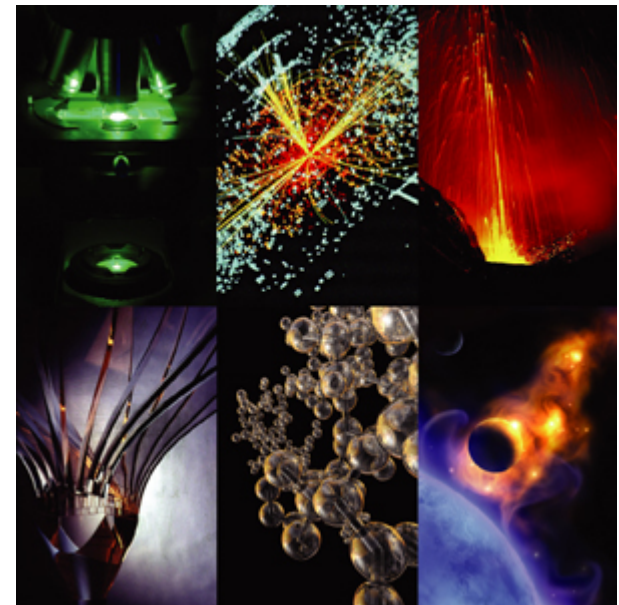# egee

Enabling Grids for E-sciencE

The EGEE Project - Overview

Mike Mineter
National e-Science Centre

Trinity College, Dublin
14th March 2006
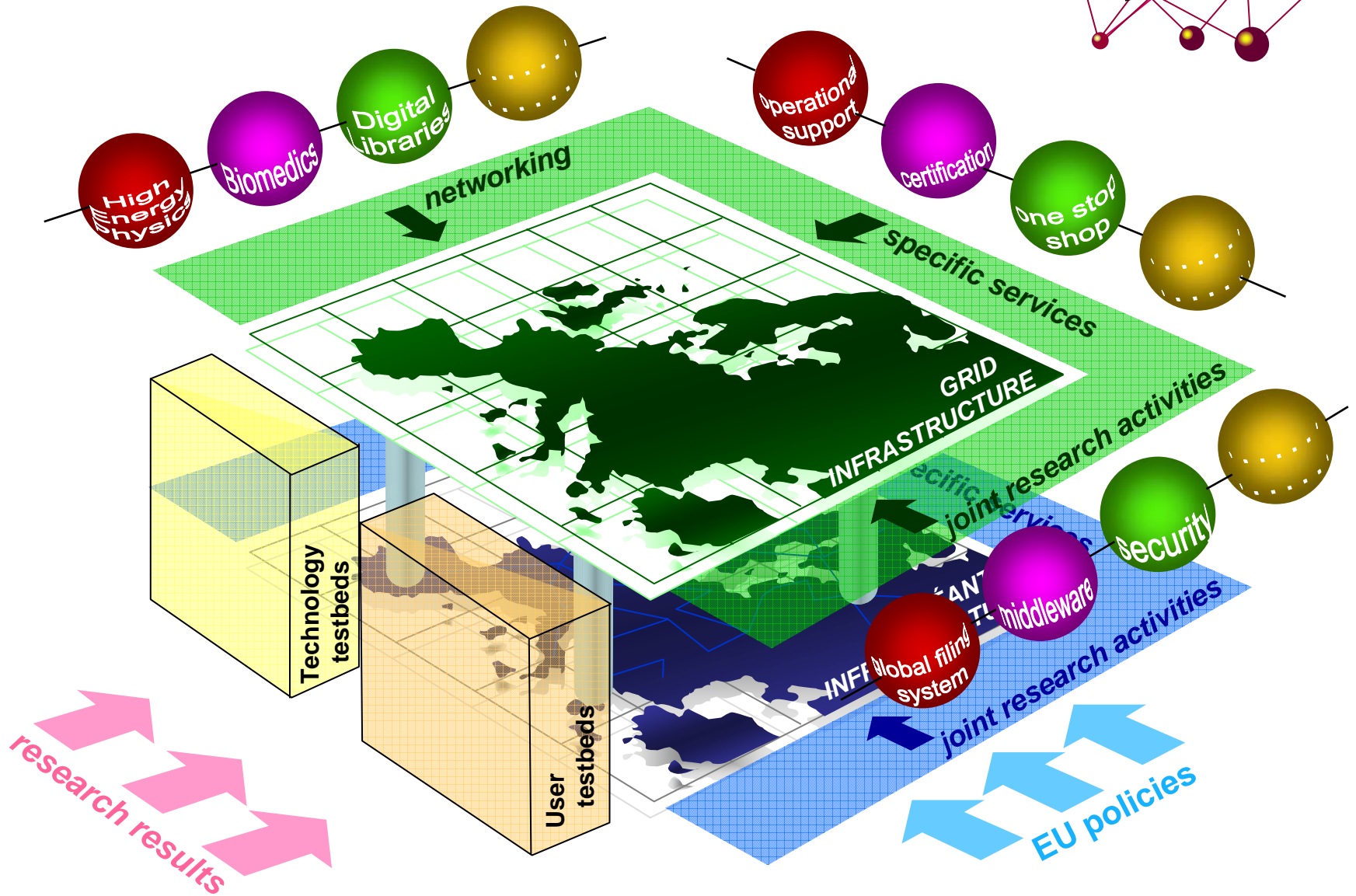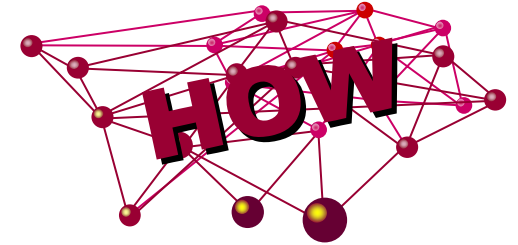
www.eu-egee.org

