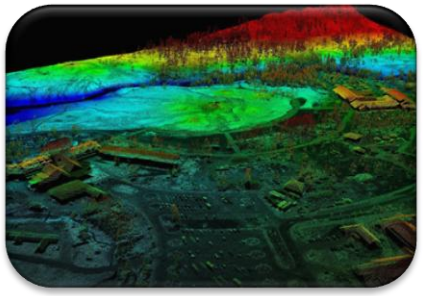
Εισαγωγή στο Grid,EGEE και το HellasGrid

Introduction to Grid, EGEE and HellasGrid

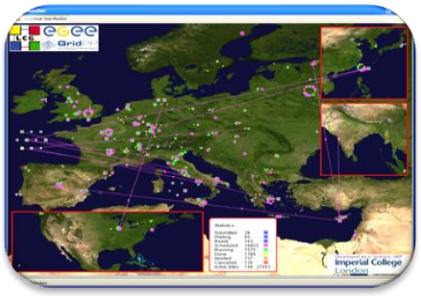
Athanasia Asiki
aassiki@cslab.ece.ntua.gr
Computing Systems Laboratory,
National Technical University of Athens

Contents

What is the Grid?



Grid paradigms



Enabling Grid for E-sciencE (EGEE)

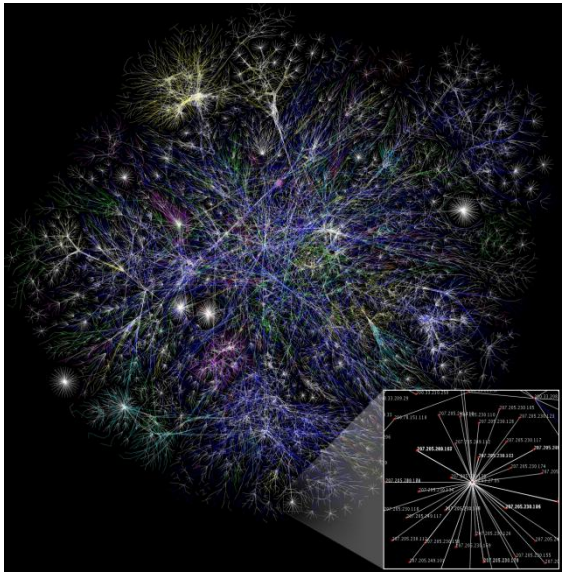
gLite middleware

HellasGrid Taskforce

Grid Technologies (Grids)

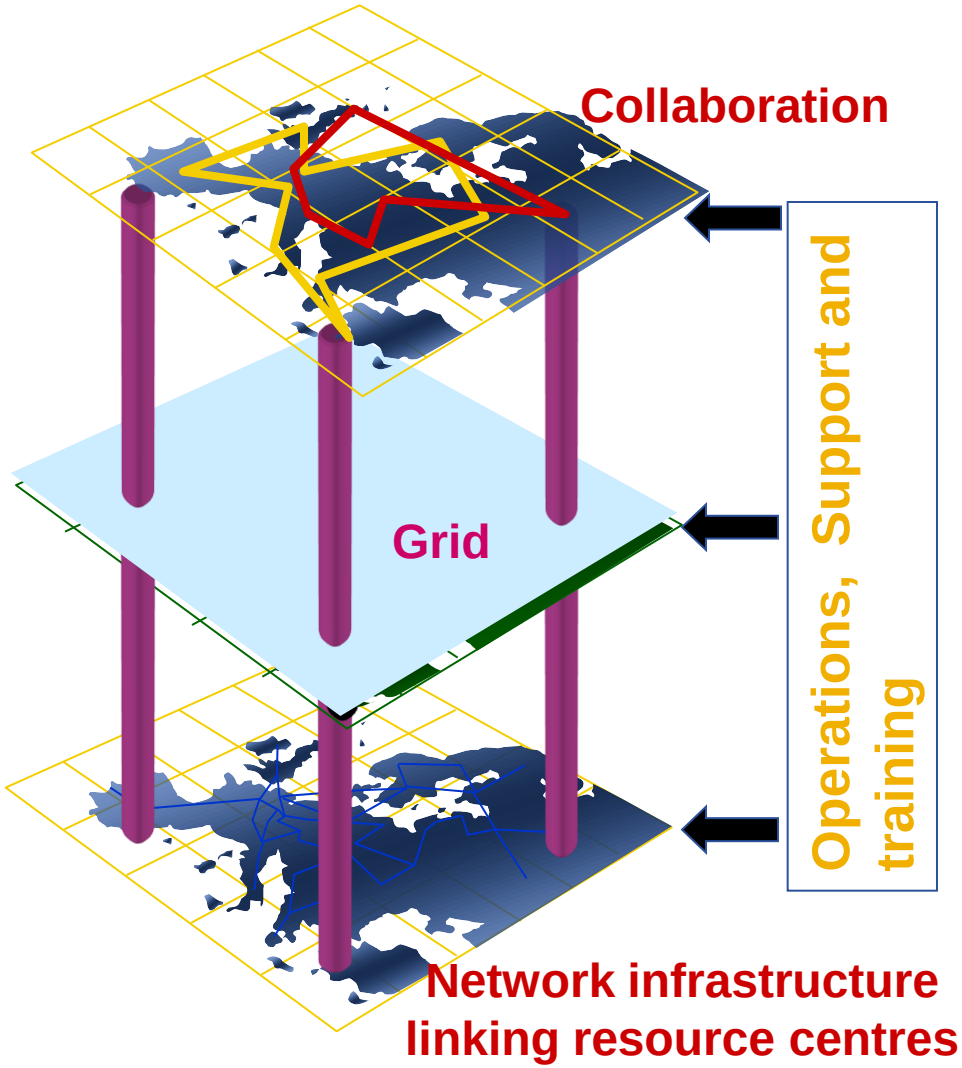
What is the Grid?

- The *World Wide Web* provides seamless access to information that is stored in many millions of different geographical locations



- The *Grid* is an emerging infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe

The Grid


Collaboration
Grid
Operations, Support and training
Network infrastructure
linking resource centres

Definition of Grid systems

- **Collection of geographically distributed heterogeneous resources**
  "Most generalized, globalized form of distributed computing"
- "An infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources"
  Ian Foster and Carl Kesselman
- **Offers access to a virtual and very powerful computing system**
- **A user does not care, in which resource his / her job / jobs is going to be executed**

Defining the Grid

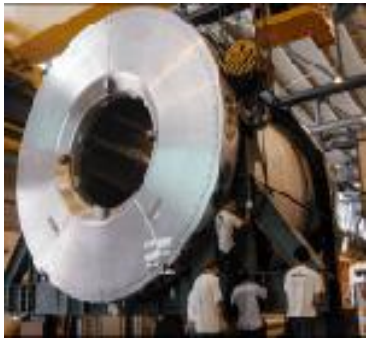
- In general terms:

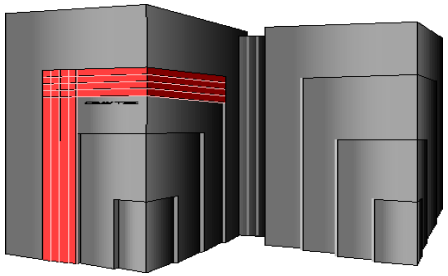
  A Grid is the combination of networked resources and the corresponding middleware, which provides services for the user.

Resources

- **An entity that is going to be shared**

  such as:
  - ✓ **computational units**
  - ✓ **storage units**
  - ✓ **sensors**
  - ✓ **visualization tools**
  - ✓ **software**
  - ✓ **licenses**
  - ✓ **experience**





Principles of grid

- **Resource sharing**
  - Geographically distributed resources offer computational power, storage capacity and bandwidth to the users
- **Secure and reliable access**
  - Authentication
  - Authorization
  - Access policy
- **Open standards**
- **Co-operation among people belonging to different organizations, institutes, groups**

Why is this Difficult or Different?

- **Lack of central control**
  - Where things run?
  - When jobs run?
  - What are the permissions of each user?
- **Shared resources**
  - Variability
- **Communication and coordination**
  - Different sites implies different administration rules, users, institutional goals and other constraints

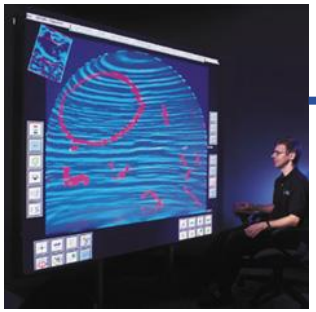
Why Grid?

