Grid Computing: enabling scientific collaboration in Europe and beyond

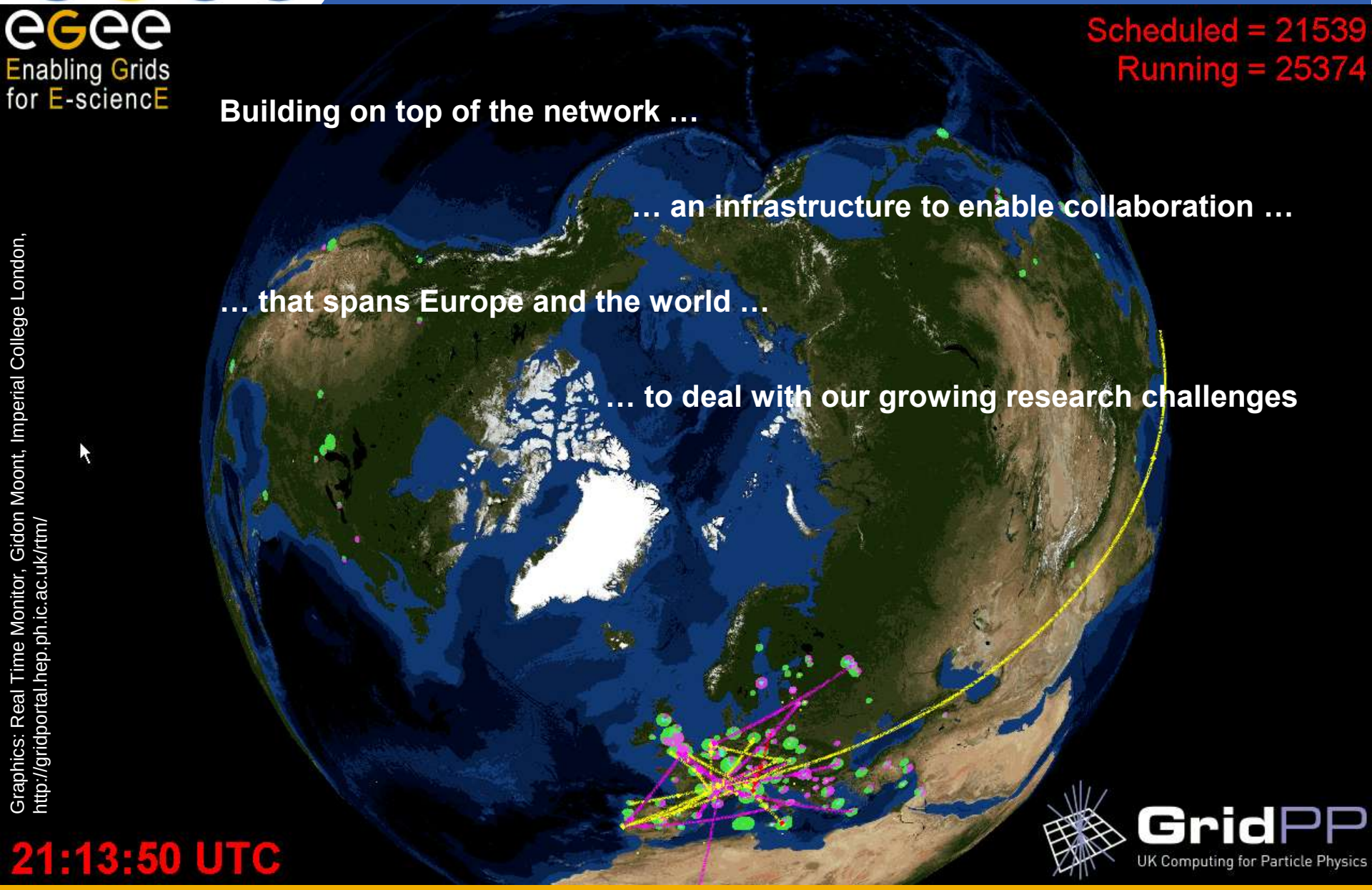
David Groep
Nikhef

Belnet Networking Conference
Brussels, 28th November 2008
partially based on Bob Jones' general EGEE presentation

www.eu-egee.org

An Infrastructure for Research

Enabling Grids for E-sciencE



GridPP
UK Computing for Particle Physics

Why do we need it?

Collected data in research and industry grows exponentially

The Bible	**5**	**MByte**
X-ray image	5	MByte/image
Functional MRI	1	GByte/day
Bio-informatics databases	500	GByte each
Refereed journal papers	1	TByte/yr
Satellite world imagery	5	TByte/yr
US LoC contents	20	TByte
Internet Archive 1996-2002	100	TByte
Particle Physics today	5	PByte/yr
LHC era physics, Astronomy, …	**20**	**PByte/yr**

And the size of the scientific collaborations grows even faster ...

with today 'Web 2.0' techniques penetrating research, like *MyExperiment*

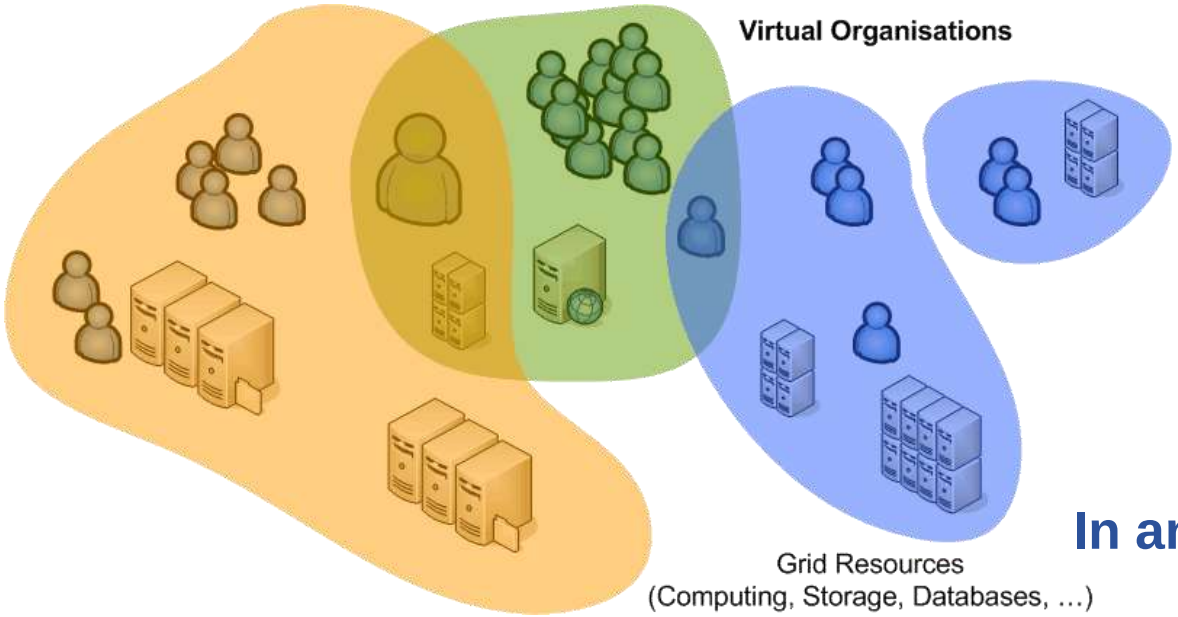
What is Grid?

Cycle scavenging
- harvest idle compute power
- improve RoI on desktops

Cluster computing and storage
- What-if scenarios
- Physics event analysis
- Improve Data Centre Utilization

About interlinking global communities

- more than one organisation
- more than one application
- more than one …



Based on
- open protocols
- collective service

In an usable, persistent manner

Grid Infrastructure

To bring about ICT-enhanced research – and sustain it – requires a *persistent infrastructure,* based on standards

Hardware infrastructure

Networks, clusters, supercomputers, databases, mass storage, visualisation, ...

