# An introduction to EGEE

**Mike Mineter**
**NeSC Edinburgh**
**mjm@nesc.ac.uk**

www.eu-egee.org

Information Society

# Acknowledgements

**This presentation includes slides and information from many colleagues in EGEE, especially from the 1st project review in February 2005, including :**

- Fabrizio Gagliardi (1st Review)
- Bob Jones (UK AHM 2004 talk)
- Ian Bird
- Frédéric Hemmer
- Roberto Barbera

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
- FAQ

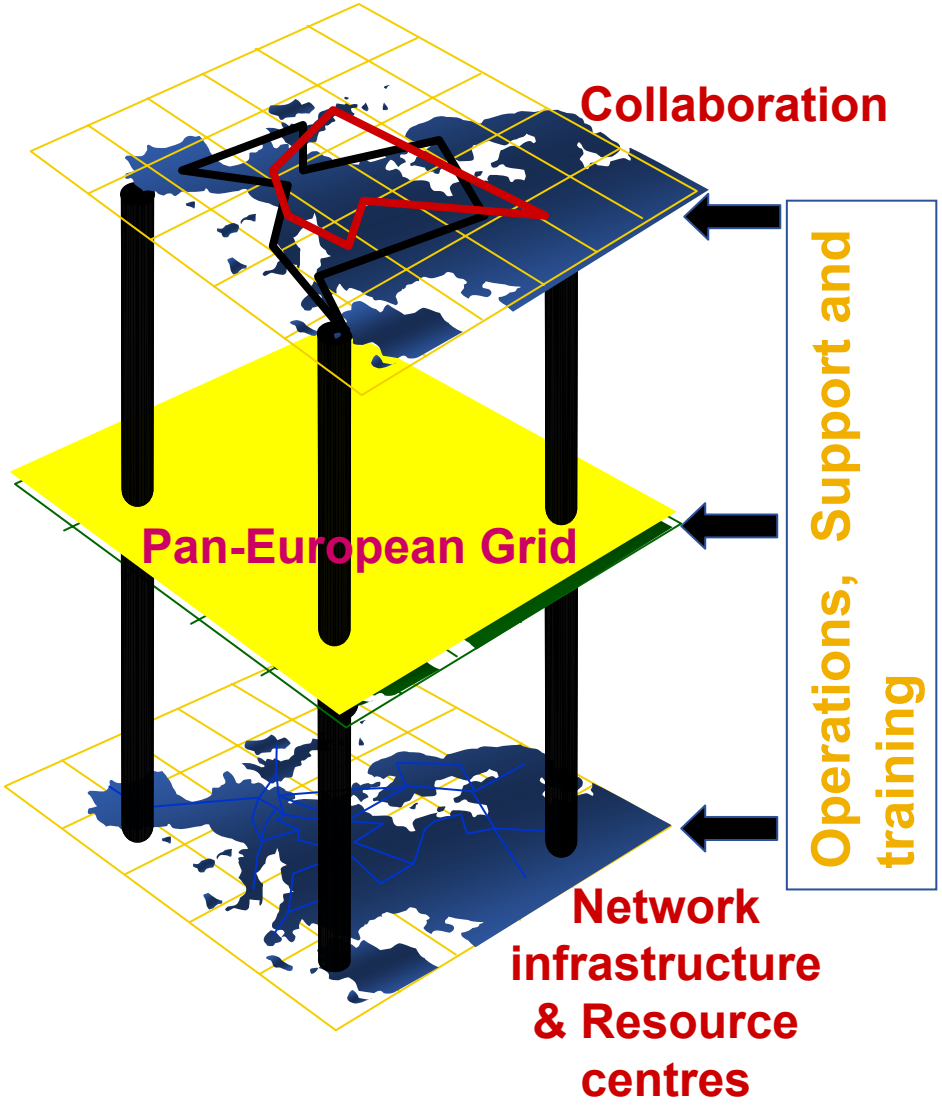
# EGEE funding…

- **Funded by the European Commission**
- **DECISION No 1513/2002/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 June 2002 concerning the Sixth Framework Programme of the European Community for research, technological development and demonstration activities, contributing to the creation of the European Research Area and to innovation (2002 to 2006)**
- **Structuring the European Research Area (ERA) http://europa.eu.int/comm/research/era/index_en.html**
- **Area 3.2.3: Communication Network Development – Grids**
  - Instrument: **I**ntegrated **I**nfrastructure **I**nitiative (I3)
  - Call launched 17 Dec 2002, proposal submitted May 2003
- **Contract number:**
  - 508833
- **Began April 2004 for 2 years**

