

Introduction to the EGEE project

Fabrizio Gagliardi
EGEE designated Project Director

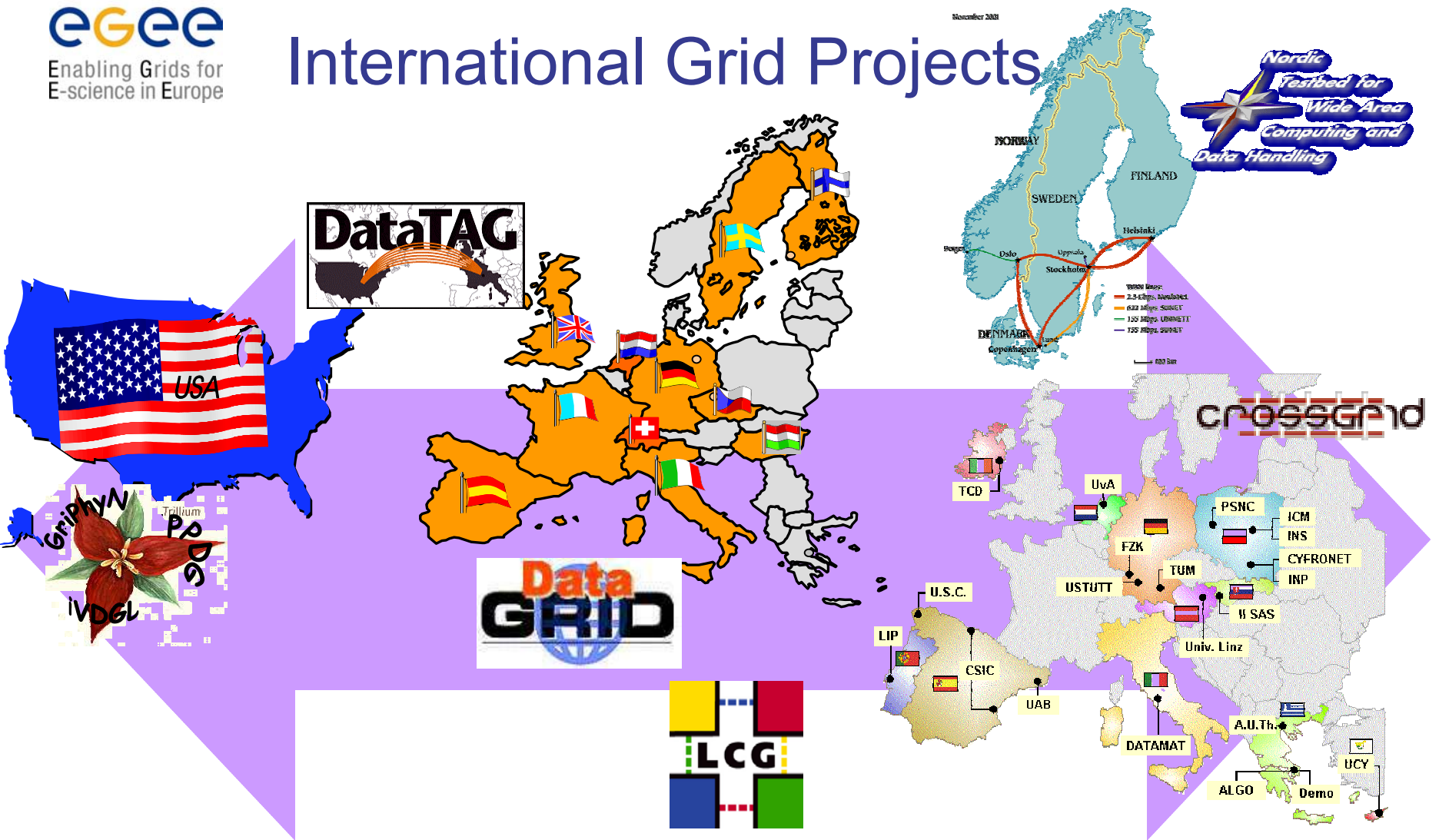


**EGEE is proposed as a project funded by the
European Union under contract IST-2003-508833**

DataGrid prototypes: DataGrid (I)