- Complex computation
- Data that can't fit on one site
- Data owned by multiple sites
- Applications that need to run faster and in a more complex manner

Grid metaphorically …

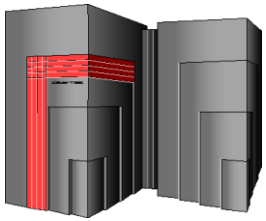
Mobile Access

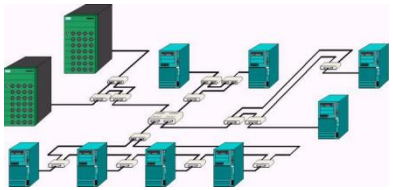
Workstation



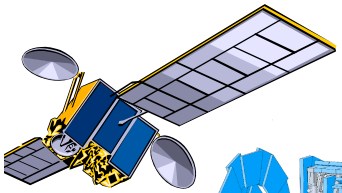
Visualising

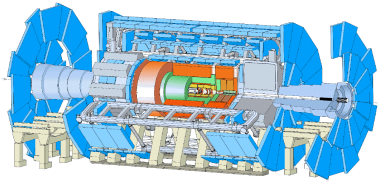
GRID MIDDLEWARE



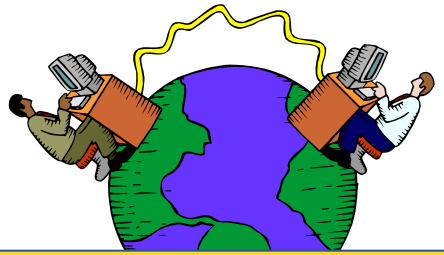


Supercomputer, PC-Cluster





Data-storage, Sensors, Experiments



Internet, networks

Why now?

- **Development of networking technology (doubling every nine months or so over the last years) and high-speed networks**
  - ✓ **widespread penetration of optical fibers**
  - ✓ **wireless connections**
  - ✓ **new Internet technologies (ADSL, WiMax)**

- **Moore's law everywhere**
  - ✓ **Instruments, detectors, sensors, scanners, …**

    ⇨ Organising their effective use is the challenge

- **Applications require a huge amount of computations to be executed and the collaboration among scientists**

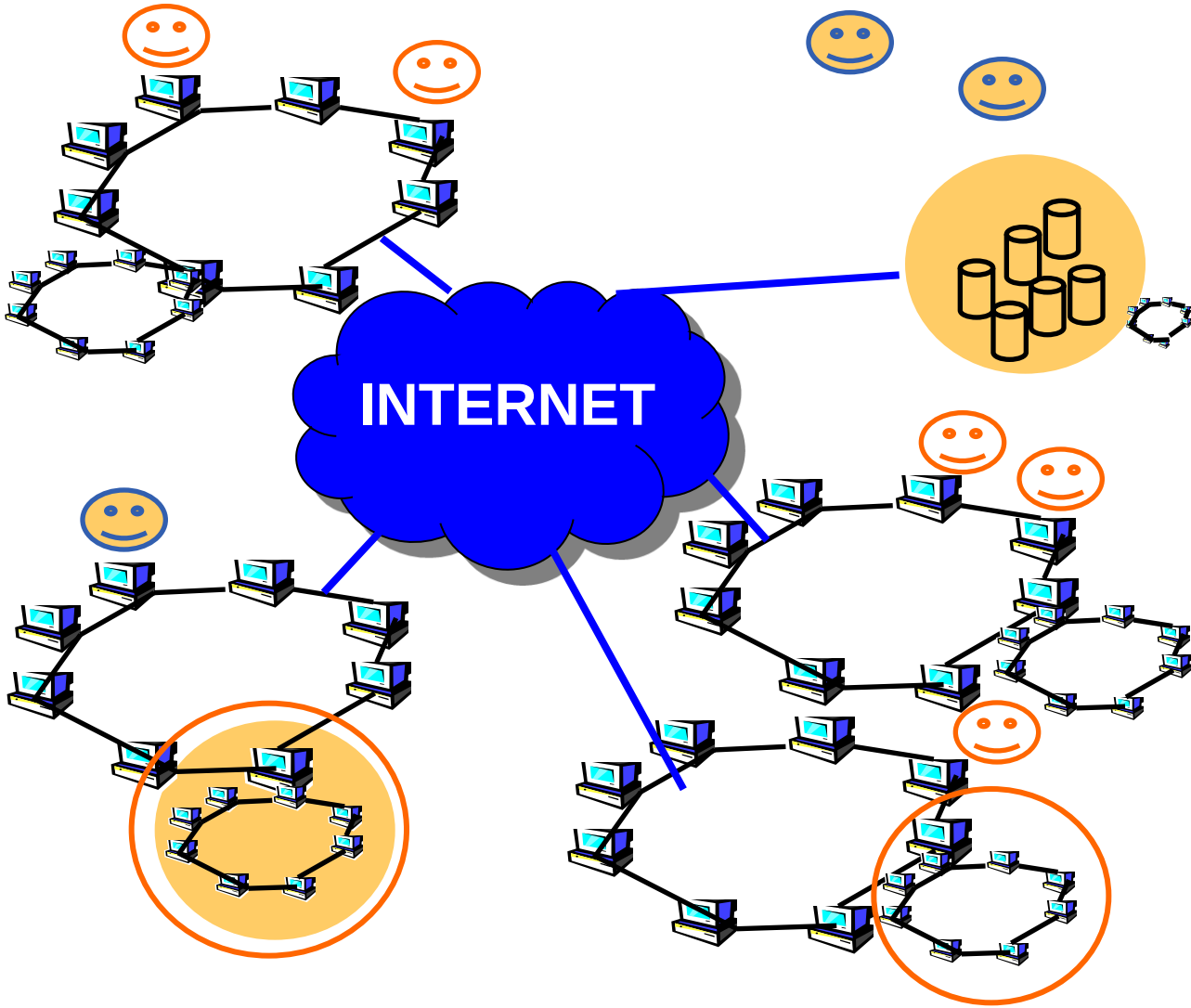
E-science

- **Science that became feasible and promiscuous by resource sharing (sharing of data, scientific instruments, computational resources, colleagues) across the Internet**

  - ✓ Often very compute intensive
  - ✓ Often very data intensive (both creating new data and accessing very large data collections) – data deluges from new technologies
  - ✓ **Crosses organisational and administrative boundaries**

VO concept

- gLite middleware runs on each shared resource to provide
  - **Data services**
  - **Computation services**
  - **Security service**
- Resources and users form Virtual organisations: basis for collaboration
- Distributed services (both people and middleware) enable the grid



Virtual Organizations

- **Virtual Organization**

  "A set of individuals and / or institutions defined by highly controlled sharing rules, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share and the conditions under which sharing occurs"

  Ian Foster

- **Abstract entities grouping users, institutions and resources in the same administrative domain**

What is going to be shared ?

- ✓ resources
- ✓ software
- ✓ special equipment
- ✓ licenses
- ✓ services
- ✓ Internet bandwidth

Virtual Organizations (VOs)

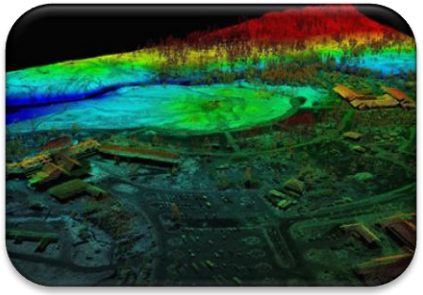
- Astrophysics, astro-particle physics
- Biomedical and Bioinformatic Applications
- Computational chemistry
- Earth sciences
- Finance
- Fusion
- Geophysics
- High-energy physics
- Infrastructure
- Other …

- Our regional VO: SEE
- VO for trainings : hgdemo

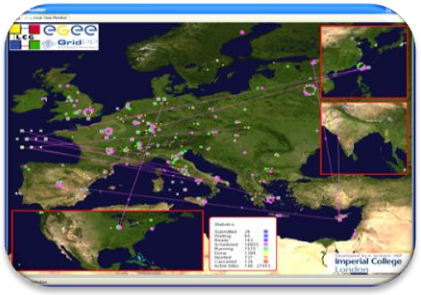
- **List of existing VOs**
  - http://cic.gridops.org/index.php?section=home&page=volist#1

Contents

What is the Grid?



Grid paradigms



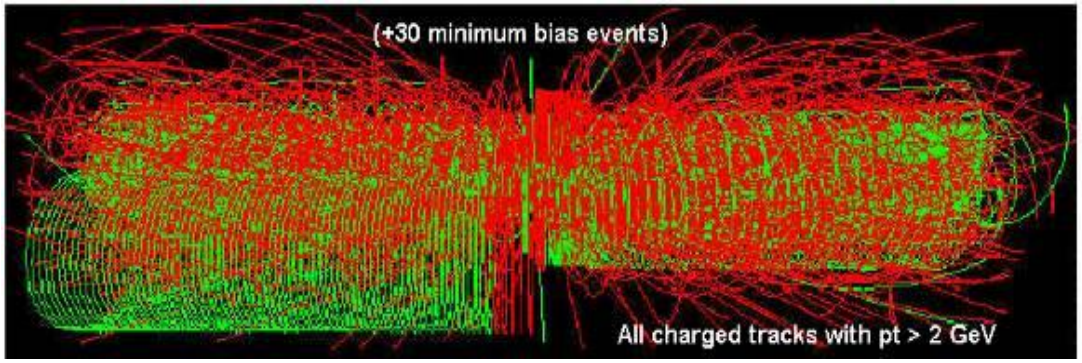
Enabling Grid for E-sciencE (EGEE)