# EGEE – towards e-infrastructure

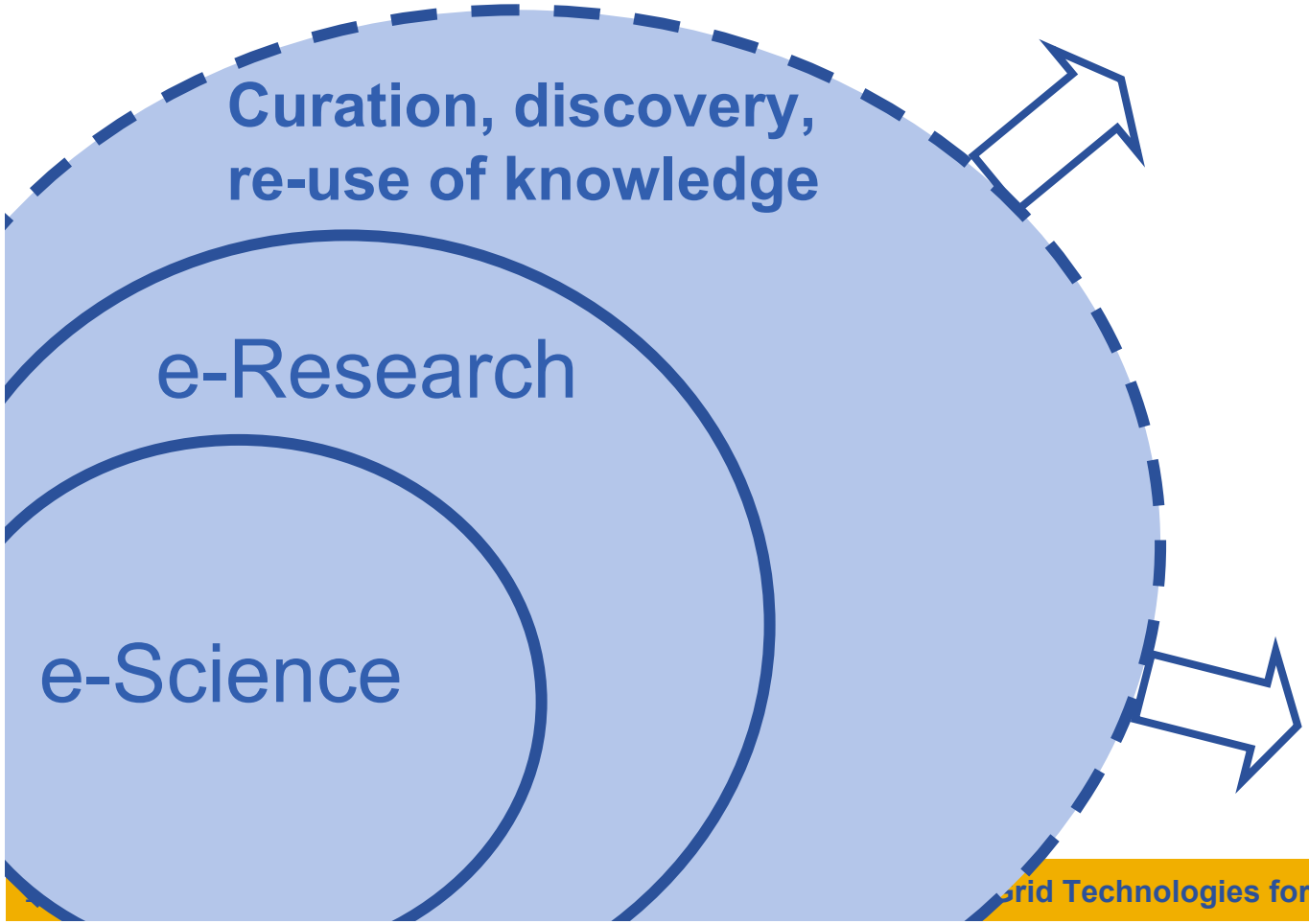
EGEE is building a large-scale production grid service to:

- Underpin research, technology and public service
- Link with and build on national, regional and international initiatives
- Foster international cooperation both in the creation and the use of the e-infrastructure



