Enabling Grids for E-science in Europe

Information Society Technologies

What is EGEE?

John Murison,
EGEE Training Team

EGEE is funded by the European Union under contract IST-2003-508833

Acknowledgements

This talk is based on presentations by

Dr Fabrizio Gagliardi, EGEE Project Director and

Dr Bob Jones, EGEE Technical Director.

Enabling Grids for E-science in Europe

Goals of this module

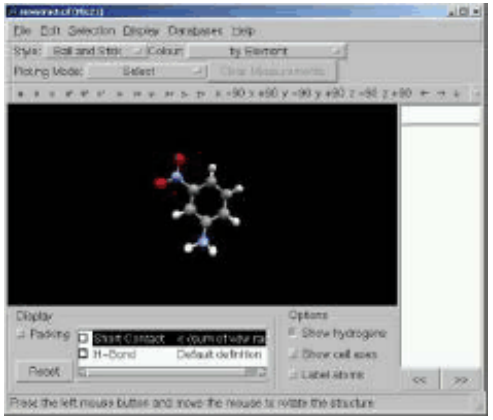
To provide a description of the aims of the Grid Research infrastructure project EGEE, and the context within which it has been undertaken.

The terms of the problem

- Technological progress produces more sophisticated digital sensors (particle physics detectors, satellites, radio-telescopes, synchrotrons…)
- Much of science is therefore becoming increasingly “data-intensive”
- Huge amounts of data need to be analyzed by large and geographically distributed scientific communities
- Consequently, single computers, clusters or supercomputers are not powerful enough for the necessary calculations and the data processing

Result:

access to large facilities is difficult and expensive for the scientific community, particularly in less favoured countries
=> increase of the “electronic divide”



Technological push?

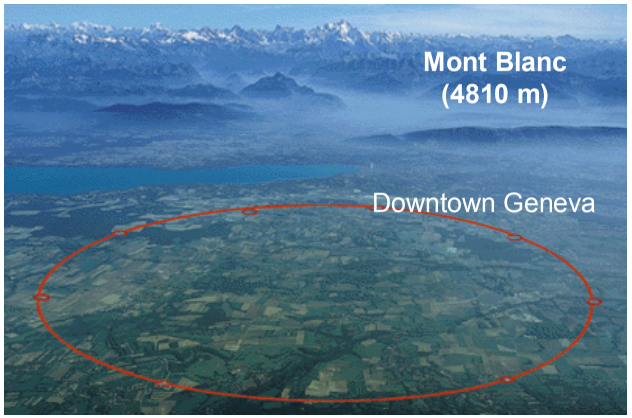
- Is Grid technology merely a ‘funding opportunity’?
- Is it an example of scientists wanting to do something because they (just about) could?
- Is it in fact a technology driven activity, without any real purpose?
- Consider this diagram, and what it implies…

CERN: an example of data intensive science and a large international facility

- CERN is building the Large Hadron Collider (LHC) the most powerful instrument ever built to investigate elementary particles physics
- Data Challenge:
  - One Megabyte of data digitised for each collision
  - 10^{10} collisions recorded each year = 10 Petabytes/year of data !!!
  - LHC data correspond to about **20 million CDs each year!**

- Simulation, reconstruction, analysis: LHC data handling requires a computing power equivalent to ~100,000 of today's fastest PC processors!

(10^6 mega; 10^9 giga; 10^{12} tera; 10^{15} peta)



The Grid: a possible solution

- The World Wide Web provides seamless access to information stored in different geographical locations
- The Grid provides seamless access to computing power and data storage capacity distributed over the globe
- Relies on advanced software, called middleware:
  - authenticates, authorizes and accounts (AAA)
  - understands and locates the data which the scientist needs
  - distributes the computing processing to wherever in the world there is available and useful capacity
  - sends the results back

The name Grid was chosen by analogy with the electric power grid

Challenges

- Must share data between thousands of scientists with multiple interests
- Must connect major computer centres, not just PCs (not P2P computing)
- Must ensure that all data is accessible anywhere, anytime
- Must grow rapidly, yet remain reliable for more than a decade
- Must cope with different computer centres access policies
- Must ensure data security

