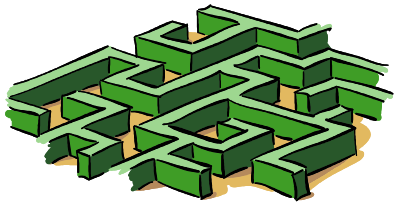
SuperB GRID: starting work

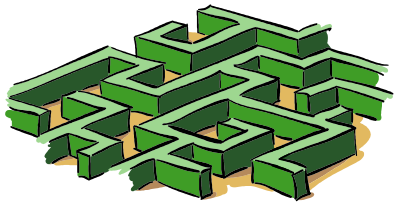
Armando Fella, INFN CNAF

SuperB Workshop, LAL, Orsay, February 15 -18 2009

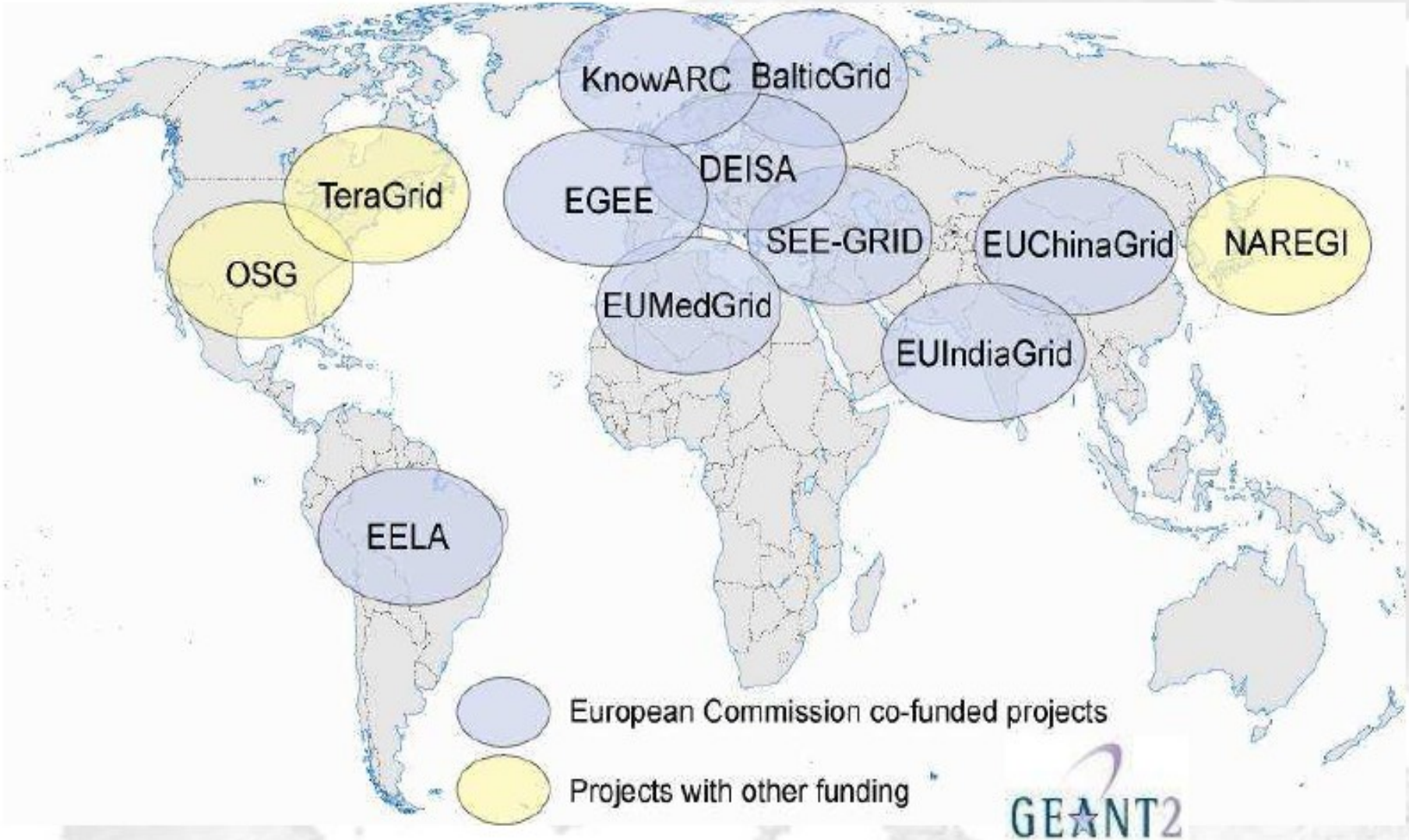


Presentation layout

