

# Introduction to the EGEE project



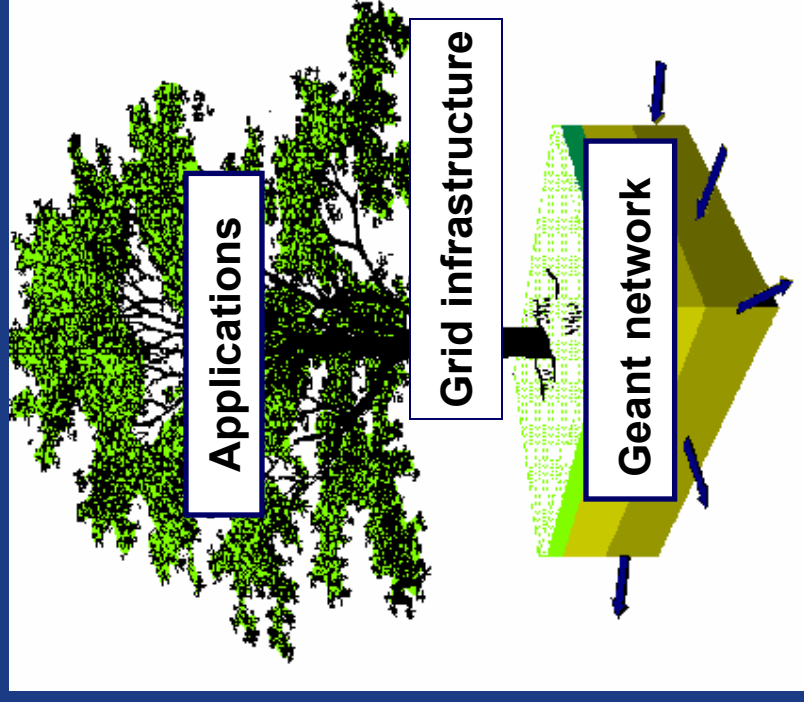
Fabrizio Gagliardi  
EGEE Project Director

**EGEE**  
Enabling Grids for  
E-science in Europe

EGAAP, 14 June 2004  
CERN, Geneva

# What is EGEE ? (I)

- EGEE (Enabling Grids for E-science in Europe) is a seamless Grid infrastructure for the support of scientific research, which:
  - Integrates current national, regional and thematic Grid efforts, especially in HEP (High Energy Physics)
  - Provides researchers in academia and industry with round-the-clock access to major computing resources, independent of geographic location



# What is EGEE ? (II)

- 70 leading institutions in 27 countries, federated in regional Grids
- 32 M Euros EU funding (2004-5), O(100 M) total budget
- an ultimate combined capacity of over 20000 CPUs (the largest international Grid infrastructure ever assembled)
- ~ 300 persons



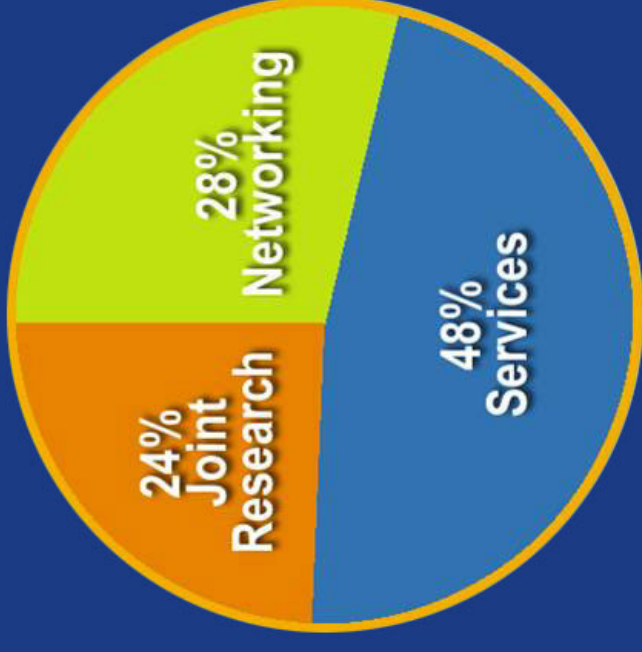
# What will EGEE provide?

