Overview of the EGEE project and the gLite middleware

Gergely Sipos
MTA SZTAKI
sipos@sztaki.hu

www.eu-egee.org

Information Society

Outline

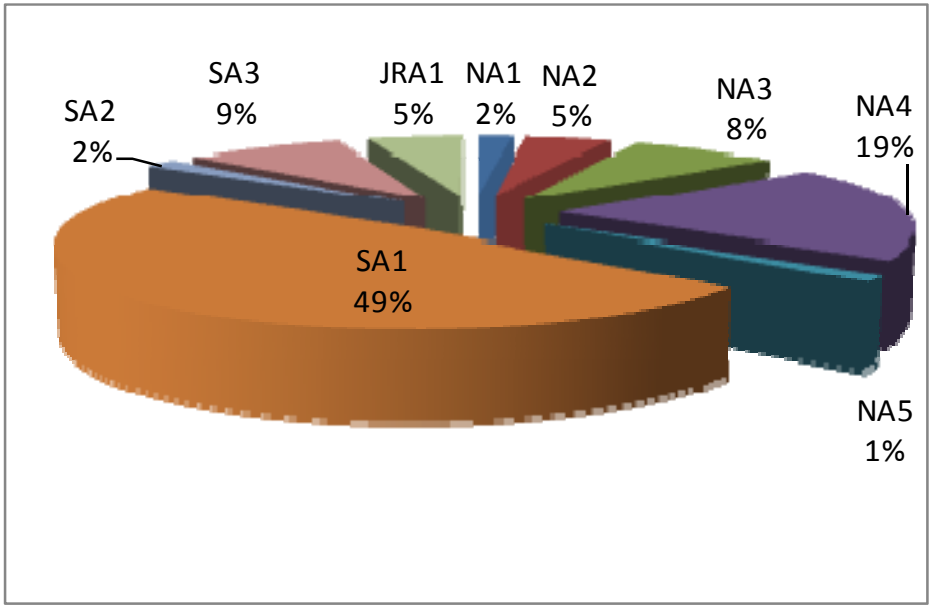
- **What is EGEE?**
  - The project
  - The infrastructure
- **gLite middleware**
- **EGEE applications**
- **Sources of further information**

The EGEE project

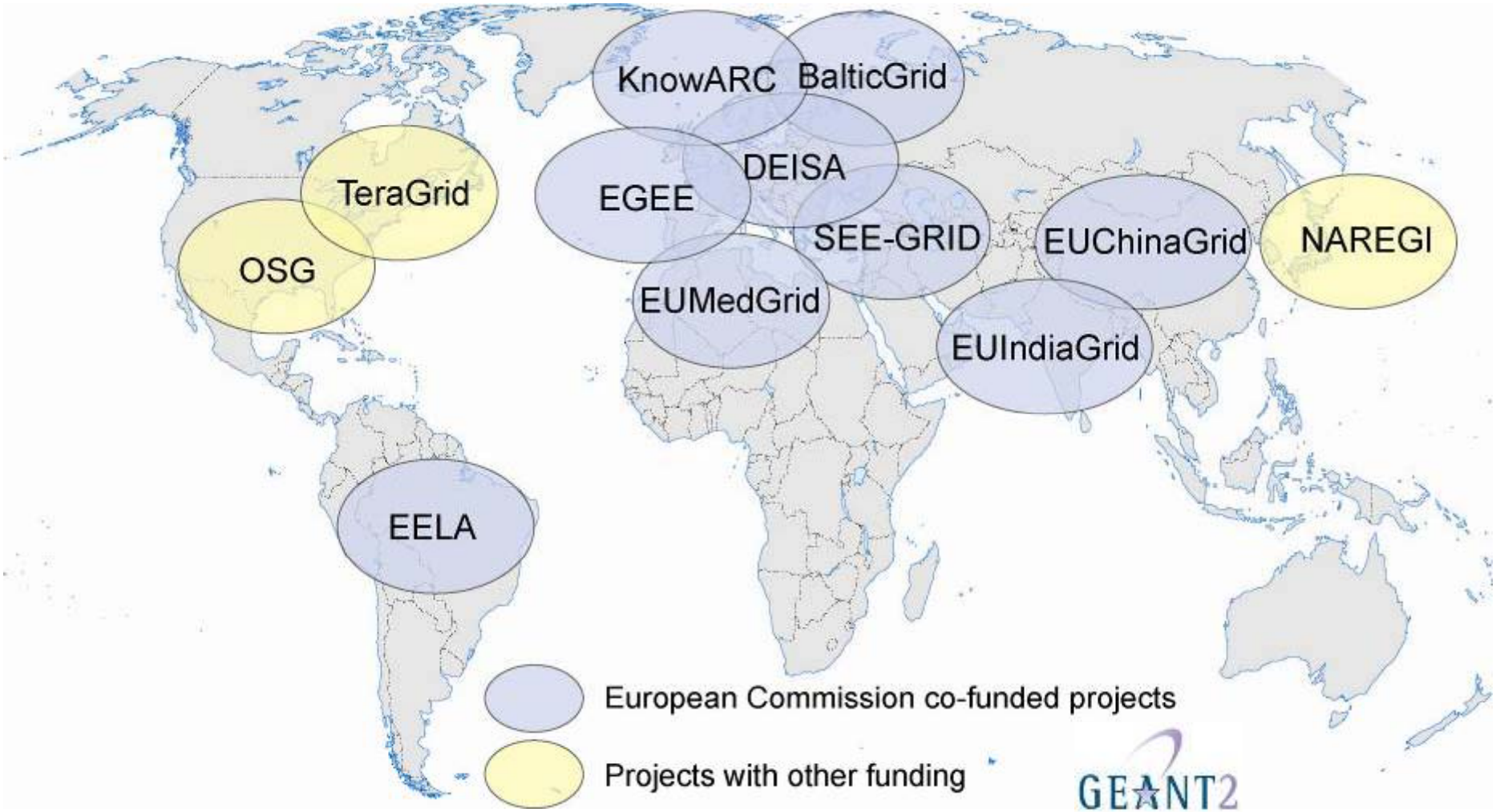
- **Flagship European grid infrastructure project, now in 3rd phase with 42 beneficiaries and 100 Joint Research Unit members**
- **Objectives**
  - Large-scale, production-quality grid infrastructure for e-Science
  - Attracting new resources and users from industry as well as science
  - Maintain and further improve gLite Grid middleware
- **Structure**