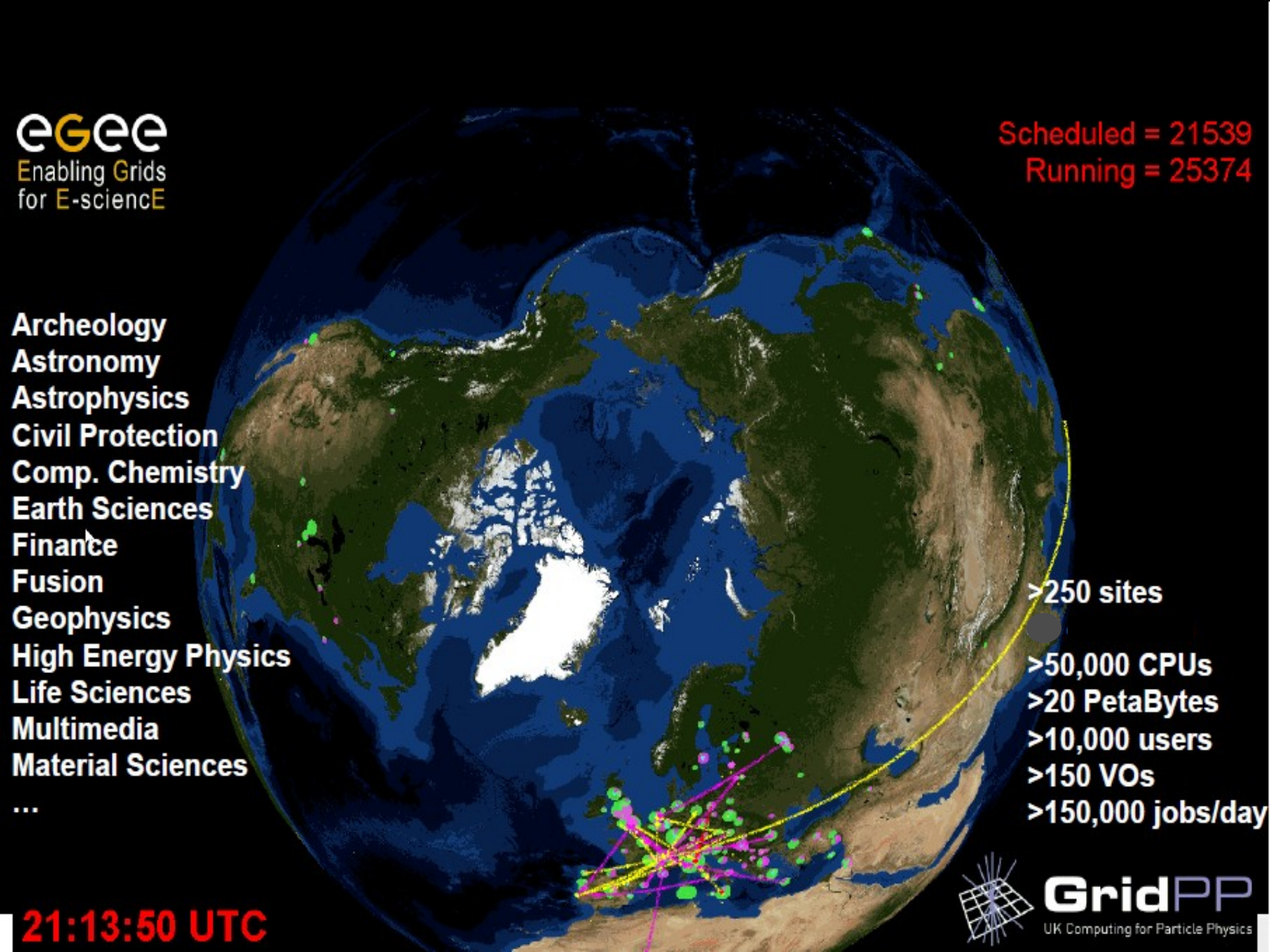
- GRID environment quick intro
  (thanks to A.Ghiselli for slides from CSFI08 presentation)
- SuperB grid remarks
- What has been done
- What to do
- Conclusion, references, training info