Trust and 'federated' infrastructure

authentication, authorization, accounting, billing and settlement, policy agreements

Software infrastructure

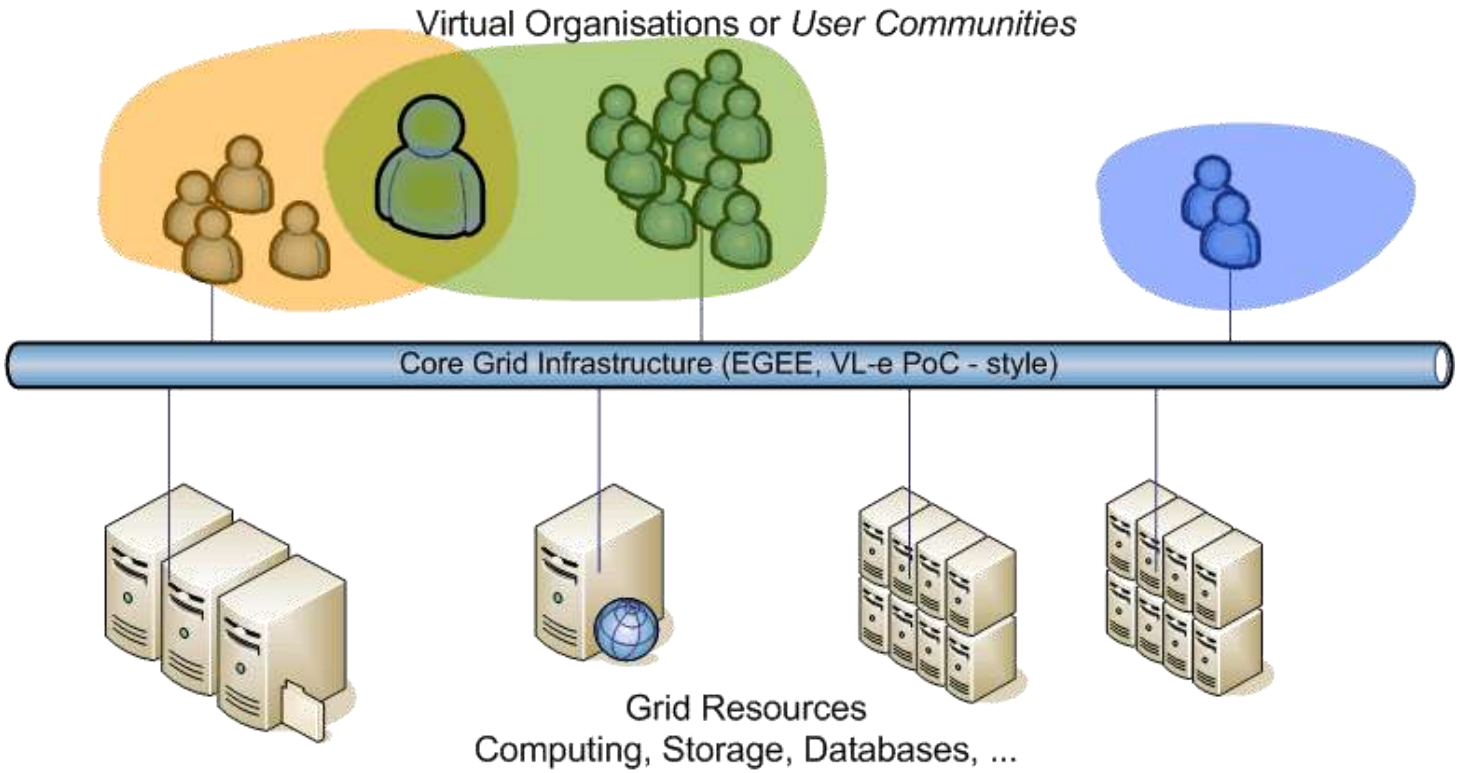
execution services, workflow, resource information systems, database access, storage management, meta-data

Application infrastructure

user support, training, integration in domain specific software, ...

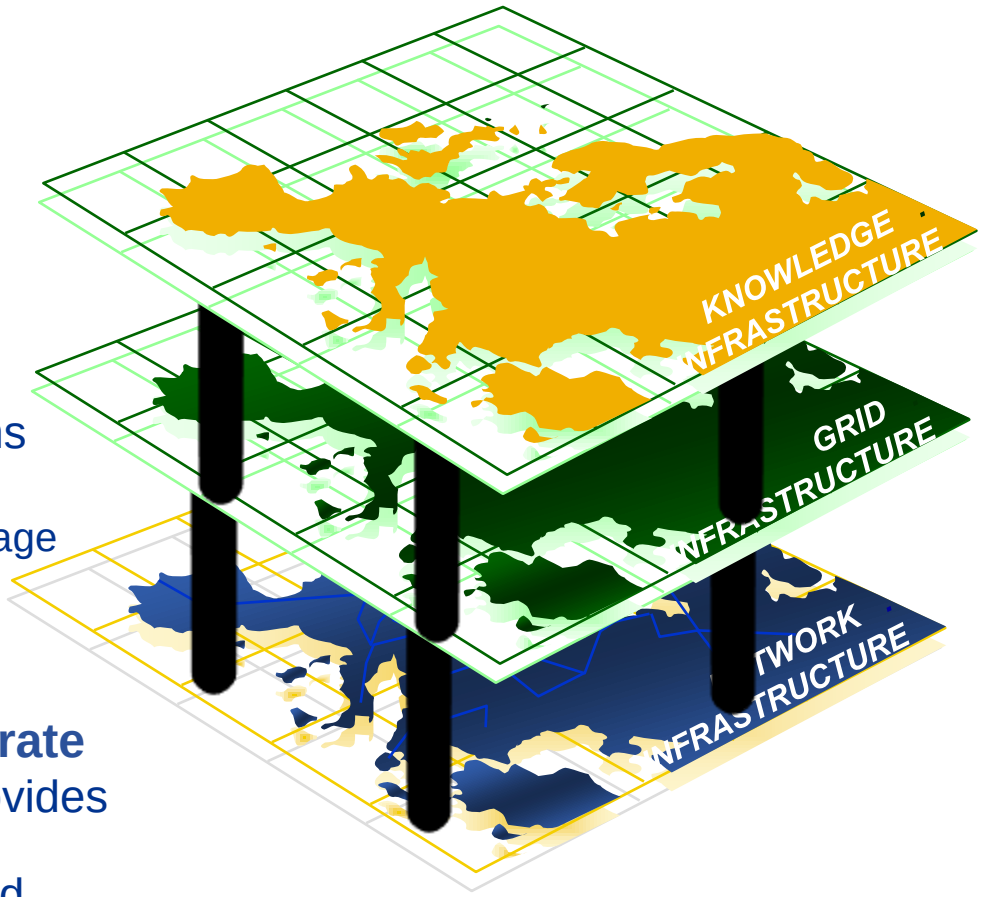
A Grid Network

- Europe, through its national grids and coordination, is building a **multi-disciplinary** grid for research
- Based on the high-bandwidth network
- As a **sustainable**, 'always-on', **standards-based** service



How e-Infrastructures help e-Science

- **e-Infrastructures provide easier access for**
  - Small research groups
  - Scientists from many different fields
  - Remote and still developing countries
- **… to new technologies**
  - Produce, store and search massive amounts of data
  - Transparent access to millions of files across different administrative domains
  - Low cost access to resources
    - Mobilise large amounts of CPU & storage on short notice (PC clusters)
  - High-end facilities (supercomputers)
- **And help to find new ways to collaborate**
  - Eases distributed collaborations & provides new ways of community building
  - Develops applications using distributed complex workflows
  - Gives easier access to higher education

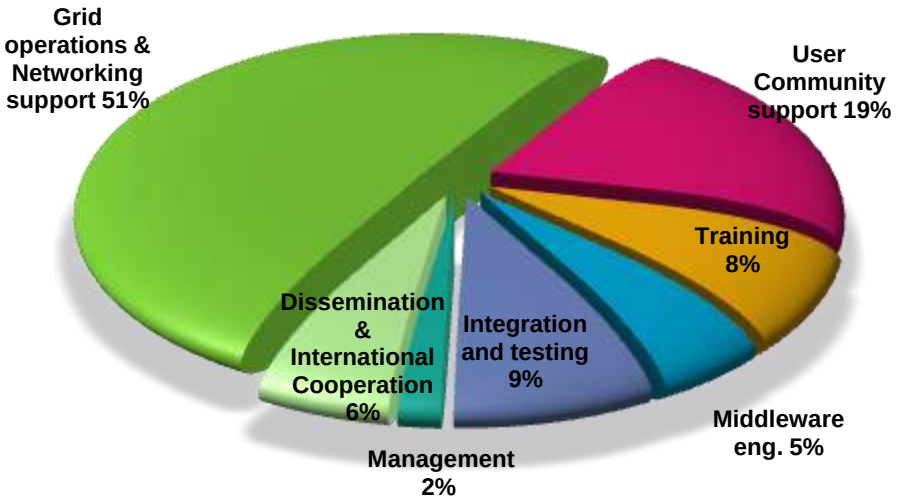


EGEE-III

Flagship Grid infrastructure project co-funded by the European Commission

Main Objectives

- Expand/optimise existing EGEE infrastructure, include more resources and user communities
- Prepare migration from a project-based model to a sustainable federated infrastructure based on National Grid Initiatives



Duration: 2 years
Consortium: ~140 organisations across 33 countries
EC co-funding: 32Million €

EGEE – What do we deliver?