  EGEE: 1 April 2004 – 31 March 2006

  EGEE-II: 1 April 2006 – 31 March 2008
- **EGEE-III:** 1 May 2008 – 30 April 2010
  - 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



Collaborating e-Infrastructures

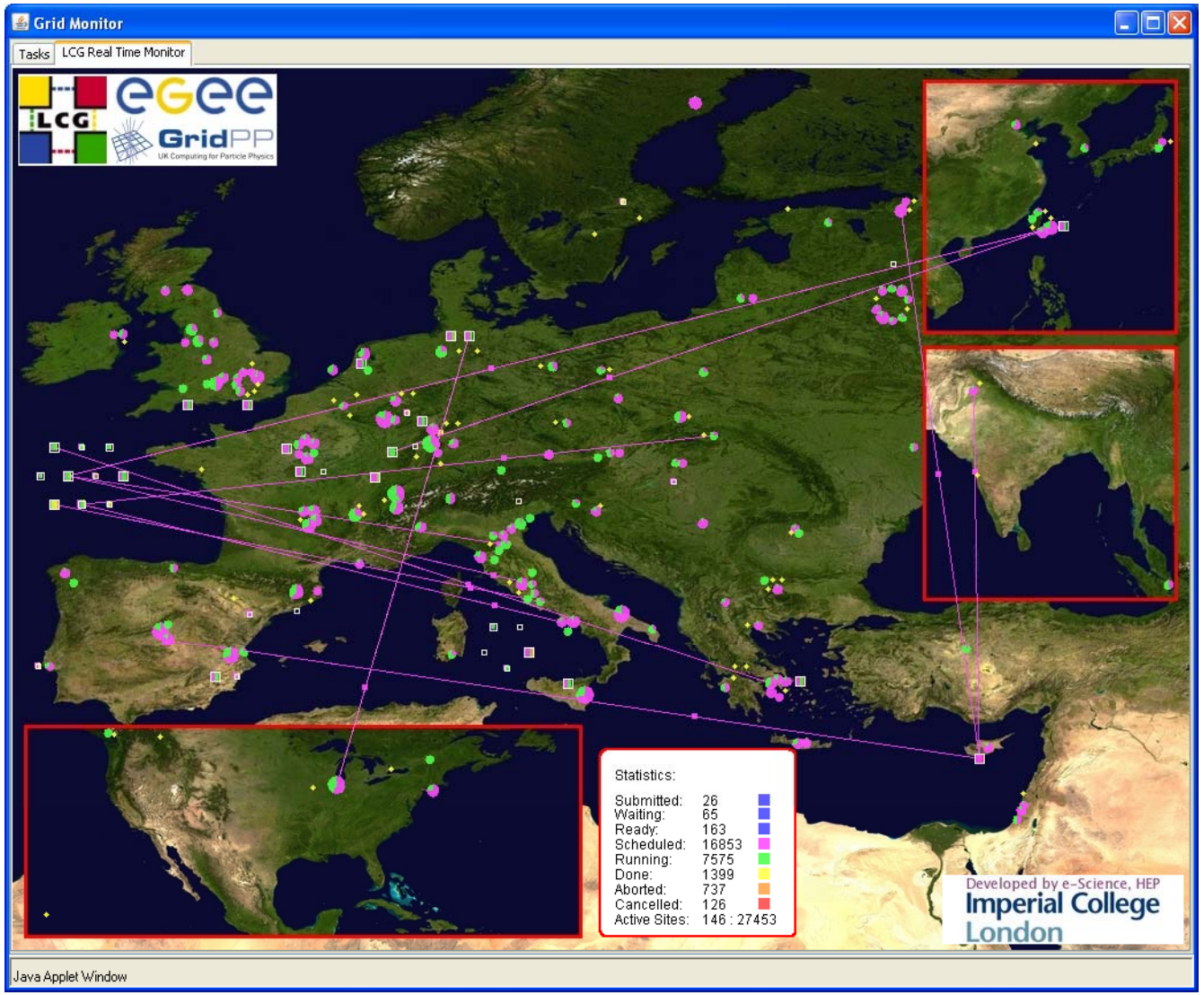


Potential for linking ~80 countries by 2008

What is happening now?