# Why DILIGENT / DELOS are important to EGEE -1

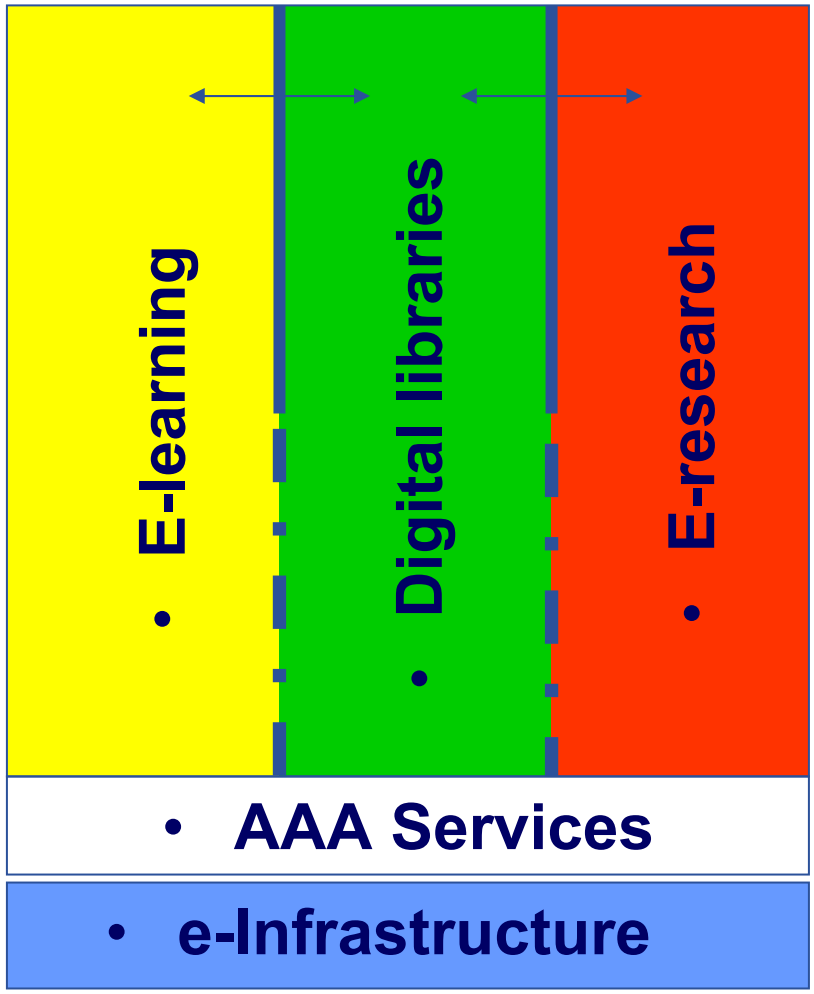
## The impact of grids when they support…



# Why DILIGENT / DELOS are important to EGEE -2

- **Centrality of curation, preservation**
- **Under-recognised by many researchers**
- **VDL needed for research as well as learning**

**Derived from a slide by the UK's JISC**



# Background

- **By 2003:**
  - Grid technology shown to be viable
  - Large amount of functional middleware
  - …thanks to:
    - FP5 : DataGrid, DataTAG, CrossGrid, etc…
    - USA: VDT, Globus, Condor, etc.
    - … and others

- **Next step - *major production infrastructure***
  - EGEE was proposed to the EU in 2003

- **2 year project began in April 2004, with a 4-year vision.**

# Grids for e-Infrastructure…

- In 2003, what was missing?
  - **Production-quality (stable, mature) Grid middleware**
  - **Production-quality operational support**
    - Grid Operation Centres, Helpdesks, etc.
  - **Multi-discipline grid-enabled application environment**
    - Now led by HEP, Bio-info
  - **Administrative and policy decision framework in order to share resources at pan-European scale (and beyond)**
    - Areas such as AAA (Authentication, Authorisation, Accounting)
    - End-to-end issues (Network related)
    - Funding Policies (Grid economics)
    - Resource Sharing Policies
    - Usage Policies