- **Infrastructure operation**
  - **Sites distributed across many countries**
    - **Large quantity of CPUs and storage**
    - **Continuous monitoring of grid services & automated site configuration/management**
    - **Support multiple Virtual Organisations from diverse research disciplines**

- **Middleware**
  - **Production quality middleware distributed under business friendly open source licence**
    - Implements a service-oriented architecture that virtualises resources
    - Adheres to recommendations on web service inter-operability and evolving towards emerging standards

gLite Lightweight Middleware for Grid Computing

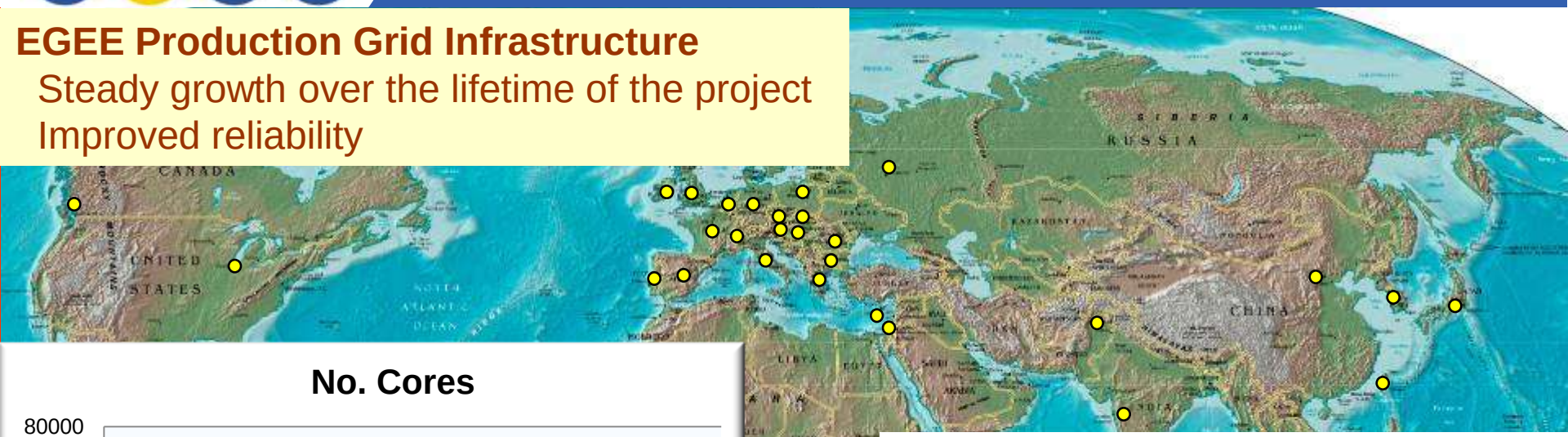
- **User Support - *Managed process from first contact through to production usage***
  - **Training**
  - **Expertise in grid-enabling applications**
  - **Online helpdesk**
  - **Networking events (User Forum, Conferences etc.)**

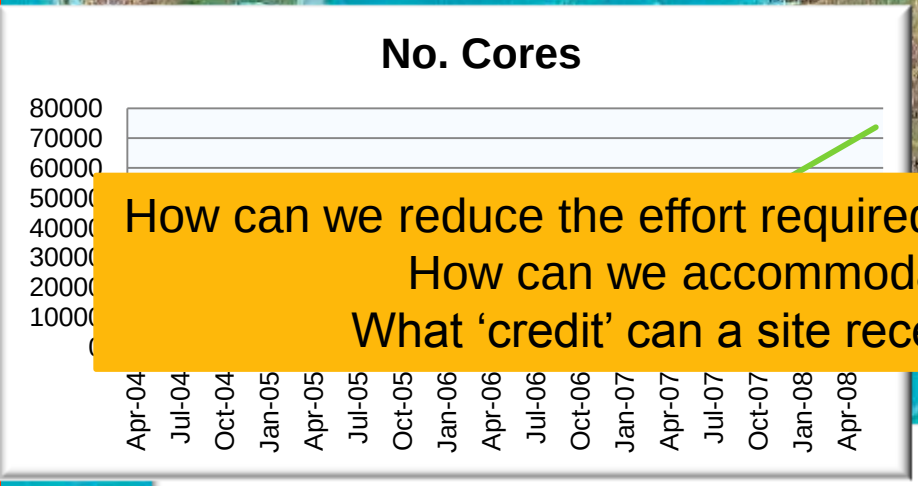
GGUS Global Grid User Support

EGEE Achievements - Infrastructure

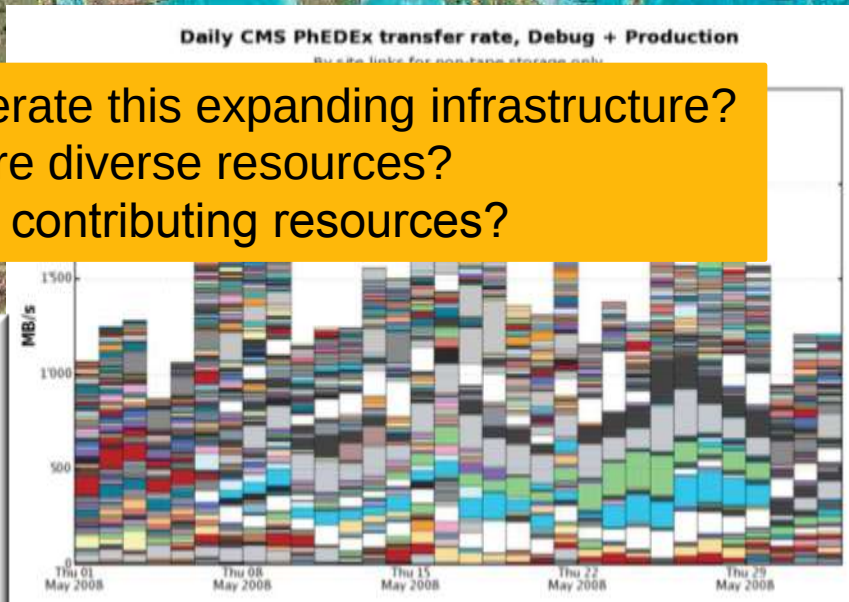
EGEE Production Grid Infrastructure

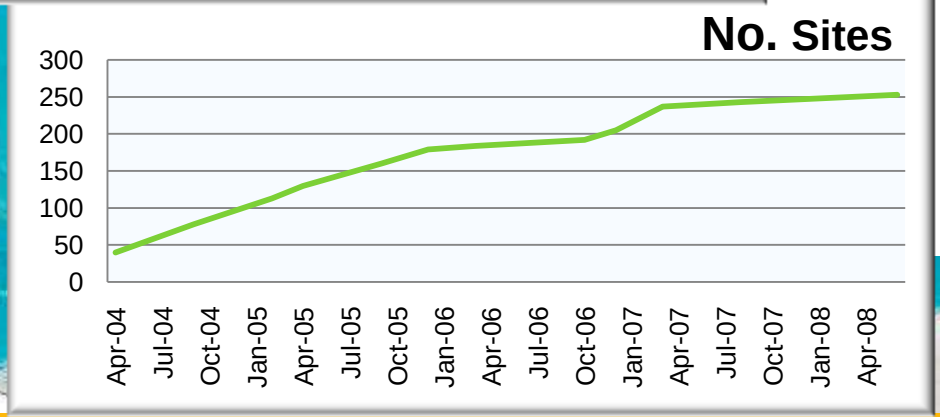
- Steady growth over the lifetime of the project
- Improved reliability