gLite middleware

HellasGrid Taskforce

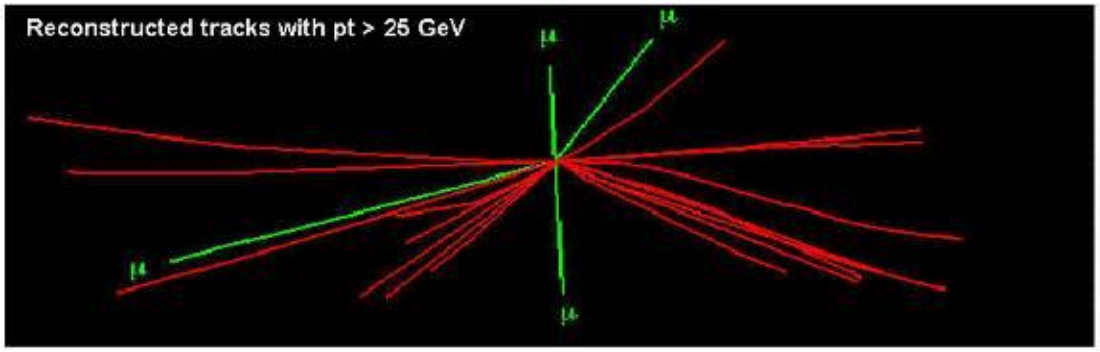
LHC Data Challenge

Starting from this event (particle collision) …



- ✓ Data Collection
- ✓ Data Storage
- ✓ Data Processing

You are looking for this "signature"…

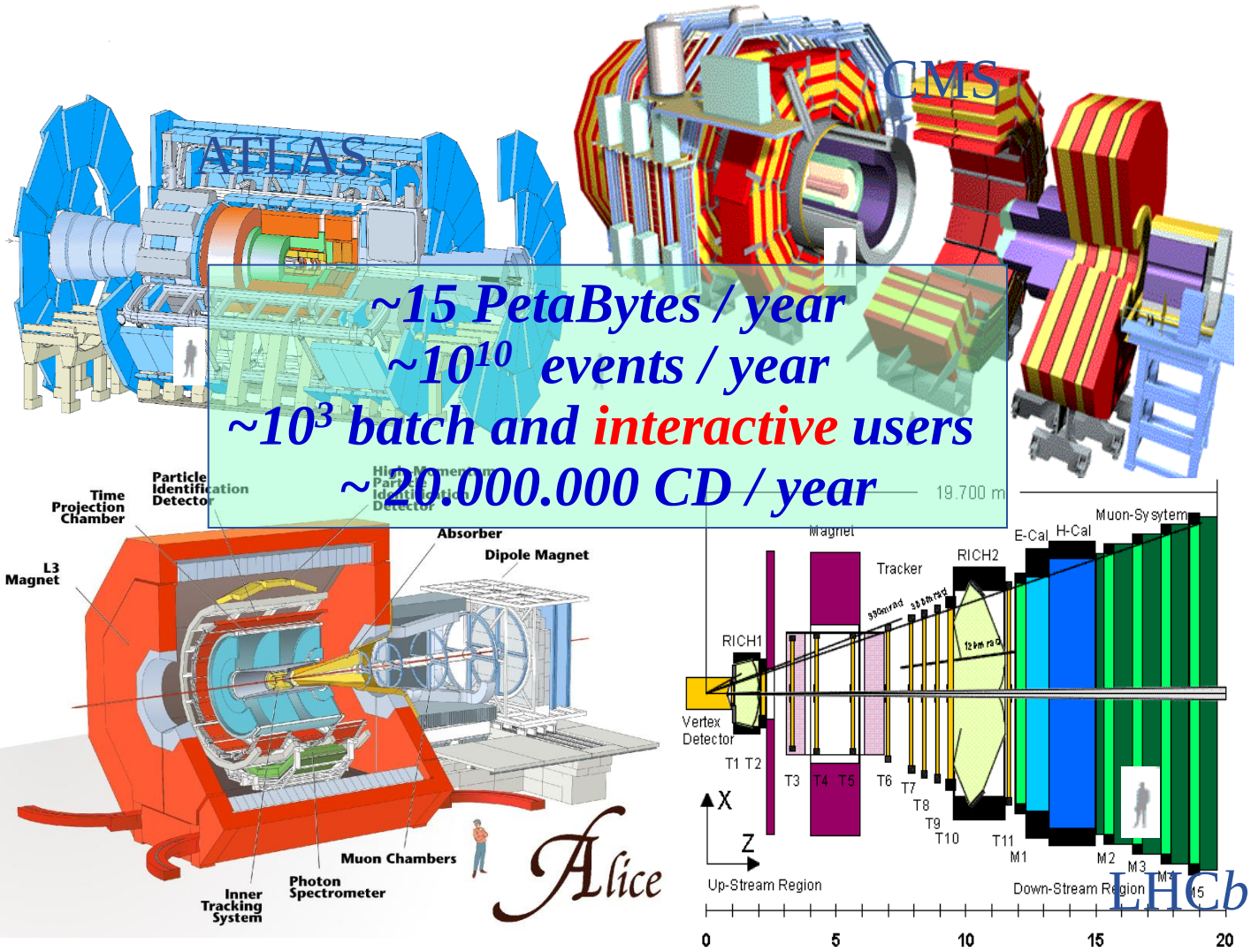


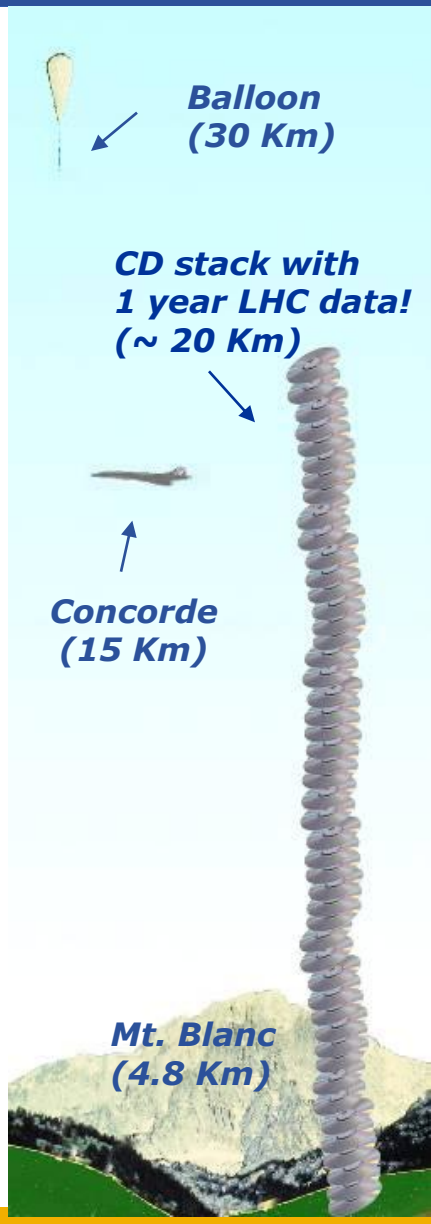
- ***Selectivity: 1 in 1013***

✓Like looking for 1 person in a thousand world populations!

✓Or for a needle in 20 million haystacks!

Amount of data from the LHC detectors





LHC Computing Grid

- **The LHC Computing Grid Project (LCG) was born to prepare the computing infrastructure for the simulation, processing and analysis of the data of the Large Hadron Collider (LHC) experiments.**

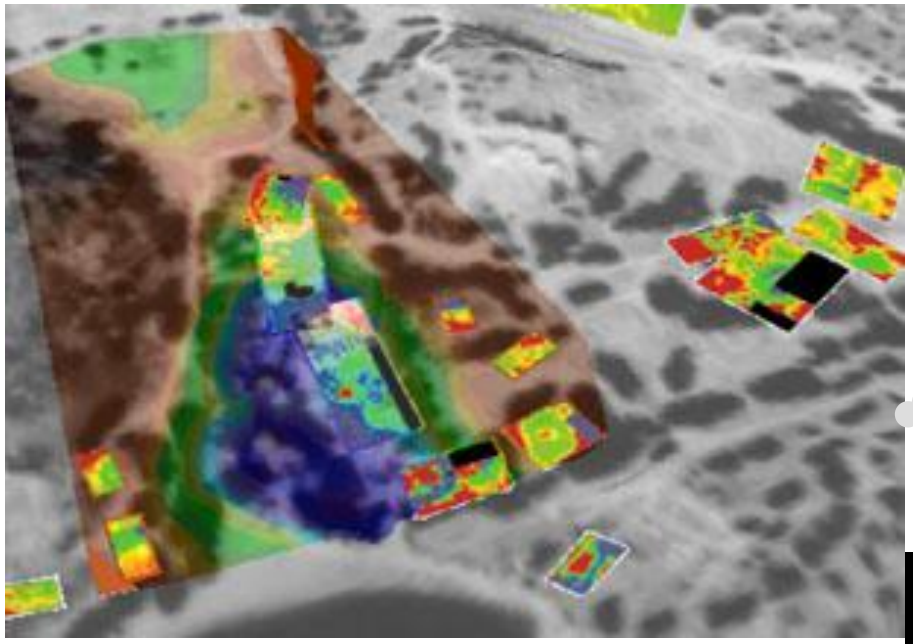
⇨ **The processing of the enormous amount of data, that will be generated, will require large computational and storage resources and the associated human resources for operation and support.**