- **Simplified access** (access to all the operational resources the user needs)
- **On demand computing** (fast access to resources by allocating them efficiently)
- **Pervasive access** (accessible from any geographic location)
- **Large scale resources** (of a scale that no single computer centre can provide)
- **Sharing of software and data** (in a transparent way)
- **Improved support** (use the expertise of all partners to offer in-depth support for all key applications)



# EGEE Activities

- Emphasis on operating a production grid and supporting the end-users
- **48 % service activities** (Grid Operations, Support and Management, Network Resource Provision)
- **24 % middleware re-engineering** (Quality Assurance, Security, Network Services Development)
- **28 % networking** (Management, Dissemination and Outreach, User Training and Education, Application Identification and Support, Policy and International Cooperation)





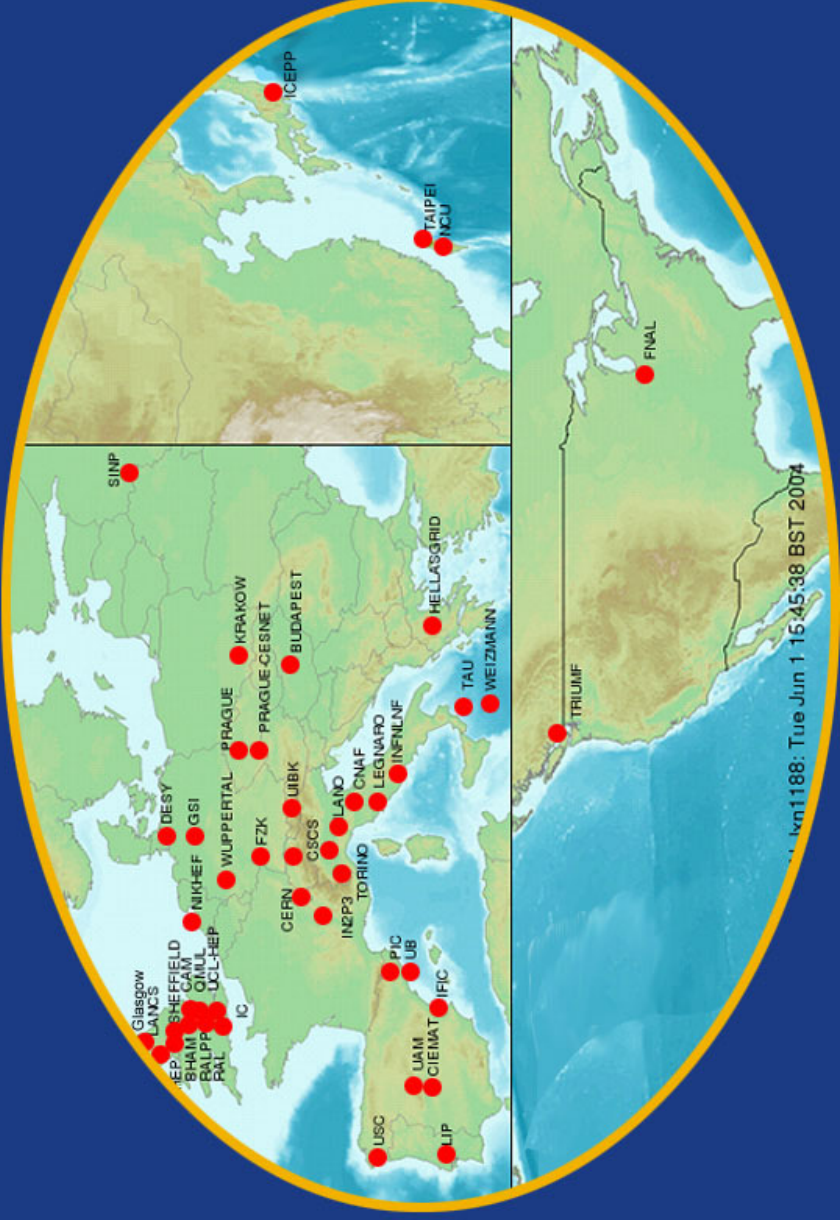
# EGEE infrastructure

- Access to networking services provided by **GEANT** and the **NRENs**
- Production Service:
  - in place (based on HEP LCG-2)
  - for production applications
  - MUST run reliably, runs only proven stable, debugged middleware and services
  - Will continue adding new sites in EGEE federations
- Pre-production Service:
  - For middleware re-engineering
- Certification and Training/Demo testbeds