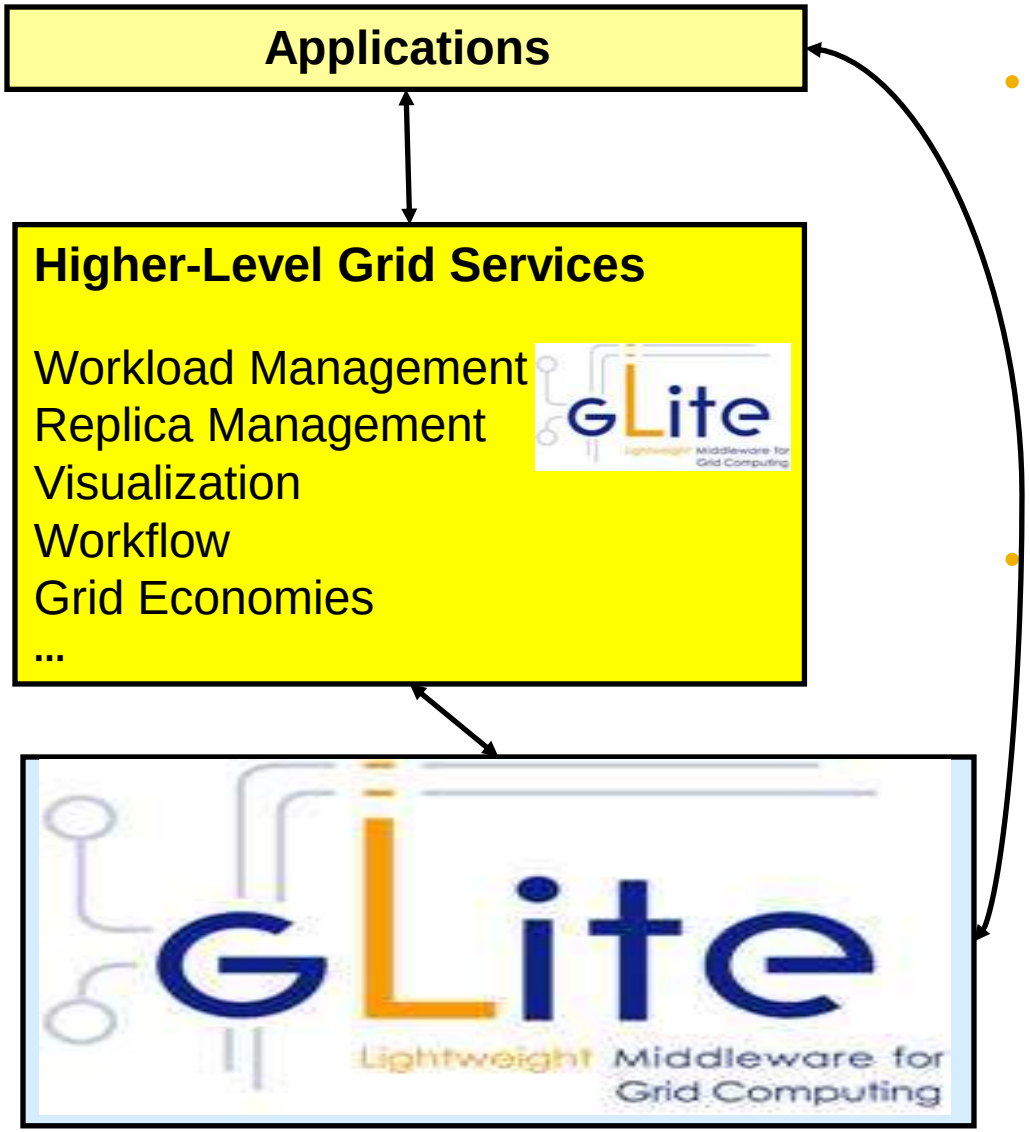


How can we reduce the effort required to operate this expanding infrastructure?
How can we accommodate more diverse resources?
What 'credit' can a site receive for contributing resources?





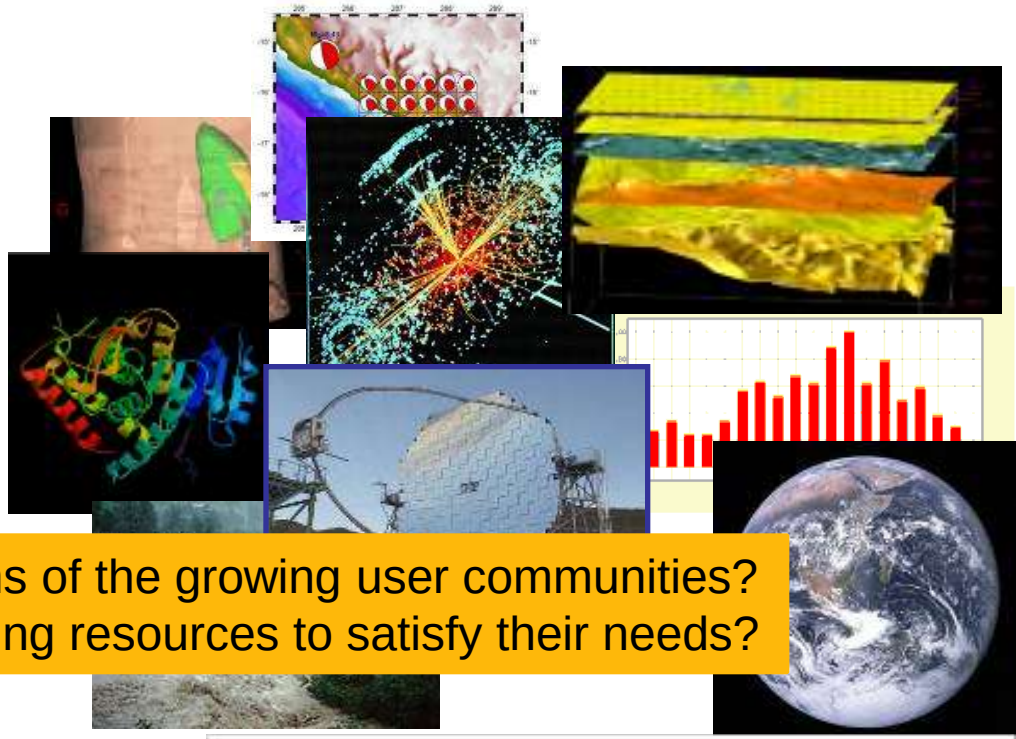
Grid Middleware



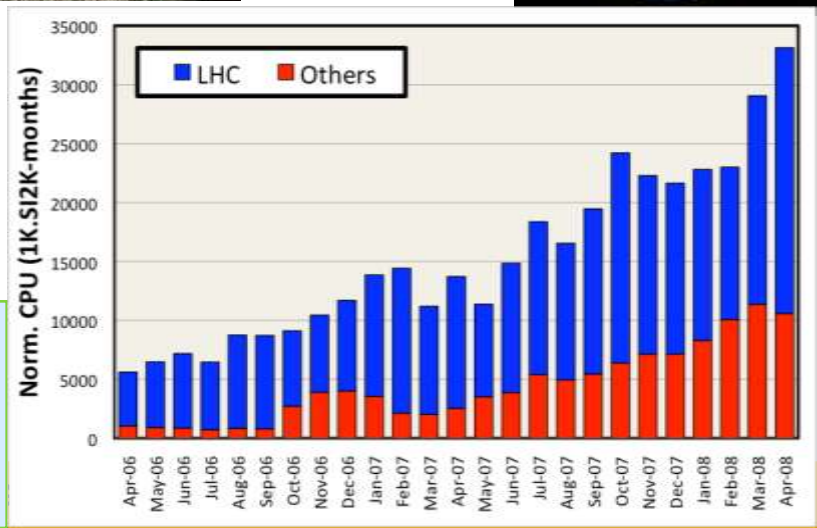
- **Applications access both Higher-level Grid Services and Foundation Grid Middleware**
  - Application code
  - Frameworks
  - Community Portals
- **VOs complement gLite with other high-level services via the RESPECT programme**
  - Rec. External Software Pkgs. for the EGEE Community
  - Identify useful, 3rd-party software that works with gLite
  - Make users aware of that software to avoid duplicated efforts

EGEE Achievements - Applications

- **>270 VOs from several scientific domains**
  - Astronomy & Astrophysics
  - Civil Protection
  - Computational Chemistry
  - Comp. Fluid Dynamics
  - Computer Science/Tools
  - Condensed Matter Physics
  - High Energy Physics
  - Life Sciences
- **Further applications under evaluation**

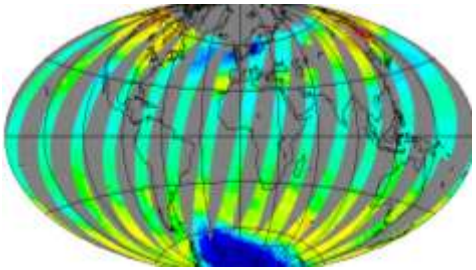


How do we match the expectations of the growing user communities?
Will we have enough computing resources to satisfy their needs?

