egee

Enabling Grids for E-sciencE

EGEE and EGEE-II

Fabrizio Gagliardi

EGEE Project Director

Peter Kunszt

EGEE Middleware Activity (JRA1)
Data Management Cluster Leader

Bob Jones

EGEE Technical Director

www.eu-egee.org

Information Society

Enabling Grids for E-sciencE

- **EU Lisbon Declaration March 2000:**

  "The European Union (EU) should become the world's most competitive knowledge-based economy by 2010."

  Based on 3 elements
  - Single market
  - Single currency
  - Single European approach for research
    - → European Research Area (ERA)

- **Policy Forums**
  - The e-Infrastructures Reflection Group (**eIRG**)
  - European Strategy Forum on Research Infrastructures (**ESFRI**)

- **EU 6th Framework Programme (2002-2006):**
  - Funding: ~275 M€
  - Two major lines of funding:
    - Grid Technologies
    - Grid Infrastructures

- **Additional National Funding in the EU:**
  - ~300-500 M€
  - UK e-Science (very much supporting Grid) 250 M £ (2001-2005)

eGee

- **implementation blocks**
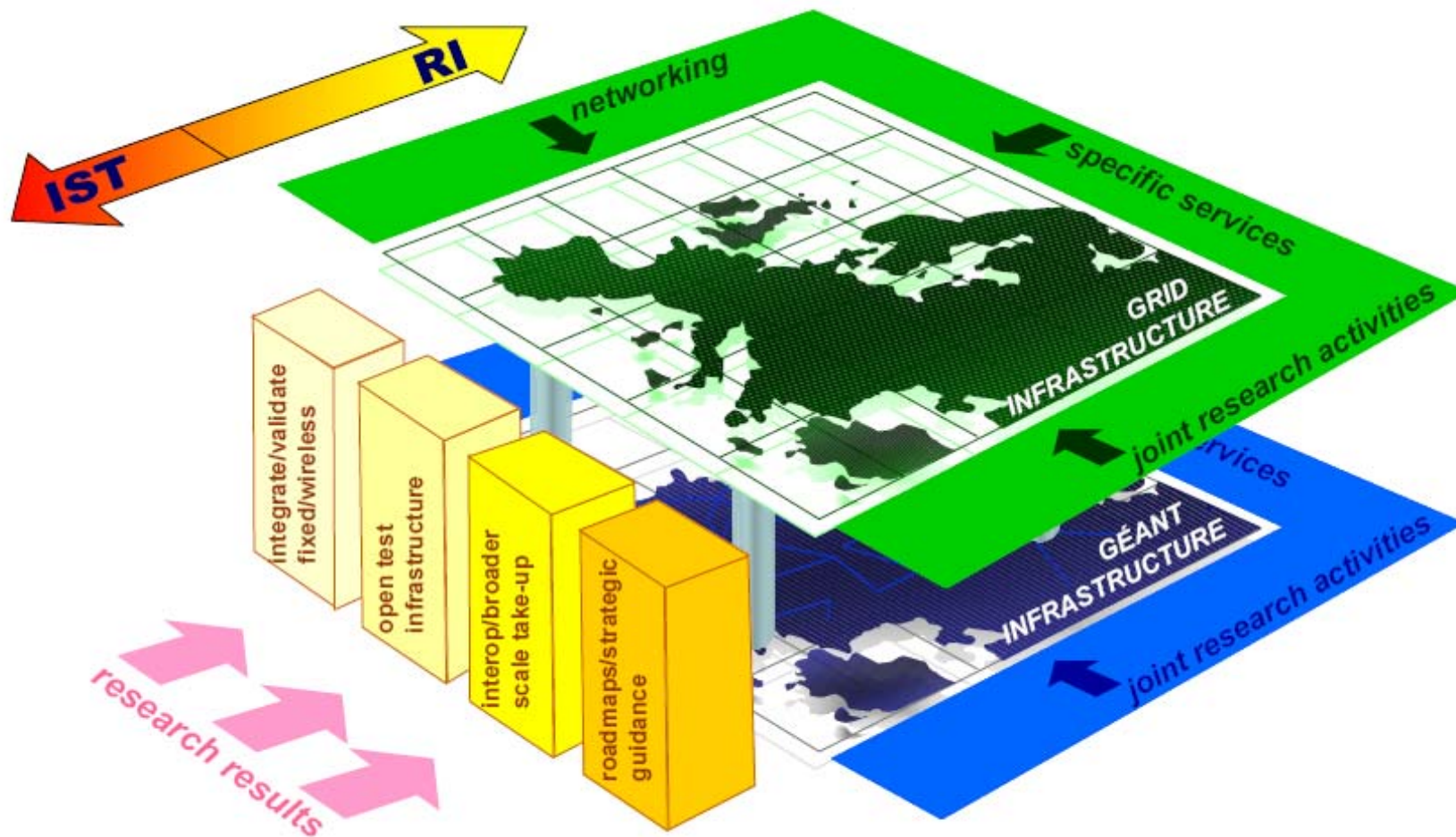
From a talk by Mario Campolargo, Brussels, 30 May 2005