Information Society
and Media

Enabling Grids for E-sciencE
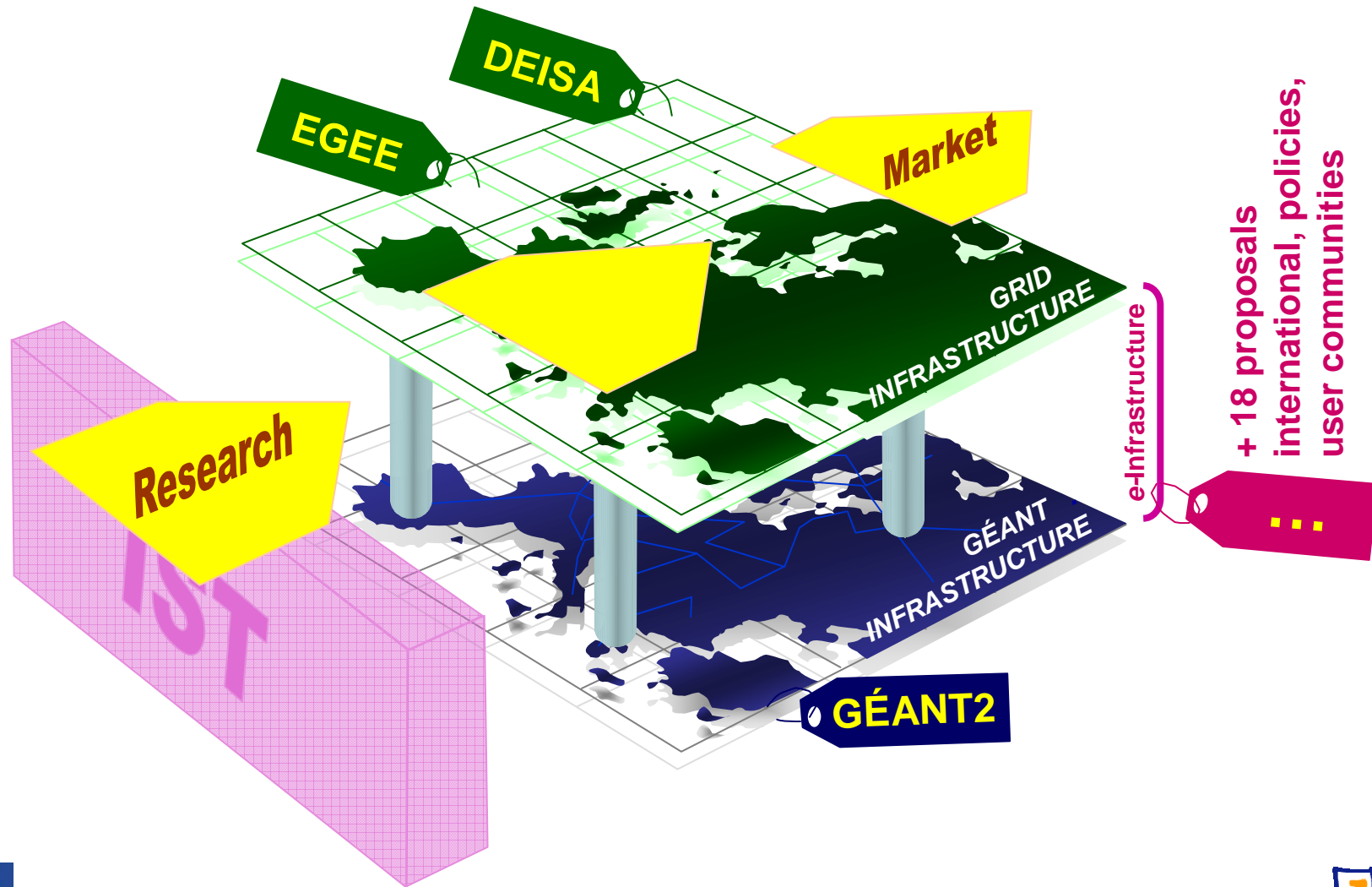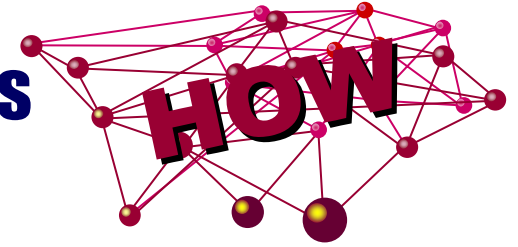
EGEE – "Enabling Grids for e-Science"

- **Why does EGEE exist?**

- **What has EGEE done?**

- **What is EGEE doing?**

- **What will EGEE do?**

- **Lisbon strategy: Research and Innovation will be the most important factors in determining Europe's success through the next decades**

- **Entering the "knowledge society" from the "industrial society"**
    - Industrial society: transportation, energy infrastructures
    - Knowledge society: e-Infrastructure
      built with Information and Communication Technologies

- **Result: EU investing in people and in the technology**

- **New infrastructures = New ways for people to cooperate**

e-Infrastructure - Implementation blocks

HOW

High Energy Physics
Biomedics
Digital libraries

Operational support
certification
One stop shop

networking

specific services

GRID INFRASTRUCTURE

joint research activities

specific services

security

middleware

Global filing system

Technology testbeds

User testbeds

research results

joint research activities

EU policies

Mário Campolargo DG INFSO F3, Pisa 24th October 2005

e-Infrastructure - Strategic building blocks

HOW

EGEE

DEISA

Market

Research

IST

GRID INFRASTRUCTURE

GÉANT INFRASTRUCTURE

GÉANT2

e-Infrastructure

+ 18 proposals international, policies, user communities

...

European Commission

Mário Campolargo DG INFSO F3, Pisa 24th October 2005

Information Society and Media

- **To underpin collaboration**

- **Link with and build on national, regional and international initiatives**

- **Foster world-wide international cooperation both in the creation and the use of grid infrastructure**

Collaboration

Grid

Operations, Support and training

Network infrastructure (*GÉANT***)**

eGee

Enabling Grids for E-sciencE
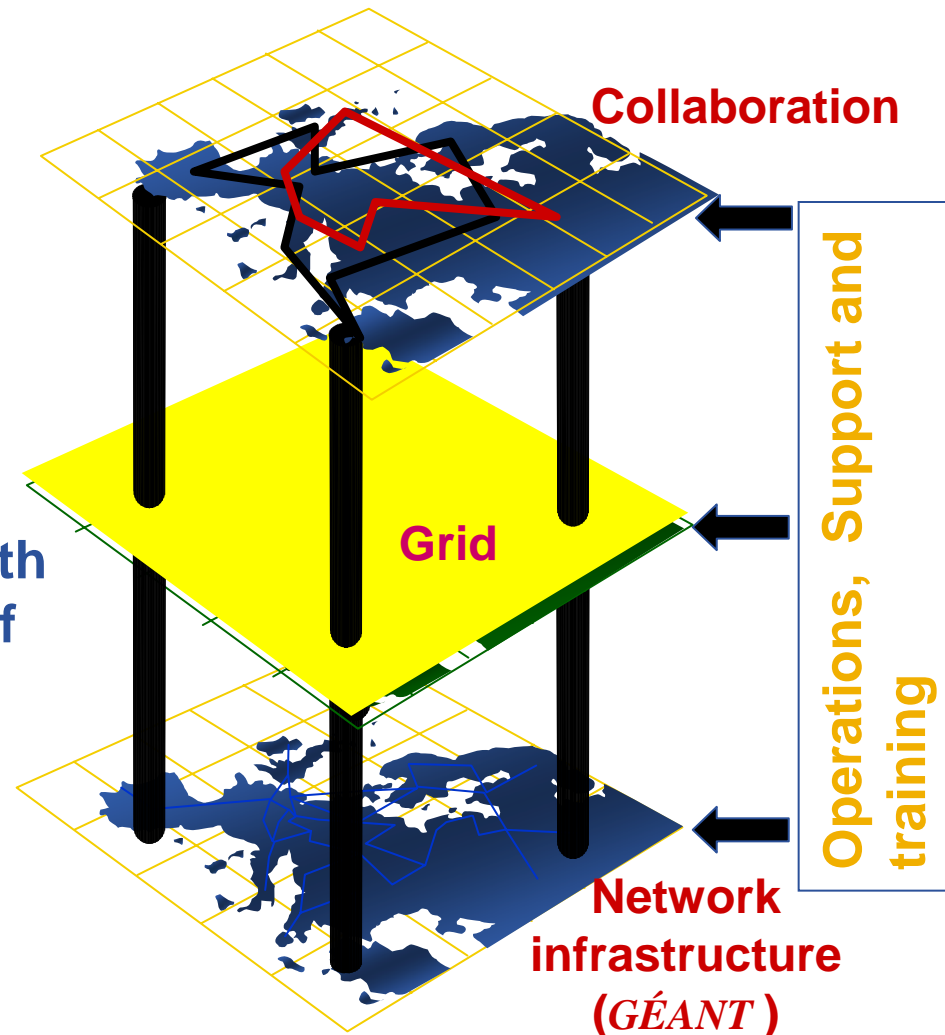
- **Infrastructure**
  - Manage and operate production Grid for European Research Area
  - Interoperate with e-Infrastructure projects around the globe
  - Contribute to Grid standardisation efforts