Collaborating e-Infrastructures



Potential for linking ~80 countries by 2008

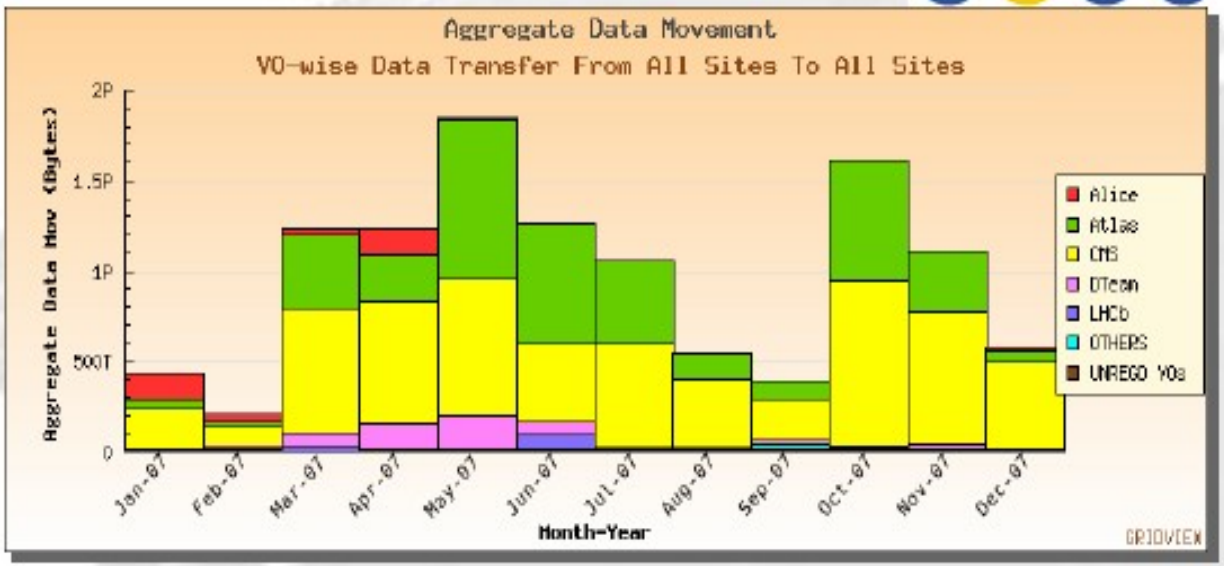

eGee
Enabling Grids
for E-sciencE
Scheduled = 21539
Running = 25374
Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences
...
>250 sites
>50,000 CPUs
>20 PetaBytes
>10,000 users
>150 VOs
>150,000 jobs/day
21:13:50 UTC
GridPP
UK Computing for Particle Physics

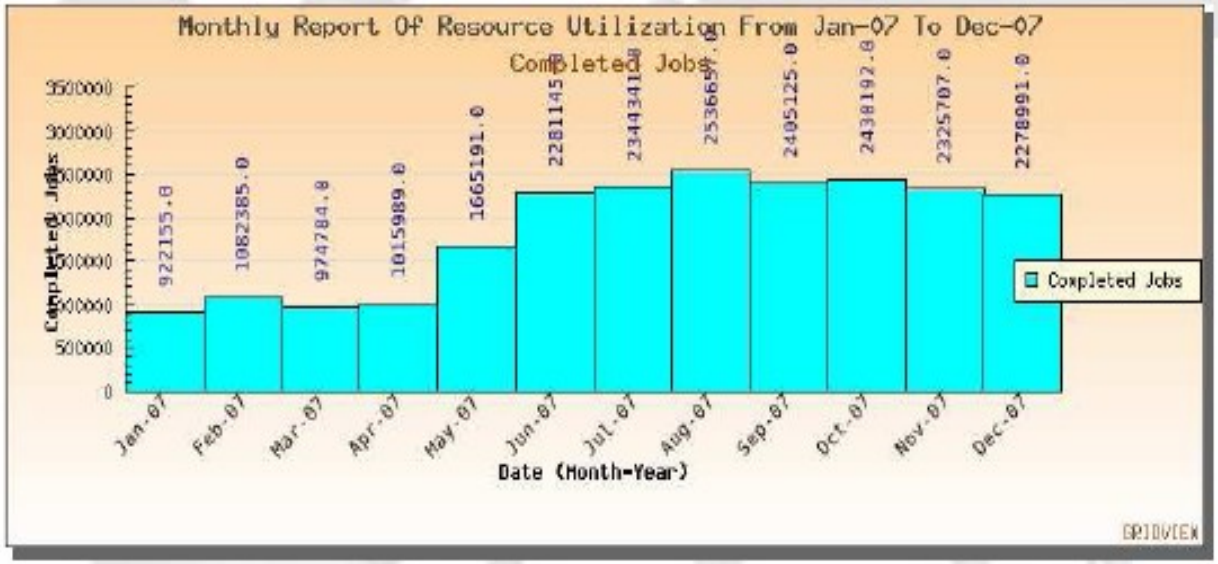
EGEE workload in 2007

Data:

