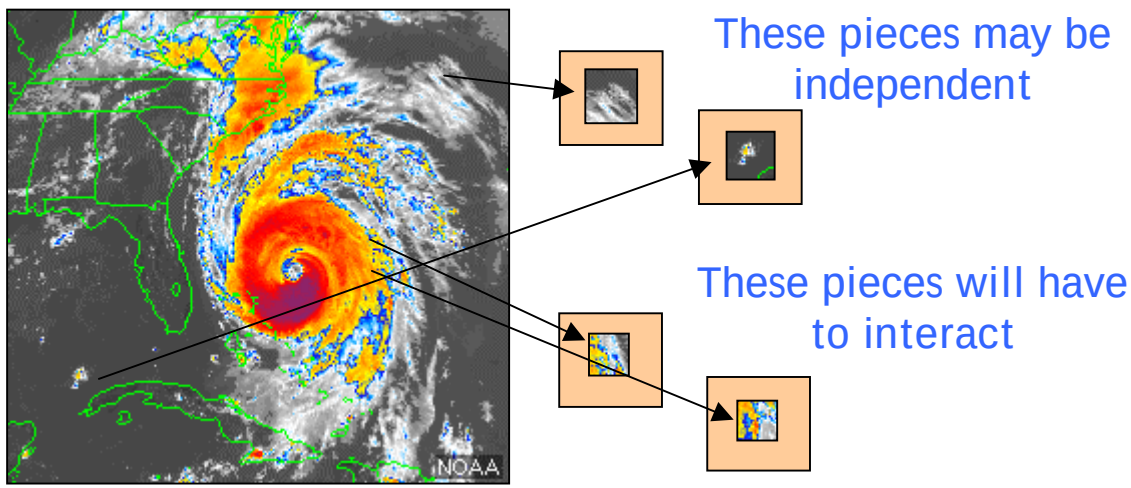


Simulation using different parameters:



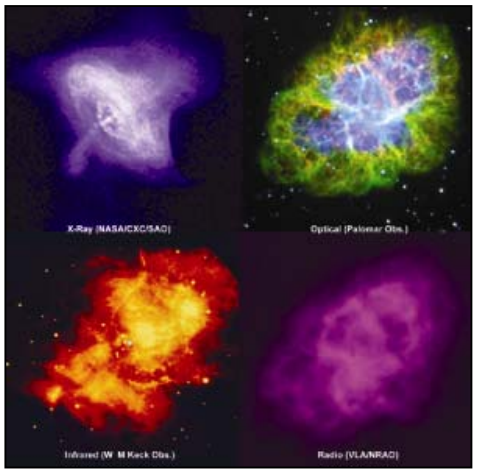


Not so good for closely coupled problems

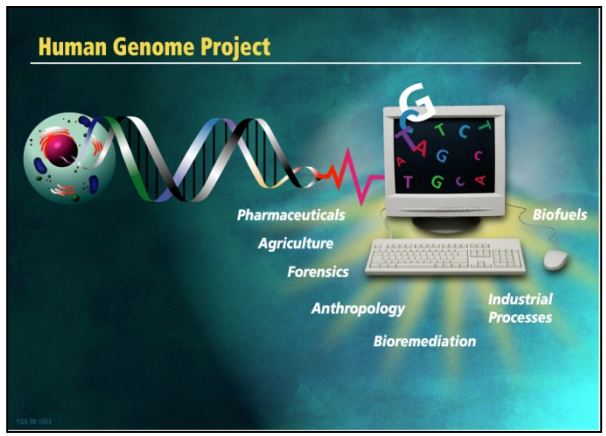


Other Uses for a Grid

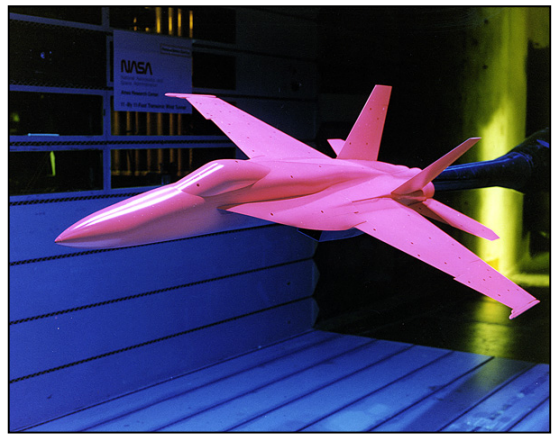
- Astronomy



- Bioinformatics



- Engineering

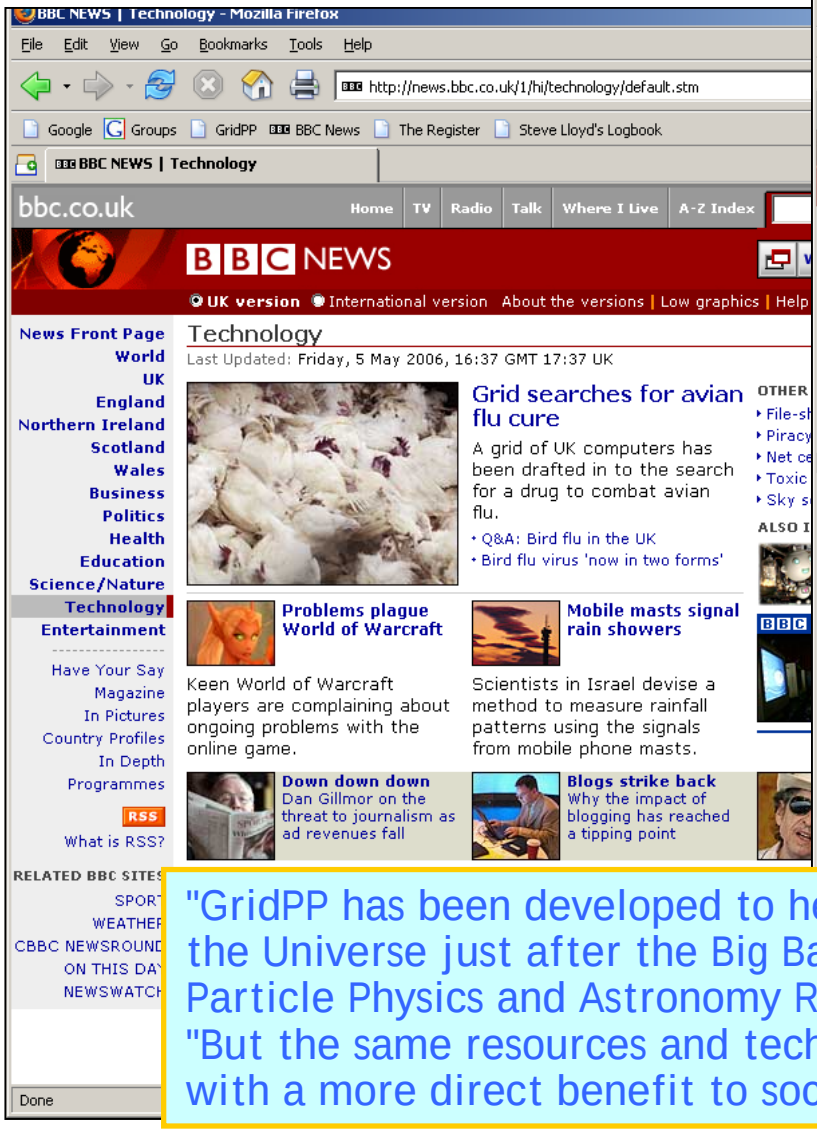


- Healthcare

- Commerce

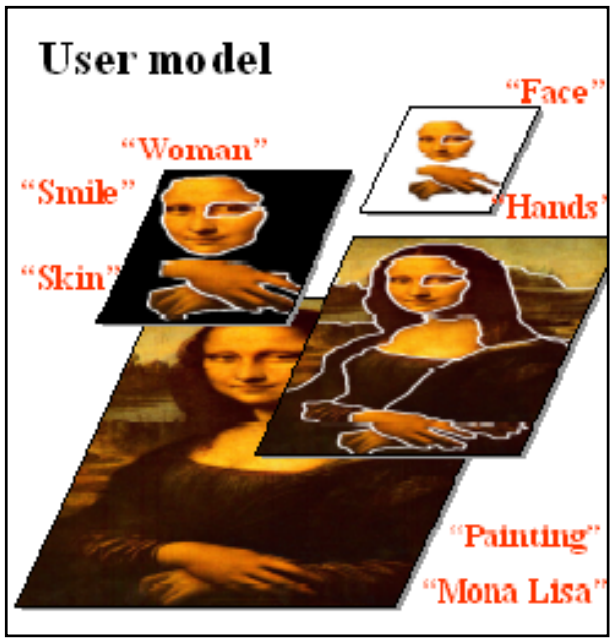
- Gaming

Avian Flu Studies



"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