- **Support applications from diverse communities**
  - High Energy Physics
  - Biomedicine
  - Earth Sciences
  - Astrophysics
  - Computational Chemistry
  - Fusion
  - Geophysics
  - Finance, Multimedia
  - …

- **Business**
  - Forge links with the full spectrum of interested business partners
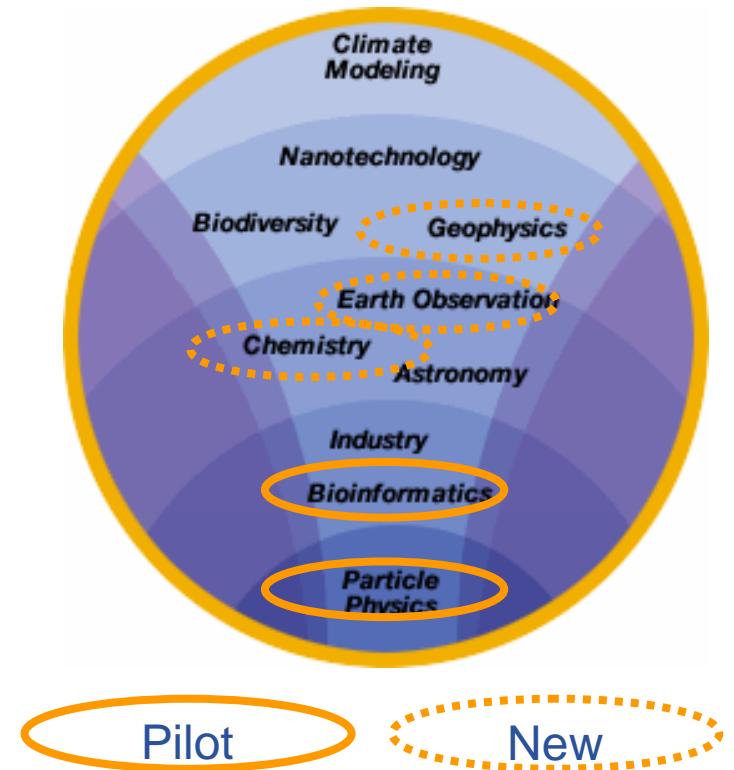
+ **Disseminate knowledge about the Grid through training**
+ **Prepare for sustainable European Grid Infrastructure**

Enabling Grids for E-sciencE
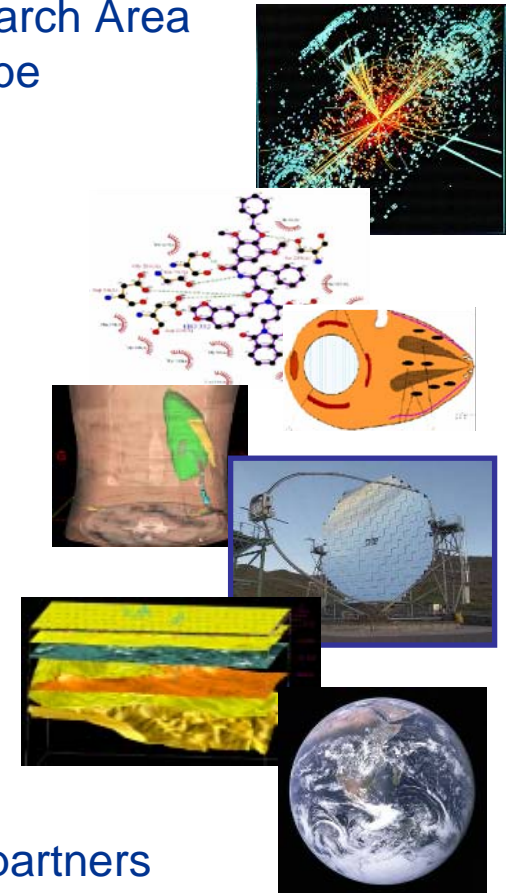
- **70 leading institutions in 27 countries, federated in regional Grids**

- **~32 M Euros EU funding for first 2 years starting April 2004 (matching funds from partners)**

- **Leveraging national and regional grid activities**

- **Promoting scientific partnership outside EU**

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

- **Goal: establish production quality sustained Grid services with**
  - 3000 users from at least 5 disciplines
  - integrate 50 sites into a common infrastructure
  - offer 5 Petabytes (10^{15}) storage
- **Achieved**
  - **> 180 sites in 39 countries**
  - **~ 20 000 CPUs**
  - **> 5 PB storage**
  - **> 10 000 concurrent jobs per day**
  - **> 60 Virtual Organisations**

Climate Modeling

Nanotechnology

Biodiversity   Geophysics

Earth Observation

Chemistry   Astronomy

Industry

Bioinformatics

Particle Physics

Pilot   New

eGee

- **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)**

24% Joint Research

28% Networking

48% Services

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

egee

Enabling Grids for E-sciencE

- … the largest multi-VO production grid in the world!