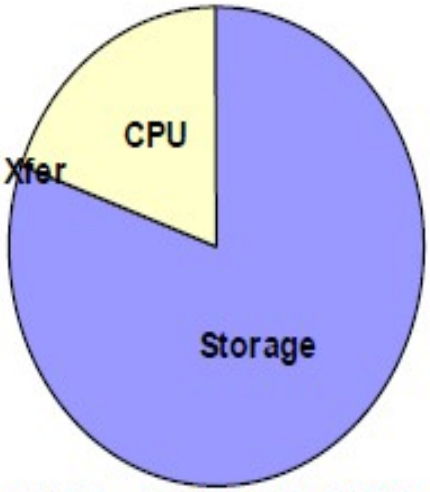
25Pb stored

11Pb transferred





CPU: 114 Million hours



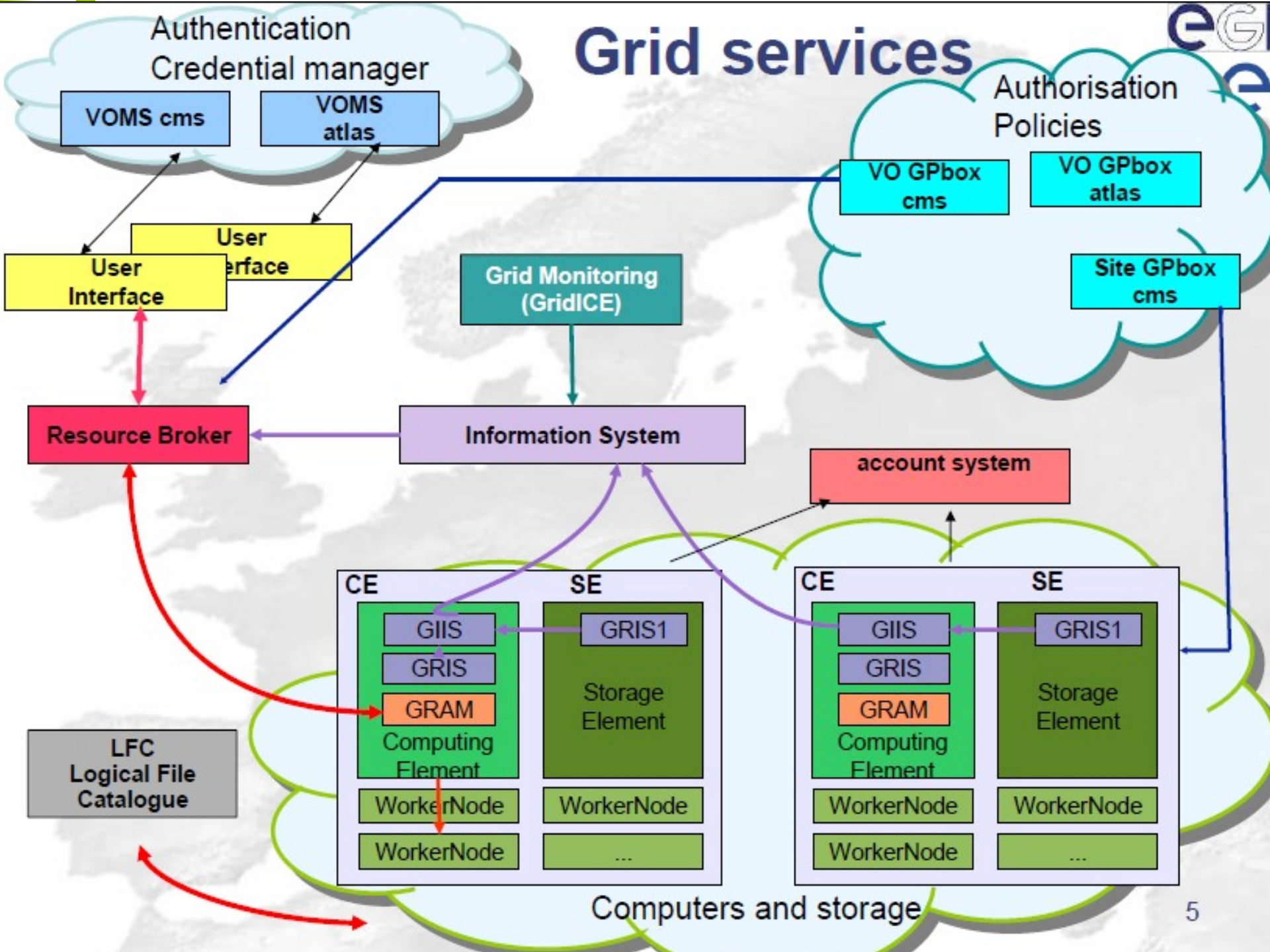
Estimated cost if performed with Amazon's EC2 and S3: £ 29,977,003