⇨ **Preparation of a common infrastructure of**

- ✓ **libraries**
- ✓ **tools**
- ✓ **frameworks**

required to support the physics application software

Some examples

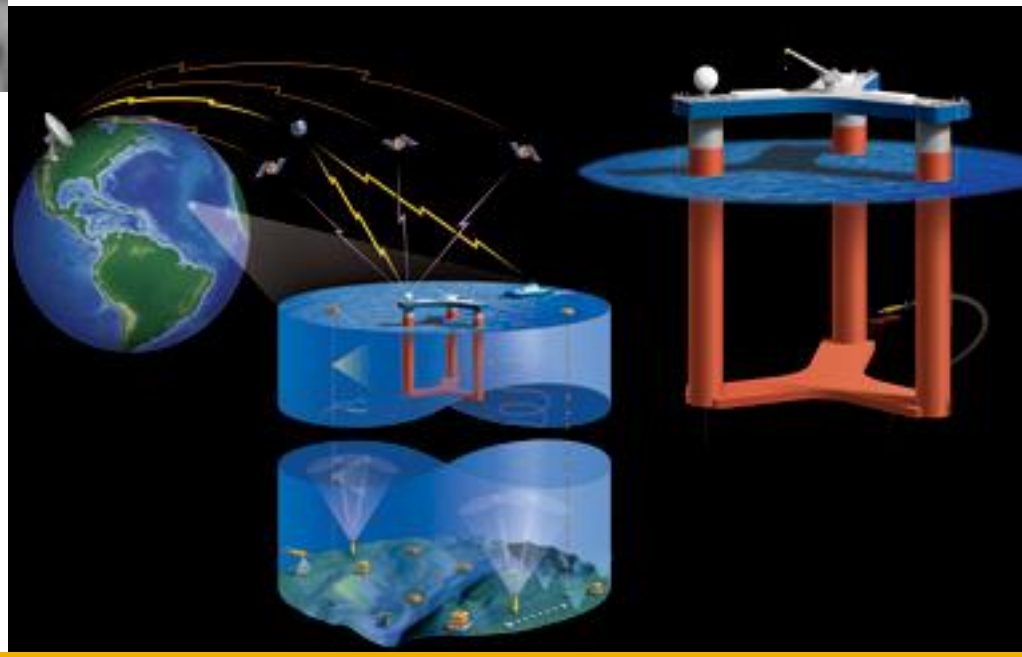


ArchaeoGrid

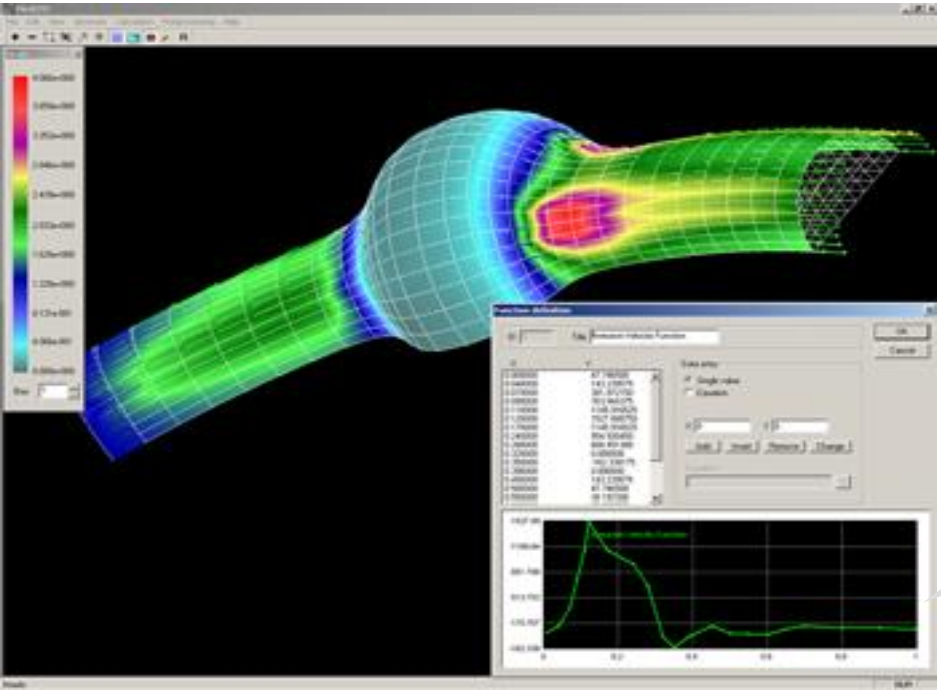
Create a computer model that weaves together data from many sources and predicts feedback interaction

LOOKING

Observe and analyze data streams in real time. A sensor grid with thousand of different sensors providing real time data and measurements from ocean-going researchers enabling an enormous data grid infrastructure.



Some examples

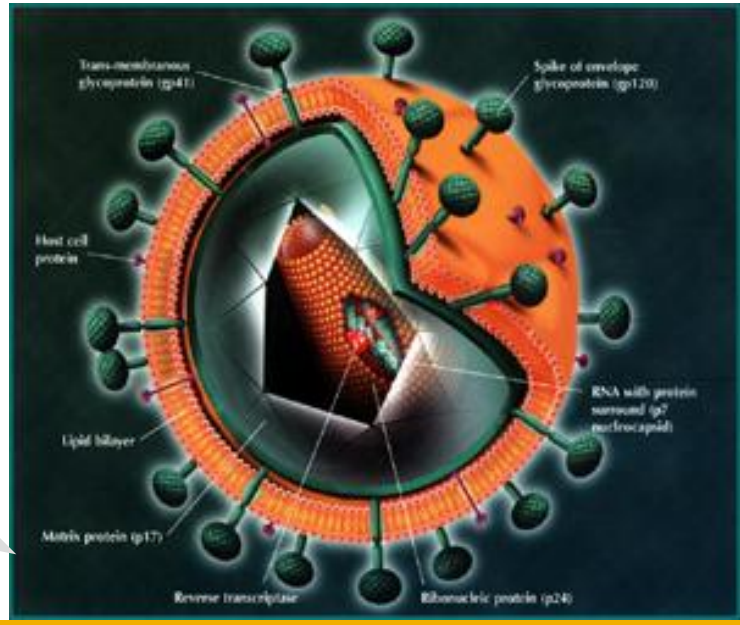


Parallel Blood Flow Simulation

Allows surgeons to perform virtual stent surgery until they get it just right. It combines parameters such as blood velocity and pressure with a series of medical images to automatically create a 3D computational model.

ViroLab

Aims to create a collaborative virtual labaratory for grid-based decision support for viral disease treatment. HIV treatment in the increasingly common case of HIV drug resistance is mainly studied. Virolab “vertically” integrates biomedical information relating to viruses, patients and literature resulting in a rule-based decision support system for drug ranking.



Some examples

Polar Grid

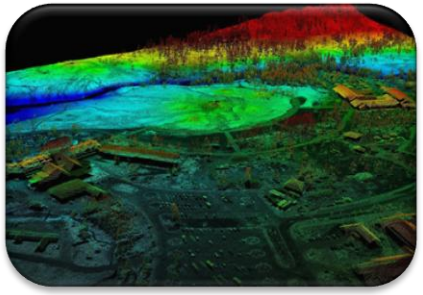
A planned project for an advance cyberinfrastructure, empowering smaller universities, and provide scientists with a gateway to teraflops of power: enough to drive new and improved high-performance simulations and enable measurement and prediction of ice sheet response to climate change and effect on ocean levels.

MoSES (Modelling and Simulation for e-Social Science)

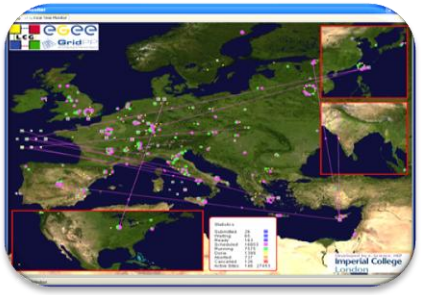
Runs predictive models integrating real Census data, survey data, healthcare data of UK population. Determine the impact of different policy decisions and various social aspects like increasing life expectancy, immigration, aging population.

Contents

What is the Grid?