- What's happening now?
  http://gridportal.hep.ph.ic.ac.uk/rtm/

- What resources are connected?
  http://goc.grid-support.ac.uk/gridsite/monitoring/

- **EGEE committed to "hit the ground running"**
- **EGEE profits from the resources - no funded computing/data resources in EGEE**
  - Provided by the VOs
- **LCG obtains additional production and operation efforts**
- **LCG experiments now comprise several of the many VOs in EGEE**
- **Current service ("LCG-2") based on work done in LCG**
- **Middleware components to be upgraded by "gLite" services as they are proven**
  - "gLite 3" will be forged from LCG 2.7 + gLite services

LCG : Large Hadron Collider Compute Grid
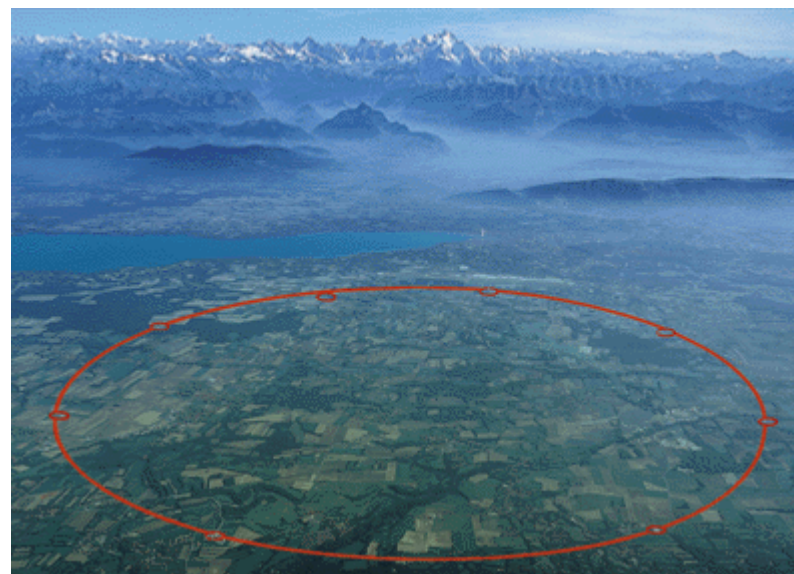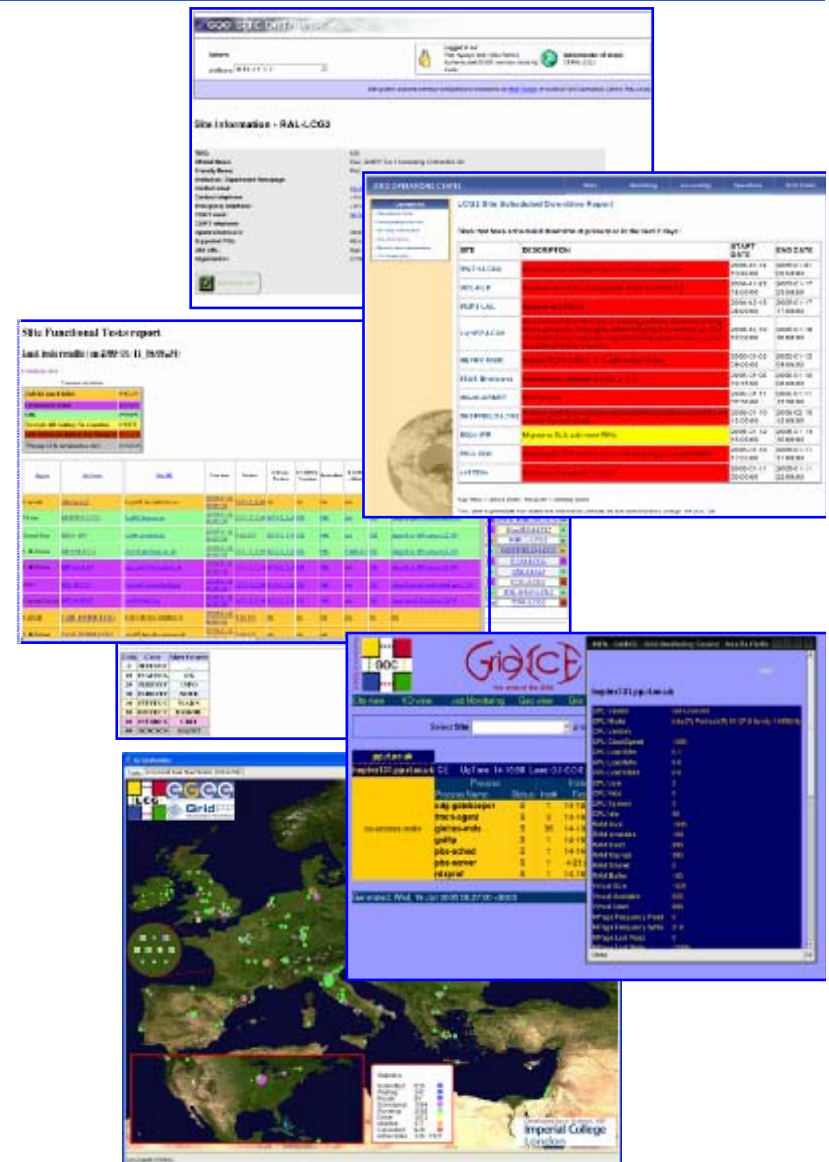
Operations
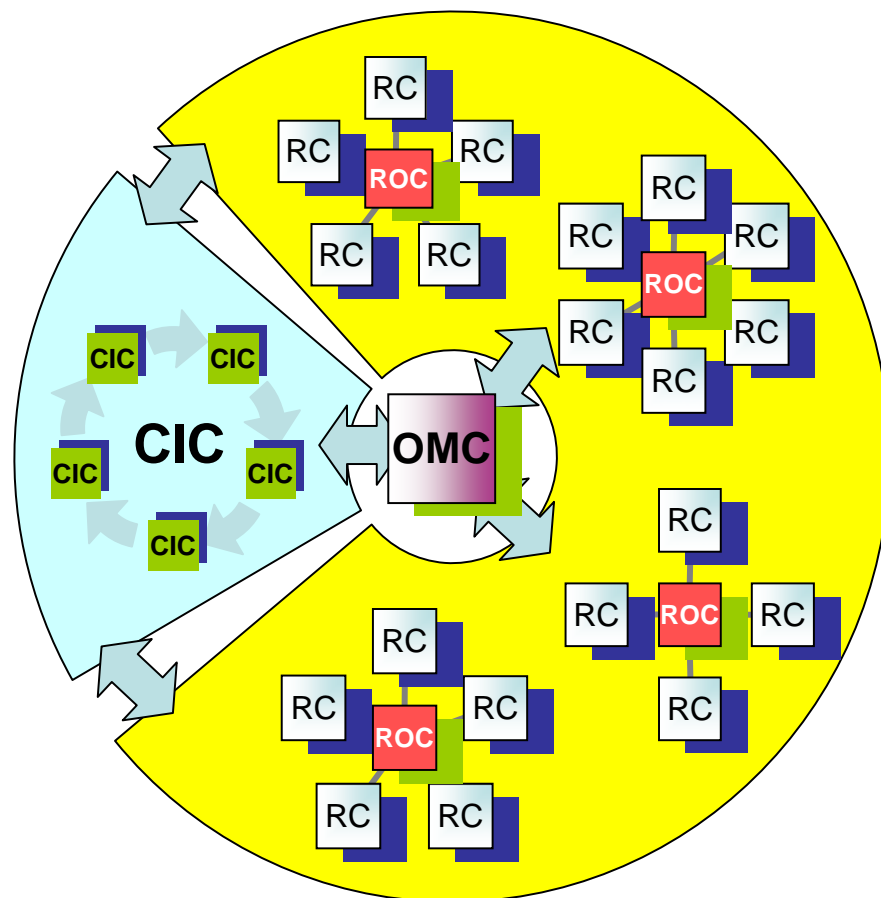
Grid Operations