Daytime the Italian national GRID in EGEE

IT federation

- 1 country
- 5 Partners:
  - INFN
  - ENEA
  - SPACI-UNILE
  - SPACI-UNINA
  - SPACI-UNICAL
- 5 more RCs:
  - Istituto Tecnologie Biomediche – CNR/BARI (LIBI Project)
  - PERUGIA University
  - Istituto Linguistica Computazionale CNR-PISA
  - Scuola Normale Superiore – PISA
  - ESA-ESRIN

Grid services



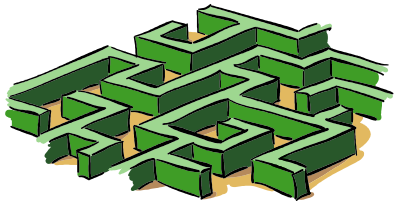
SuperB grid remarks

- The experiment is entering in grid world in quite a mature scenario in terms of main services readiness as job life management, data handling, monitor systems and mass data transfer.

- Several tools were developed and mantaind by LHC experiments (most) and LCG/EGEE organization to ensure the correct exploitation of resources.

- A new grid “client” now should start exploring in deep what has been already done, the choices, the reasons, measuring all this with the own specific necessities.

- Use the results of the study TO select the right tools/procedure and TO collaborate with “grid institutions” for the improvement of the ones leaking in some aspects. (see following slides on software management)



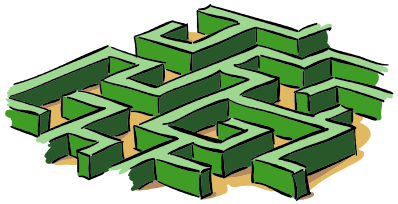
What has been done

- Official request of Virtual Organization (VO) creation → superb.org

***Virtual Organization**. A set of users sharing the same application environment. They need application-specific software installed on their testbed, a specific user/machine environment and access to a given set of resources.*

- Resources evaluation with CNAF Tier 1 centre:
  - during the startup period BaBar and SuperB collaborations will share the CPU time and Disk space resouces.

- The BaBar resources @ CNAF are configured to permit the access to SuperB software ($SBROOT). During this workshop we'll discuss and try to submit the first simulation job to CNAF farm.



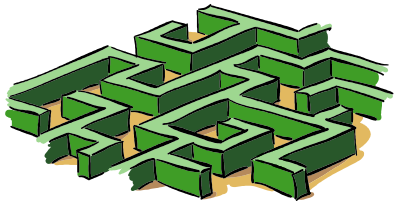
VO Creation: superb.org

The effective enabling of base grid services is expected during March 09. It includes the following specific services shared with other VOs:

- Virtual Organisation Membership System (**VOMS**) activation that will manage the user authentication via Public Key Infrastructure.
- Enabling User Interface (**UI**) access machine.
- Enabling the Logical File Catalog (**LFC**) service for data handling tasks involving Storage Elements (**SE**).
- Update the Grid Information System (**GIS**) of all the INFNGRID sites involved in SuperB.

Virtual Organisation roles:

VO Manager: Armando Fella, Luca Tomassetti
Software Manager: Roberto Stroili, Armando Fella
Resource Manager: Eleonora Luppi, Fabrizio Bianchi



VO Creation: superb.org II

The “reletionship” between the VO and the EGEE project is ruled by precise Statements. The following is extrected by https://edms.cern.ch/document/503245

“Participation in the EGEE project and use of the EGEE infrastructure is a collaborative effort. A VO using this infrastructure is required to:

- Contribute computational resources to the EGEE/LCG infrastructure corresponding approximately to the average needs of the VO for large-scale, production use.
- Help drive the evolution of the infrastructure and the middleware through use of the system and by providing feedback.