Benefits

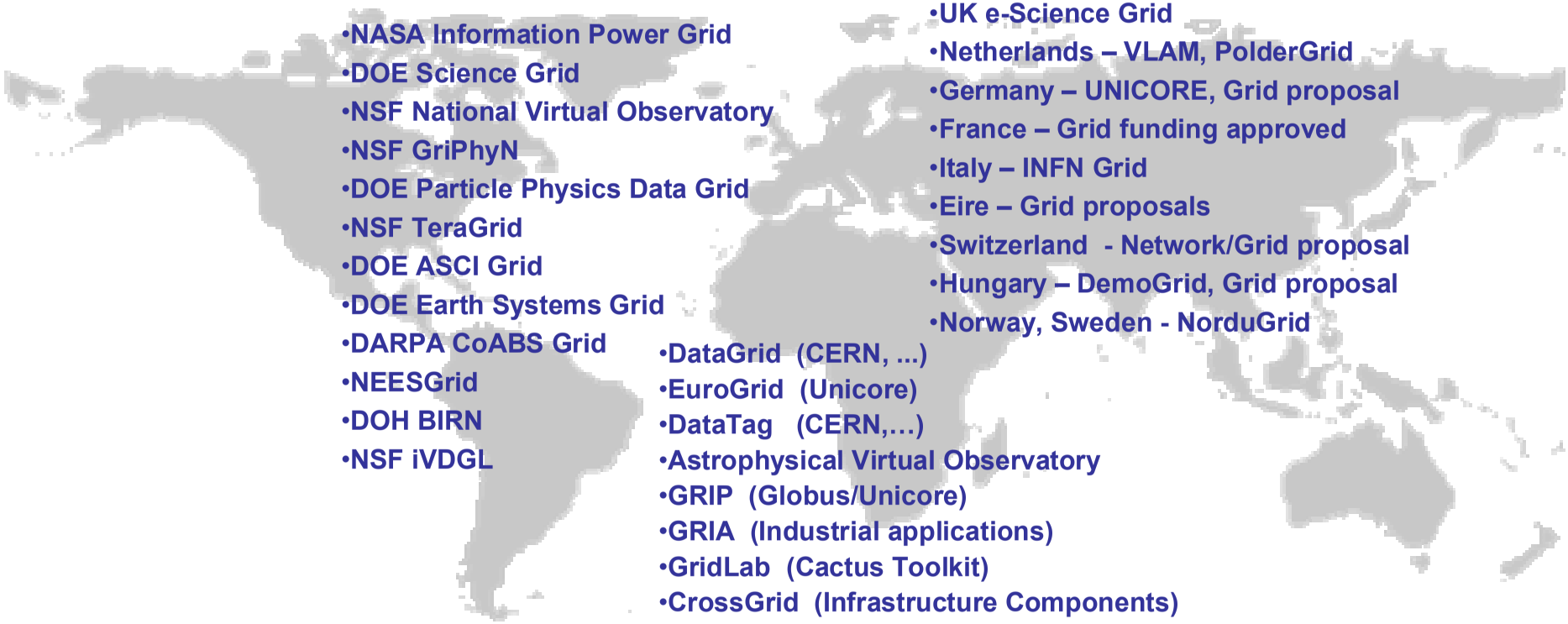
- Effective and seamless collaboration of dispersed communities, scientific first and then industrial
- Ability to run large-scale applications aggregating thousands of computers, for very wide range of applications
- Transparent access to distributed resources from your desktop
- The term "e-Science" has been coined to express these benefits
- In the vision of the "Knowledge Grid", the Grid can act as unifying agent between applications and non homogeneous data

Possible Applications

- Medical/Healthcare *(imaging, diagnosis and treatment)*
- Bioinformatics *(study of the human genome and proteome to understand genetic diseases)*
- Nanotechnology *(design of new materials from the molecular scale)*
- Engineering *(design optimization, simulation, failure analysis and remote Instrument access and control)*
- Natural Resources and the Environment *(weather forecasting, earth observation, modeling and prediction of complex systems)*