# First EGEE infrastructure

- Based on HEP-LCG testbed: more than 50 sites worldwide



# EGEE Operations

- **Operation Management Centre**
  - located at CERN, coordinates operations and management
  - coordinates with other grid projects
- **Core Infrastructure Centres**
  - behave as single organisations
  - operate infrastructure services
- **Regional Operation Centres**
  - first point of contact for new users and user support





# EGEE Middleware Activity

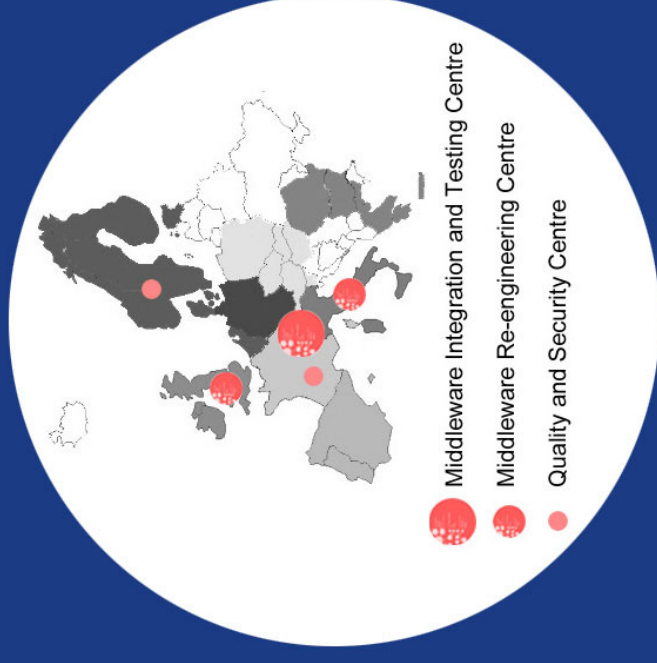
- Middleware selected based on requirements of Applications and Operations
- Harden and re-engineer existing middleware functionality, leveraging the experience of partners
- Provide robust, supportable components
- Support components evolution (WS-RF)



Enabling Grids for  
E-science in Europe

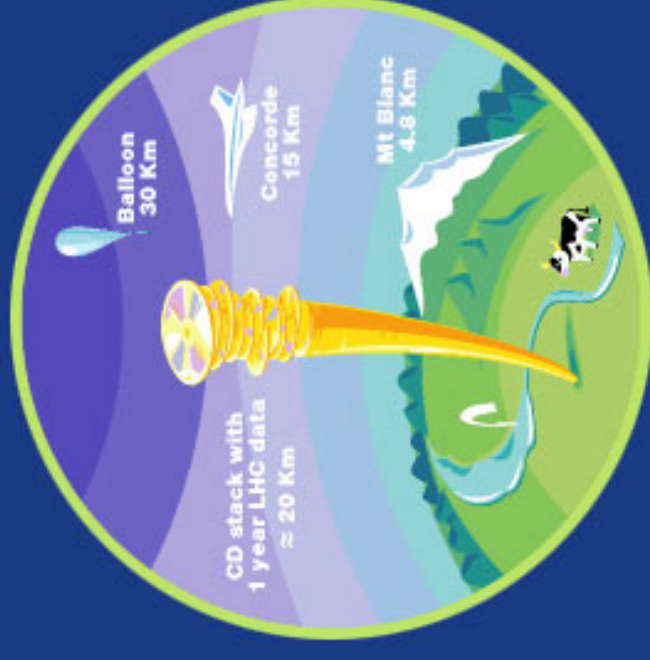
# EGEE Middleware Implementation

- Activity concentrated in few major centers and organized in “Software clusters”
- Production grid service running HEP LCG-2 grid middleware
- In parallel develop a “next generation” grid facility
- migrate LCG-2 to new middleware in 2005