- **EGEE project is tackling most of the above issues**

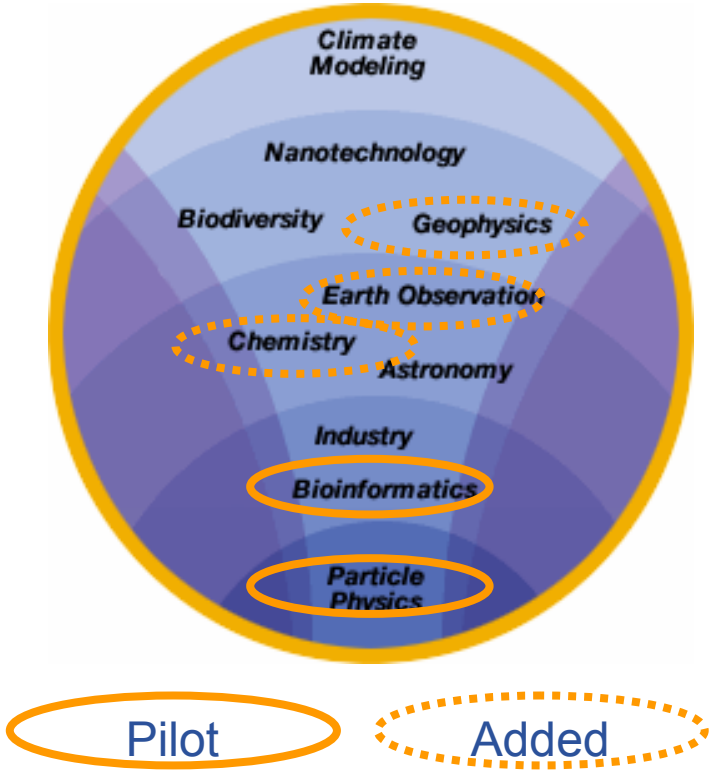
# Project Goals

## A four year programme:

- Build, deploy and operate a consistent, robust and secure grid that attracts new computing resources
- Improve and maintain the middleware in order to deliver a reliable service to users
- Attract new users from science and industry and ensure training and support for them

# In the first 2 years EGEE will

- **Establish production quality sustained Grid services**
  - 3000 users from at least 5 disciplines
  - integrate 50 sites into a common infrastructure
  - offer 5 Petabytes ($10^{15}$) storage
- **Demonstrate a viable general process to bring other scientific communities on board**
- **Propose a second phase in mid 2005 to take over EGEE in early 2006**



# EGEE has a scope much wider than…

- **Europe:**
  **it is an International project with partners world-wide, and funding from the EU**

- **e-Science:**
  **It will support non-scientific research and collaborations in industry, the public sector, … (health, virtual digital libraries…)**

- **Its end date of March 2006:**
  **the goal of EGEE is to create an infrastructure that will be sustainable, far beyond the end of its initial phase of funding. (LHC comes on line in 2007)**

# Contents

## EGEE: Enabling Grids for E-sciencE

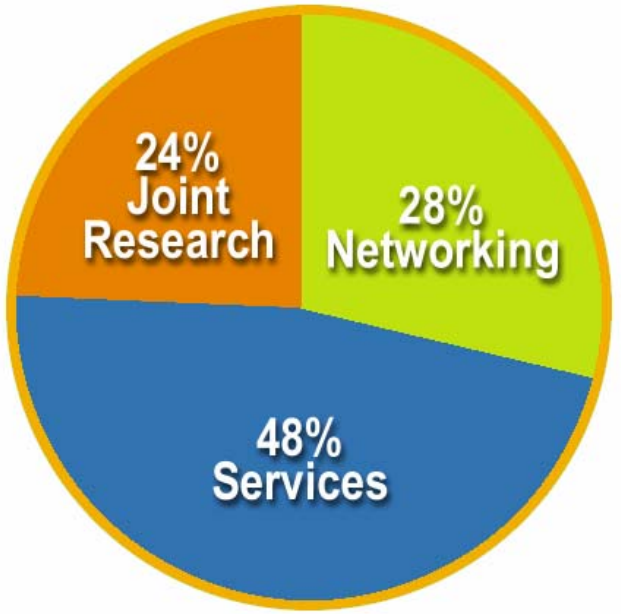
- Goals
- **Organisation**
- Activities and status
- FAQ

# EGEE Organisation

- 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 partnership outside EU

# Activities Definition

- **Network Activities**
  - NA1: **Project Management**
  - NA2: **Dissemination and Outreach**
  - NA3: **User Training and Induction**
  - NA4: **Application Identification and Support**
  - NA5: **Policy and International Cooperation**
- **Service Activities**
  - SA1: **Grid Support, Operation and Management**
  - SA2: **Network Resource Provision**
- **Joint Research Activities**
  - JRA1: **Middleware Reengineering + Integration**
  - JRA2: **Quality Assurance**
  - JRA3: **Security**
  - JRA4: **Network Services Development**