Grid paradigms



Enabling Grid for E-sciencE (EGEE)

gLite middleware

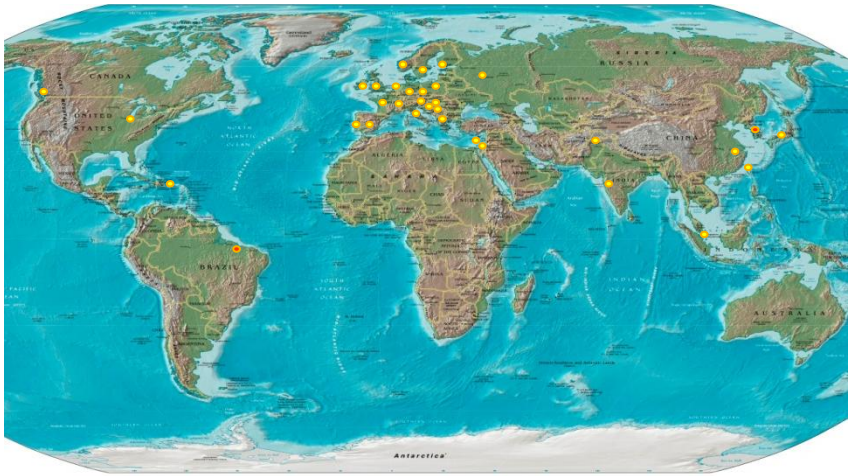
HellasGrid Taskforce

From EGEE to EGEE III

- **EGEE objective:**

  "to establish a seamless European Grid infrastructure for the support of the European Research Area (ERA)"

- **EGEE:**
  - Accomplished all of its objectives
  - Scope expanded beyond Europe



- **EGEE-II :**
  - → Full capacity from day one
  - → Large-scale, production-quality infrastructure across the European Research Area and beyond
  - → Supporting a wide range of applications
  - → Staff with extensive knowledge of Grid technology

EGEE Mission

- ***Mission:***

✓ **Manage and operate production Grid infrastructure for the European Research Area**

✓ **Interoperate with e-Infrastructure projects around the globe (Open Standards-GGF) and Contribute to Grid standardisation efforts**

✓ **Incorporate new users from the industry and from the research community as well assuring the best possible training and support**

- **Support applications deployed from diverse scientific communities:**
  - ✓ High Energy Physics
  - ✓ Earth Sciences
  - ✓ Computational Chemistry
  - ✓ **Fusion**
  - …
  - ✓ Biomedicine
  - ✓ Astrophysics
  - ✓ Finance, Multimedia
  - ✓ Geophysics . . .
- **Prepare for a permanent/sustainable European Grid Infrastructure (in a GÉANT2-like manner)**

- EGEE III objectives:
  - expand and optimize EGEE, by continuous operation of the infrastructure
  - support for more user communities
  - add of further computational and data resources
  - prepare the migration of the existing production European Grid from a project-based model to a sustainable federated infrastructure based on National Grid Initiatives for multi-disciplinary use

- Available infrastructure to the Research and Academic community 24 hours per day and 7 days per week

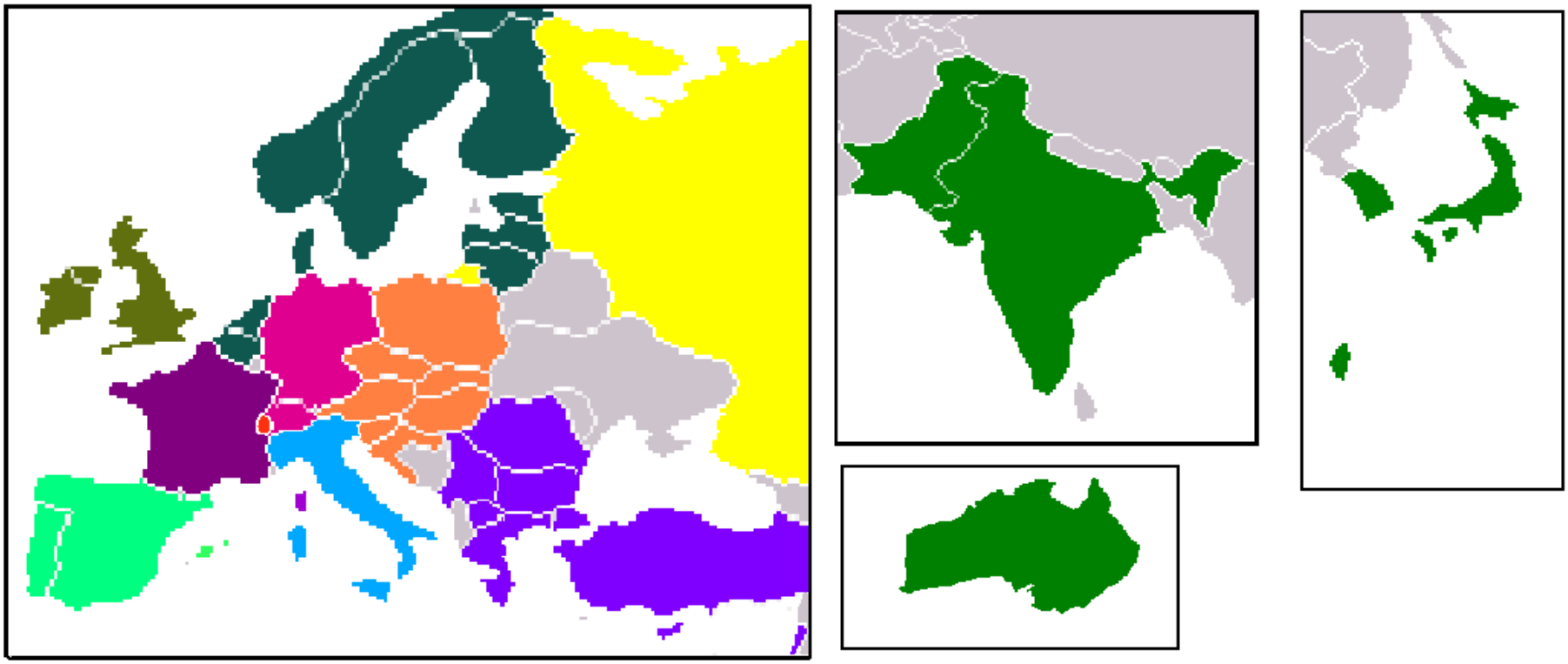
- **http://www.eu-egee.org/**

- Participants:
  - ✓ 50 countries

- Consists of:
  - ✓ 250 sites
  - ✓ ~ 68000 CPUs
  - ✓ ~ 20 PB
  - ✓ ~140 VOs
  - ✓ Massive data transfers > 1.5 GB/s

(October 2009)

Operations centres in EGEE



Regional Operations Centres (ROC)

- Front-line support for user and operations issues
- Provide local knowledge and adaptations
- One in each region – many distributed

User Support Centre (GGUS)

- In FZK: provide single point of contact (service desk), portal
- https://gus.fzk.de/pages/home.php

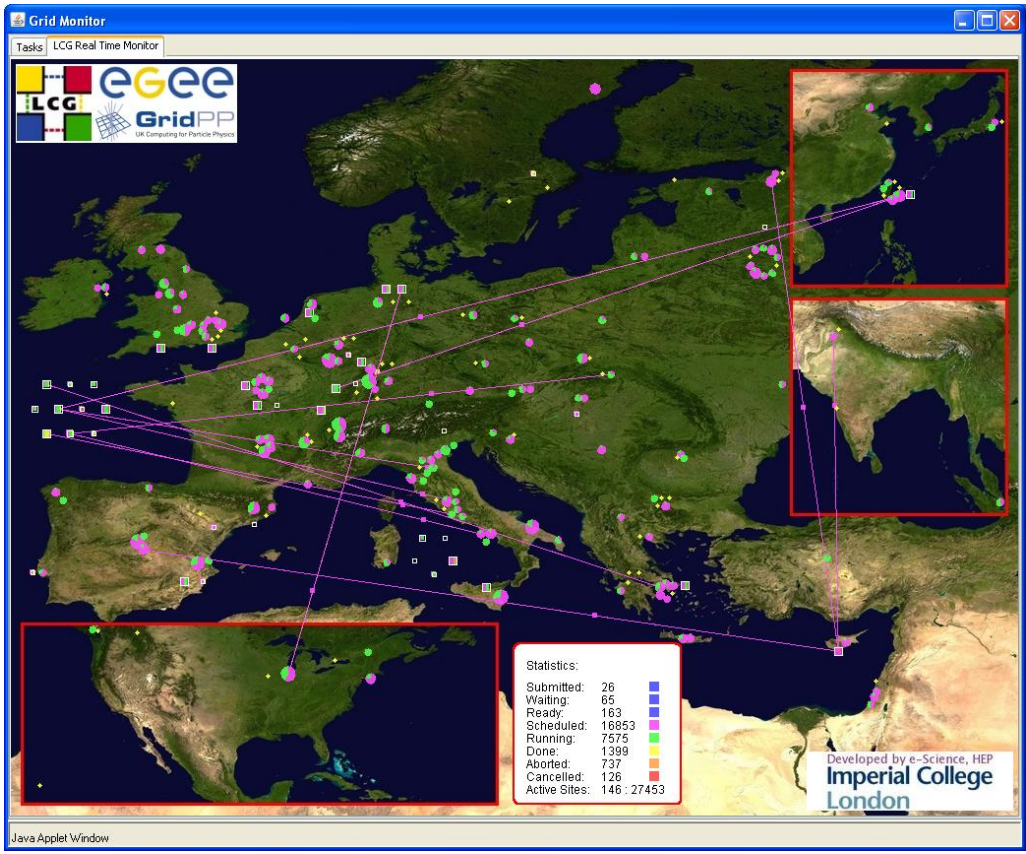
What is happening now?