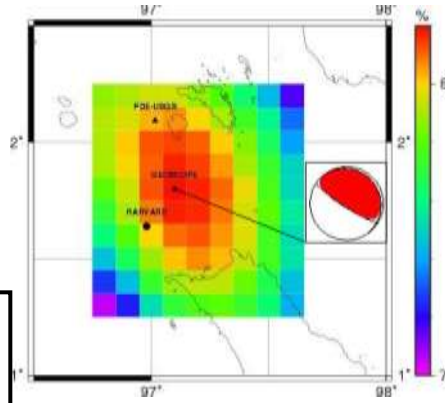


Applications have moved from testing to routine and daily usage
~80-95% efficiency

Earth Science



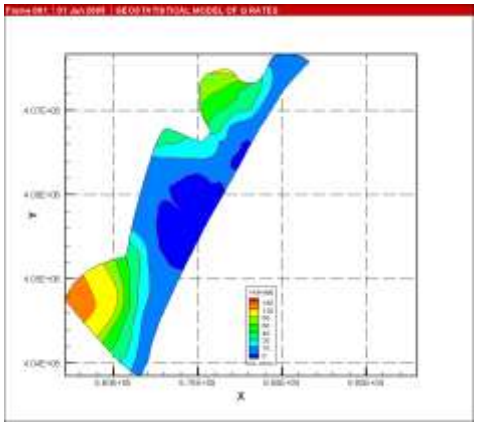
ESA, UTV(IT), KNMI(NL), IPSL(FR)- Production and validation of 7 years of Ozone profiles from GOME



Rapid Earthquake analysis (mechanism and epicenter)
50- 100CPUs
IPGP(FR)

Geocluster for Academy and industry CGG(FR)

Flood of a Danube river- Cascade of models (meteorology,hydraulic ,hydrodynamic….) UISAV(SK)



Modelling seawater intrusion in costal aquifer (SWIMED) CRS4(IT),INAT(TU), Univ.Neuchâtel(CH)

Data mining Meteorology & Space Weather (GCRAS, RU)

Air Pollution model- BAS(BG)

DKRZ(DE)- Data access studies, climate impacts on agriculture

Specfem3D: Seismic application. Benchmark for MPI (2 to 2000 CPUs) (IPGP,FR)

Mars atmosphere CETP (FR)

Astronomy & Astrophysics

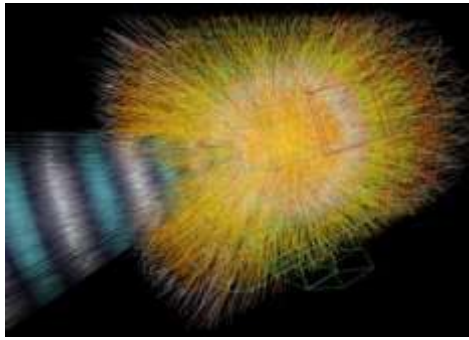
LOFAR large distributed radio telescope

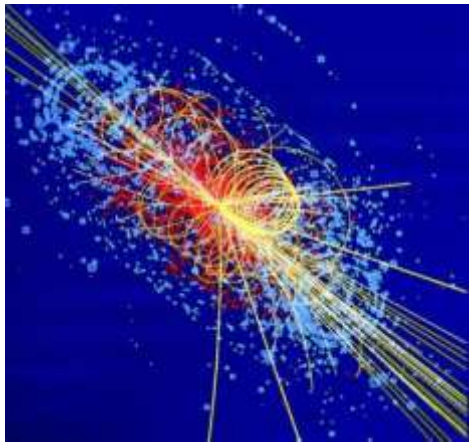
AUGER & ARGO Cosmic Ray Observatories

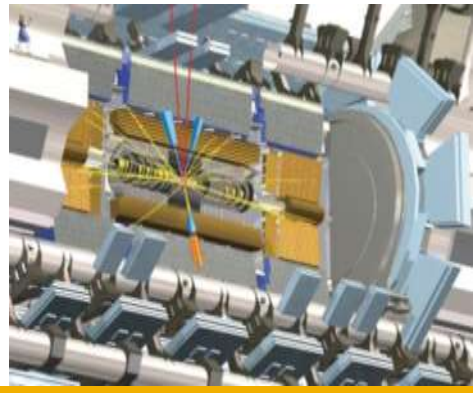
CERN's Large Hadron Collider

The LHC Computing Challenge

- **The scale and complexity of the data**
  - ➔ **15 PetaBytes of new data each year**
- **The computing capacity to support 7,000 researchers all actively analysing the data**
  - ➔ **60’000 of (today's) fastest CPUs**
- **The way in which the data is accessed will depend on the physics that emerges**

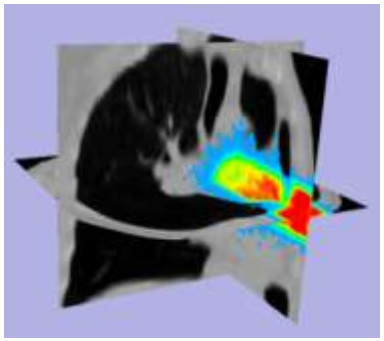






Medical Imaging - ThIS (Therapeutic Irradiation Simulator)

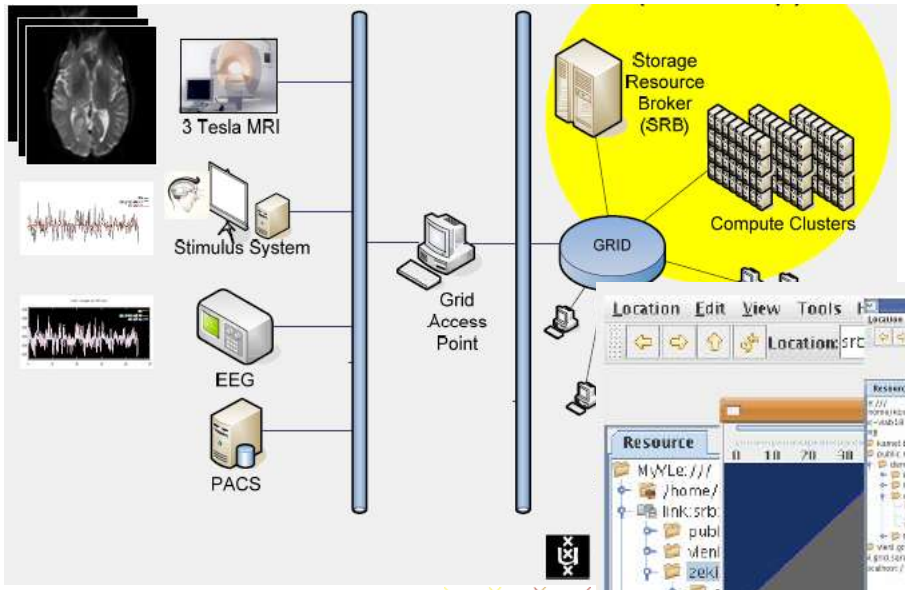
- Monte-Carlo simulation of irradiations of living tissues with photons, protons or light ions beams for cancer therapy



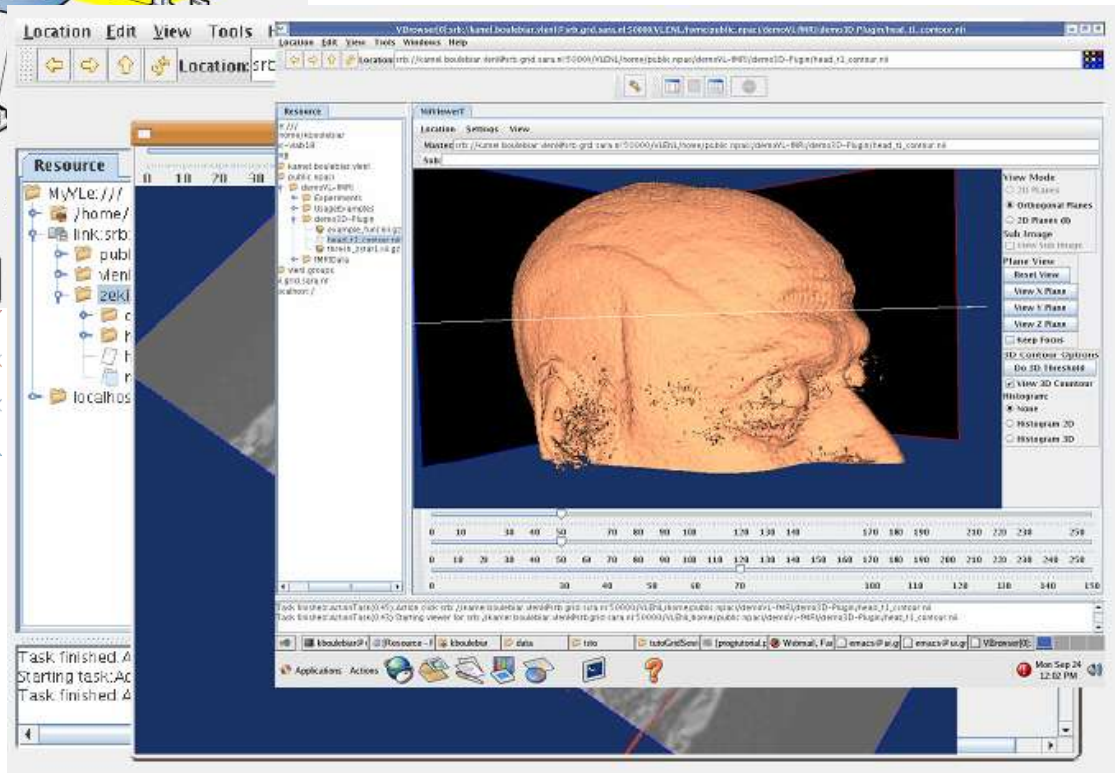
Bioinformatics - Grid Protein Sequence Analysis