# EGEE Pilot Applications (I)

- HEP
  - Have been running large distributed computing systems for many years
  - Now focus on computing for LHC → hence LCG (LHC computing grid project)
  - other current HEP experiments use grid technology (Babar, CDF, D0..)
  - LHC experiments are currently executing large scale data challenges (DCs)



# EGEE Pilot Applications (II)

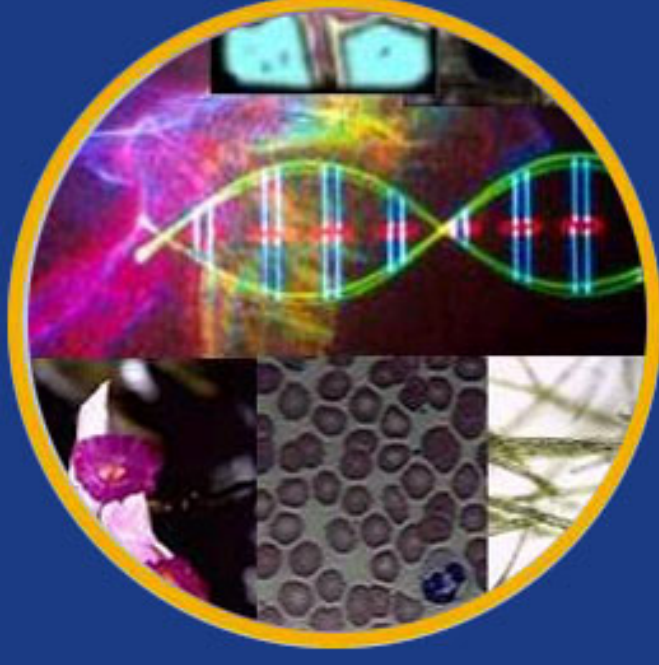
- **Biomedics**
  - Bioinformatics (gene/proteome databases distributions)
  - Medical applications (screening, epidemiology, image databases distribution, Parallel algorithms for medical image processing, simulation, etc)
  - Interactive application (human supervision or simulation)
  - Security/privacy constraints
    - Heterogeneous data formats (genomics, proteomics, image formats)
    - Frequent data updates
    - Complex data sets (medical records)
    - Long term archiving requirements





# How to access EGEE (I)

- 0) Review information provided on the EGEE website ([www.eu-egee.org](http://www.eu-egee.org))
- 1) Establish contact with the EGEE applications group lead by Vincent Breton ([breton@clermont.in2p3.fr](mailto:breton@clermont.in2p3.fr))
- 2) Provide information by completing a questionnaire describing your application
- 3) Applications selected based on scientific criteria, Grid added value, effort involved in deployment, resources consumed/contributed etc.



# How to access EGEE (II)

- 4) Follow a training session
- 5) Migrate application to EGEE infrastructure with the support of EGEE technical experts
- 6) Initial deployment for testing purposes
- 7) Production usage (contribute computing resources for heavy production demands)



# Moving your application to EGEE (I)

- **Data Intensive**
  - Access to diverse data sources (format, read/write, location etc.)
  - Quantity of data
- **Compute Intensive**
  - EGEE attracts mostly farms of commodity PCs
  - MPI available for distributed applications at many sites
  - Interface to DEISA for application migration is under discussion
- **Interfaces**
  - Standard interfaces provided (e.g. APIs, GENIUS portal)
  - Application specific interfaces can be linked to the infrastructure (DEVASPIIM, HKIS, BioGrid)
  - Interactivity



Enabling Grids for  
E-science in Europe

# Moving your application to EGEE (II)

- Security
  - Infrastructure can help control access to sites, data, network and information
- EGEE sites are administered/owned by different organisations
  - Sites have ultimate control over how their resources are used
  - Limiting the demands of your application will make it acceptable to more sites and hence make more resources available to you