Grid projects

Many Grid development efforts — all over the world



- NASA Information Power Grid
- DOE Science Grid
- NSF National Virtual Observatory
- NSF GriPhyN
- DOE Particle Physics Data Grid
- NSF TeraGrid
- DOE ASCI Grid
- DOE Earth Systems Grid
- DARPA CoABS Grid
- NEESGrid
- DOH BIRN
- NSF iVDGL

- DataGrid (CERN, ...)
- EuroGrid (Unicore)
- DataTag (CERN,…)
- Astrophysical Virtual Observatory
- GRIP (Globus/Unicore)
- GRIA (Industrial applications)
- GridLab (Cactus Toolkit)
- CrossGrid (Infrastructure Components)
- EGSO (Solar Physics)

- UK e-Science Grid
- Netherlands – VLAM, PolderGrid
- Germany – UNICORE, Grid proposal
- France – Grid funding approved
- Italy – INFN Grid
- Eire – Grid proposals
- Switzerland - Network/Grid proposal
- Hungary – DemoGrid, Grid proposal
- Norway, Sweden - NorduGrid

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

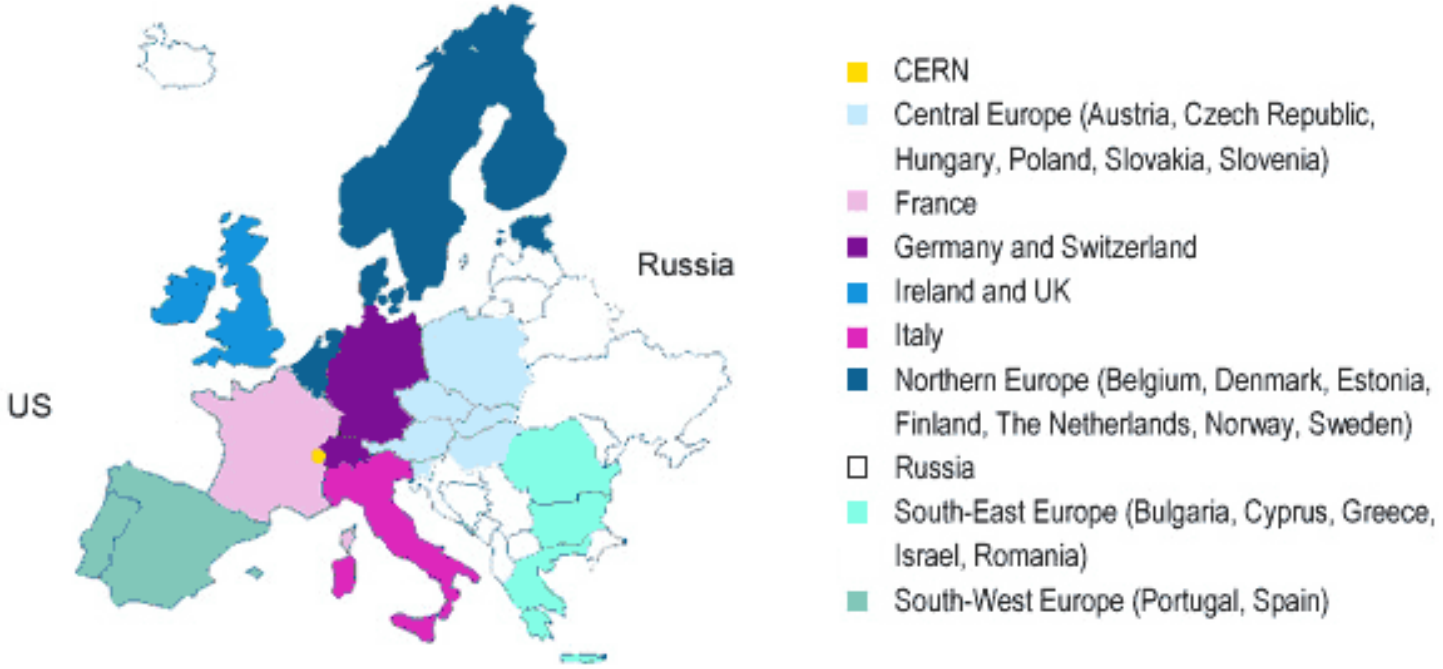
from Grid development to Grid deployment