Real Time Monitor

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

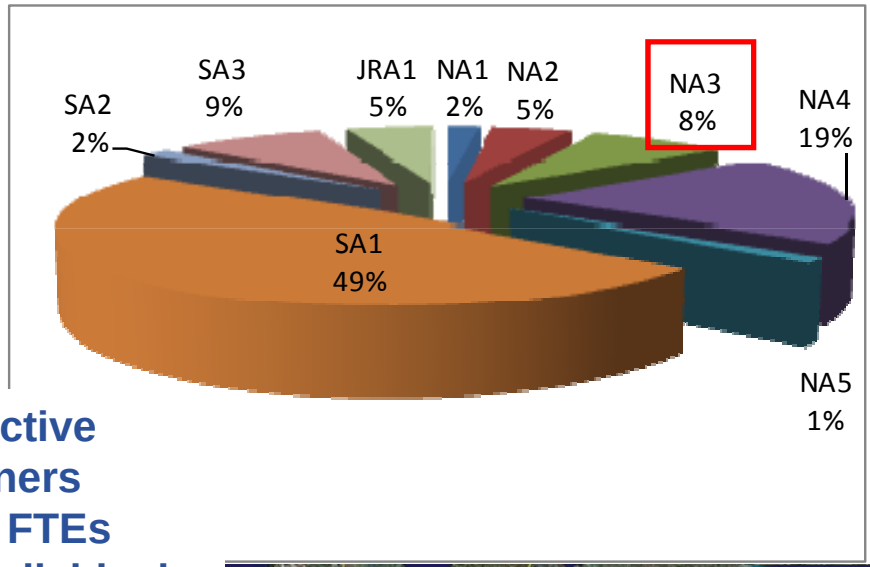


http://gridportal.hep.ph.ic.ac.uk/rtm/

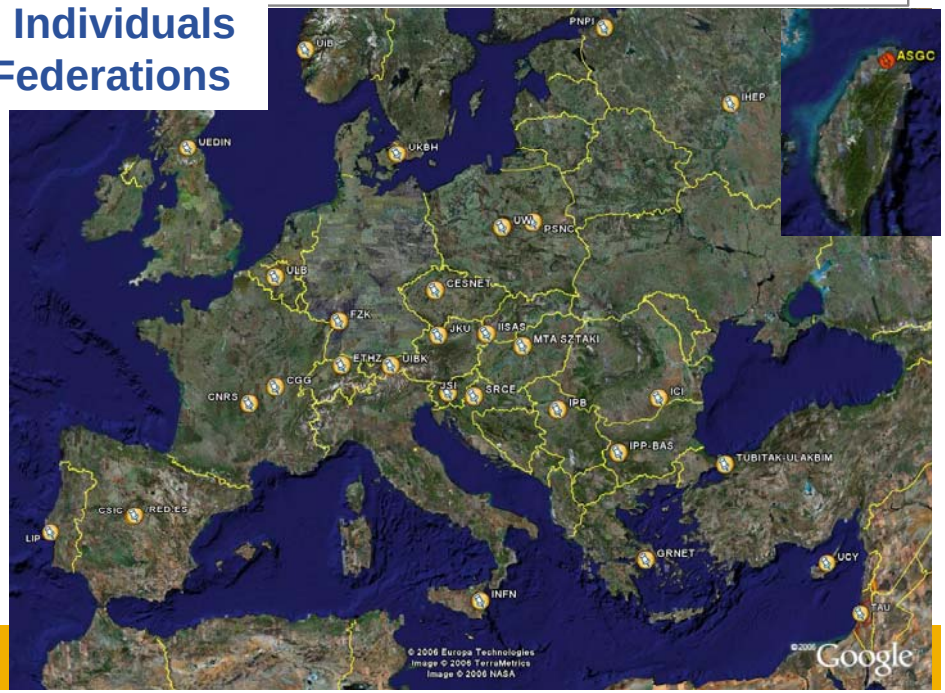
(snapshot 16 January 2007)

NA3 activity: User training and induction

- **Expand portfolio of training materials & courses**
- **Train a wide variety of EGEE users (internal/external)**
- **Develop effective mechanisms for training end-users of the EGEE infrastructure**
- **Collaborate in cross-activity initiatives**
  - ICEAGE Project Digital Library
  - http://library.iceage-eu.org/
  - Videos, MP3 talks on grid computing
- **http://www.egee.nesc.ac.uk/**
  - Training events
  - Training material repository
- **http://egee.lib.ed.ac.uk/**
  - EGEE Digital Library
  - Repository of training materials