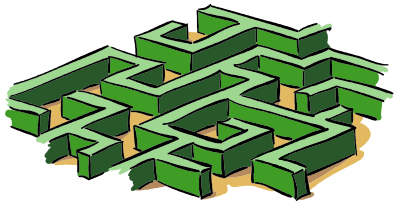
The EGEE project will provide:

- generic user support system,
- generic application porting support, and
- standard training sessions and tutorials

to all registered VOs. Those VOs with significant scientific interest may negotiate higher levels of resources and support with EGEE.”

What to do

- Short term:
  - Test the software setup at CNAF to permit the exploiting of resources via standard batch system
  - Tuning/test the systems: we'll use the GILDA dedicated VO to test all the grid functionalities
  - Cooperate with GRID people in R&D and deployment/test new solutions
  - First SuperB GRID Simulation job should work!
  - Computing model definition (begin)
- Mid term
  - Training session for users! → GILDA school (see references slide)
  - Data Handling issues: how to move physics data around the grid
  - Cooperate with GRID people in R&D and deployment/test new solutions
  - Computing model definition (finalize)



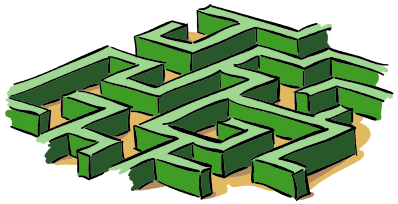
Collaboration @ CNAF

A discussion with CNAF farming group is in progress about experiment's software management in grid environment.

The problem: job initialization (loading libraries) results to impact the Wall Clock Time, the network and storage infrastructure with a potential denial of service effect.

The proposed solution regards the functionally address separation of the software components.

The goal is to define a standard to be adopted by INFNGRID/EGEE project.



References

Main projects:

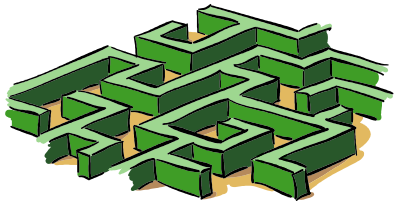
INFN production grid http://grid-it.cnaf.infn.it/
LHC computing grid http://lcg.web.cern.ch/LCG/
EGEE portal http://www.eu-egee.org/
How to create a VO http://grid-it.cnaf.infn.it/index.php?id=865&type=1

Training, e-learning:

Gilda https://gilda.ct.infn.it/
Grid dictionary http://grid.infn.it/modules/myiframe/index.php?iframeid=30

Examples of tools by experiments:

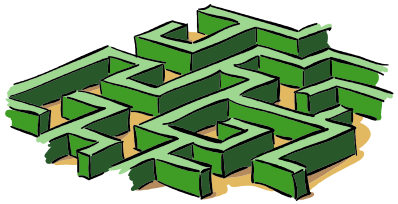
CMS Crab https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideCrab
CMS Phedex http://cmsweb.cern.ch/phedex/
ALICE MonALISA http://pcalimonitor.cern.ch/map.jsp
ATLAS Panda https://twiki.cern.ch/twiki/bin/view/Atlas/PanDA



Questions ?

BACKUP Slides:

- What is a Grid
- EU e-infrastructure: a success story
- Supported job types
- Users and resources distribution



What is a Grid

In 1998, I. Foster and C. Kesselman wrote in The Grid:

Blueprint for a New Computing Infrastructure:

"A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive access to highend computational capabilities."