Cambridge Ontology

Retrieval Requirements

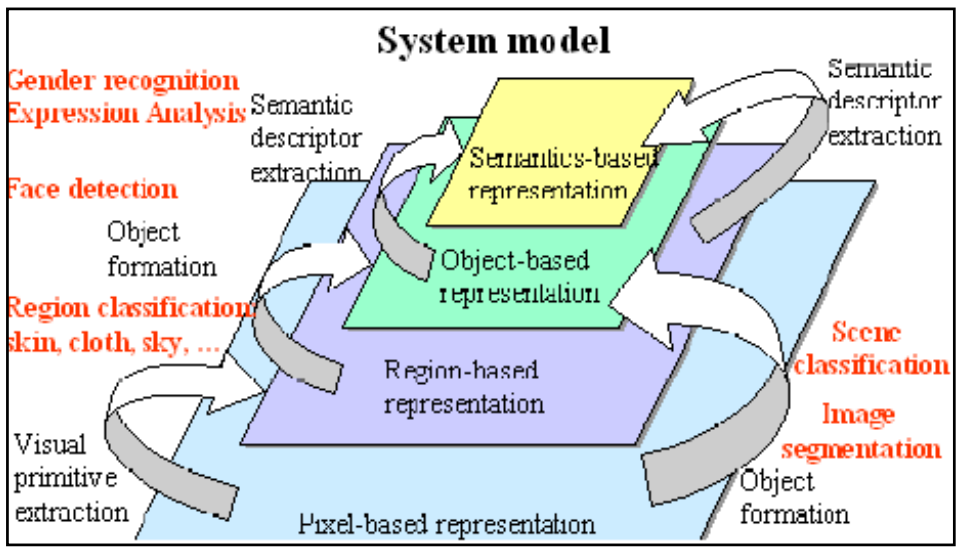
Retrieval Results

Ontological Query Language

Query Evaluation

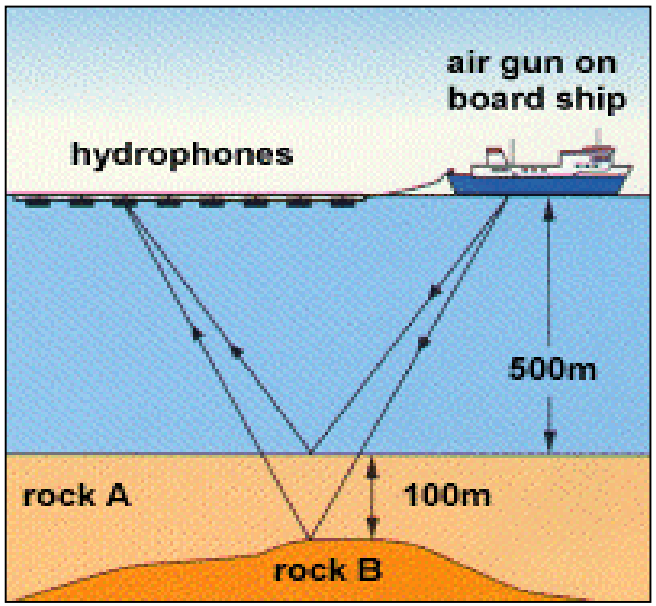
Relevance Assessment

Semantic Gap

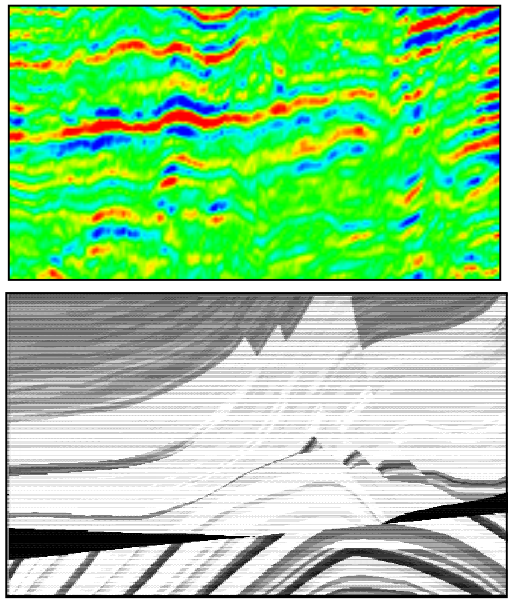


Cambridge Ontology – Startup looking at content based image retrieval. Search picture content without using meta-data or image annotations.