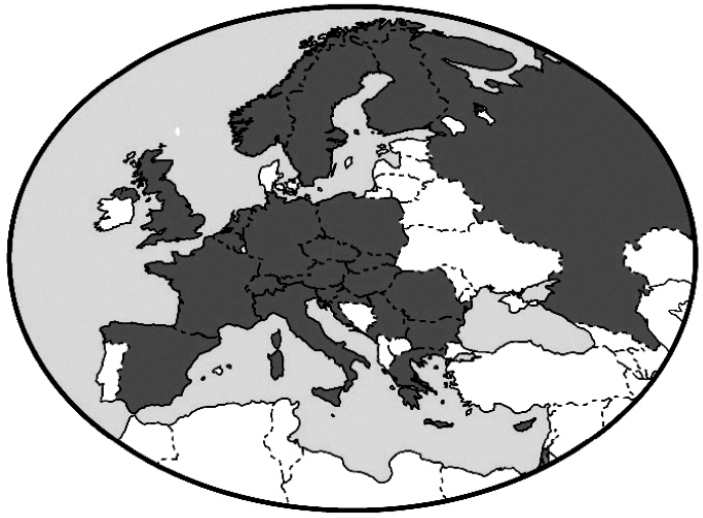


29 Active partners
~ 29 FTEs
89 Individuals
6 Federations

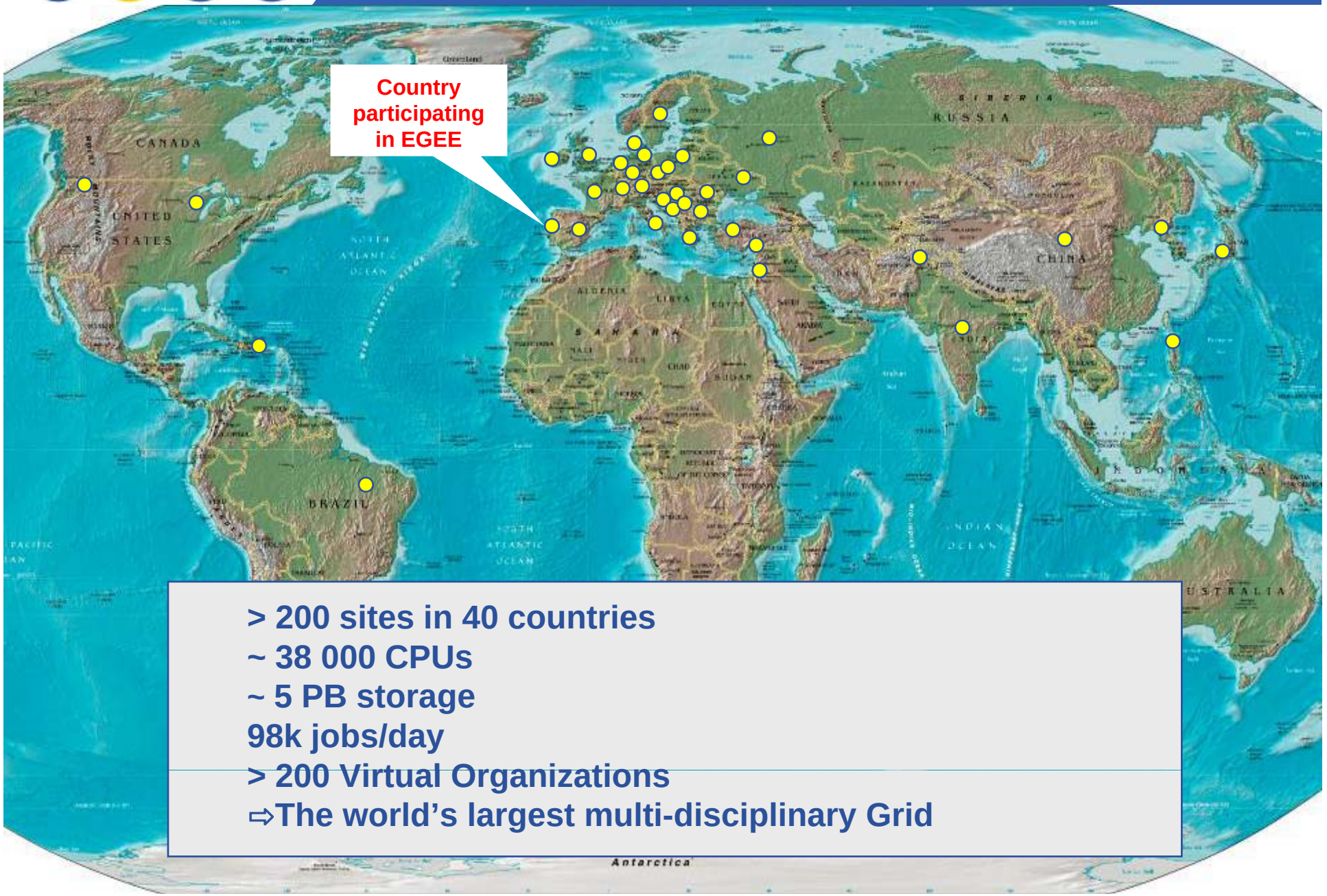


NA4 Activity: Application identification and support

- **Application Identification and Support (NA4)**
  - 25 countries, 40 partners, 280+ participants, 1000s of users
- **Support the large and diverse EGEE user community:**
  - **Promote dialog**: Users' Forums & EGEE Conferences
  - **Technical Aid**: Porting support, procedural issues
  - **Liaison**: Software and operational requirements
- **Main activities:**
  - 5 application clusers: HEP, Life sciences, Astronomy & astrophysics, Earth science, Computational chemistry, Fusion, Grid observatory
  - Support:
    - Application porting support
      www.lpds.sztaki.hu/gasuc
    - VO support
    - Direct user support
      www.ggus.org
    - Regional support
- **http://egeena4.lal.in2p3.fr**