In 2002, I. Foster wrote in *What is the Grid? A Three*

***Point Checklist*:**

"... The essence of the [definition] can be captured in a simple checklist,... a Grid is a system that:

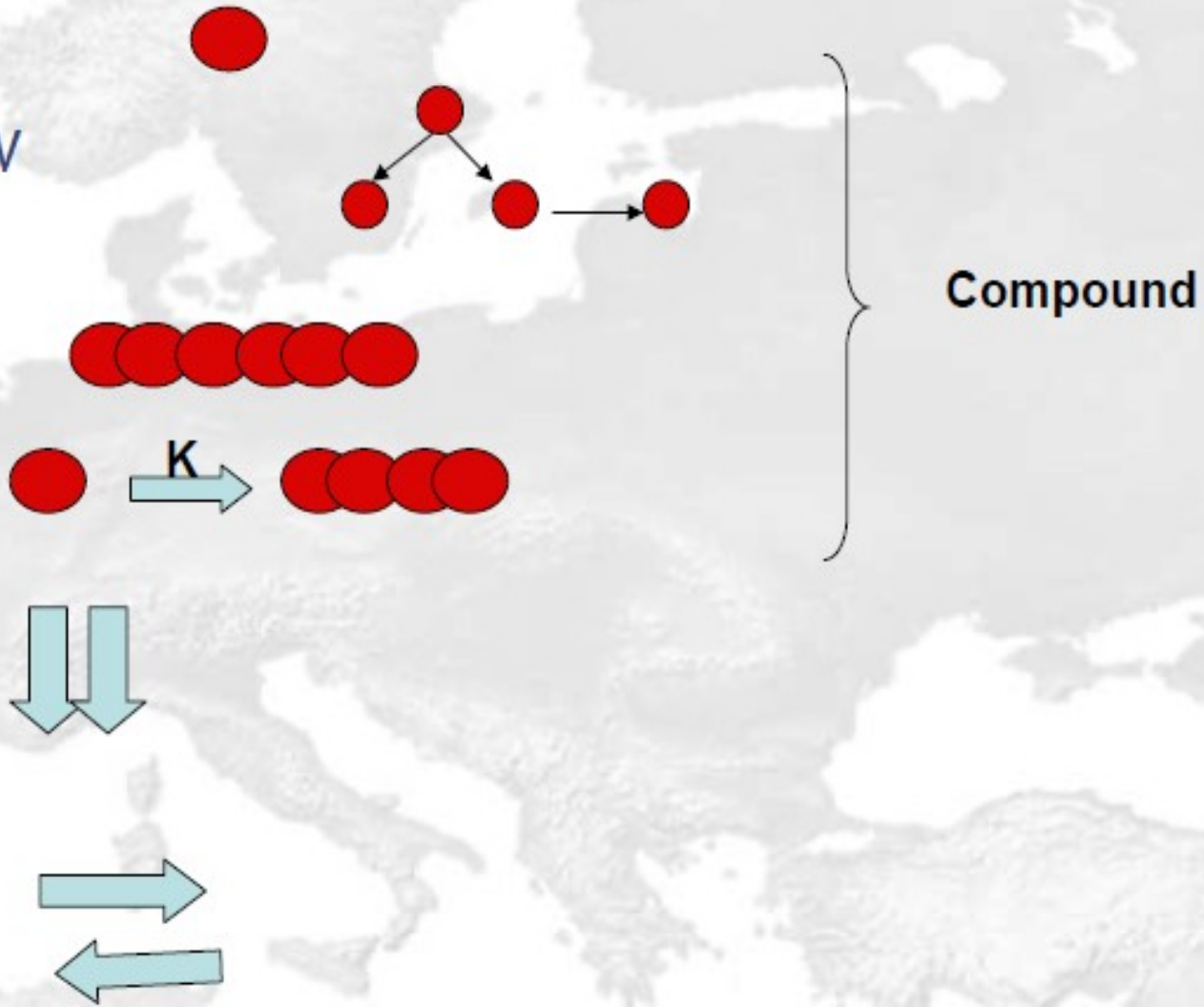
1. *coordinates resources that are not subject to centralized control…*
2. *… using standard, open, general purpose protocols and interfaces...*
3. *… to deliver nontrivial qualities of service.*