Functional MRI analysis

VL-e Medical Applications on the EGEE Infrastructure



Storage of fMRI research data for sharing between groups and processing of image alignments



Research work by:

Silvia Olabarriaga (AMC, UvA)

Tristan Glatard (IvI,UvA)

Abdullah Ozsoy (IvI,UvA)

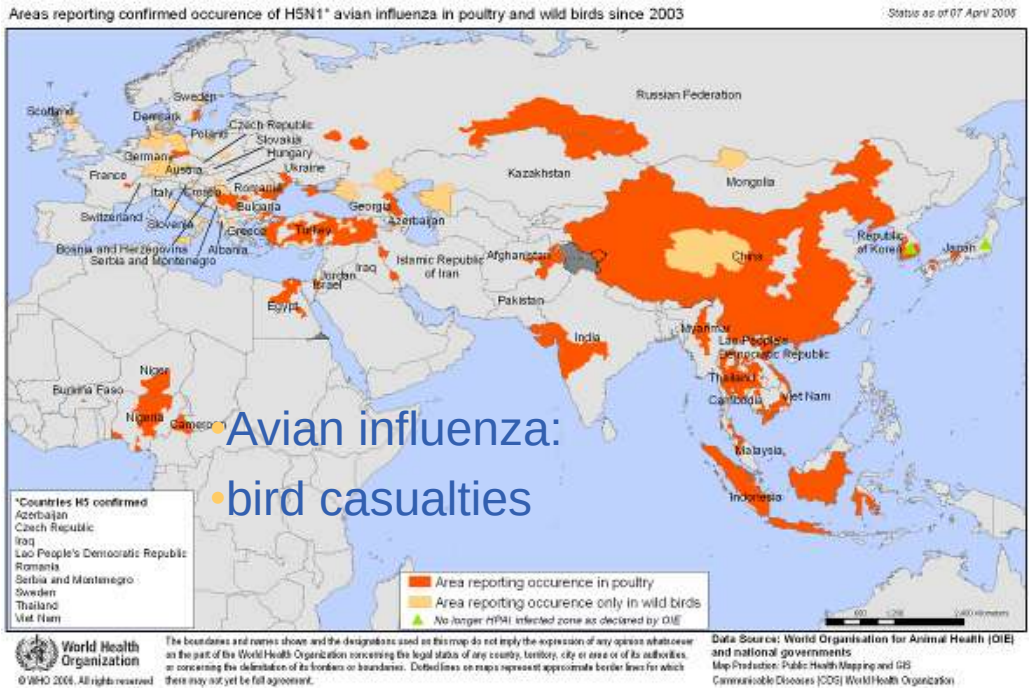
In silico drug discovery

- **Diseases such as HIV/AIDS, SRAS, Bird Flu, Malaria etc. are a threat to public health due to world wide exchanges and circulation of persons**

- **Grids open new perspectives to *in silico* drug discovery**
  - Reduced cost and adding an accelerating factor in the search for new drugs

International collaboration is required for:

- Early detection
- Epidemiological watch
- Prevention
- Search for new drugs
- Search for vaccines

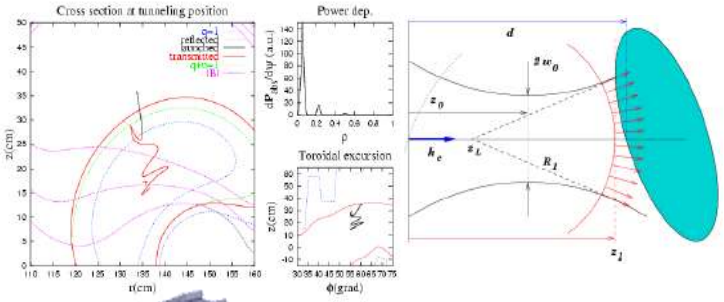


- Avian influenza:
- bird casualties

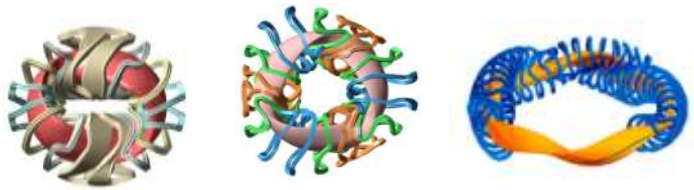
Fusion

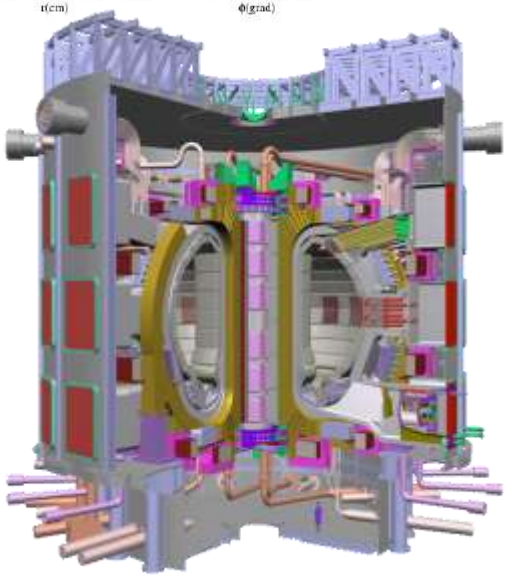
Commercial exploitation of fusion energy still needs to solve several outstanding problems requiring exceptional computing facilities including supercomputers and cluster-based grids

- Ion Kinetic Transport
- Massive Ray Tracing
- Stellarator Optimization



Interworking course-grained clusters and MPP systems across both the EGEE and DEISA grids