The New York Times

Technology

Europe Exceeds U.S. in Refining Grid Computing

By JOHN MARKOFF and JENNIFER L. SCHENKER

Published: November 10, 2003

Cellphone networks are an example of the difference. Although the technology was invented in the United States, the current European digital cellular networks are generally acknowledged to offer superior service. But Europe's telecom companies have wasted tens of billions of dollars buying the rights to deliver third generation, or 3-G, cellphone services that have generated little interest. With grid computing, Europe may have as much as an 18-month lead in deploying the advances in practical ways, European scientists and government officials said.

While the United States is beginning to respond to a report in February from the National Science Foundation Advisory Panel on Cyberinfrastructure urging coordinated investment in grid technologies, the European Union is preparing to start two major initiatives in early 2004.

One, called Enabling Grids for E-science in Europe, aims to build the largest international grid infrastructure to date, operating in more than 70 institutions throughout Europe, providing 24-hour grid service and a computing capacity comparable to 20,000 of today's most powerful personal computers.

Enabling Grids for E-sciencE

- **Objectives**
  - consistent, robust and secure service grid infrastructure
  - improving and maintaining the middleware
  - attracting new resources and users from industry as well as science

- **Structure**
  - 71 leading institutions in 27 countries, federated in regional Grids
  - leveraging national and regional grid activities worldwide
  - funded by the EU with ~32 M Euros for first 2 years starting 1st April 2004

CERN
Central Europe
France
Germany and Switzerland
Ireland and UK
Italy
Networking
Northern Europe
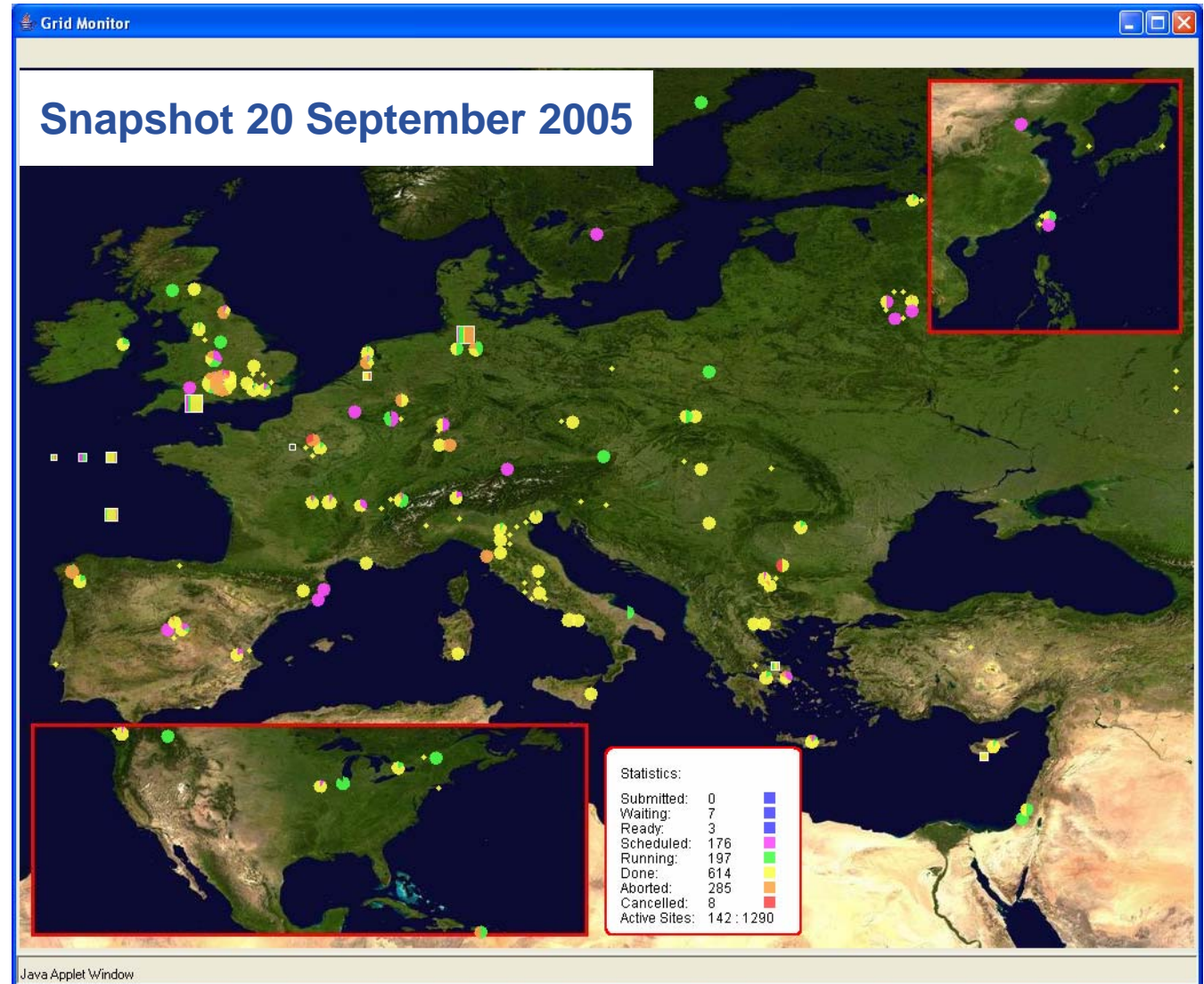Russia
South-East Europe
South-West Europe
USA