- 9.8 M Euros EU funding over 3 years (twice as much from partners)
- 90% for middleware and applications (High Energy Physics, Earth Observation, Genomic Exploration)
- Total of 21 partners, over 150 programmers from research and academic institutes as well as industrial companies
- Three year phased developments & demos (2001-2003)
- Several improved versions of middleware software (final release end 2003)
- Several components of software integrated in the large Particle Physics Production LHC Computing Project (LCG)
- Software used by partner projects: DataTAG, CROSSGRID, GRACE

International Grid Projects



Through links with sister projects, there is the potential for a truly global scientific applications grid

DataGrid prototypes: DataGrid (II)

- DataGrid testbed: more than 1000 CPUs at more than 15 sites (up to 40)
- Connections made possible by the EU-funded **GEANT project**
 - connecting more than 30 countries across Europe
 - speeds of up to 10 Gbit/s
 - high data throughput
 - quality of Service



Site	Country	CPUs	Storage
CC-IN2P3*	FR	620	192 GB
CERN*	CH	138	1321 GB
CNAF*	IT	48	1300 GB
Ecole Poly.	FR	6	220 GB
Imperial Coll.	UK	92	450 GB
Liverpool	UK	2	10 GB
Manchester	UK	9	15 GB
NIKHEF*	NL	142	433 GB
Oxford	UK	1	30 GB
Padova	IT	11	666 GB
RAL*	UK	6	332 GB
SARA	NL	0	10000+ GB
TOTAL	5	1075	14969 GB

*also Dev. TB; +200 TB including tape

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

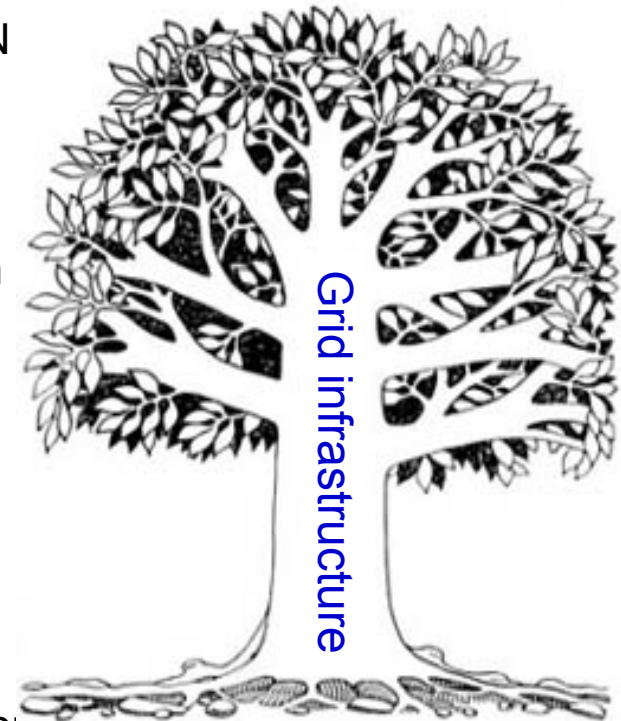
- **Build on**

- EU and EU member states major investments in Grid Technology
- International connections (US and AP)
- 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

ERA Applications



Geant Research Network

EGEE: Why? (I)