EGEE manifesto:
Enabling Grids for E-science in Europe

- **Goal**
  - Create a wide European Grid production quality infrastructure on top of present and future EU RN infrastructure
- **Build on**
  - EU and EU member states major investments in Grid Technology
  - International connections
  - Several pioneering prototype results
  - Large Grid development teams in EU
  - Requires major EU funding effort
- **Approach**
  - Leverage current and planned national and regional Grid programmes
  - Work closely with relevant industrial Grid developers, NRENs and US-AP projects

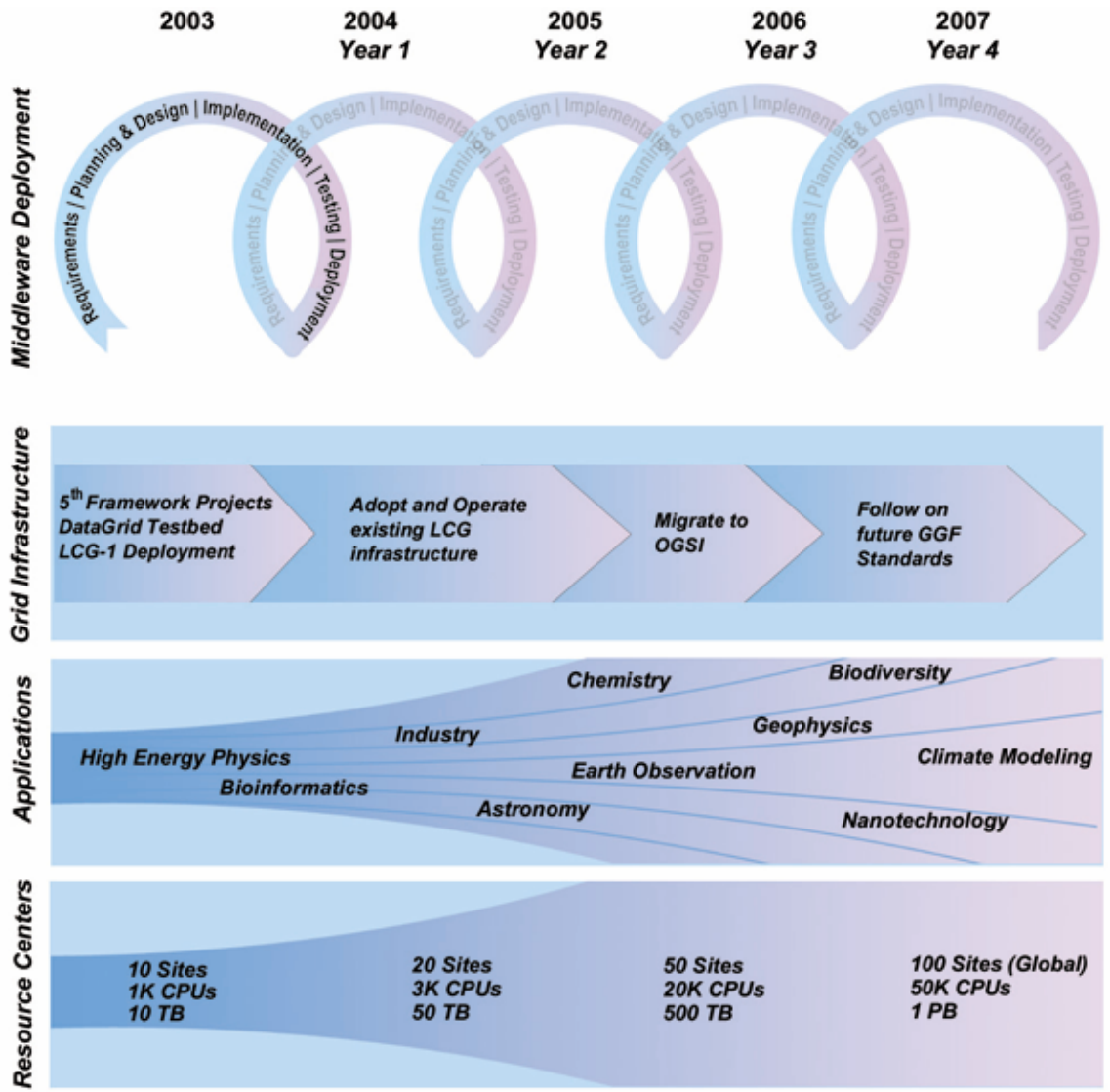
EGEE Partners

- Leverage national resources in a more effective way for broader European benefit
- 70 leading institutions in 27 countries, federated in regional Grids



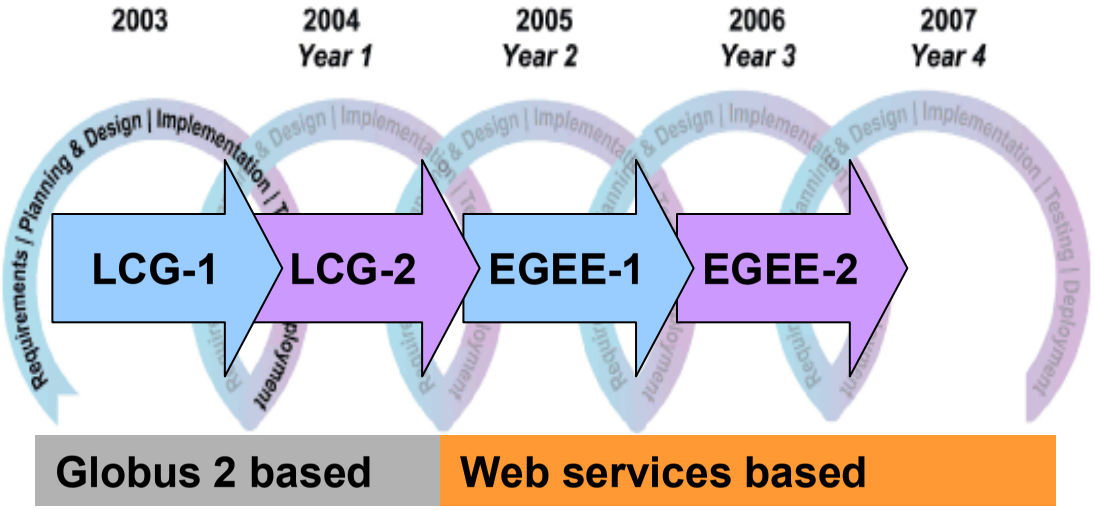
EGEE timeline

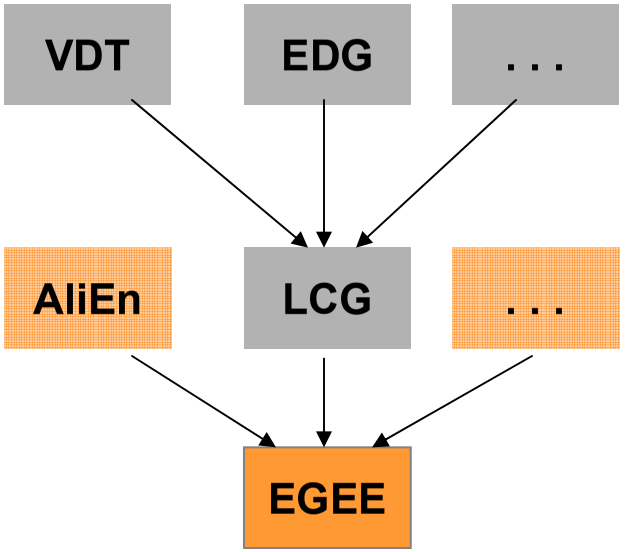
- **May 2003**: proposal submitted
- **July 2003**: proposal accepted
- **September 2003**: start negotiation
- **April 2004**: start project