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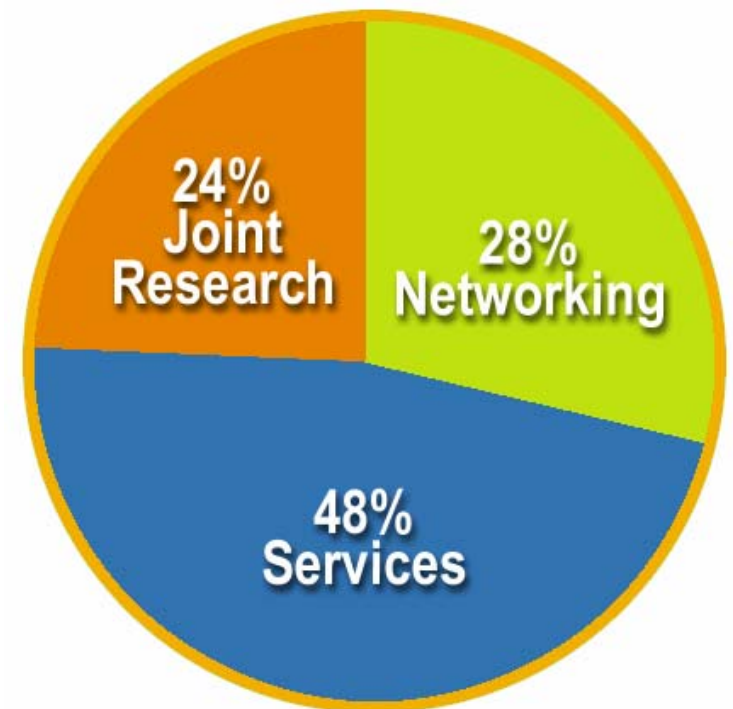
Emphasis in EGEE is on operating a production grid and supporting the end-users

Enabling Grids for E-sciencE

- >180 sites
- >15 000 CPUs (with peaks >18 000 CPUs)
- ~14 000 jobs successfully completed per day
- 20 VOs
- >800 registered users, representing thousands of scientists

Grid Monitor

Snapshot 20 September 2005

Statistics:
Submitted:	0	
Waiting:	7	
Ready:	3	
Scheduled:	176	
Running:	197	
Done:	614	
Aborted:	285	
Cancelled:	8	
Active Sites:	142 : 1290	

Java Applet Window

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

- **Hardening and re-engineering of existing middleware functionality, leveraging the experience of partners**
  - Fill eventual gaps in services
  - Provide a stack of middleware useful to EGEE applications (currently HEP, Biomedicine, Earth Sciences etc.)

- **Activity concentrated in few major centers and organized in "Software clusters"**

- **Key services:**
  - Data Management (CERN)
  - Information and Monitoring (UK)
  - Resource Brokering, Accounting (Italy-Czech Republic)
  - Quality Assurance (CERN, France)
  - Grid Security (Northern Europe)
  - Middleware Integration (CERN)
  - Middleware Testing (CERN)