Real Time Monitor

- http://gridportal.hep.ph.ic.ac.uk/rtm/

- Java tool
- Displays jobs running (submitted through RBs)
- Shows jobs moving around world map in real time,

  along with changes in status



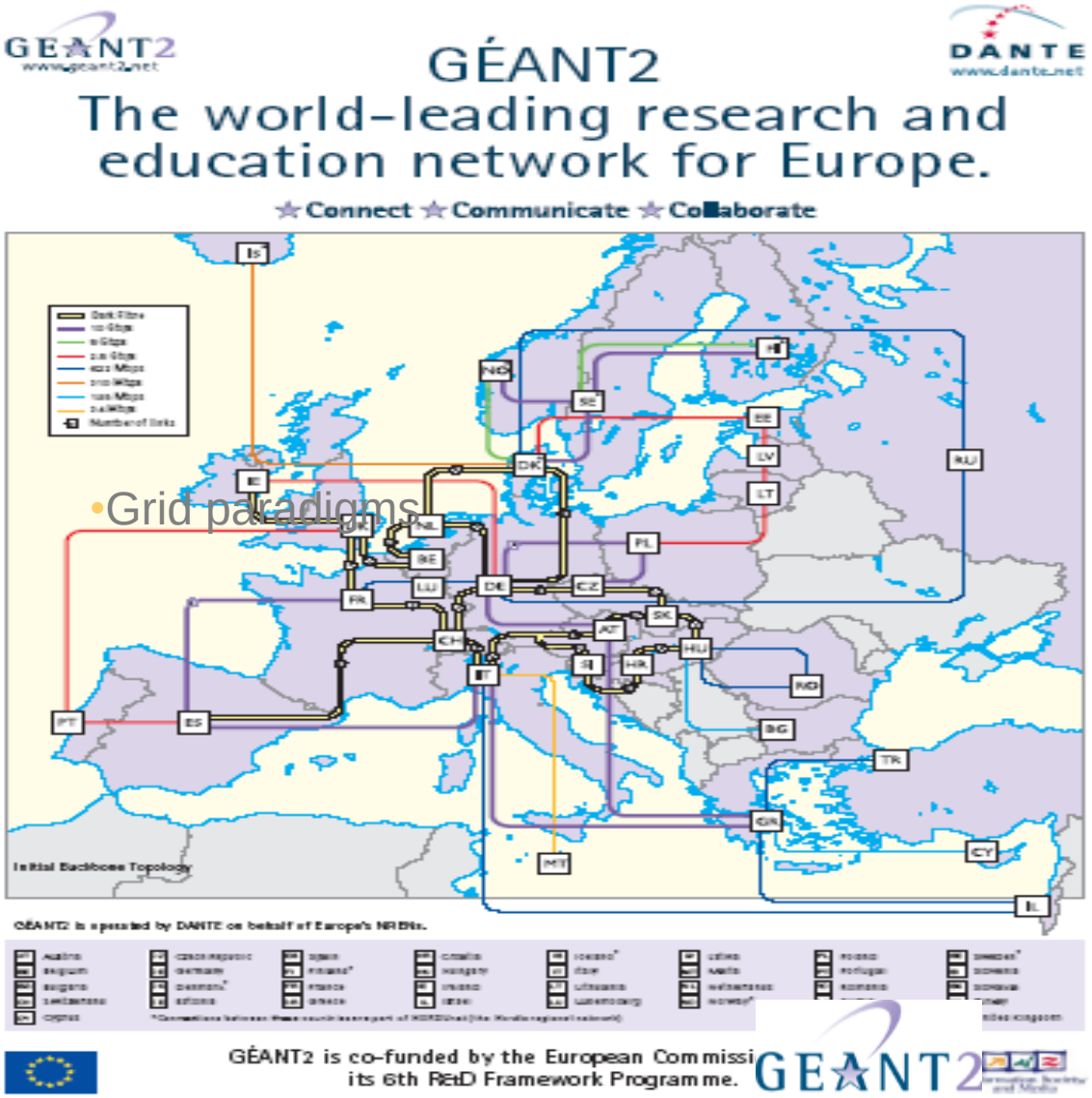
GÉANT 2

- **GÉANT2 is the seventh generation of pan-European research and education network, successor to the pan-European multi-gigabit research network GÉANT**

- 30 European National Research and Education Networks (NRENs) in
- 34 countries
- administrated by DANTE (*Delivery of Advanced Network Technology to Europe)*

- **It provides:**
  - Basic IPs services
  - Quality of service levels

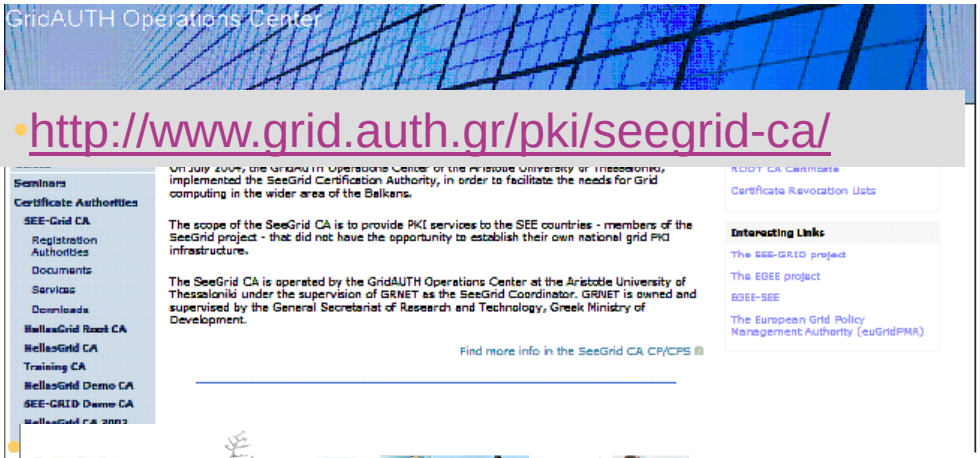
- **Greece NREN**
  - GRNET



Grid systems' related software

- **Operating system:**
  - Linux (+GNU utilities), usually a RHEL3-like, for example Scientific Linux
- **Middleware:**
  - gLite v3.x
- **Libraries and Applications**
  - Defined by the system and VOs administrators' foresight
  - The user can install and execute its own programmes

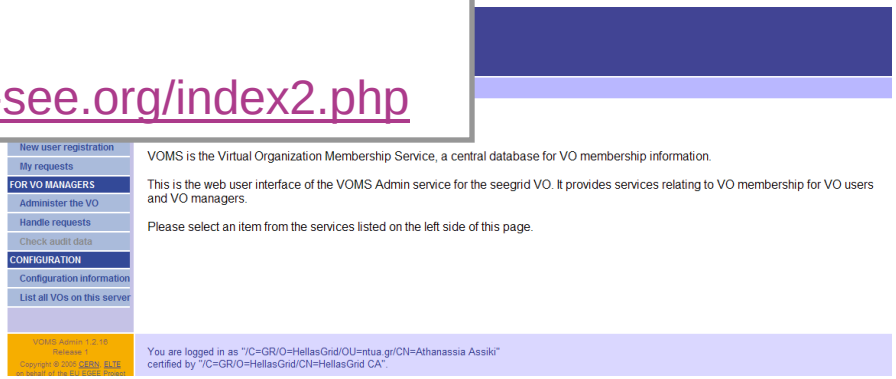
Infrastructure Sites



- http://www.grid.auth.gr/pki/seegrid-ca/

http://wiki.egee-see.org/index.php/SEE-GRID_Wiki

- **EGEE Helpdesk:**
  https://helpdesk.egee-see.org/index2.php

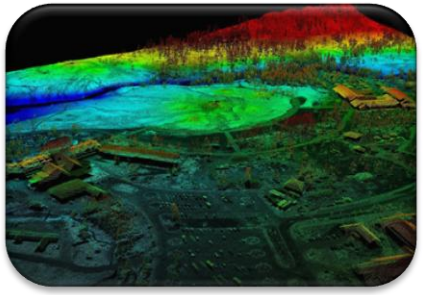


- http://www.egee-see.org/User_documentation.php

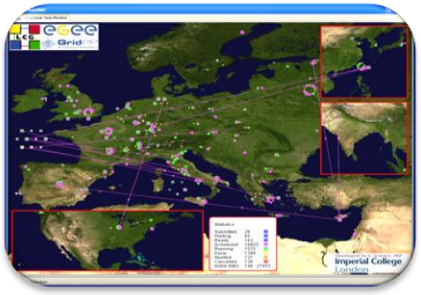
https://voms.irb.hr:8443/edg-voms-admin/seegrid/index.html

Contents

What is the Grid?