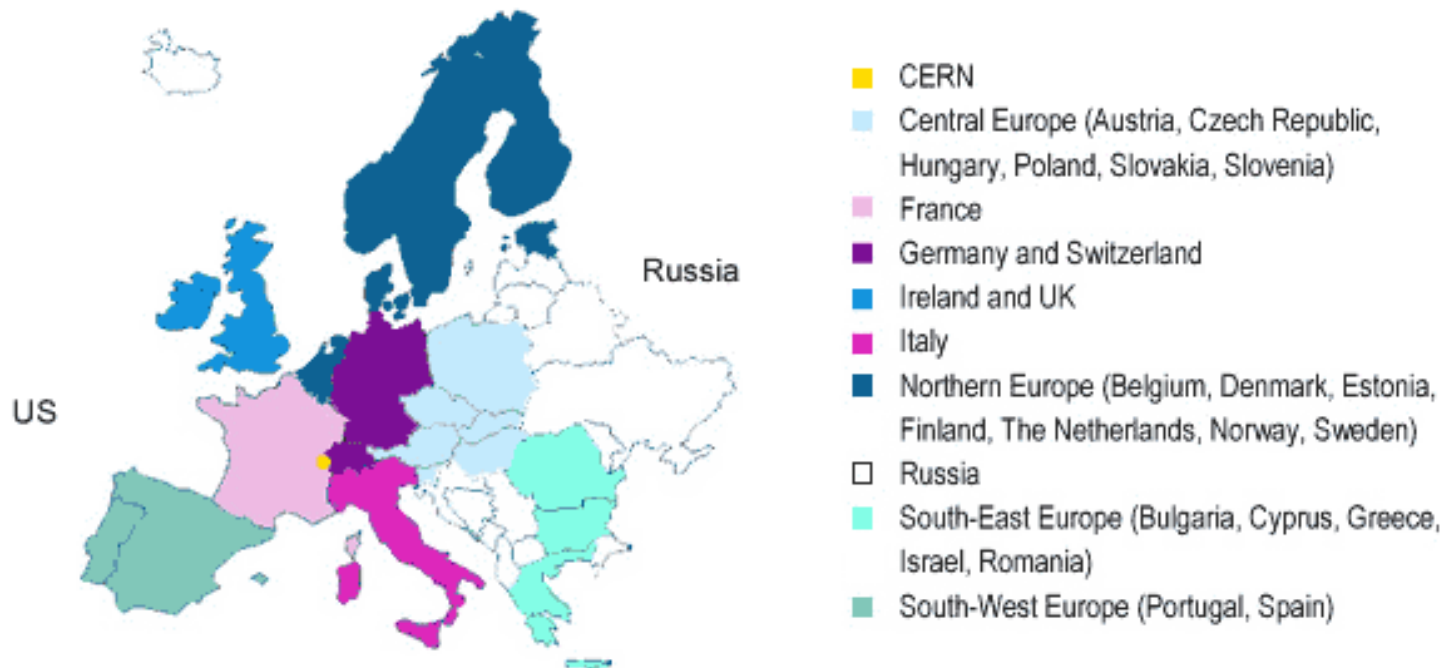
- Access to a production quality grid will change the way science and business is done in Europe
 - More effective and seamless collaboration of dispersed communities, both scientific and commercial
 - Ability to run large-scale applications comprising thousands of computers, for wide range of applications
 - Transparent access to distributed resources from your desktop
- Benefits for several application fields:
 - Bioinformatics (*study of the human genome and proteome to understand genetic diseases*)
 - Engineering (*design optimization, simulation, failure analysis and remote Instrument access and control*)
 - Medical/Healthcare (*imaging, diagnosis and treatment*)
 - Natural Resources and the Environment (*weather forecasting, earth observation, modeling and prediction of complex systems*)

EGEE: Why? (II)

- Current Grid R&D projects run to completion within the next few months or next year (e.g. X#)
- The EGEE partners have already made major progress in aligning national and regional Grid R&D efforts, in preparation for EGEE
- EGEE will preserve the current strong momentum of the European Grid community, and the enthusiasm of the hundreds of young European researchers already involved in EU Grid projects (>150 in EDG alone)