- Middleware Integration and Testing Centre
- Middleware Re-engineering Centre
- Quality and Security Centres
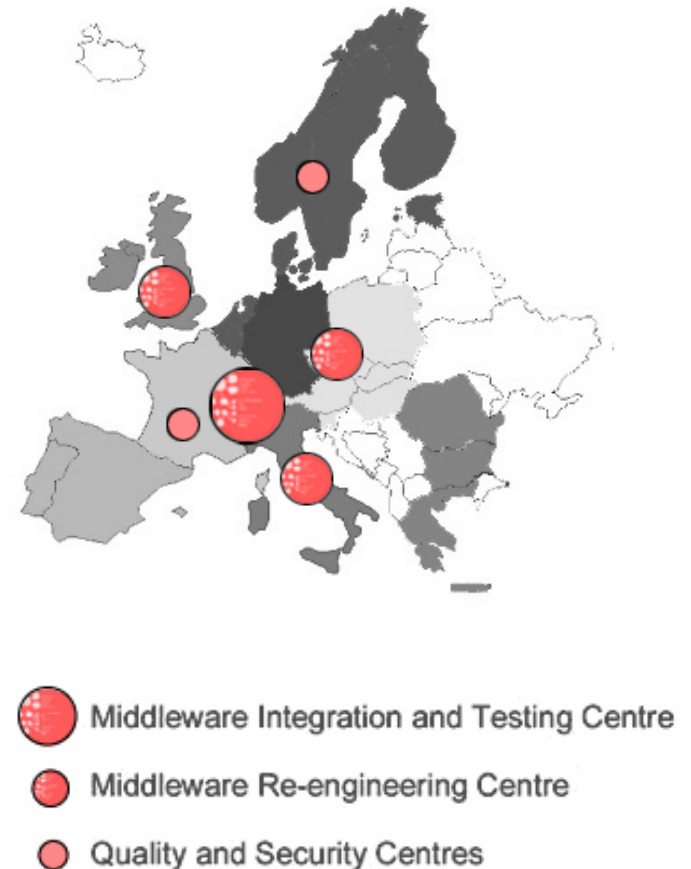
- **Clusters collaborate with US partners**
  - University of Chicago
  - University of Southern California
  - University of Wisconsin Madison

eGee

Application Requirements

- Generate/analyze a large amount of data 10PetaByte / year
- Distribute the data across 100s of sites
- Support a distributed user community all over the world
- Secure access to Grid services, resources and data

User Requirements

- Ease of use
- Need to be able to run with legacy applications
- Good support (helpdesk)

Site Requirements

- Easy to manage both on small and large sites, good support for admins
- Access to site resources has to be ultimately controllable by the site.
- Accountability and audit logs need to be made available, monitoring

Other Aspects

- Many projects involved both in multiple projects
- Inconsistent priorities on development and deployment from application projects
- Some products have a very limited lifetime, little support or different priorities

Uniform Security Architecture

Uniform Semantics

X509 and WS-Security

Separate VO and Site Management

MyProxy   VOMS   Delegation

Service Oriented Architecture

Interoperability

Portability

Modularity

Scalability

Web Services

Building on existing components in a lightweight manner

EDG   LCG   Condor

Globus   AliEn   PPDG

...

Applications

VO (Application) Middleware

| Portals User Interfaces | Data Catalogs | Application Services |

Grid Middleware
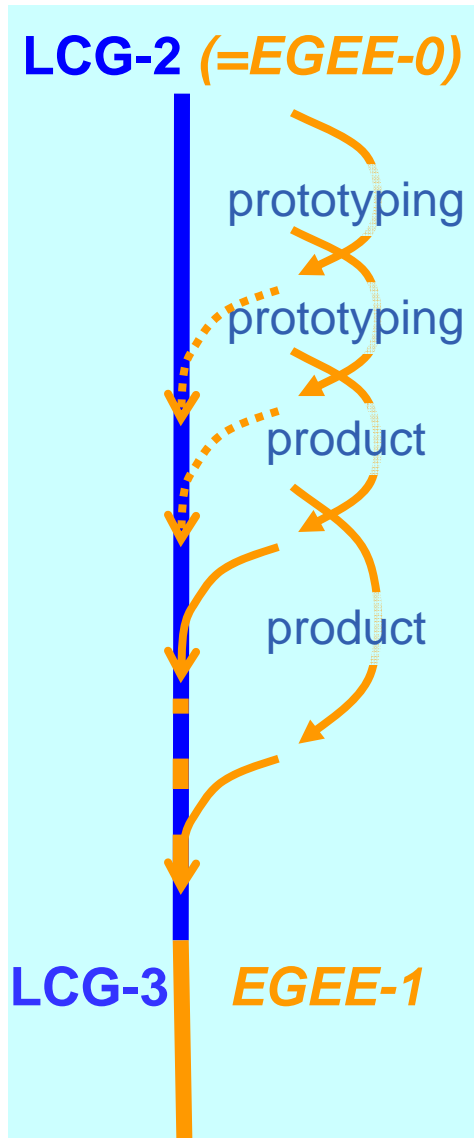
| Workload Management | Data Catalogs | Data Transfer |