RC = Resource Centre
ROC = Regional Operations Centre
CIC = Core Infrastructure Centre
OMC = Operations Management Centre

- **The *grid* is flat, but**
- ***Hierarchy* of responsibility**
  - Essential to scale the operation
- **CICs act as a single Operations Centre**
  - Operational oversight *(grid operator)* responsibility
  - rotates weekly between CICs
  - Report problems to ROC/RC
  - ROC is *responsible* for ensuring problem is resolved
  - ROC oversees regional RCs
- **ROCs responsible for organising the operations in a region**
  - Coordinate deployment of middleware, etc
- **CERN coordinates sites not associated with a ROC**

EGEE Operations Process

- **Grid operator on duty**
  - 6 teams working in weekly rotation
    - CERN, IN2P3, INFN, UK/I, Ru,Taipei
  - Crucial in improving site stability and management
- **Operations coordination**
  - Weekly operations meetings
  - Regular ROC, CIC managers meetings
  - Series of EGEE Operations Workshops
    - Nov 04, May 05, Sep 05, (June 06?)
- **Geographically distributed responsibility for operations:**
  - There is no "central" operation
  - Tools are developed/hosted at different sites:
    - GOC DB (RAL), SFT (CERN), GStat (Taipei), CIC Portal (Lyon)
- **Procedures described in Operations Manual**
  - Introducing new sites
  - Site downtime scheduling
  - Suspending a site
  - Escalation procedures
  - etc

Global Grid User Support - first contact for users
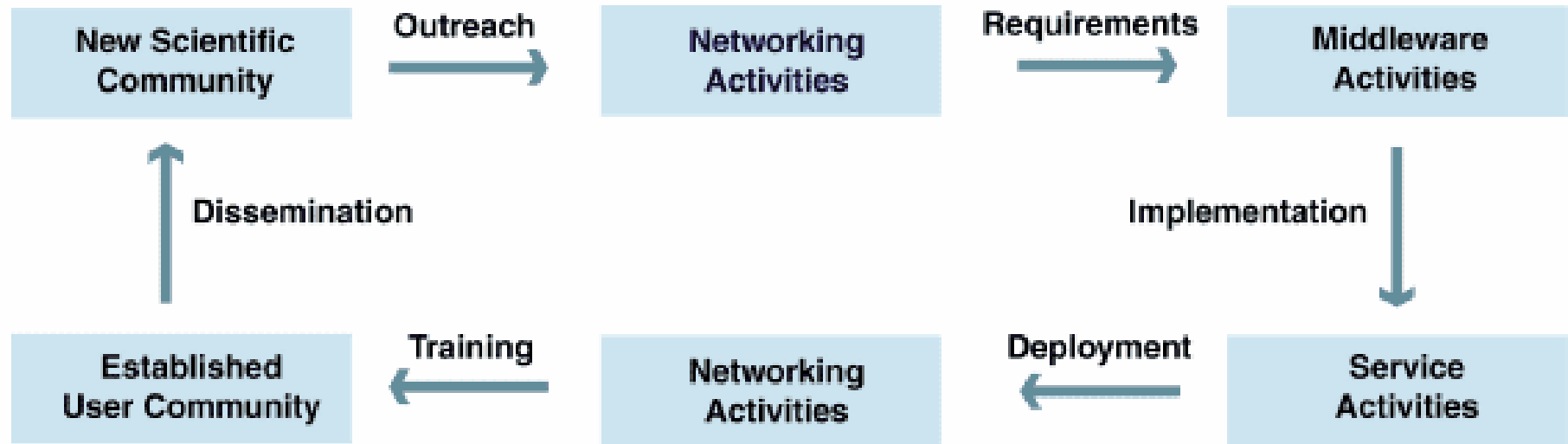

http://www.ggus.org

You need to **register**
in order to be able
To use this portal
(**GSI** or password based)

You can register
as **User** or as
Supporter.

Supporter ?
If you
have a good
knowledge in Grid
and have time
to provide support
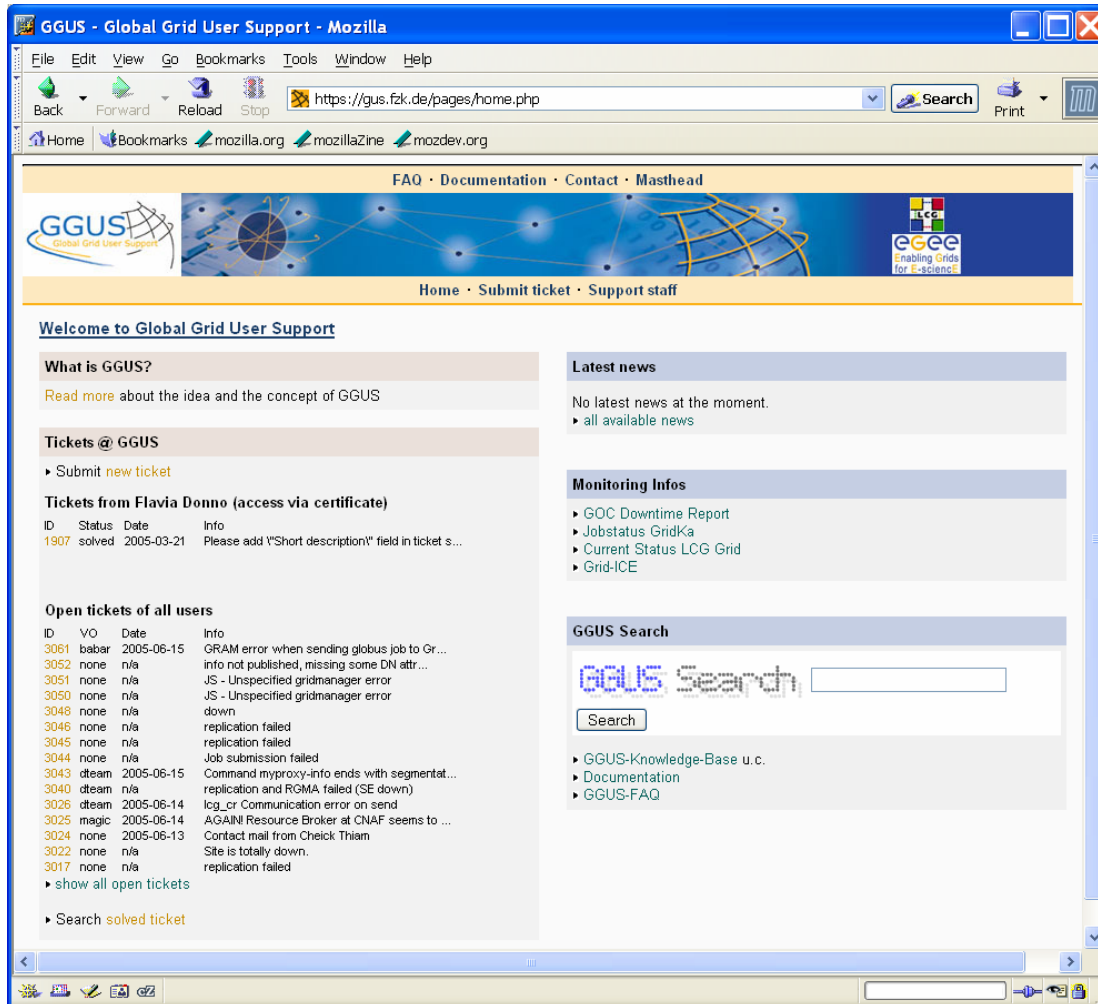
Building user communities