EGEE: Partners

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



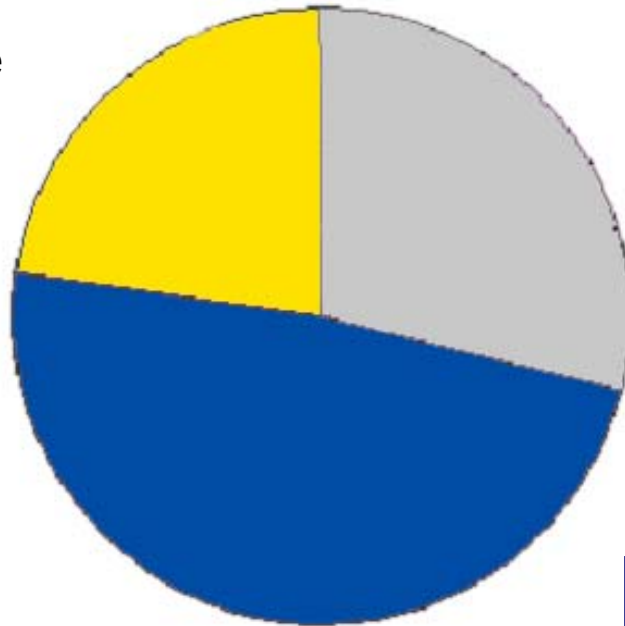
24% Joint Research

JRA1: Middleware Engineering and Integration

JRA2: Quality Assurance

JRA3: Security

JRA4: Network Services Development



48% Services

SA1: Grid Operations, Support and Management

SA2: Network Resource Provision

28% Networking

NA1: Management

NA2: Dissemination and Outreach

NA3: User Training and Education

NA4: Application Identification and Support

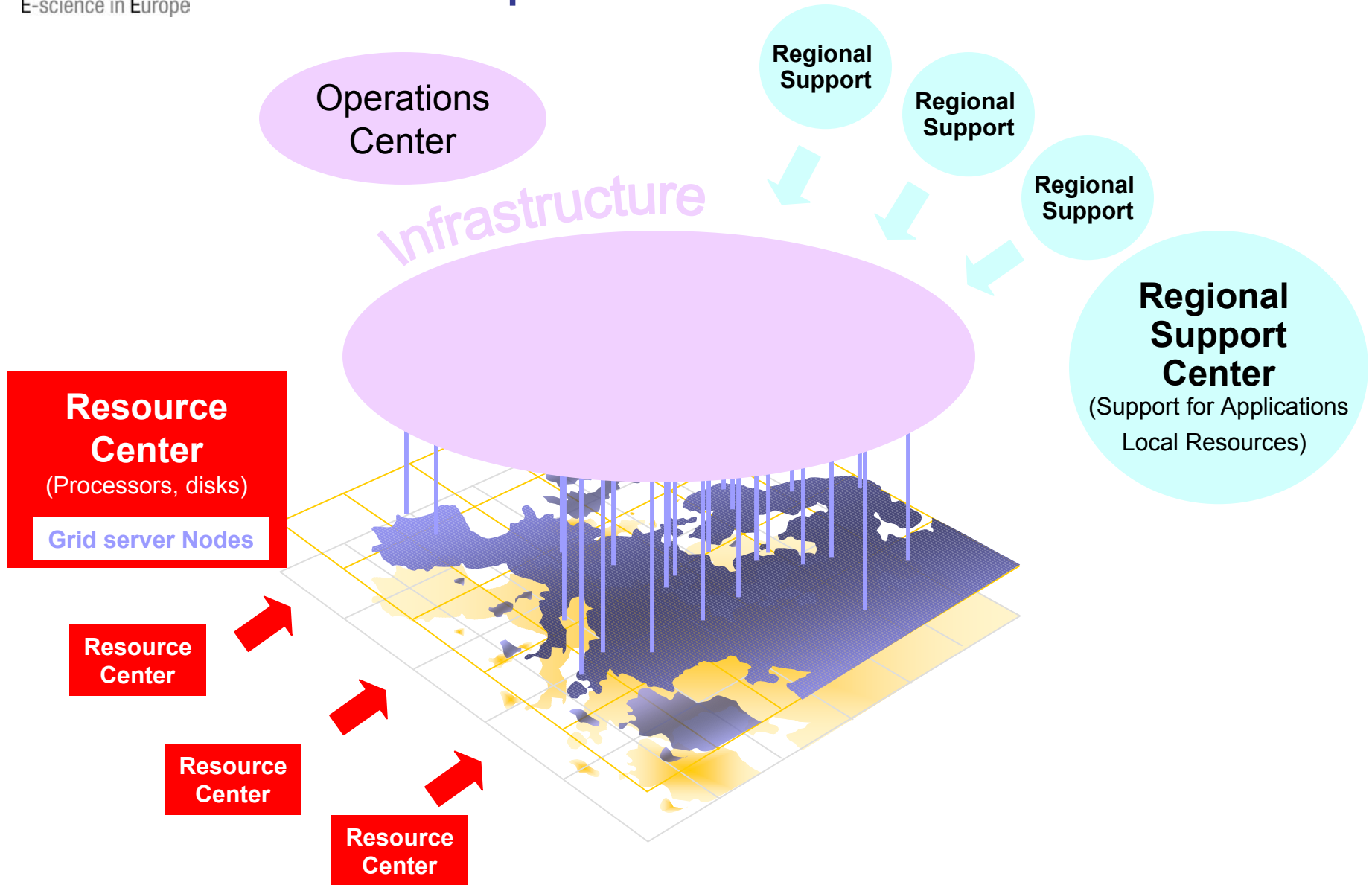
NA5: Policy and International Cooperation