Computing Element

Storage Element

Network Provisioning

| Computing | Storage | Network |

Basic Resources

Security Infrastructure

Enabling Grids for E-sciencE

LCG-2 *(=EGEE-0)*

prototyping

prototyping

product

product

LCG-3   *EGEE-1*
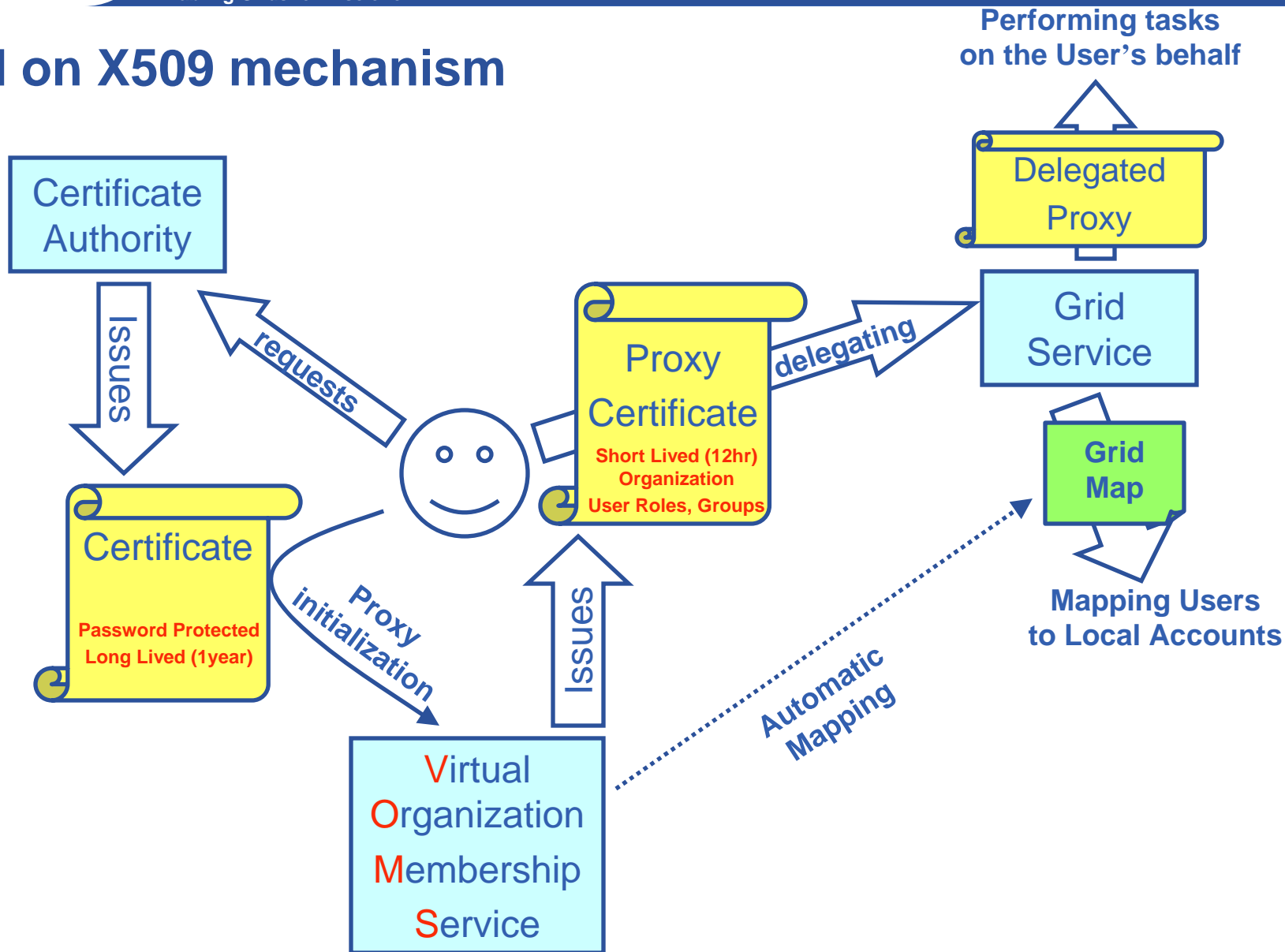
- **Fast prototyping** approach
  - Small scale prototype for a few selected users for quick feedback

- **Single out individual components for deployment on pre-production service**

- **These components are going through integration and testing**
  - Integration of configuration and setup
  - Testing of deployment and functionality

Enabling Grids for E-sciencE

- **JRA1 Software Process is based on an iterative method using industry-standard concepts**
  - Modular Architecture
  - Iterative Design based on experience and enhancement requests
  - Fully controlled software build environment
- **Regular releases and Quick Fixes**
  - Regular releases roughly at a monthly cycle since April 2005
  - Quick Fix releases for critical bug fixes in between regular releases
- **The process is fully documented in a number of standard documents:**
  - Software Configuration Management Plan
 - Test Plan
  - Quality Assurance Plan
  - Developer's Guide