Collaborating e-Infrastructures

Enabling Grids for E-sciencE

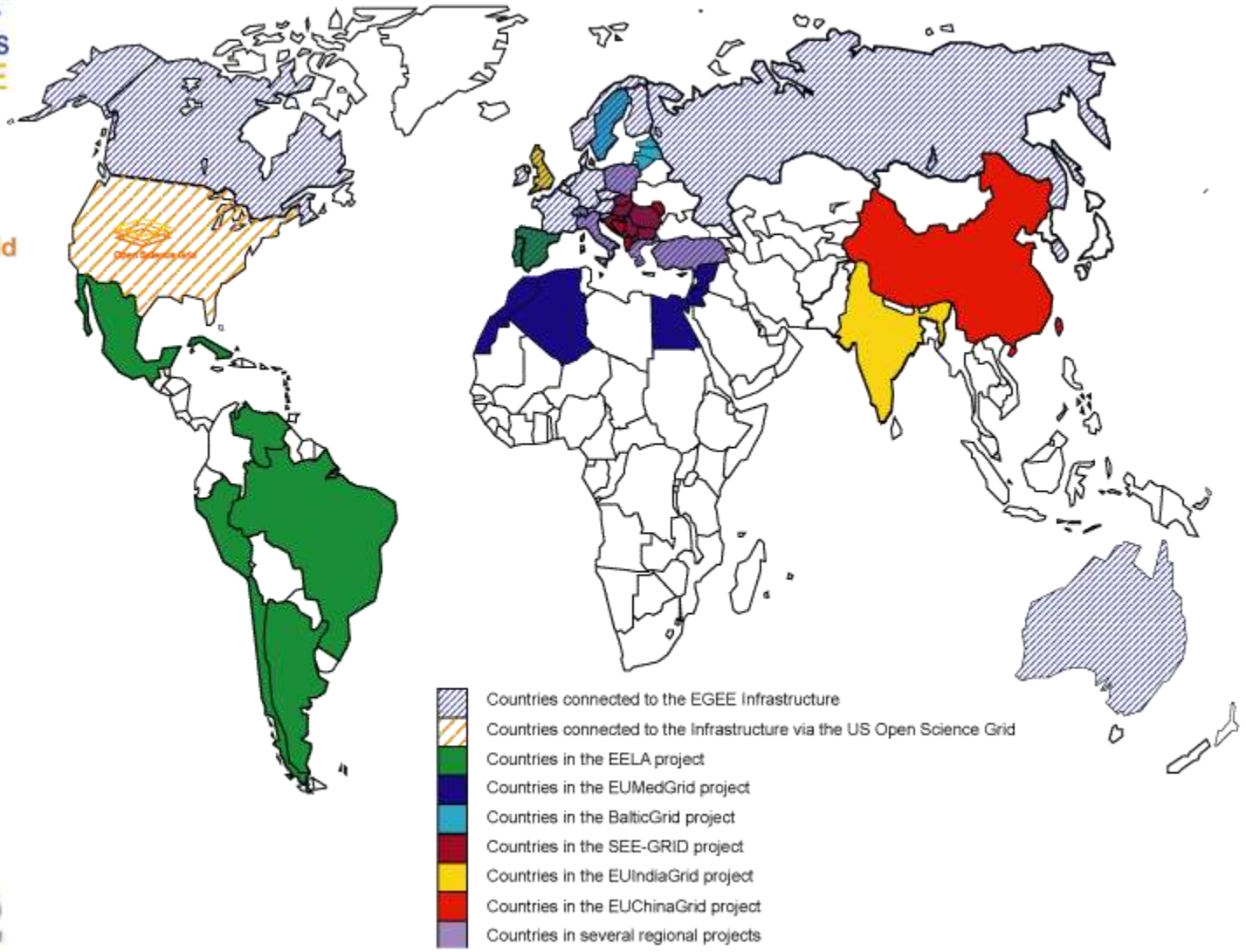
Open Science Grid

TeraGrid

Distributed European Infrastructure for Supercomputing Applications

NAREGI

GÉANT2



Global Trust

Why would I trust you? How do I know who you are?

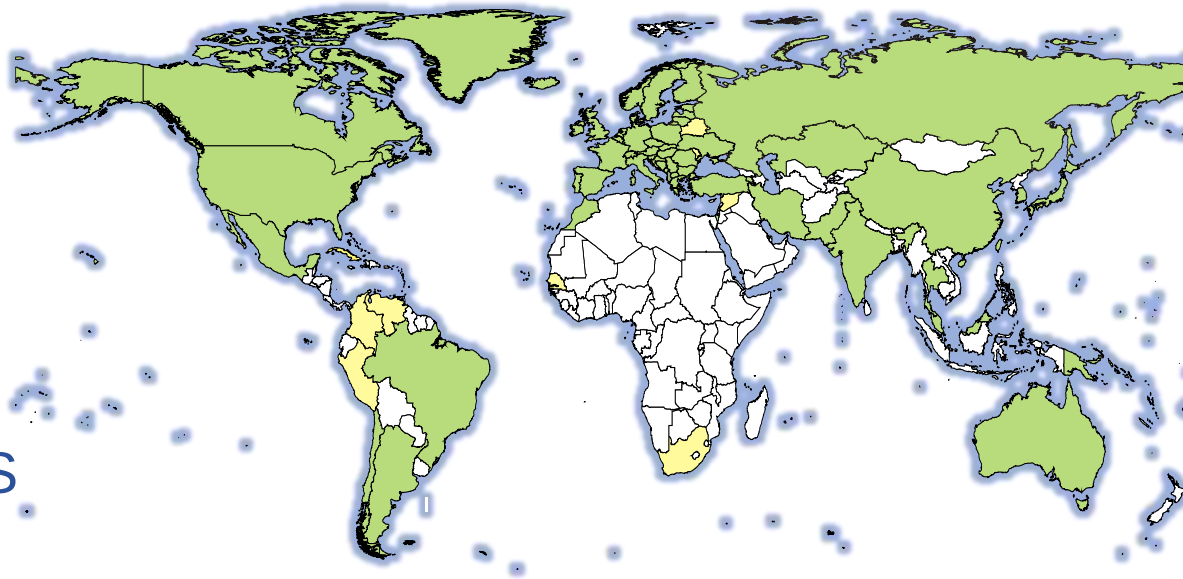
IGTF International Grid Trust Federation AP EU TAG

'digital signatures and certificates used as digital identities'

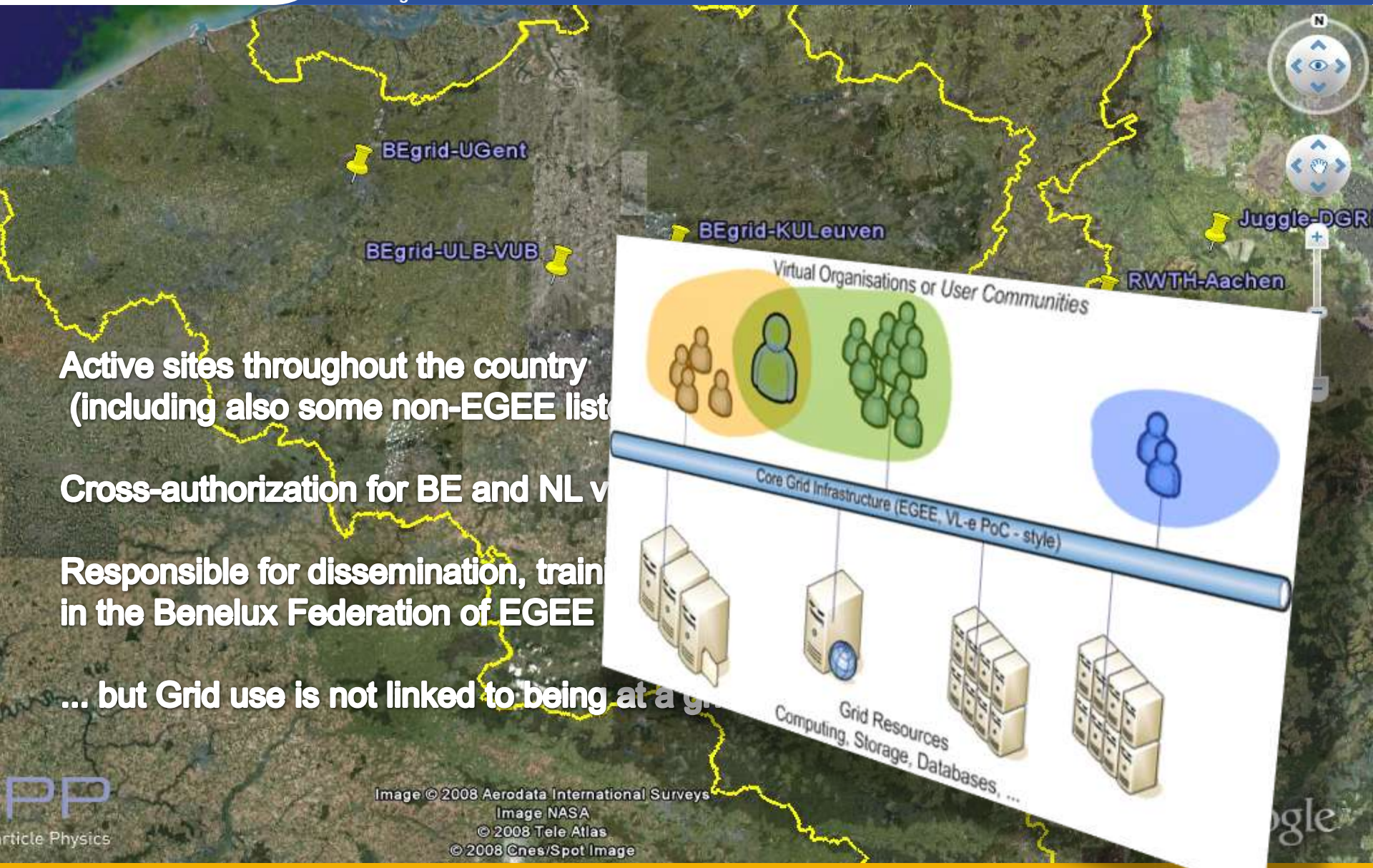
For the Grid a truly global identity is needed
— so we built the International Grid Trust Federation

- supported by the EU and e-IRG policy makers
- with over 80 member Authorities a global PKI

- New generation of CAs leverages national HE federations
  - SWITCH AAI SLCS
  - DFN
  - Joint Northern SLCS (NO,DK,SE,FU,NL)



Belgium & EGEE



Connecting to the Grid?