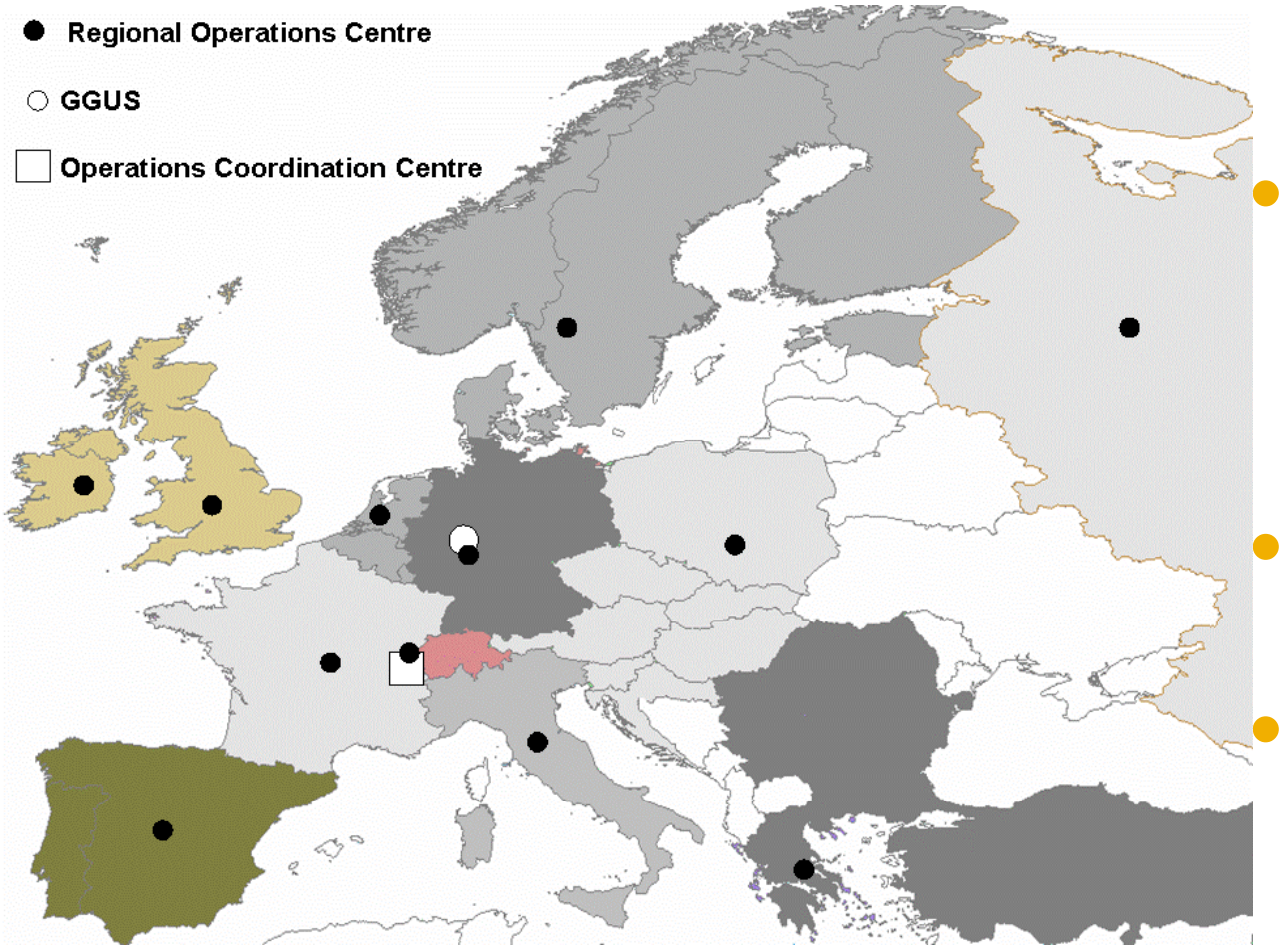


EGEE Infrastructure



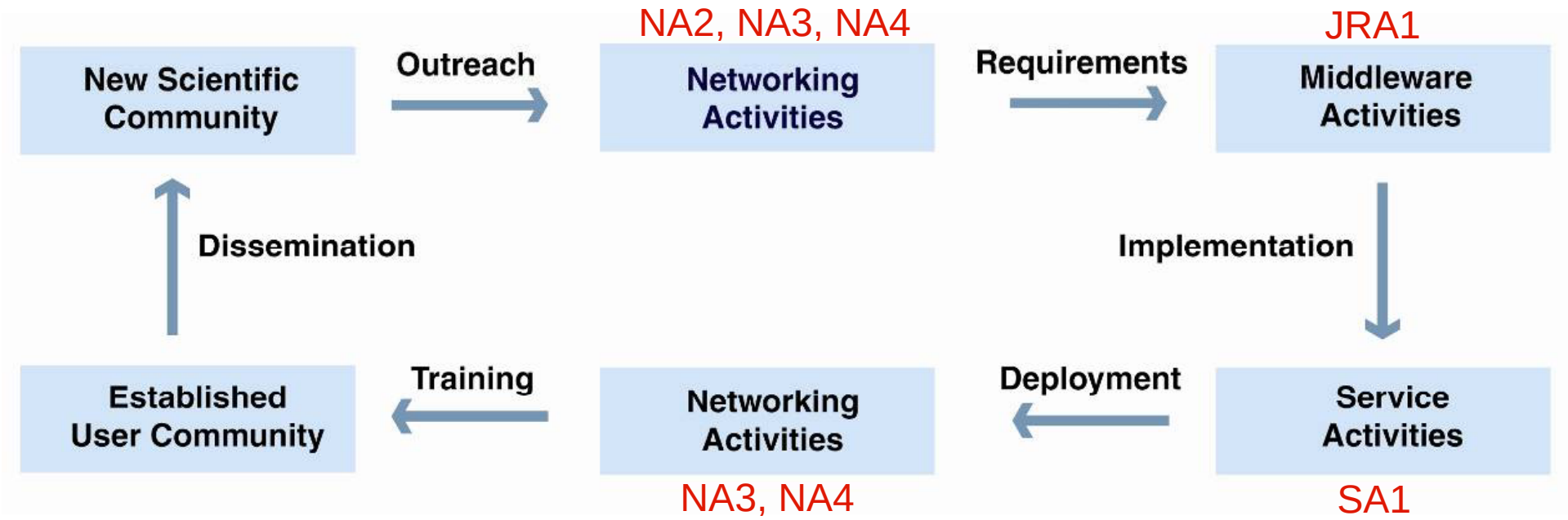
> 200 sites in 40 countries
~ 38 000 CPUs
~ 5 PB storage
98k jobs/day
> 200 Virtual Organizations
⇨The world's largest multi-disciplinary Grid

Resource management: structure



- **Operations Coordination Centre (OCC)**
  - management, oversight of all operational and support activities
- **Regional Operations Centres (ROC)**
  - providing the core of the support infrastructure, each supporting a number of resource centres within its region
  - **Grid Operator on Duty**
- **Resource centres**
  - providing resources (computing, storage, network, etc.);
- **Grid User Support (GGUS)**
  - At FZK, coordination and management of user support, single point of contact for users