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

EGEE Operations

- Operate essential grid services { Grid monitoring and control
- Proactively monitor the operational state and performance
- initiate corrective action { Core Infrastructure Services
- Validate and deploy middleware releases
- Set up operational procedures for new resources { Middleware deployment and resource induction
- Coordinate the resolution of problems from both Resource Centres and users
- Filter and aggregate problems, providing or obtaining solutions { Resource provider and user support

EGEE Operations Structure



EGEE Service Activity (I)

- Create, operate, support and manage a production quality infrastructure
- Structure:
 - EGEE Operations Management at CERN
 - EGEE Core Infrastructure Centres in the UK, France, Italy and CERN (leveraging HEP LCG at the start), responsible for managing the overall Grid infrastructure
 - Regional Operations Centres, responsible for coordinating regional resources, regional deployment and support of services in all other countries
- Offered services:
 - Middleware deployment and installation
 - Software and documentation repository
 - Grid monitoring and problem tracking
 - Bug reporting and knowledge database
 - VO services
 - Grid management services

EGEE Service Activity (II)

- **Operations Management Centre – OMC**
 - Coordinator for CICs and for ROCs
 - Team to oversee operations – problems resolved, performance targets, etc.
 - Operations Advisory Group to advise on policy issues, etc.
- **Core Infrastructure Centres – CIC**
 - Day-to-day operation management– implement operational policies defined by OMC
 - Monitor state, initiate corrective actions, eventual 24x7 operation of grid infrastructure
 - Provide resource and usage accounting, security incident response coordination, ensure recovery procedures
- **Regional Operations Centres – ROC**
 - Provide front-line support to users and resource centres
 - Support new resource centres joining EGEE in the regions



- Operations Management Centre
- Core Infrastructure Centre
- Regional Operations Centre

EGEE Service Activity (III)

- Resource Centers

Month 1: 10

Month 15: 20

Region	CPU nodes	Disk (TB)	CPU Nodes Month 15	Disk (TB) Month 15
CERN	900	140	1800	310
UK + Ireland	100	25	2200	300
France	400	15	895	50
Italy	553	60.6	679	67.2
North	200	20	2000	50
South West	250	10	250	10
Germany + Switzerland	100	2	400	67
South East	146	7	322	14
Central Europe	385	15	730	32
Russia	50	7	152	36
Totals	3084	302	8768	936

EGEE Middleware Activity

- Hardening and re-engineering of existing middleware functionality, leveraging the experience of partners
- Activity concentrated in few major centers
- Key services: Resource Access
 - Data Management (CERN)
 - Information Collection and Accounting (UK)
 - Resource Brokering (Italy)
 - Quality Assurance (France)
 - Grid Security (Northern Europe)
 - Middleware Integration (CERN)
 - Middleware Testing (CERN)






- Middleware Integration and Testing Centre
- Middleware Re-engineering Centre
- Quality and Security Centres

EGEE Networking Activity (I)

- Dissemination and outreach
 - Lead by TERENA
- User training and induction
 - Lead by Unv Edin. (NeSC)
- Application identification and support
 - Two pilot application centers (for high energy physics and biomedical grids)
 - One more generic component dealing with longer term recruitment and support of other communities
- Policy and International cooperation
 - Establish Grid policy forum
 - Coordinate relations with other projects (EU and beyond)