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

# Contents

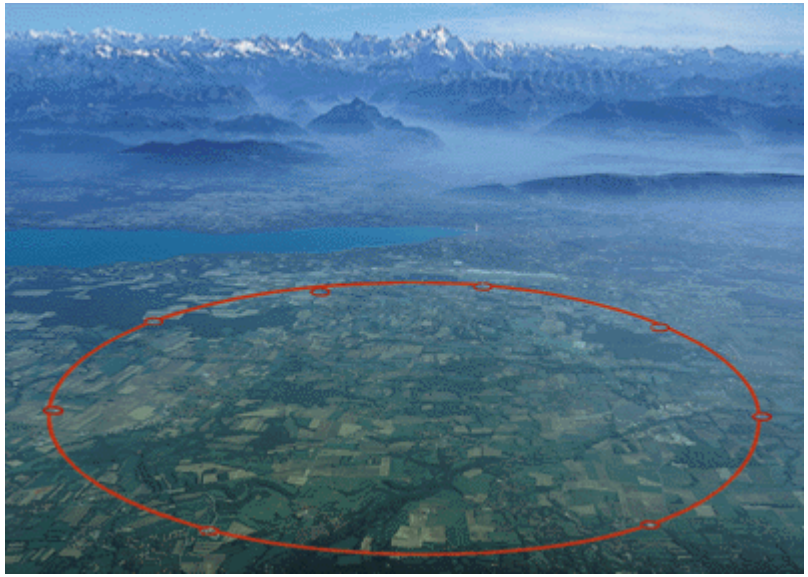
## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - JR activities: Middleware Re-engineering
  - Operations
  - (Human) Networking
- FAQ