Grid paradigms



Enabling Grid for E-sciencE (EGEE)

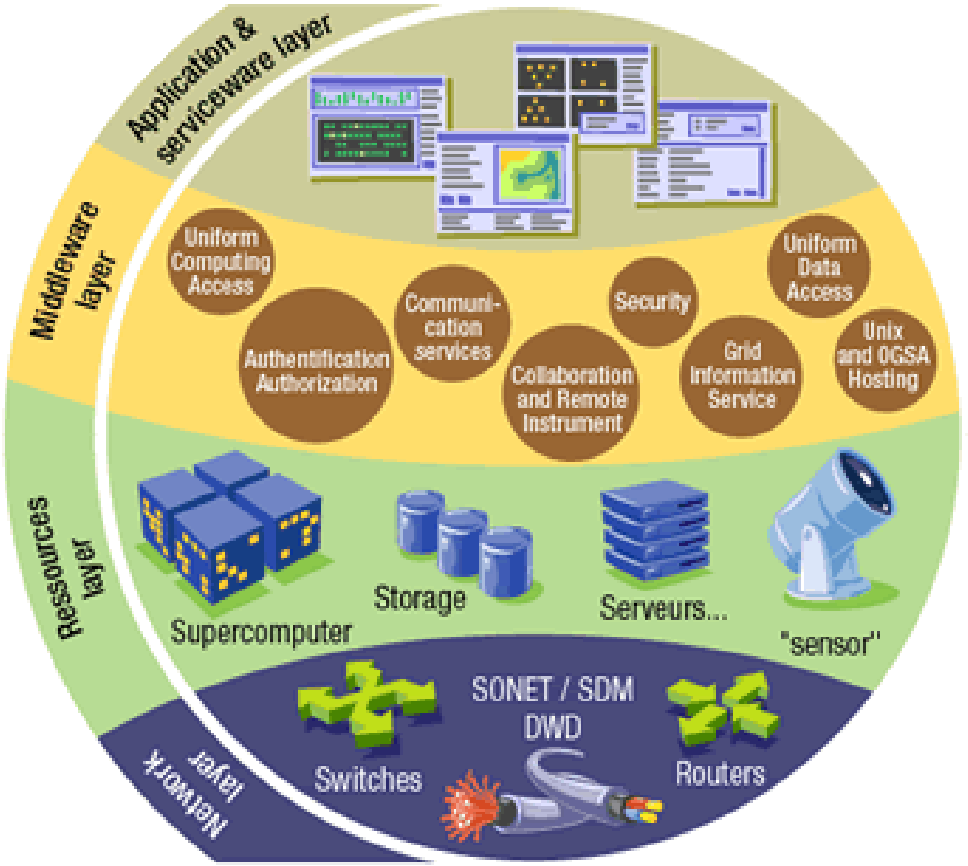
gLite middleware

HellasGrid Taskforce

Grid middleware

- **The Grid relies on advanced software, called middleware, which interfaces between resources and the applications**

- **The Grid middleware:**
  - Basic services
    - Secure and effective access to resources
  - High level services
    - Optimal use of resources
    - Authentication to the different sites that are used
    - Job execution & monitoring of progress
    - Problem recovery
    - Transfer of results back to the user



gLite
Lightweight Middleware for Grid Computing

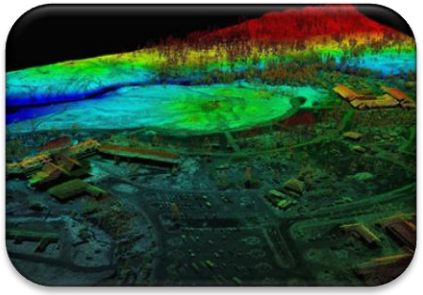
- **Part of the EGEE project**
- **Next generation middleware for grid computing**
- **In its development participate from different academic and industrial European centers**
- *Provides services for computing element, data management, accounting, logging and bookeping, information and monitoring, service discovery, security, workload management*

"Gridifying" an application

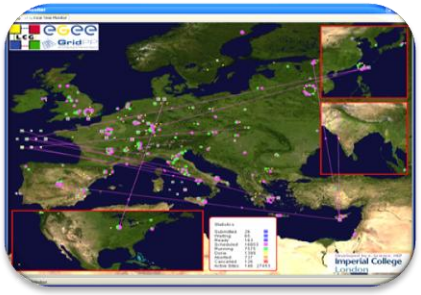
- **Applications running on a PC need to be adapted to run on a Grid**
  - Include new layers of grid-enabled software
- **Aspects to consider:**
  - Acquiring authentication credentials
  - Locating available data
  - Structure of jobs for computational tasks
  - Initiate computations
  - Monitor progress of computations and data transfers
  - Collection of output results

Contents

What is the Grid?



Grid paradigms



Enabling Grid for E-sciencE (EGEE)

gLite middleware

HellasGrid Taskforce

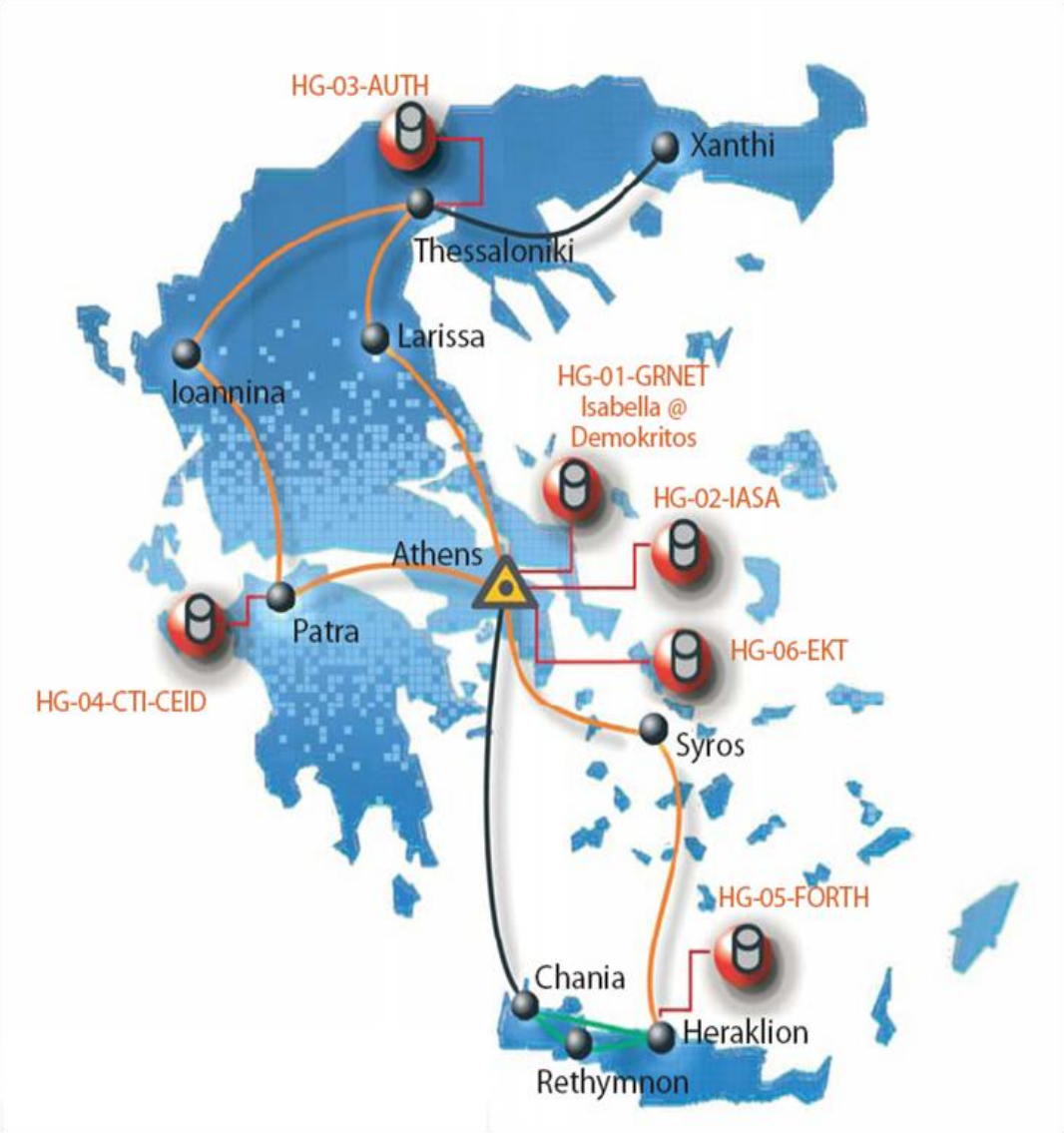
HellasGrid Infrastructure, Phase I & II

- **HellasGrid I**
  - Located at N.C.S.R. Demokritos (a.k.a. Isabella)
  - 34 dual Intel **P4 Xeon @ 2.8GHz, 1GB RAM, 2x 70GB SCSI HDD**, 2x Gbit
  - IBM FAStT900 Storage Area Network
    - 2x Redundant Fiber Channel Controllers with 1Gbyte Cache each
    - 70x146.8GB= **10,276TB raw storage capability,** over 5 disk shelves
  - Tape Library ~30 TBytes, integrated monitoring
  - December 2004

- **HellasGrid II**
  - 5 sites: ΕΚΤ (>220), ΙΕΣΕ (48), ΑΠΘ (128), ITE (128), ΙΤΥ (128)
  - ~700 CPUs **x86_64, 2 GB RAM, 1x 80GB SATA HDD**, 2x Gbit
  - ~20 TBytes storage space in SAN (5x 4TBs)
  - ~50 TBytes Tape Library in National Documentation Center

http://www.hellasgrid.gr/

HellasGrid e-Infrastructure



HellasGrid structure

- **Main site: HG-01-GRNET (Isabella, cslab@ICCS/NTUA)**
- **HG-02…HG-06 sites @ (NDC, IASA, AUTH, FORTH, CTI)**