- **Infrastructure**
  - Extend and consolidate EGEE e-Infrastructure
  - Linking national, regional and thematic Grid efforts
  - → High-capacity infrastructure
    - greatly surpasses capabilities of local clusters and individual centres
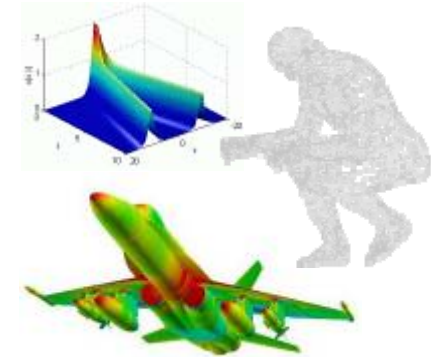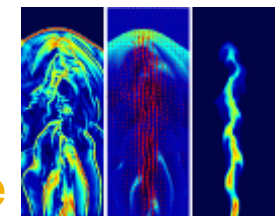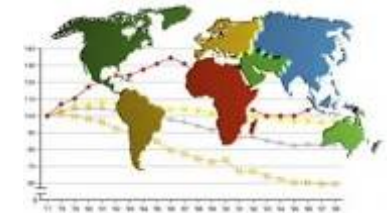
- **e-Science**
  - Support distributed research communities
  - that agree on common access policies
  - → Unique tool for collaborative compute-intensive science
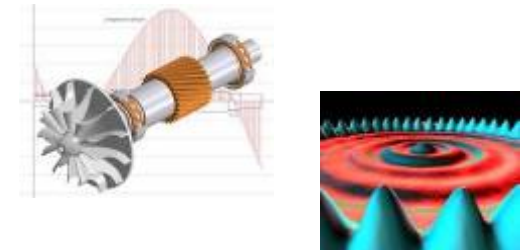
- **Coordination**
  - Interoperability with other Grid infrastructures
  - Structuring effect: Related Projects
  - Participating in developing Grid standards
  - → Paving the way for a long-term sustainable Grid infrastructure

eGee

Enabling Grids for E-sciencE

- **Manage and operate Grid infrastructure for the ERA**
- **Interoperate with e-Infrastructure projects around the globe**
- **Contribute to Grid standardisation efforts**

- **Support applications deployed from diverse scientific communities**
  - High Energy Physics
  - Biomedicine
  - Earth Sciences
  - Astrophysics
  - Computational Chemistry
  - Fusion
  - Geophysics (supporting the Industrial application, EGEODE)

- **Reinforce links with the full spectrum of interested industrial partners**
- **Disseminate knowledge about the Grid through training**

- **Prepare for a permanent European Grid Infrastructure (in a GÉANT2-like manner)**
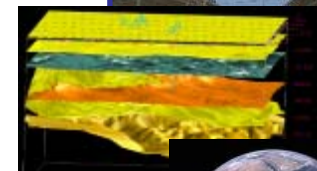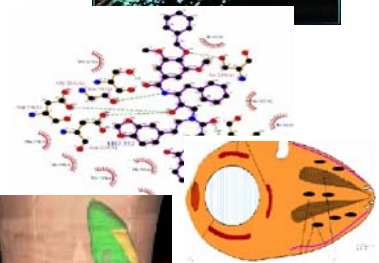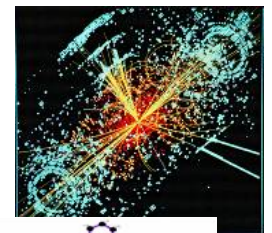
- **More than 90 partners**
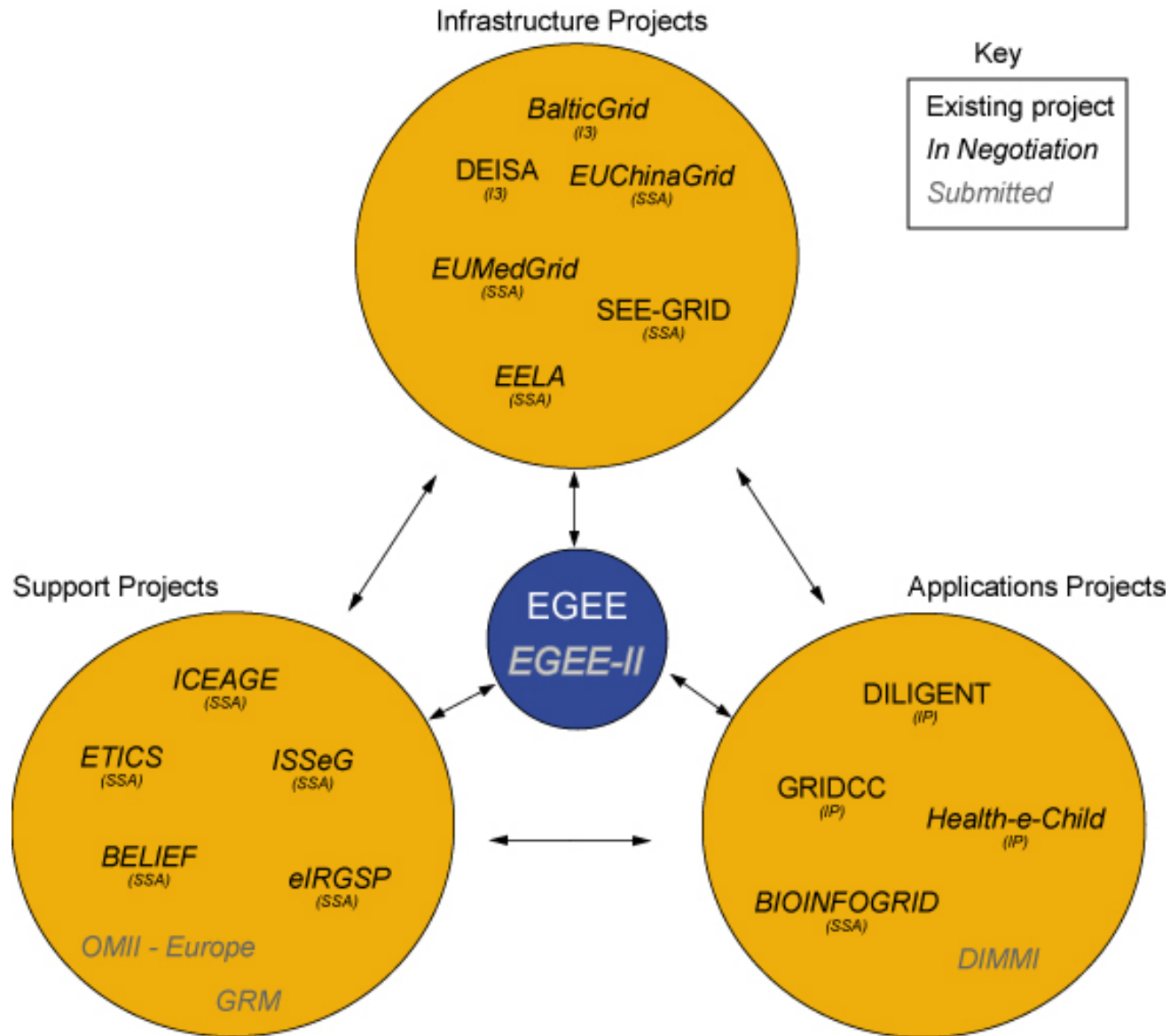- **32 countries**
- **12 federations**
→ **Major and national Grid projects in Europe, USA, Asia**