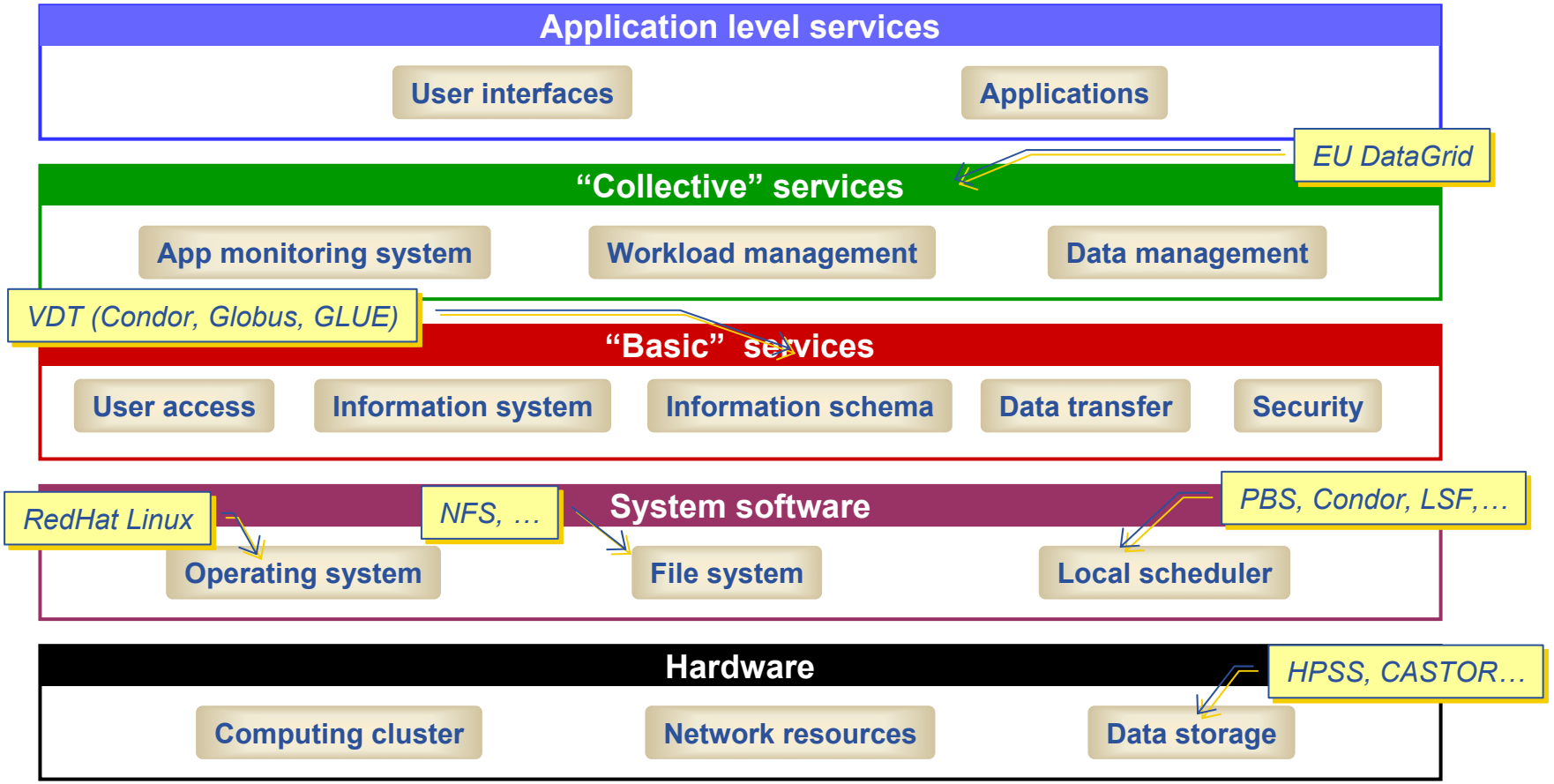
# LCG and EGEE

- EGEE committed to “hit the ground running” in the proposal
- Current service (“LCG-2”) based on work done in LCG
- EGEE profits from the resources - no funded computing/data resources in EGEE
- LCG obtains additional production and operation efforts

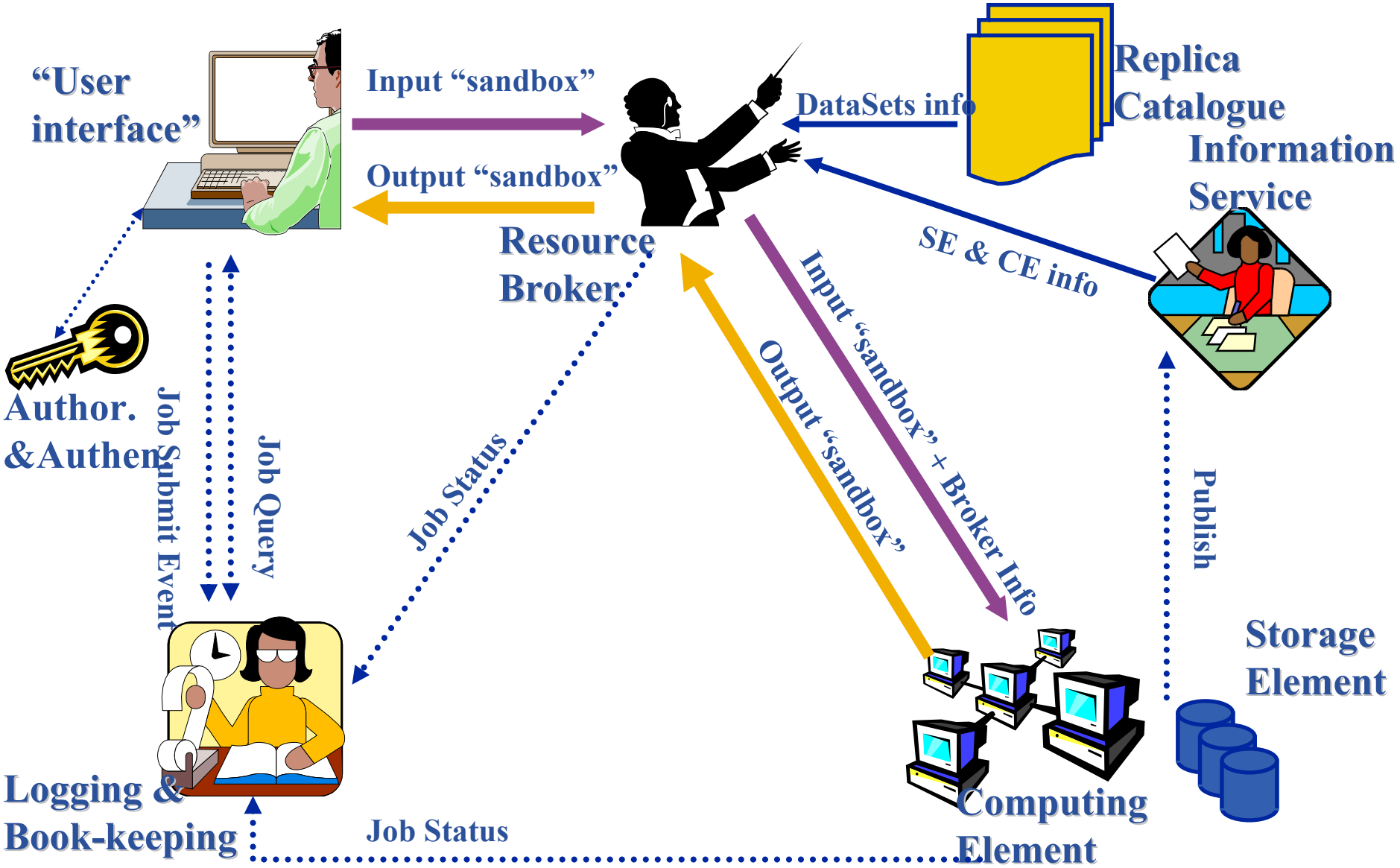
## LCG : Large Hadron Collider Compute Grid