Enabling Grids for E-sciencE

- ## Why t-infrastructure?
  - e-Infrastructure for production
  - t-Infrastructure for training

- ## Need guaranteed response for tutorials; limit the vulnerability of production systems
  - use training grid
  - have training CA
  - able to change middleware to prepare participants for future releases on production system

- ## Also:
  - need safe resources for installation training
  - easy entry point for new communities

19 sites in 3 continents !

International cooperation

- **Cooperation between EGEE and other Grid activities**
  - **Globus Alliance, Condor**
  - **Training/workshop events (International Summer School of Grid Computing, July… 2 intense weeks… registrations open! http://www.dma.unina.it/~murli/ISSGC06/)**
  - eInfrastructure reflection group in Europe http://www.e-irg.org/
- **Standard setting through attendance at global standard bodies such as the Global Grid Forum.**
  - Grid Storage Management GGF working group - http://sdm.lbl.gov/gsm/
  - Security, Authentication: US – EU cooperation
- **Mutual recognition of Certificate Authorities**
  - Requires collaboration to establish policy - and mutuality

Enabling Grids for E-sciencE

- **Joint Security Policy Group**
  - EGEE with strong input from OSG
  - Policy Set:

```
                    Incident
                    Response
  Certification                          Audit
  Authorities                         Requirements

  Usage          Security & Availability      VO
  Rules               Policy               Security

      User Registration      Application Development
      & VO Management        & Network Admin Guide
```
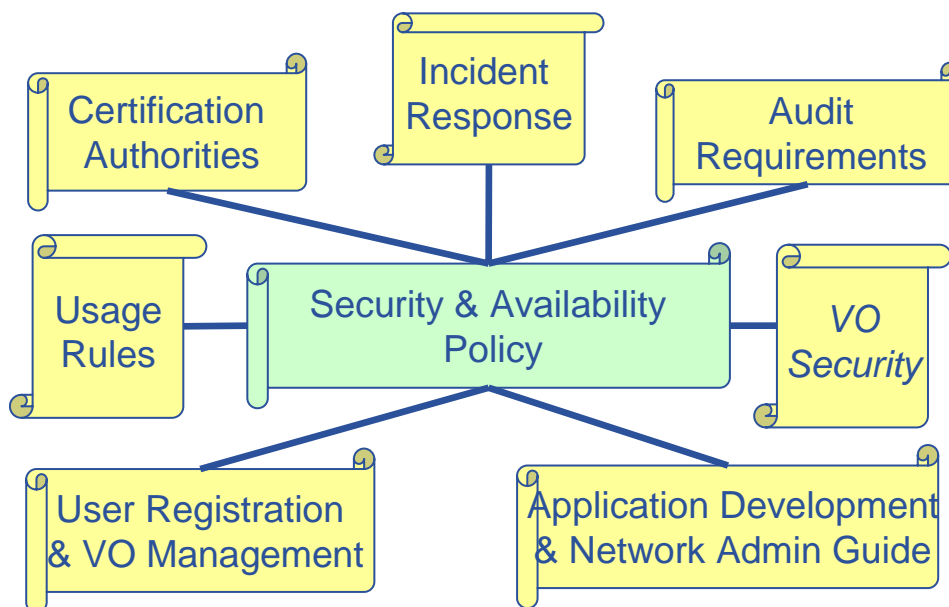
- **Policy Revisions**
  - Grid Acceptable Use Policy (AUP)
    - https://edms.cern.ch/document/428036/
    - common, general and simple AUP
    - for all VO members using many Grid infrastructures
      - *EGEE, OSG, SEE-GRID, DEISA, national Grids…*
  - VO Security
    - https://edms.cern.ch/document/573348/
    - responsibilities for VO managers and members
    - VO AUP to tie members to Grid AUP accepted at registration
  - Incident Handling and Response
    - https://edms.cern.ch/document/428035/
    - defines basic communications paths
    - defines requirements (MUSTs) for IR
      - *reporting*
      - *response*
      - *protection of data*
      - *analysis*
    - not to replace or interfere with local response plans

EGEE-II

Enabling Grids for E-sciencE

- **EGEE-II proposal submitted to the EU**
  - Proposed start 1 April 2006

- **Natural continuation of EGEE**
  - Expanded consortium
  - Emphasis on providing an infrastructure
    - → increased support for applications
    - → interoperate with other infrastructures
    - → more involvement from Industry
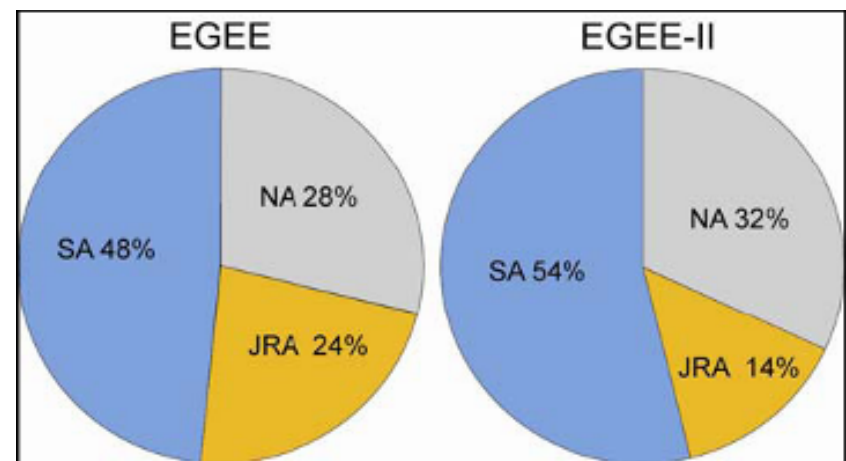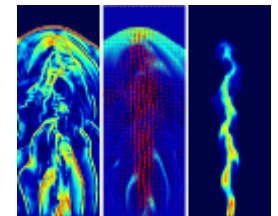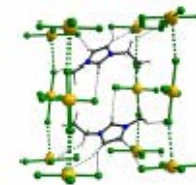
- **More than 90 partners**
- **32 countries**
- **12 federations**
→ **Major and national Grid projects in Europe, USA, Asia**