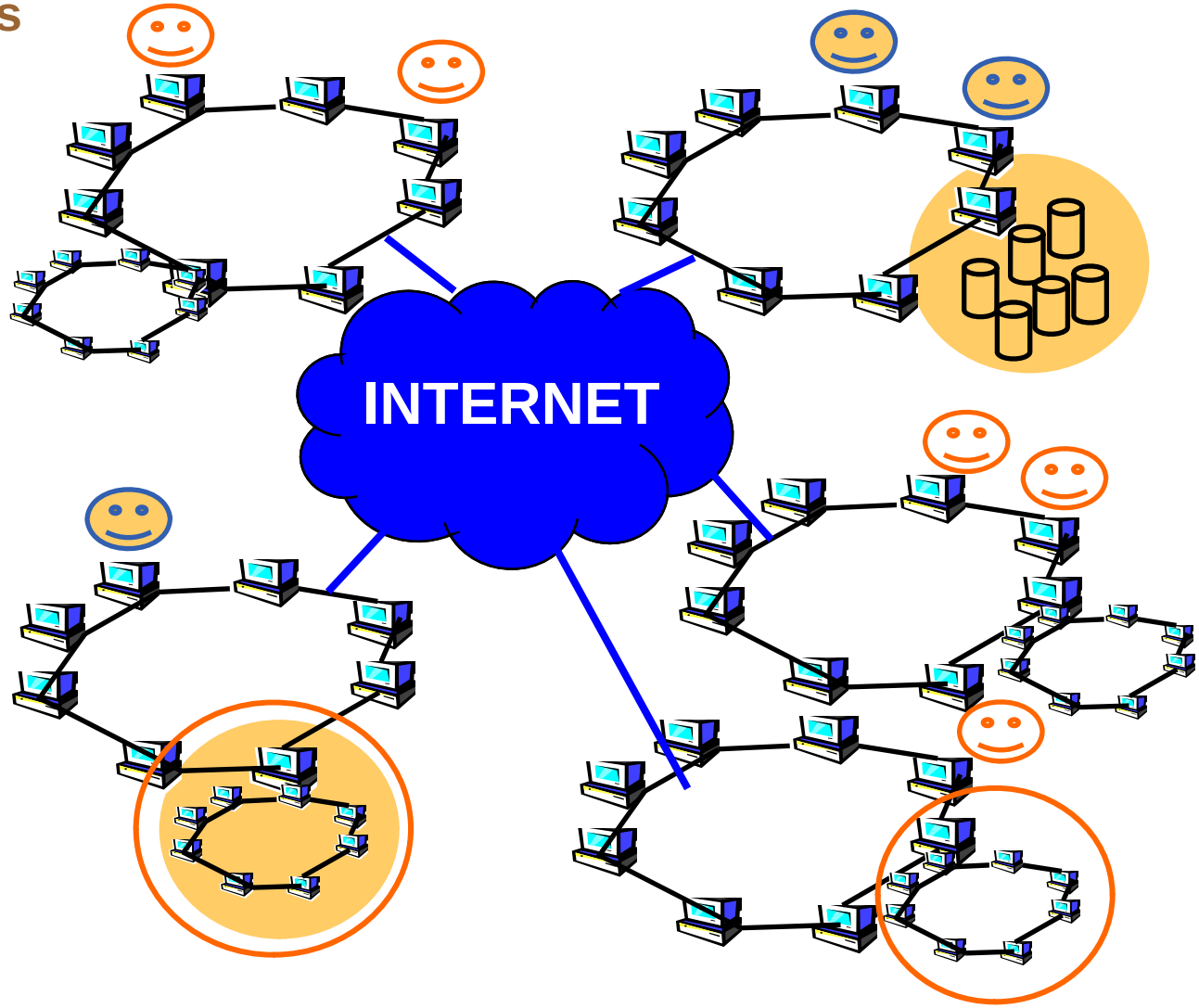
The EGEE Virtuous Cycle



Building effective user communities

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**



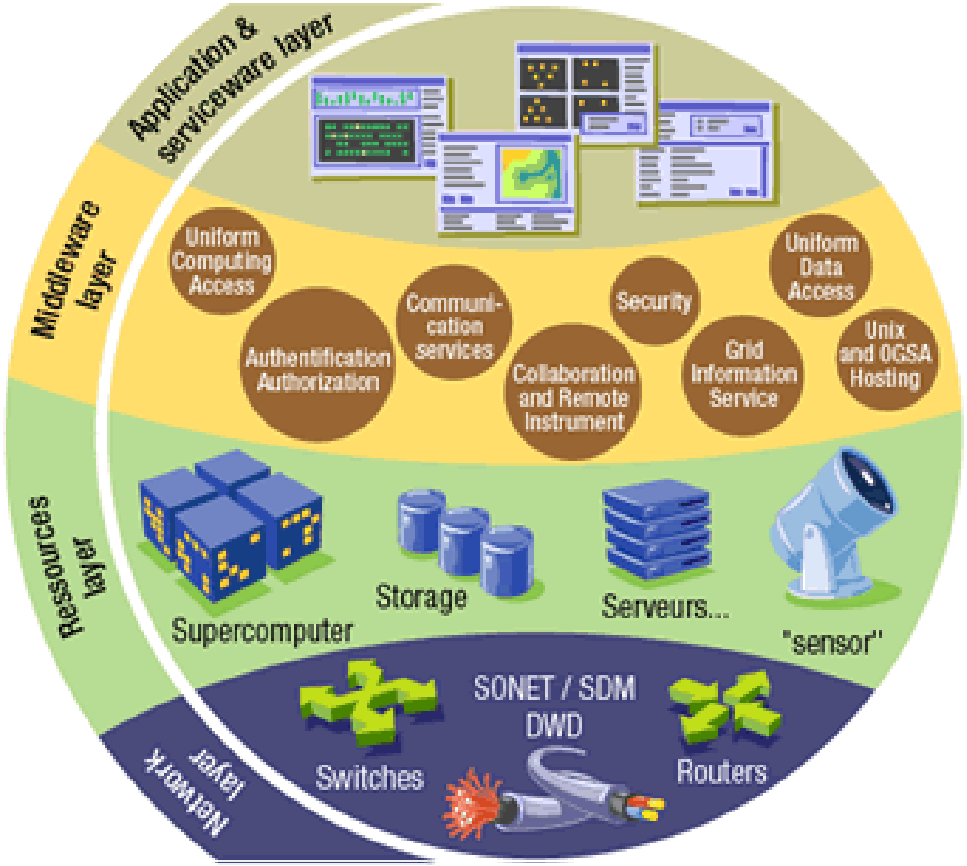
Outline

- **What is EGEE?**
  - The project
  - The infrastructure
- **gLite middleware**
- **EGEE applications**
- **Sources of further information**

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



Grid Middleware

- **When using a PC or workstation you**
  - Login with a username and password (“Authentication”)
  - Use rights given to you (“Authorisation”)
  - Run jobs
  - Manage files: create them, read/write, list directories
- **Components are linked by a bus**
- **Operating system**
- **One admin. domain**

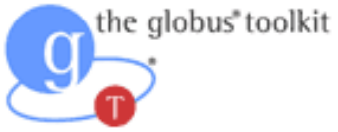
- **When using a Grid you**
  - Login with digital credentials – single sign-on (“Authentication”)
  - Use rights given you (“Authorisation”)
  - Run jobs
  - Manage files: create them, read/write, list directories
- **Services are linked by the Internet**
- **Middleware**
- **Many admin. domains**