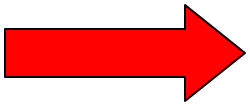
EGEE Implementation

- From day 1 (1st April 2004)
  - Production grid service based on the LCG infrastructure running LCG-2 grid middleware (*SA*)
  - LCG-2 will be maintained until the new generation has proven itself (fallback solution)
- In parallel develop a "next generation" grid facility
  - Produce a new set of grid services according to evolving standards (Web Services)
  - Run a development service providing early access for evaluation purposes





Deployment Status

Core Sites already integrated 

Site	CPU	
CERN (CH)	324	
FZK (D)	144	
PIC (E)	160	
FNAL (US)	4	
CNAF(I)	715	
Nikhef (NL)	250	
Taipei (AP)	98	
RAL (UK)	146	
Total	**1841**	

With the other sites (currently running LCG-1), the expected capacity will exceed the provisions foreseen for 2004:

around 4,000 CPUs at about 30 sites

Introduction to EGEE components

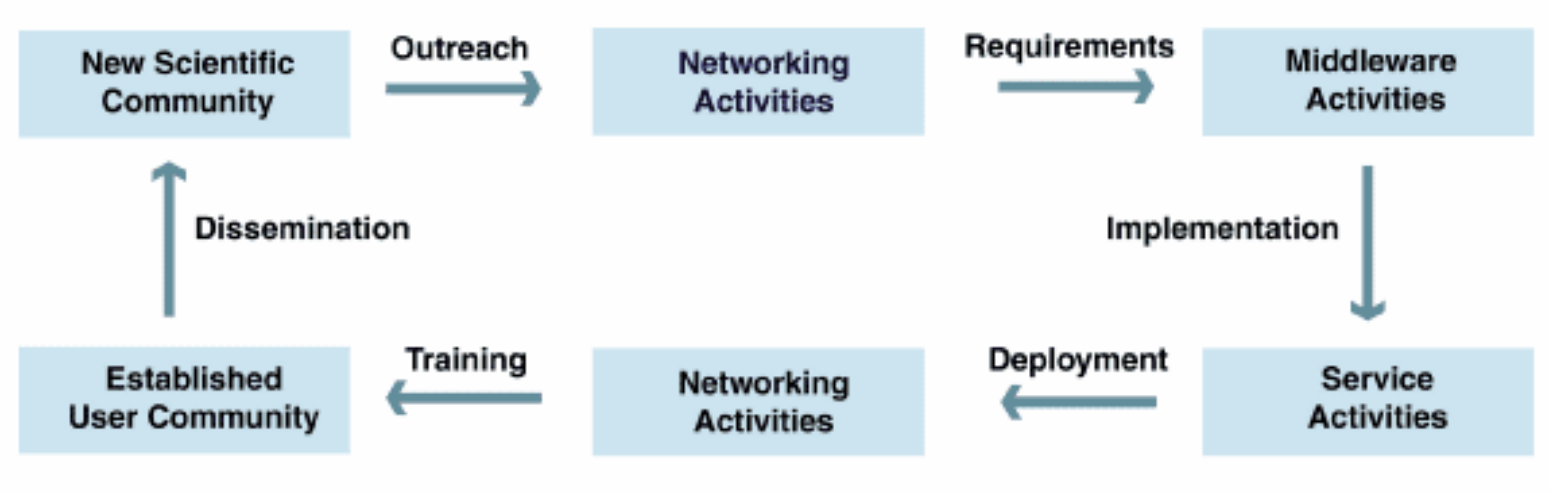
All the work required to be completed within EGEE is divided into different types of Activity:

- **Service Activies (SA)**
- **Joint Research Activities (JRA)**
- **Network Activities (NA)**

These activities have to cooperate as they work, in order to keep focussed on the aims of the project as a whole.