+ 27 countries through related projects:
- BalticGrid
- SEE-GRID
- EUMedGrid
- EUChinaGrid
- EELA

Enabling Grids for E-sciencE

Enabling Grids for E-sciencE

- **Similar structure to EGEE**
- **Increased focus on**

- **Services:**
  – New countries joining
  – New activity for sw integration

- **Networking**
  – Increase support for applications
  – Reinforce outreach, dissemination and training
    - Closer links to industry (CERN openlab project)
    - Extend coverage to all regions

- **Coordination:**
  – Technical Coordination Group (TCG)
  – User Information Group (UIG)

EGEE

SA 48%
NA 28%
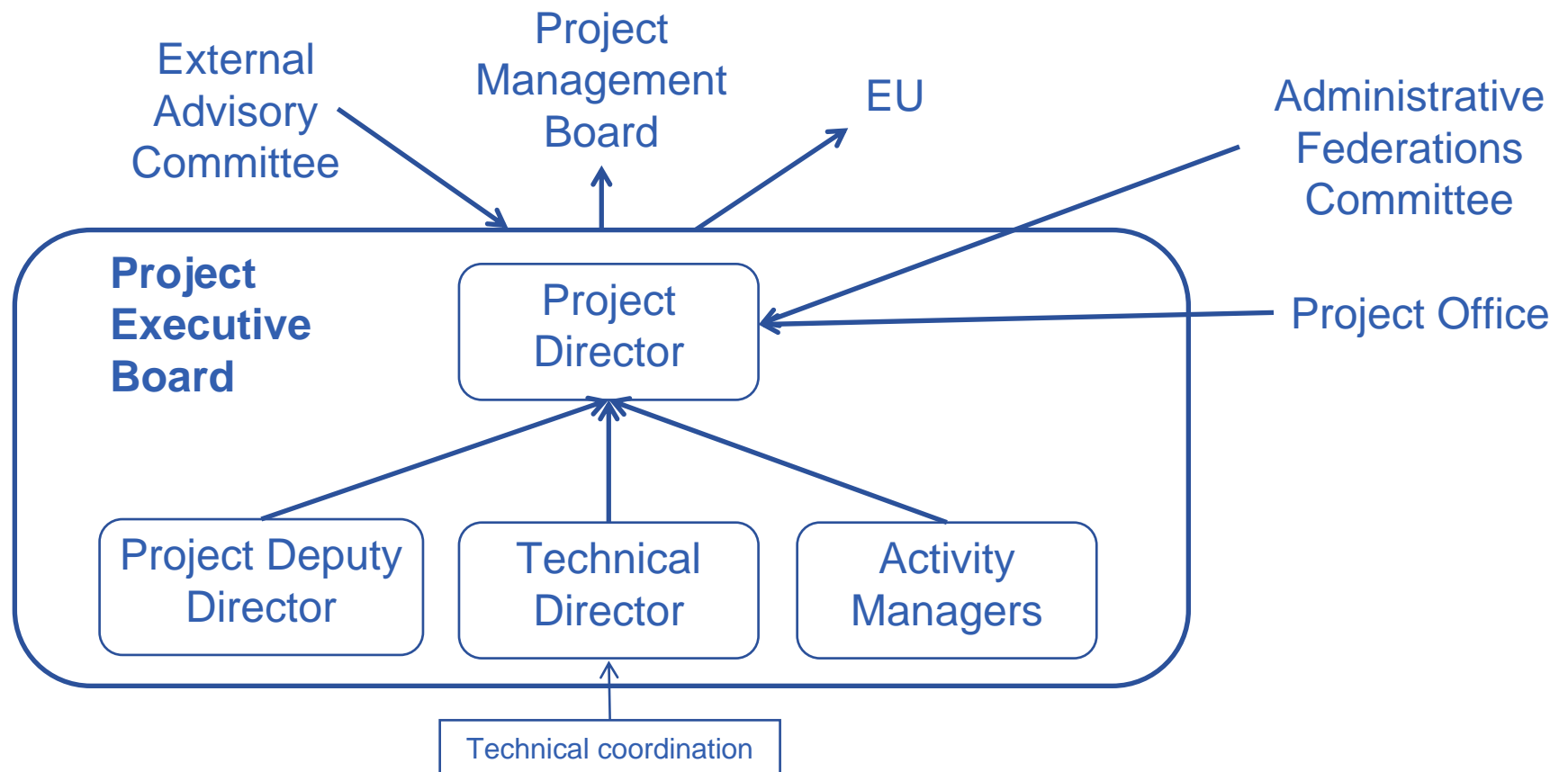JRA 24%

EGEE-II

SA 54%
NA 32%
JRA 14%
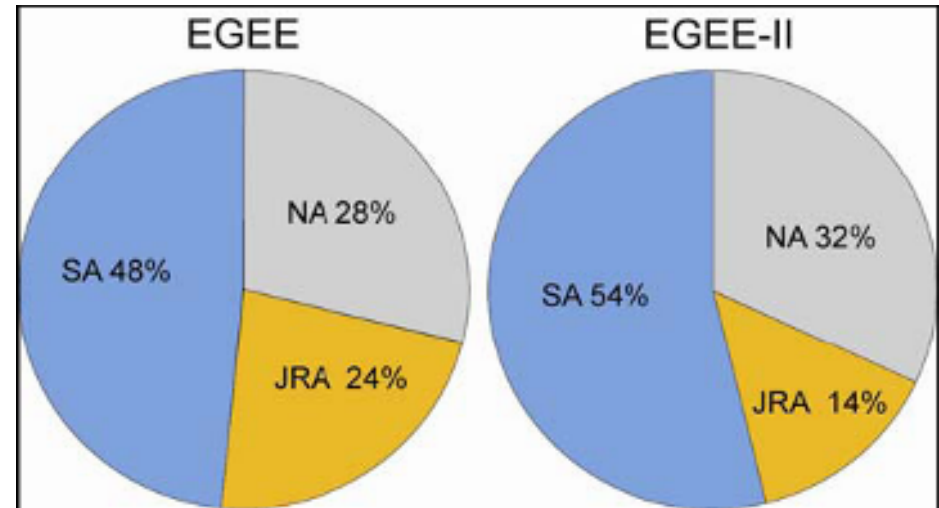
- **JRA1 – Re-engineering of Grid middleware**
  – Shift from development to incremental improvement mode

- **JRA2 – Quality assurance**
  – Includes managing of security efforts through Security Coordination Group

→ **Reflects the more mature status of the middleware**
→ **Integrate proportion of required software from other projects and external providers**

egee

Enabling Grids for E-sciencE

- **Challenge: large consortium**
- → **Clearly defined management structure**
- → **Clear reporting lines**

External Advisory Committee

Project Management Board

EU

Administrative Federations Committee

Project Executive Board

Project Director

Project Office

Project Deputy Director

Technical Director

Activity Managers

Technical coordination

Enabling Grids for E-sciencE

- **For Users**
  - Simplified & pervasive access
  - On-demand computing
  - Large scale resources
  - Sharing of software and data
  - Comprehensive support systems

- **For Resource Providers**
  - Large scale operations
  - Specialist competence
  - User contacts
  - Collaboration among resource partners

- **For European Industry**
  - As a partner
    - Tune the project to industrial needs and pass on knowledge
  - As a user
    - R&D
  - As a provider
    - Transfer knowledge on how to provide Grid services
  - As affiliate
    - Industry Forum: feedback, exchange of knowledge

- **Long-term benefits to the public**
  - Raising awareness of the Grid
  - Early establishment of a European Grid infrastructure