'If you belong to the Belgian Research World then you most probably qualify to make use of the BEgrid services'

http://www.begrid.be/

In 4 easy steps:

1. Go to the BEGrid or EGEE web site
   - Ask the Benelux Regional Helpdesk
   - Follow an EGEE training course at http://www.eu-egee.org/
2. Get a globally trusted *certificate* to identify you
   - BELNET CA is IGTF accredited
   - Trusted by all production grids in the world
3. Join (or set up) a Virtual Organisation
   - BEtest will get you started
4. **Use the grid to accelerate your research**

EGEE'08 Istanbul

545 participants from 48 countries

Goal:

- **Long-term sustainability of grid infrastructures in Europe**

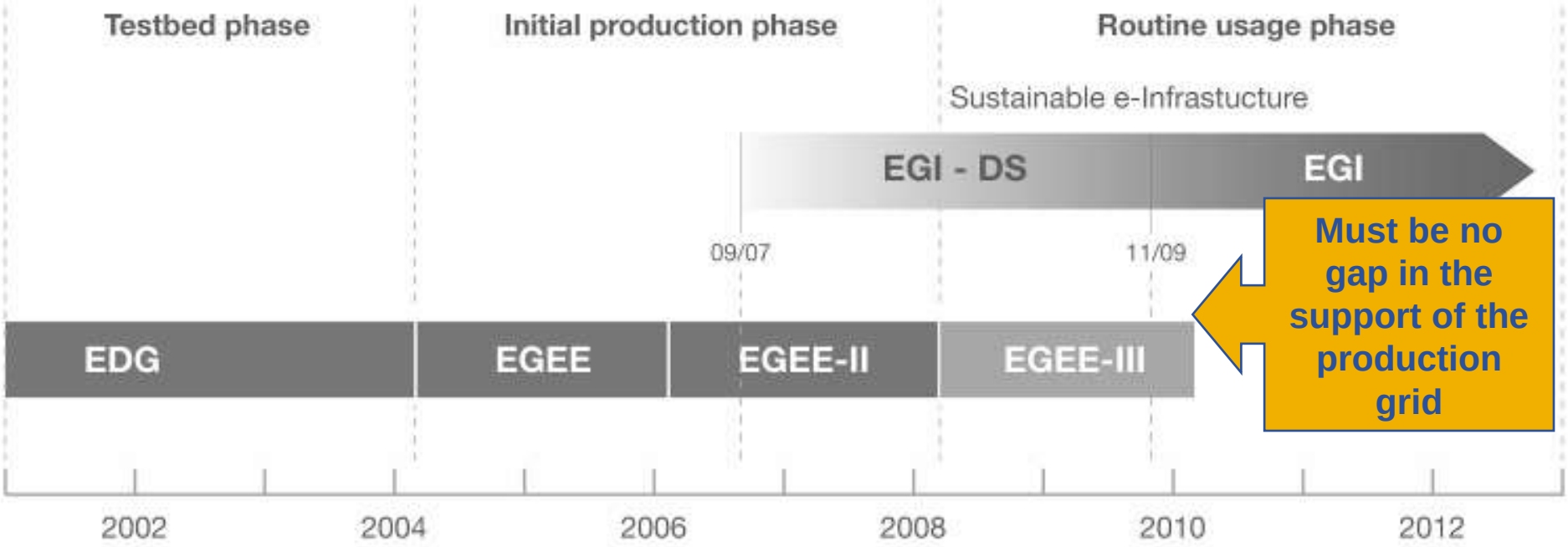
Approach:

- **Establishment of a new federated model bringing together NGIs to build the EGI Organisation**

EGI Organisation:

- **Coordination and operation of a common multi-national, multi-disciplinary Grid infrastructure**
  - To enable and support international Grid-based collaboration
  - To provide support and added value to NGIs
  - To liaise with corresponding infrastructures outside Europe

European Grid Initiative timeline



European Grid Initiative

- EGI Design Study proposal approved by the European Commission (started 1st September'07)
- Supported by 35+ National Grid Initiatives (NGIs)

  http://web.eu-egi.eu/partners/ngi/

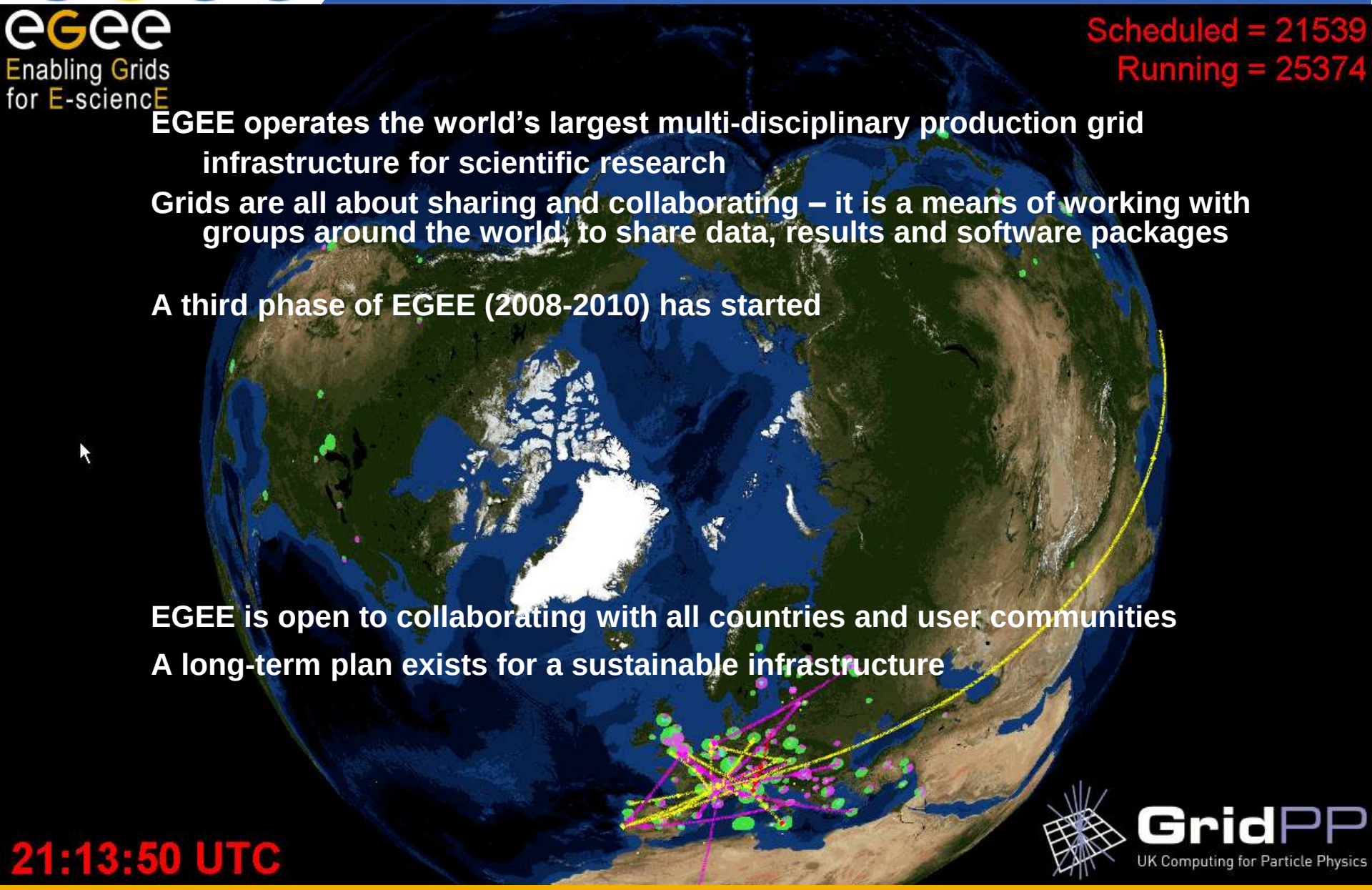
- 2 year project to prepare the setup and operation of a new organizational model for a sustainable pan-European grid infrastructure
- Draft EGI Blueprint produced:

  Blueprint Proposal http://www.eu-egi.eu/blueprint.pdf
  Functions Description http://www.eu-egi.eu/functions.pdf

http://www.eu-egi.org

Summary

Enabling Grids for E-sciencE



EGEE operates the world's largest multi-disciplinary production grid infrastructure for scientific research

Grids are all about sharing and collaborating – it is a means of working with groups around the world, to share data, results and software packages

A third phase of EGEE (2008-2010) has started

EGEE is open to collaborating with all countries and user communities

A long-term plan exists for a sustainable infrastructure

GridPP
UK Computing for Particle Physics

Connecting & Catalysing User Groups for Best Practices & Standardisation

OGF25/EGEE User Forum, 2-6 March 2009, Catania, Italy

Hosted by INFN

Multi-disciplinary infrastructures | **Scientific applications** | **Distributed Computing** | **Business** | **Innovation & Sustainability** | **Interoperation & Interoperability**