EGEE Middleware: gLite

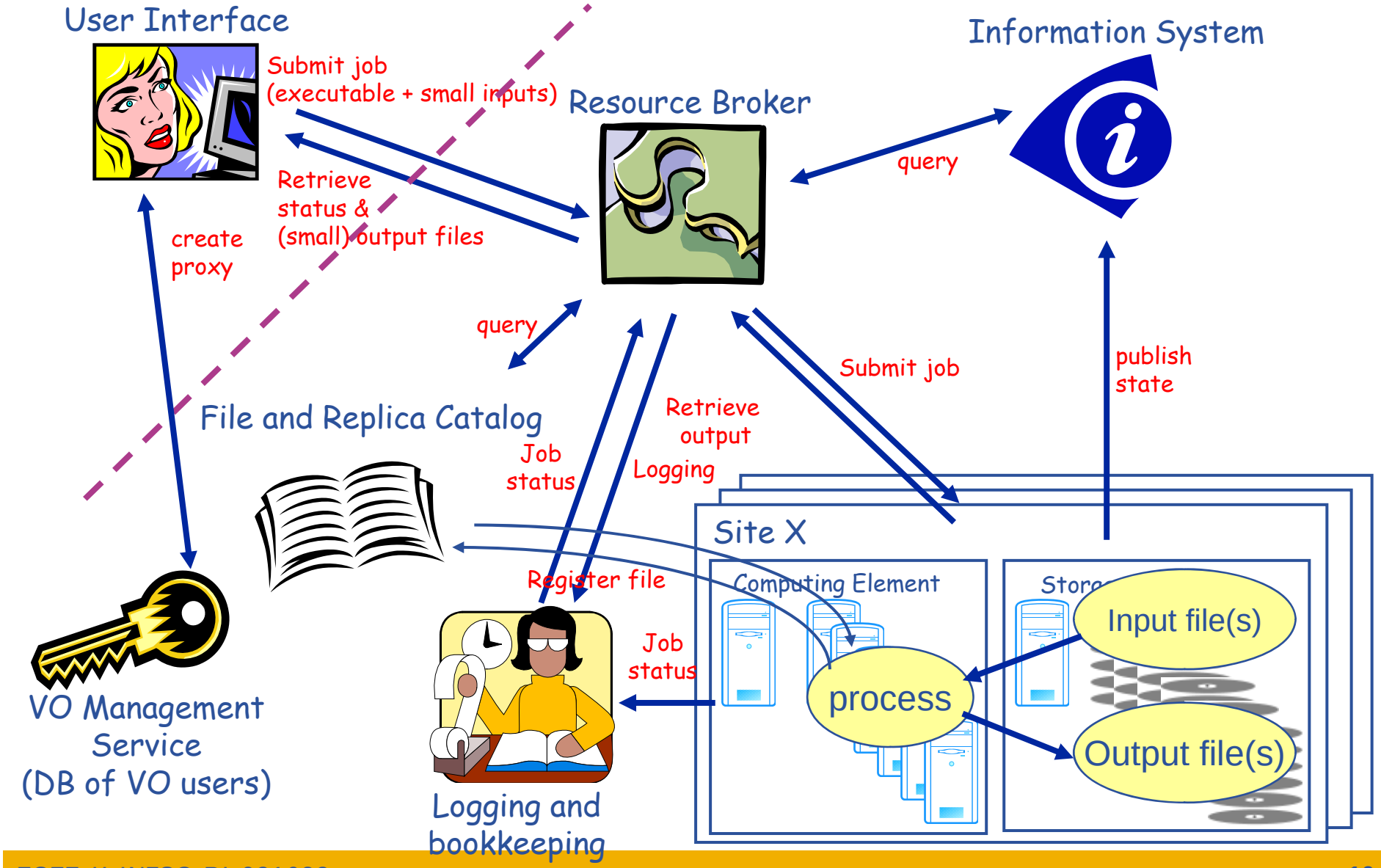
- **gLite 3.0**, **gLite 3.1**

⇨ Merger of LCG 2.7 and GLite 1.5

  - **Exploit experience and existing components from VDT (Condor, Globus), EDG/LCG, and others**
  - Develop a lightweight stack of generic middleware useful to EGEE applications (HEP and Biomedics are pilot applications).
    - Should eventually deploy dynamically (e.g. as a globus job)
    - Pluggable components – cater for different implementations
  - Focus is on providing a stable and usable infrastructure



Basic gLite use case: Job submission



Main components

User Interface (UI): The place where users logon to the Grid

Resource Broker (RB) (Workload Management System (WMS):
Matches the user requirements with the available resources on the Grid

Information System: Characteristics and status of CE and SE

File and replica catalog: Location of grid files and grid file replicas

Logging and Bookkeeping (LB): Log information of jobs

Computing Element (CE): A batch queue on a site's computers where the user's job is executed

Storage Element (SE): provides (large-scale) storage for files

Main components

User Interface (UI): The place where users logon to the Grid

Resource Broker (RB) (Workload Management System (WMS):
Matches the user requirements with the available resources on the Grid

Inf SE

Fil eplicas

Lo

All built upon authorisation, authentication, security

Computing Element (CE): A batch queue on ere
the user's job is executed

Storage Element (SE): provides (large-scale) storage for files

Who provides the resources?!

Service	Provider	Note
User interface	User / institute / VO	Computer with client SW
Resource Broker (WMS)	VOs - EGEE does not fund RBs	
Information System	Grid operations - EGEE funded effort	
File and replica catalog	VOs - EGEE does not fund catalogs	
Logging and Bookkeeping	VOs - EGEE does not fund LB servers	
Computing Element (CE)	VOs - EGEE does not fund CEs	VOs provide resources to match average need
Storage Element (SE)	VOs - EGEE does not fund SEs	VOs provide resources to match average need
External services	User / institute / VO	To extend the capabilities of the core infrastructure

Empowering VOs

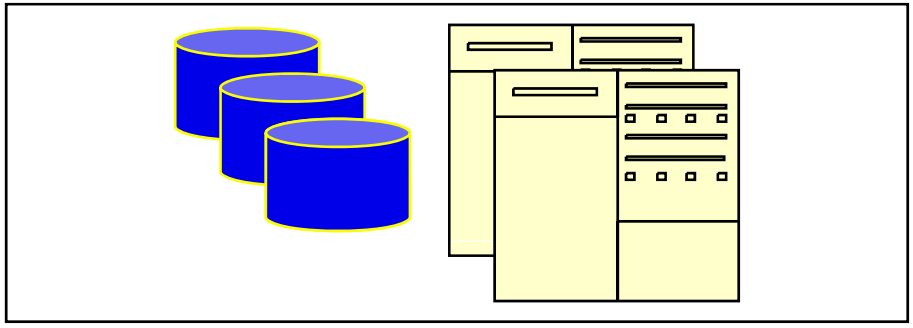
Application

Application toolkits

Command line & APIs

Higher-level gLite services (WMS,…)

Basic gLite services: CE, SE, info, security



Where computer science meets the application communities!