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

Enabling Grids for E-sciencE

- **EGEE is running the largest multi-VO grid in the world!**
- **Creating the "grid layer" in e-Infrastructure for research, public service and industry**
- **Key concepts for EGEE**
    - Sustainability – planning for the long-term
    - Production quality
    - And…

- **Grids are fundamentally about …**
- **… how diverse communities and organisations commit to cooperate**
- **… and how these communities can be supported by operations, training, support, and (most transient of all!) middleware**
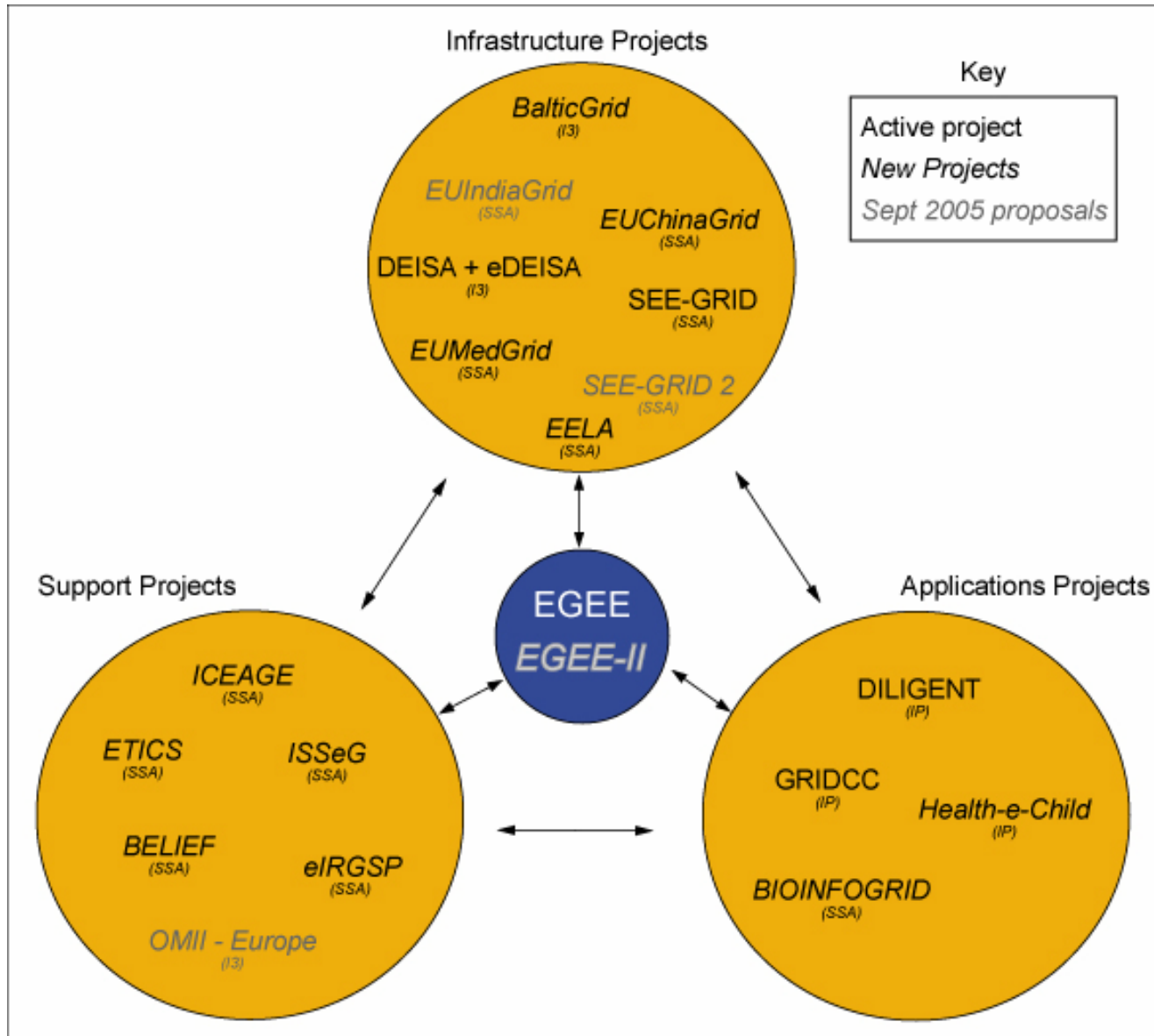
Name	Description	Common partners with EGEE
BalticGrid	EGEE extension to Estonia, Latvia, Lithuania	KTH – PSNC – CERN
EELA	EGEE extension to Brazil, Chile, Cuba, Mexico, Argentina	CSIC – UPV – INFN – CERN – LIP – RED.ES
EUChinaGRID	EGEE extension to China	INFN – CERN – DANTE – GARR – GRNET
EUMedGRID	EGEE extension to Malta, Algeria, Morocco, Egypt, Syria, Tunisia, Turkey	INFN – CERN – DANTE – GARR – GRNET – RED.ES
ISSeG	Site security	CERN – CSSI – FZK – CCLRC
eIRGSP	Policies	CERN – GRNET
ETICS	Repository, Testing	CERN – INFN – UWM
ICEAGE	Repository for Training & Education, Schools on Grid Computing	UEDIN – CERN – KTH – SZTAKI
BELIEF	Digital Library of Grid documentation, organisation of workshops, conferences	UWM
BIOINFOGRID	Biomedical	INFN – CNRS
Health-e-Child	Biomedical – Integration of heterogeneous biomedical information for improved healthcare	CERN