EU e-Infrastructure: a success story

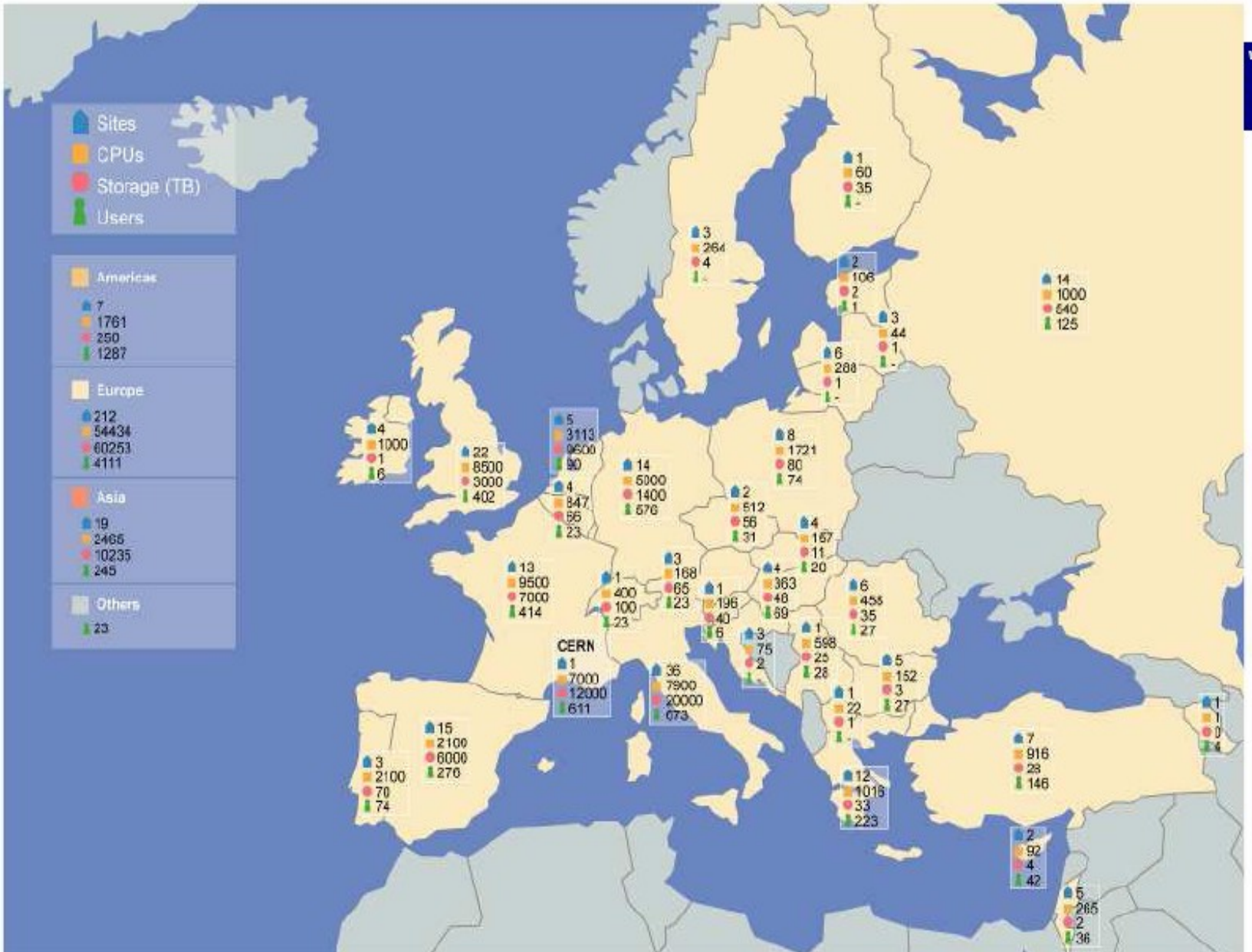
- **Hundreds of M€ of EC and National funds (2001-08) have got a set of top level robust Open Source Grid Middleware Services (e.g.the gLite release) providing a solid foundation for the Researchers activities**
- **Deployed in the world largest Grid e-Infrastructure (= Internet + Grid): the one of EGEE (Enabling Grid and E-sciencE ->EGEE II -> EGEE III),**
- **Real Services, not prototypes, daily offered by federations of well managed resource centers to thousands of users of several research communities**
  - gLIte today: 150K Jobs/day succesfull completed in 250 centers,
  - up to >10 GByte/sec of aggregated sustained access to scientific data
  - transparent access to ~ 20 PB of Storage
  - International Standards (Web Services, W3C, OGF, OASIS), high level of security (PKI, X509), interoperability of different implementations (ARC UNICORE)
- **EU e-Infrastructure is ahead in this domain**
- **Members States and EC are now guaranteeing the long term sustainibility of the baseline Grid MW and e-Infrastructures with the**
  - European Grid Initiative EGI (FP7 Project: EGI-Design Study) and the National Grid Initiatives (NGIs)

Supported Job Types

- Batch-like
- Dag workflow
- Collection
- Parametric
- MPI
- Interactive



Users and resources distribution



Architettura e servizi di Grid e confronto con l'architettura di Internet