# Current production grid: LCG-2

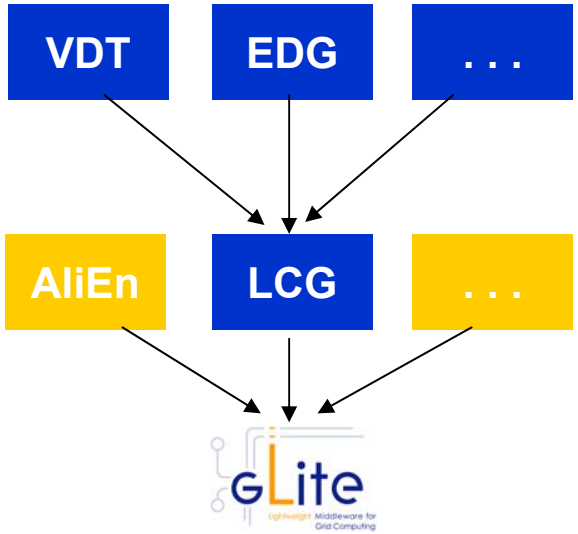


# Current production m'ware: LCG-2



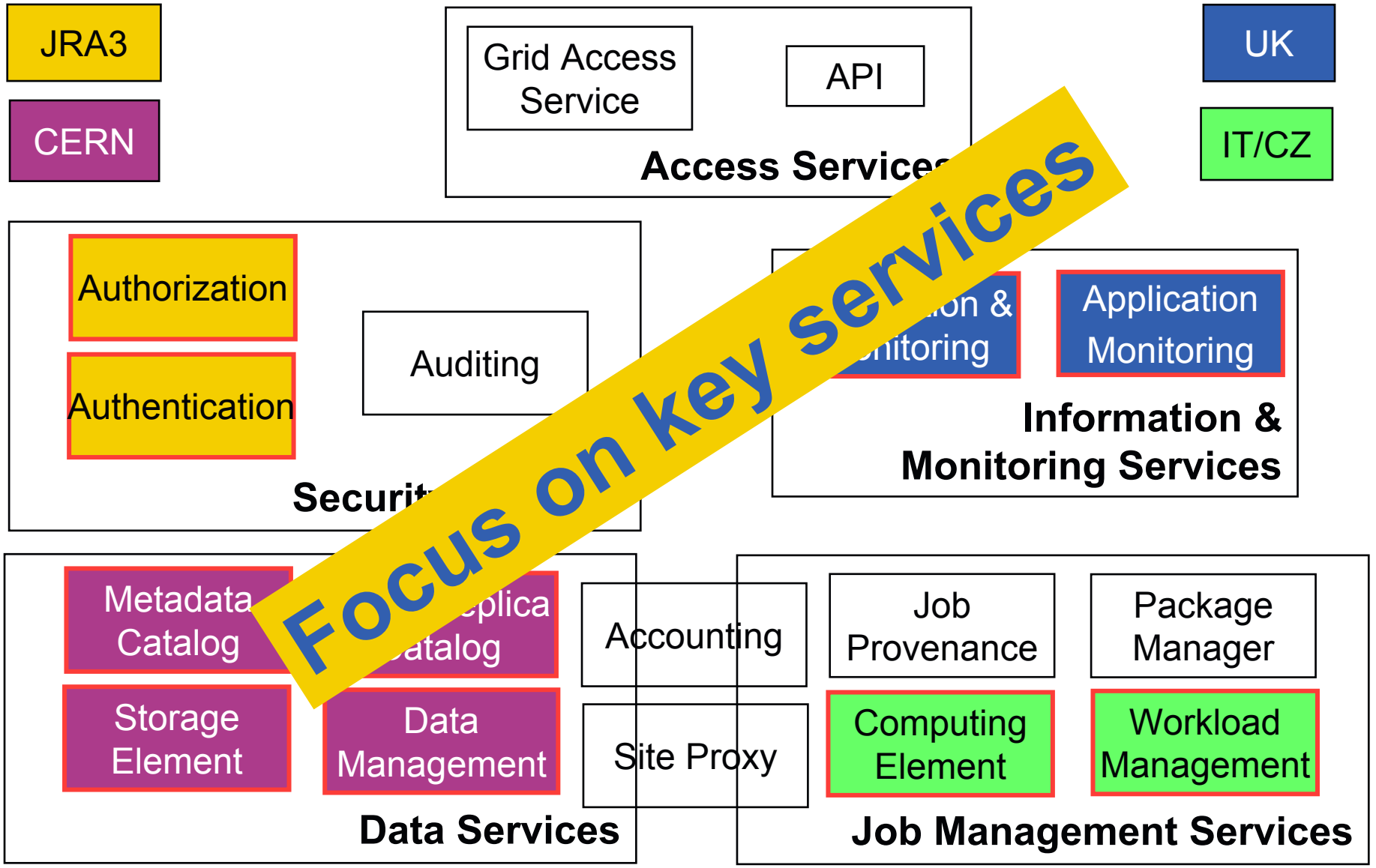
# gLite: Guiding Principles

- **Service oriented approach**
  - Allow for multiple interoperable implementations
- **Lightweight (existing) services**
  - Easily and quickly deployable
  - Use existing services where possible
    - Condor, EDG, Globus, LCG, …
- **Portable**
  - Being built on Scientific Linux and Windows
- **Security**
  - Sites and Applications
- **Performance/Scalability & Resilience/Fault Tolerance**
  - Comparable to deployed infrastructure




- **Co-existence with deployed infrastructure**
  - Co-existence with LCG-2 and OSG (US) are essential for the EGEE Grid services
- **Site autonomy**
  - Reduce dependence on ‘global, central’ services
- **Open source license**

# gLite Services for Release 1



# gLite Services for Release 1
# Software stack and origin (simplified)

- **Computing Element**
  - Gatekeeper (Globus)
  - Condor-C (Condor)
  - CE Monitor (EGEE)
  - Local batch system (PBS, LSF, Condor, …)
- **Workload Management**
  - WMS (EDG)
  - Logging and bookkeeping (EDG)
  - Condor-C (Condor)
- **Storage Element**
  - File Transfer/Placement (EGEE)
  - glite-I/O
  - GridFTP (Globus)
  - SRM: Castor (CERN), dCache (FNAL, DESY), ….
- **Catalog**
  - File and Replica Catalog (EGEE)
  - Metadata Catalog (EGEE)
- **Information and Monitoring**
  - R-GMA (EDG)
- **Security**
  - VOMS (DataTAG, EDG)
  - GSI (Globus)

# Middleware

- **February**
  - Installation and basic functionality testing with JRA1, SA1 and NA4
- **March**
  - Further testing for robustness and scalability
  - Official release of gLite (release 1) on pre-production service
- **April**
  - Start of incremental releases
    - leading to gLite v2.0 at PM21
  - Evaluation of OMII
    *** *see tomorrow's talk* ***