- **Fundamental activity in preparation of LHC start up**
  - Physics
  - Computing systems
- **Examples:**
  - LHCb: **~700 CPU/years in 2005** on the EGEE infrastructure
  - ATLAS: **over 10,000 jobs per day**
    - Comprehensive analysis: see S.Campana et al., "Analysis of the ATLAS Rome Production experience on the EGEE Computing Grid", e-Science 2005, Melbourne, Australia
  - A lot of activity in all involved applications (including as usual a lot of activity within non-LHC experiments like BaBar, CDF and D0)
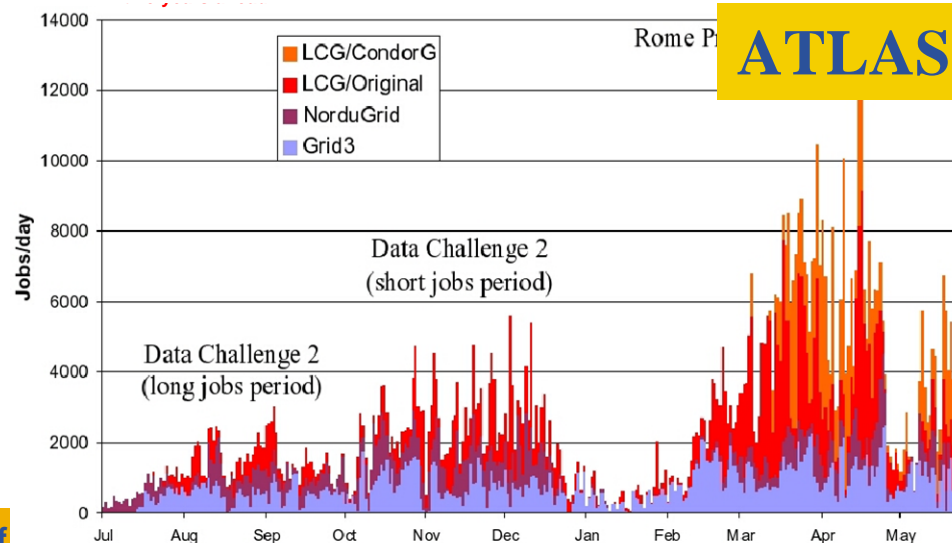
Roberto Barbera, User Forum, Feb.'06

CPU used: 6,389,638 h
Data Output: 77 TB
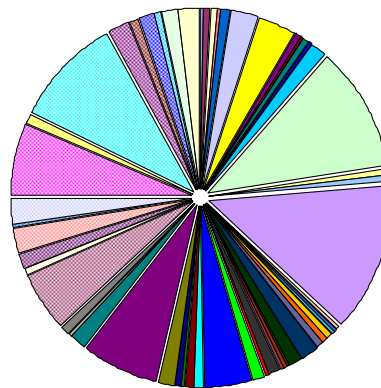
LHCb

DIRAC.Barcelona.es 0.214%	
DIRAC.CERN.ch 0.571%	
DIRAC.CracowAgu.pl 0.001%	
DIRAC.LHCBONLINE.ch 0.779%	
DIRAC.PNPI.ru 0.000%	
DIRAC.ScotGrid.uk 3.068%	DIRAC.Zurich-spz.ch 0.003%
DIRAC.Zurich.ch 0.756%	LCG.ACAD.bg 0.106%
LCG.BHAM-HEP.uk 0.705%	LCG.Barcelona.es 0.281%
LCG.Bari.it 1.357%	LCG.Bologna.it 0.032%
LCG.CERN.ch 10.960%	LCG.CESGA.es 0.528%
LCG.CGG.fr 0.676%	LCG.CNAF-GRIDIT.it 0.012%
LCG.CNAF.it 13.196%	LCG.CNB.es 0.385%
LCG.CPPM.fr 0.242%	LCG.CSCS.ch 0.282%
LCG.CY01.cy 0.103%	LCG.Cagliari.it 0.515%
LCG.Cambridge.uk 0.010%	LCG.Catania.it 0.551%
LCG.Durham.uk 0.476%	LCG.Edinburgh.uk 0.031%
LCG.FZK.de 1.708%	LCG.Ferrara.it 0.073%
LCG.Firenze.it 1.047%	LCG.GR-01.gr 0.349%
LCG.GR-02.gr 0.226%	LCG.GR-03.gr 0.171%
LCG.GR-04.gr 0.056%	LCG.GRNET.gr 1.170%
LCG.HPC2N.se 0.001%	LCG.ICI.ro 0.088%
LCG.IFCA.es 0.022%	LCG.IHEP.su 1.245%
LCG.IN2P3.fr 4.143%	LCG.INTA.es 0.076%
LCG.IPP.bg 0.033%	LCG.ITEP.ru 0.792%
LCG.Imperial.uk 0.891%	LCG.Iowa.us 0.287%
LCG.JINR.ru 0.472%	LCG.KFKI.hu 1.436%
LCG.Lancashire.uk 6.796%	LCG.Legnaro.it 1.569%
LCG.Manchester.uk 0.285%	LCG.Milano.it 0.770%
LCG.Montreal.ca 0.069%	LCG.NIKHEF.nl 5.140%
LCG.NSC.se 0.465%	LCG.Napoli.it 0.175%
LCG.Oxford.uk 1.214%	LCG.PIC.es 2.366%
LCG.PNPI.ru 0.278%	LCG.Padova.it 2.041%
LCG.Pisa.it 0.121%	LCG.QMUL.uk 6.407%
LCG.RAL-HEP.uk 0.938%	LCG.RAL.uk 9.518%
LCG.RHUL.uk 2.168%	LCG.SARA.nl 0.675%
LCG.Sheffield.uk 0.094%	LCG.Torino.it 1.455%
LCG.Toronto.ca 0.343%	LCG.Triumf.ca 0.105%
LCG.UCL-CCC.uk 1.455%	LCG.USC.es 1.853%

ATLAS

Enabling Grids for E-sciencE

- **Note the contrast between**
    - **"best-efforts" and production grids** for international collaborations… with hundreds of sites providing resources
        - Operational infrastructure (>40% of EGEE budget on operations)
        - Quality of service / policy issues
        - Focus on stability of sites
        - Support for VO's
    - **Research and production middleware**
        - procedures for upgrading middleware
            - *Pre-production grid – running many VO's applications*
    - **Project grids and international production grids**
        - Extent of international cooperation, policy agreement...
        - Multiple VO's

Enabling Grids for E-sciencE

- **Industry Task Force**
  - Group of industry partners in the project
  - Links related industry projects (NESSI, BEinGRID, …)
  - Works with EGEE's Technical Coordination Group (TCG) to place industry requirements on equal footing

- **Collaboration with CERN openlab project**
  - IT industry partnerships for hardware and software development

- **EGEE Business Associates (EBA)**
  - Companies sponsoring work on joint-interest subjects
    - Technical developments
    - Market Surveys
    - Business modelling
    - Exploitation strategies
    - Transfer of know-how and services to industry

- **Industry Forum (representatives in most European countries)**
  - Led by Industry to improve Grid take-up in Industry
  - Organises industry events and disseminates grid information

- **Join the EGEE Infrastructure**
  - A new VO (Virtual Organisation) benefits from the operations and support – the VO sites can be monitored and supported as part of the infrastructure
  - Potential access to other resources (part of a shared infrastructure)

- **Adopt EGEE technology for internal use**
  - Reuse middleware and site automation/monitoring tools
  - Operate private infrastructure

Usage requiring significant support can be discussed as potential joint ventures