EGEE Activity Areas

- **Services (SA)**
  - Deliver “production level” grid services (manageable, robust, resilient to failure)
- **Middleware (JRA)**
  - Grid middleware re-engineering activity in support of the production services
- **Networking (NA)**
  - Proactively market Grid services to new research communities in academia and industry
  - Provide necessary education



EGEE Activities

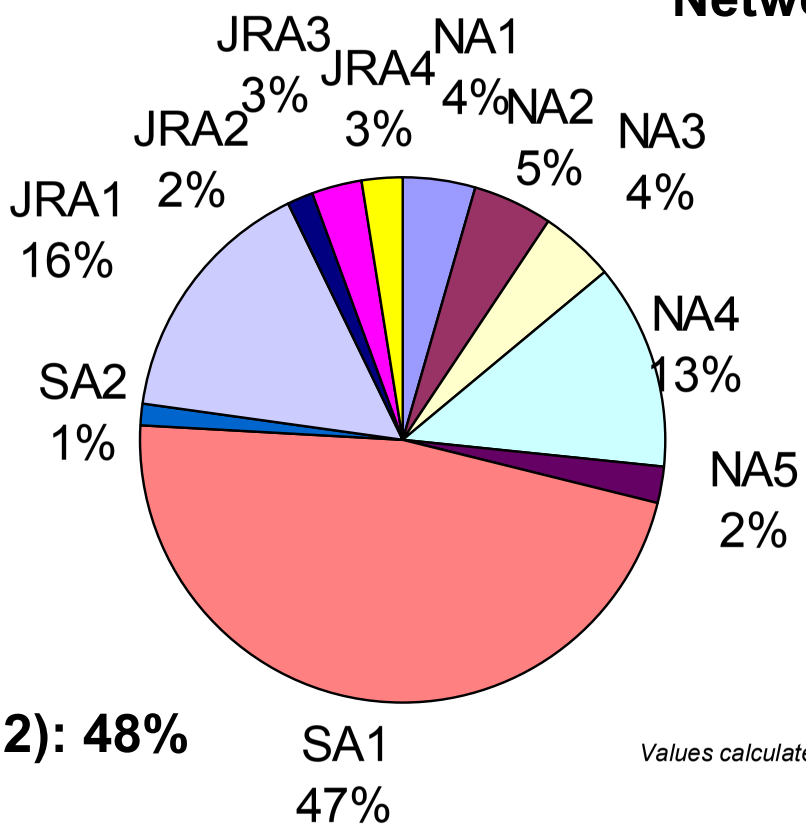
- **EGEE includes 11 activities**
- **Services**
  - **SA1**: Grid Operations, Support and Management
  - **SA2**: Network Resource Provision
- **Joint Research**
  - **JRA1**: Middleware Engineering and Integration
  - **JRA2**: Quality Assurance
  - **JRA3**: Security
  - **JRA4**: Network Services Development
- **"Networking"**
  - **NA1**: Management
  - **NA2**: Dissemination and Outreach
  - **NA3**: User Training and Education
  - **NA4**: Application Identification and Support
  - **NA5**: Policy and International Cooperation