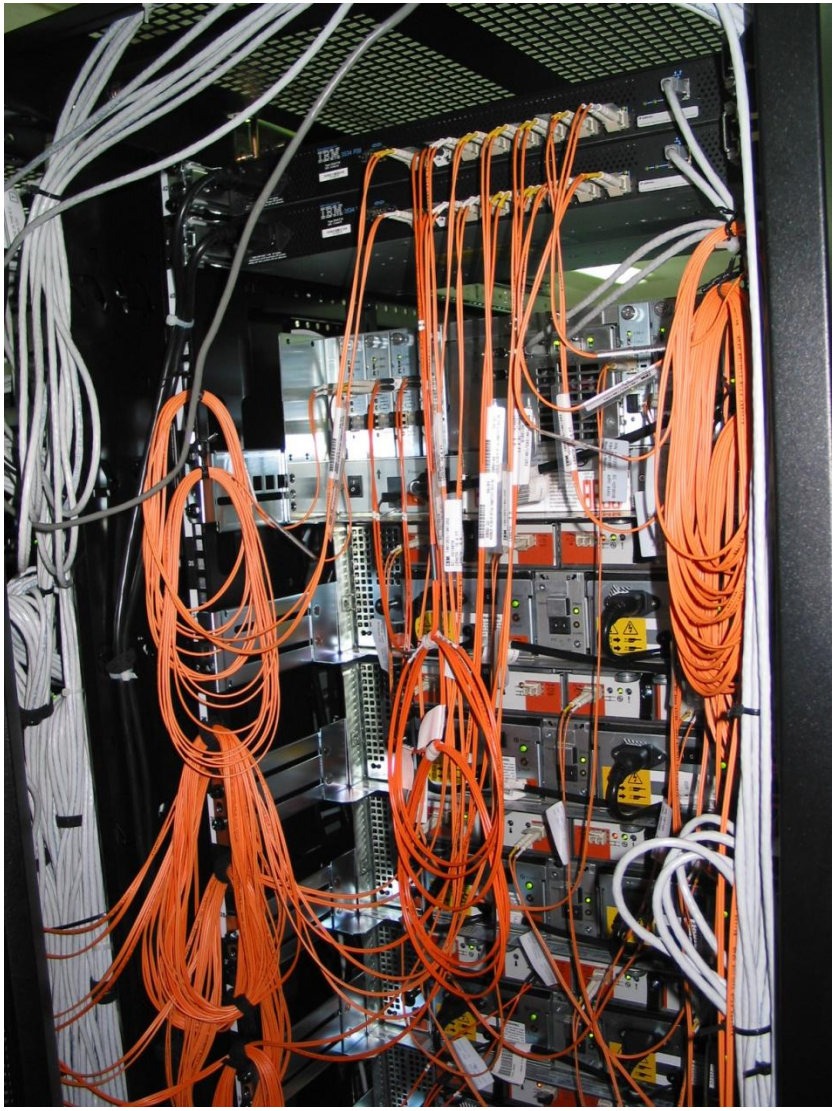
- **HG CA and VOMS** (AUTH)**:**
  http://www.grid.auth.gr/pki/seegrid-ca/

- **Helpdesk (CTI):**
  user-support@hellasgrid.gr

- **Regional monitoring tools (FORTH):**
  http://hellasgrid-ui.ics.forth.gr/acctROC/

- **Apps support** (IASA)**:**
  application-support@hellasgrid.gr

HellasGrid I Infrastructure, Isabella



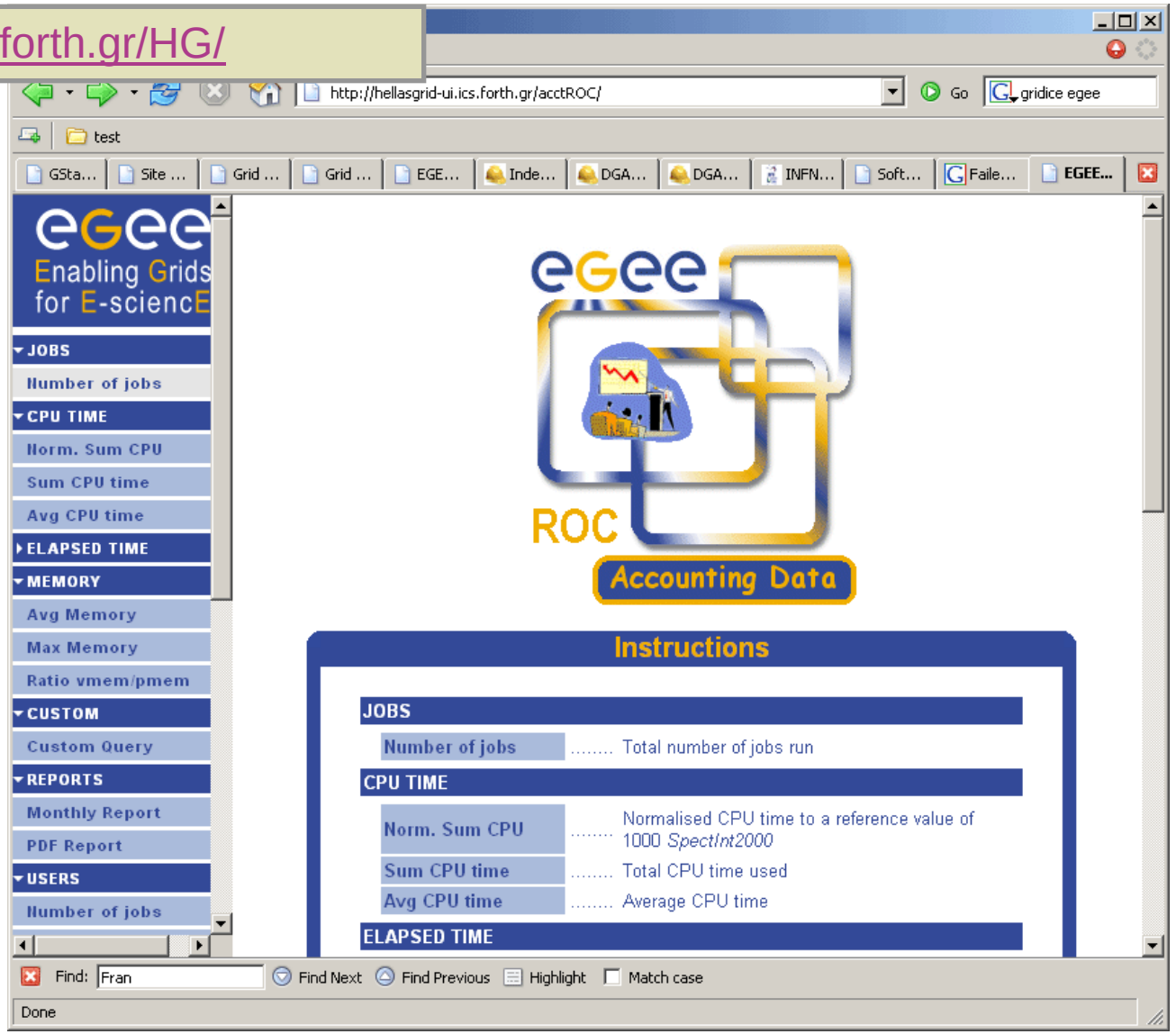
HellasGrid II Infrastructure

Registration to HellasGrid

https://access.hellasgrid.gr/register/registration_form

Accounting statistics for the HellasGrid Infrastructure

http://hellasgrid-ui.ics.forth.gr/HG/



Core Services (HG-01-GRNET)

- **Core Services**
  - Central LCG File Catalog (LFC) for the users of the VOs:
    - eumed, hgdemo, see
  - Resource Broker and Information Index (BDII) which can be accessed by the users of the VOs:
    - atlas, alice, lhcb, cms, dteam, sixt, biomed, esr, magic, compchem, see, planck, hgdemo, eumed
  - Catch-All User Interface for HellasGrid
    - Registration is handled through the Hellasgrid User-Support Team
    - UI services are offered by all HG sites

- **Service Availability Monitoring**
  - https://mon.isabella.grnet.gr/sft/lastreport.cgi (Need a valid HellasGrid Certificate)

Q&A

Thank you!

Αναφορές

- **Grid café:**
  http://www.gridcafe.org/
- **Open Grid Forum:**
  http://www.gridforum.org/
- **HellasGrid Task Force**
  http://www.hellasgrid.gr/
- **EGEE (Enabling Grids for E-science)**
  http://www.eu-egee.org/
- **The Globus Alliance**
  http://www.globus.org/
- **Grid Operations Centre**
  http://goc.grid-support.ac.uk/gridsite/gocmain/
- **gLite UserGuide**
  http://glite.web.cern.ch/glite/documentation/

Χρήσιμα web links

- **EGEE – South East Europe**
  http://www.egee-see.org/
- **SEE-GRID**
  http://www.see-grid.org/
- **GRNET**
  http://www.grnet.gr/
- **gLite**
  http://glite.web.cern.ch/glite/
- **SEE-GRID Wiki**
  http://wiki.egee-see.org/index.php/SEE-GRID_Wiki
- **GOC Wiki**
  http://goc.grid.sinica.edu.tw/gocwiki/
- **SEEREN2**
  http://www.seeren.org/