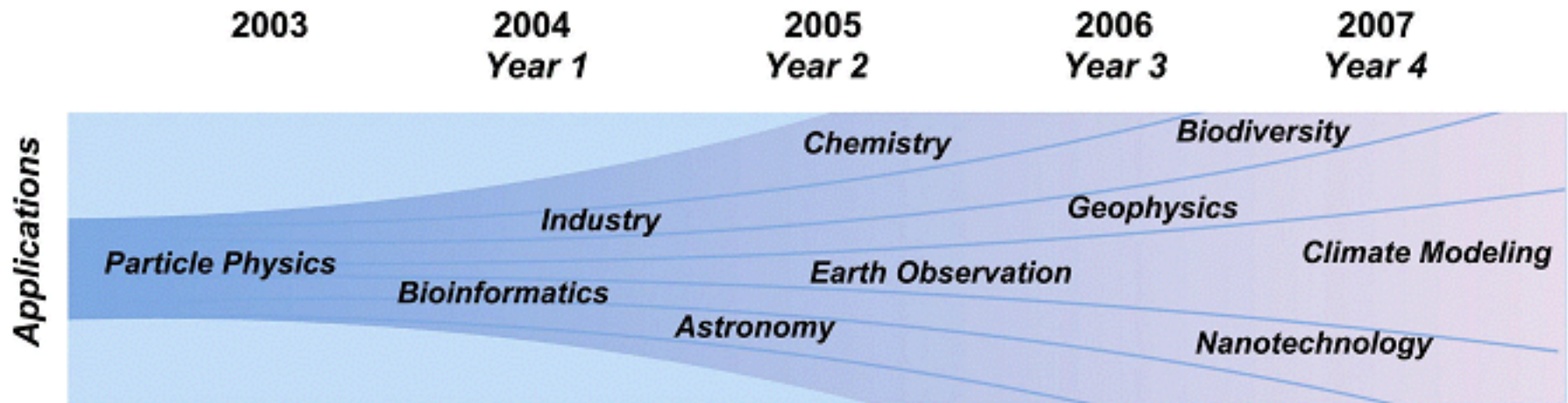


-  Lead Networking Centre
-  Pilot Application Centre
-  Regional Networking Centre

*map points indicate federations
and are not geographically precise*

EGEE Networking Activity (II)

- EGEE Scope : **ALL-Inclusive** for academic applications
- Open to industrial and socio-economic world as well
- The **major success criterion** of EGEE: how many satisfied users from how many different domains ?
- **5000 users** (3000 after year 2) from at least 5 disciplines
- 2 Pilot Application Domains: Physics & Bioinformatics



Application domains and timelines are for illustration only

EGEE “Virtuous Cycle”

A new scientific community makes first contacts to EGEE through outreach events organized by Networking Activities

Follow-up meetings by applications specialists may lead to definition of new requirements for the infrastructure



Peer communication and dissemination events featuring established users then attract new communities



If approved, the requirements are implemented by the Middleware Activities



The Networking Activities then provide appropriate training to the community in question, so that it becomes an established user

After integration and testing, the new middleware is deployed by the Service Activities

EGEE and Industry

- Industrial participation encouraged both as potential end-users and IT technology and service suppliers
- Normally through national and regional Grid EGEE federations
- EGEE will maintain an Industry Forum to keep selected Industrial and Commercial interested parties in close contact
- Services developed in first EGEE 2 years phase (2004-5) might be tendered to Industry in second phase (2006-7)

EGEE Implementation Plans

- Initial service will be based on the **LCG infrastructure** (this will be the production service, most resources allocated here)
- Also will need a certification **test-bed system**
 - For debugging and problem resolving of the production system
- In parallel must deploy a **development service**
 - Runs the candidate next software release for production
 - Treated as an reliable facility (but with less support than the production service)

Conclusions

- EGEE is expected to deliver a real production Grid infrastructure for real scientific applications in Europe
- Important to have early adopters and applications representatives in all phases of the project
- Previous EDG experience has shown the importance of dedicated Grid support people in the applications
- NA4 is going to play a fundamental role

To know more:

EU EGEE – www.eu-egee.org
EU DataGrid – www.eu-edg.org