	EGEE
EU funding	32M Euro
Duration	2 years
Partners	70
Deliverables	80

EGEE activities: relative sizes

Mware/security/QA (JRA1-4): 24%

Networking (NA1-5): 28%

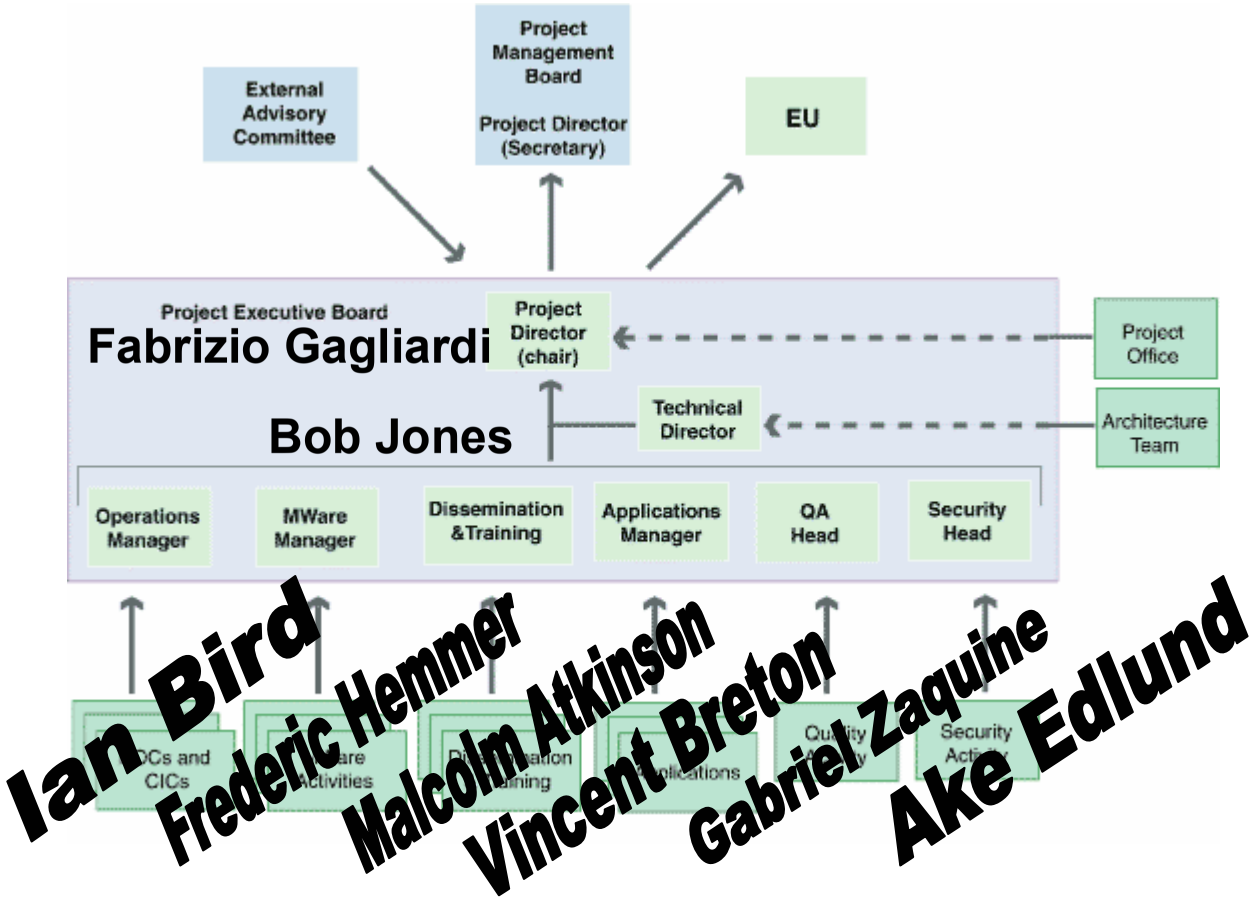


Grid operations (SA1,2): 48%

Values calculated on total activity costs in technical annex

Emphasis in EGEE is on operating a production grid and supporting the end-users

EGEE Project Management



PEB managers nominated by Project Management Board
PEB composition requested by EU

Technical Annex

- The Technical Annex is the description of the EGEE programme of work
- Built on the basis of the original proposal (submitted on May 6, 2003) and on the basis of the EU comments and guidelines
  - Maximum EU contribution for 2004-05: 31,867 K €
- Basic points:
  - build on existing Grid projects and activities
  - close synergy and exchange of responsibilities with LCG
  - structure a very wide (70 partners) infrastructure into Grid national/regional federations
  - define support process for multiple scientific communities and VOs

Enabling Grids for E-science in Europe

Conclusions

- The EU with an aggressive funding policy has fostered more than 20 Grid projects in the IST FP5, which have demonstrated the viability of Grid technology for a wide set of scientific and industrial applications
- As with the Web, this technology will first serve the scientific community and then become the new way to do computing generally
- EGEE will deploy the Grid infrastructure to move from advanced prototypes to production quality applications

Enabling Grids for E-science in Europe

To know more about EGEE:

http://www.eu-egee.org