Enabling Grids for  
E-science in Europe



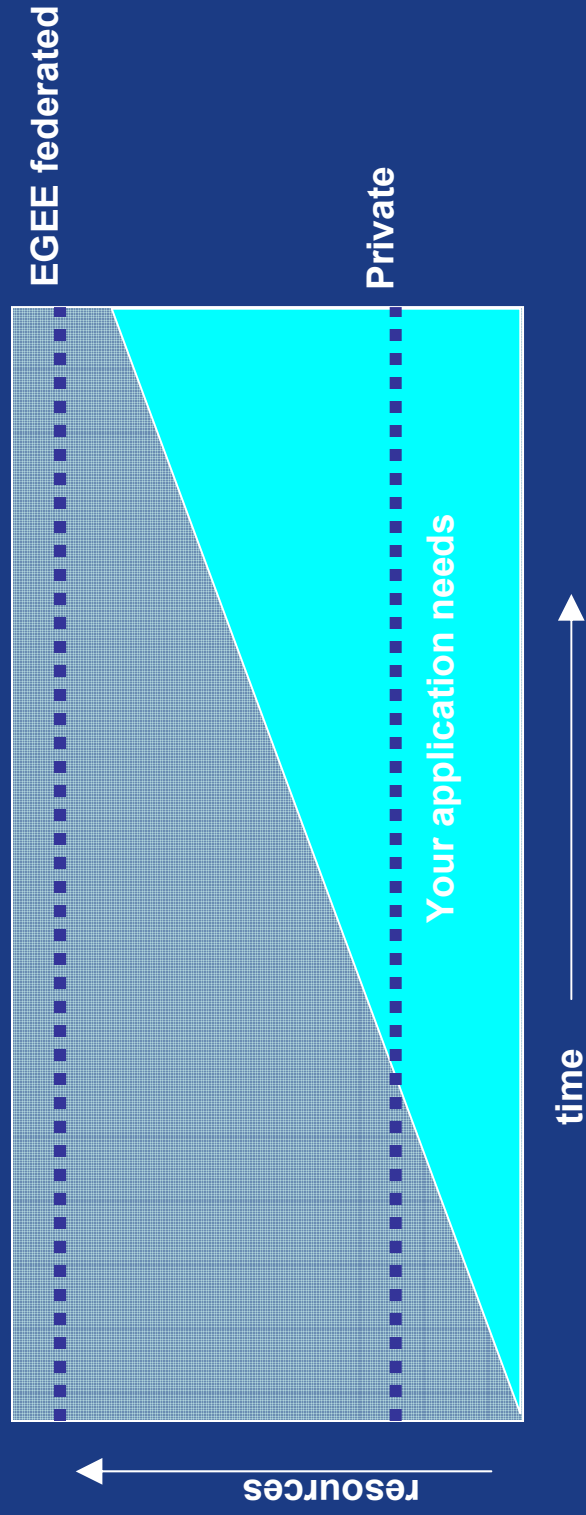
# Security & Intellectual Property (I)

- The existing EGEE grid middleware is distributed under an Open Source License developed by EU DataGrid
  - No restriction on usage (scientific or commercial) beyond acknowledgement
  - Same approach for new middleware
- Application software maintains its own licensing scheme
  - Sites must obtain appropriate licenses before installation



# Security & Intellectual Property (II)

- For applications that must operate in a closed environment EGEE middleware can be downloaded and installed on closed infrastructures



# EGEE Plans

- **two-year project** conceived as part of a four-year programme
- resources and user groups will **rapidly expand** during the course of the project
- **~3000 users** active from at least five disciplines by the end of the second year
- **from over 3000 CPUs** at the outset of the project to **over 8000** by the end of the second year
- A second two-year project is anticipated to follow on from EGEE, in which **industry** will progressively take up operations and maintenance



# Conclusions

- EGEE is the first attempt to build an international and worldwide Grid infrastructure for data intensive science
- Similar aim to NSF CyberInfrastructure initiative in the US
- EU playing a pioneering role with a substantial first two year funding
- Important to develop a long term support strategy



## Further info

- EU EGEE – [www.eu-egee.org](http://www.eu-egee.org)
- EU DataGrid – [www.eu-edg.org](http://www.eu-edg.org)
- Other Grid projects - [www.gridstart.org](http://www.gridstart.org)
- The Grid - [www.gridcafe.org](http://www.gridcafe.org)