- Recommended External Software Packages for Egee CommuniTies
- Current RESPECT tools:
  - GridWay
  - P-GRADE Portal
  - GANGA
  - i2glogin
- http://egeena4.lal.in2p3.fr/ → "Grid software" menu

Production infrastructure contains these services

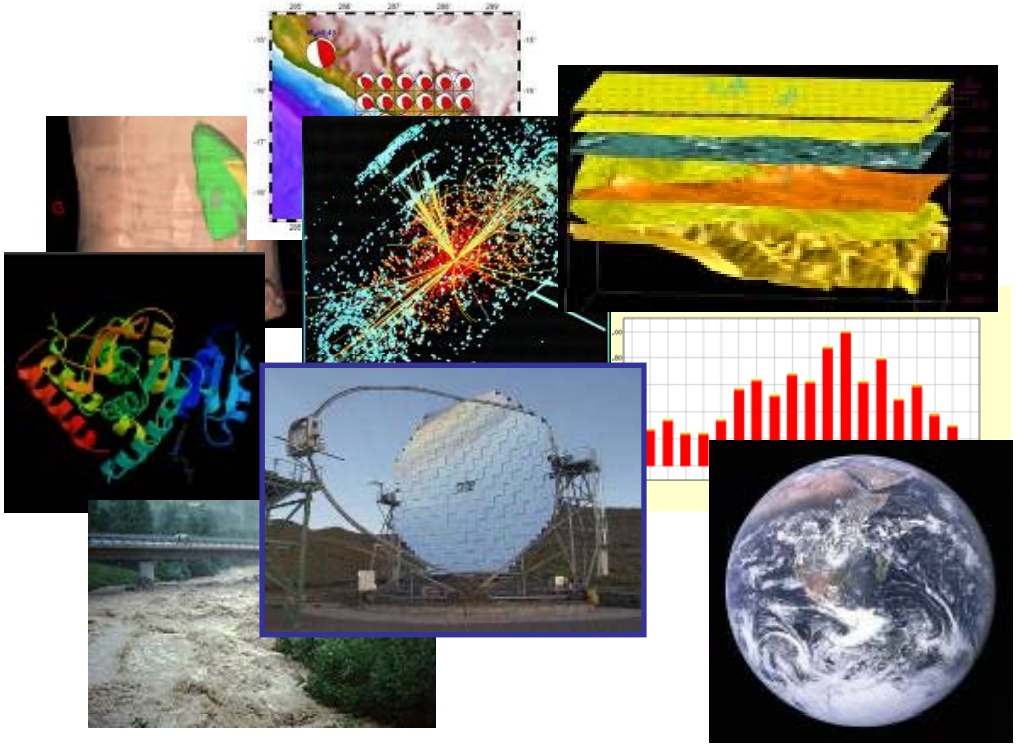
- Basic services: Must be complete and robust; Should not assume the use of Higher-Level Grid Services
- High level services: help the users building their computing infrastructure but should not be mandatory

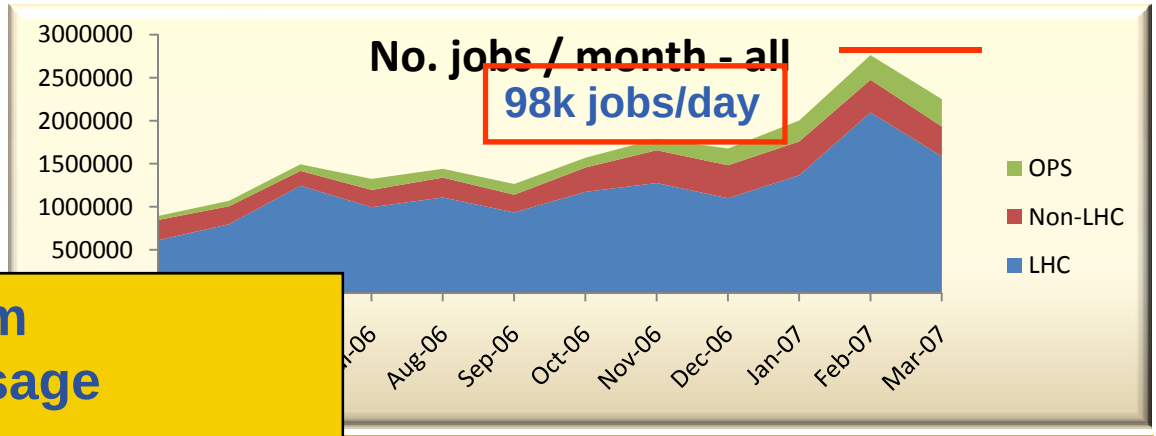
Outline

- **What is EGEE?**
  - The project
  - The infrastructure
- **gLite middleware**
- **EGEE applications**
- **Sources of further information**

Who is using EGEE?

- **>200 VOs from several scientific domains**
  - Astronomy & Astrophysics
  - Civil Protection
  - Computational Chemistry
  - Comp. Fluid Dynamics
  - Computer Science/Tools
  - Condensed Matter Physics
  - Earth Sciences
  - Fusion
  - High Energy Physics
  - Life Sciences
- **Further applications under evaluation**





Applications have moved from testing to routine and daily usage

~80-90% efficiency

Application families

- **Simulation**
  - Large number of similar, independent jobs – parameter study
- **Bulk Processing**
  - Widely-distributed input data, Sophisticated data management
- **Workflow**
  - Complex dependencies between individual tasks
- **Legacy Applications**
  - Licenses: control access to software on the grid
  - No recompilation ⇒ no direct use of grid APIs
- **Parallel Jobs**
  - Many CPUs needed simultaneously, Use of MPI libraries
  - *Limited support in gLite*: MPI configuration is not uniform
- **Responsive Apps.**
  - Short response time
  - *No real support in gLite* → Interactive Grid FP6 project

Further information, references

- EGEE
  - http://www.eu-egee.org/
- gLite middleware
  - http://www.glite.org
- gLite manuals, documentation
  - http://glite.web.cern.ch/glite/documentation/ (gLite user guide)
- Recommended External Software Packages for Egee CommuniTies (RESPECT)
  - http://egeena4.lal.in2p3.fr/
- Description of work of EGEE-III
  - https://edms.cern.ch/document/886385/4

Summary

- **EGEE is running the largest multi-VO grid in the world!**
  - Creating the “grid layer” in e-Infrastructure for research, public service and industry
- **Key concepts for EGEE**
  - Sustainability – planning for the long-term
  - Production quality
  - User support
- **EGEE’s middleware: gLite. Current version 3.0**
  - Basic middleware services
  - High level middleware services
- **External software to foster uptake of technology**