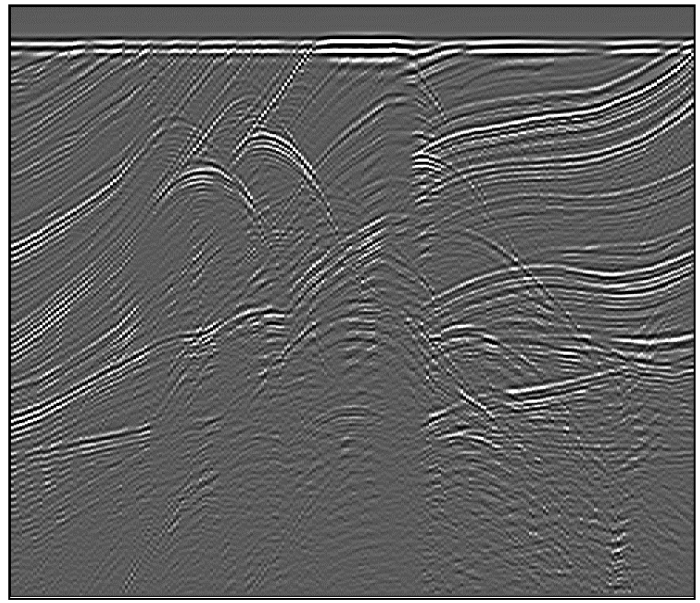
Total Exploration & Production



Marine Experiment



Results of marine experiments

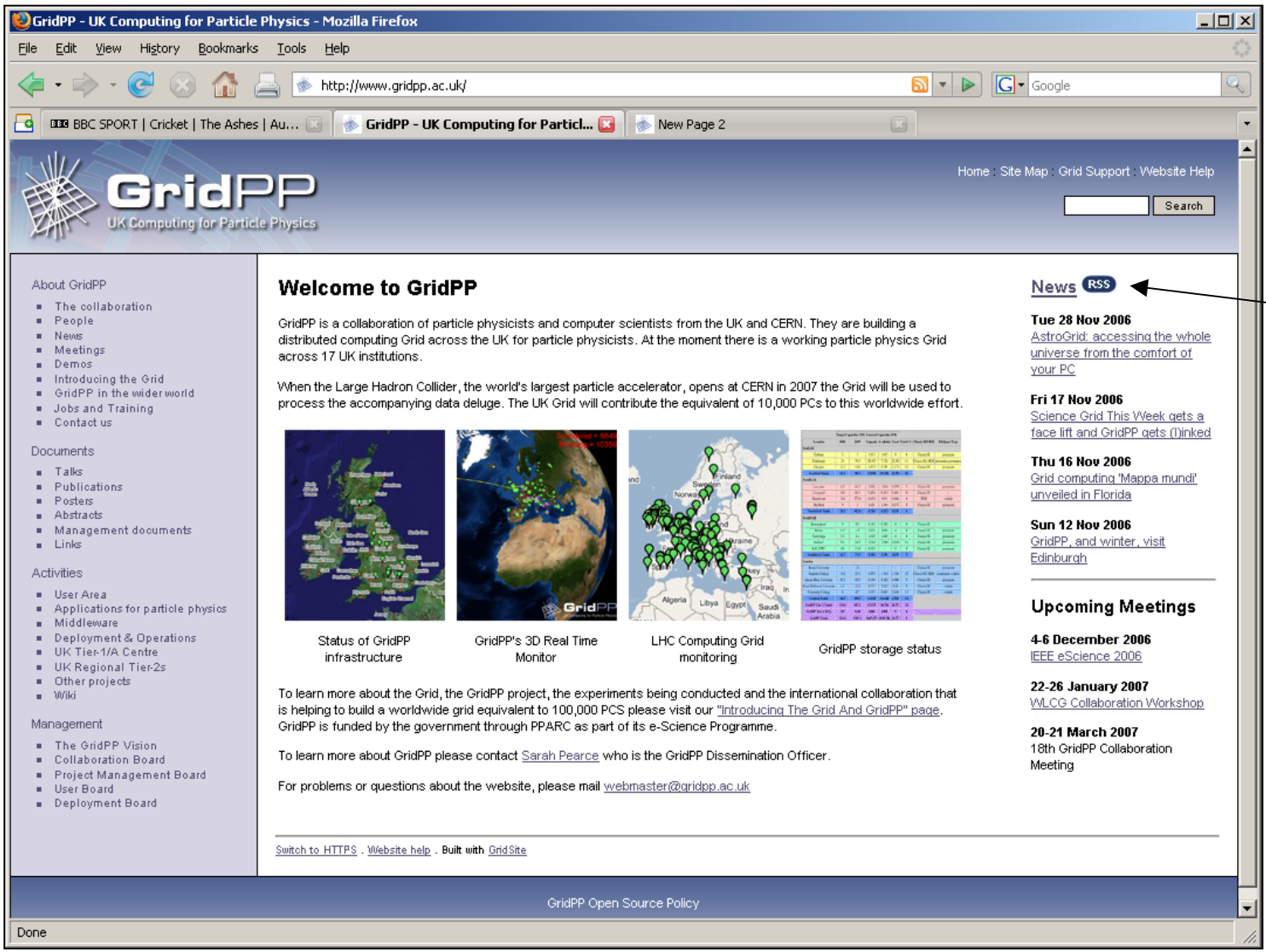


Modelled results based on bore-hole data and wave equations

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

Further Information ...

http://www.gridpp.ac.uk



RSS News feed