  - … and GT4 toolkits
- **May**
  - Update gLite architecture taking into account initial deployment and usage
  - Continue work on GGF/OGSA and provide feedback from deployment
  - Deployment of gLite components on production service
- **June**
  - Update gLite design according to architecture changes

# Middleware Summary

- **gLite 1.0 , released**
  - with key functionalities implemented
  - Moving towards Web Services

- **DJRA1.1 – EGEE Middleware Architecture (June 2004)**
  - https://edms.cern.ch/document/476451/
- **DJRA1.2 – EGEE Middleware Design (August 2004)**
  - https://edms.cern.ch/document/487871/

- **http://www.gLite.org/**
  - Links to documents

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - JR activities: Security
  - Operations
  - (Human) Networking
- FAQ

# Security

- **Support for security related software modules in gLite**
- **Continue work towards an agreed security infrastructure with other grid projects**
- **Revision of the security operational procedures**
- **Continuous evaluation of new CAs by EUGridPMA**
  - During this period it is expected that all EU member states involved in grid projects will have a national accredited Authority
- **Assessment of accounting infrastructure and analysis of what is missing to provide secure quota-based resource access**

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - JR activities: Networking
  - Operations
  - (Human) Networking
- FAQ

# Networking

- Definition of standard interface for network performance monitoring based on GGF NM-WG schema
- Specification of high-level network monitoring and diagnostic tools
- Definition of end-to-end Service Level Agreements (SLAs) between EGEE sites and GEANT
- Development and testing of a prototype bandwidth reservation service
- Further training/dissemination on IPv6 issues

# Contents

## EGEE: Enabling Grids for E-sciencE

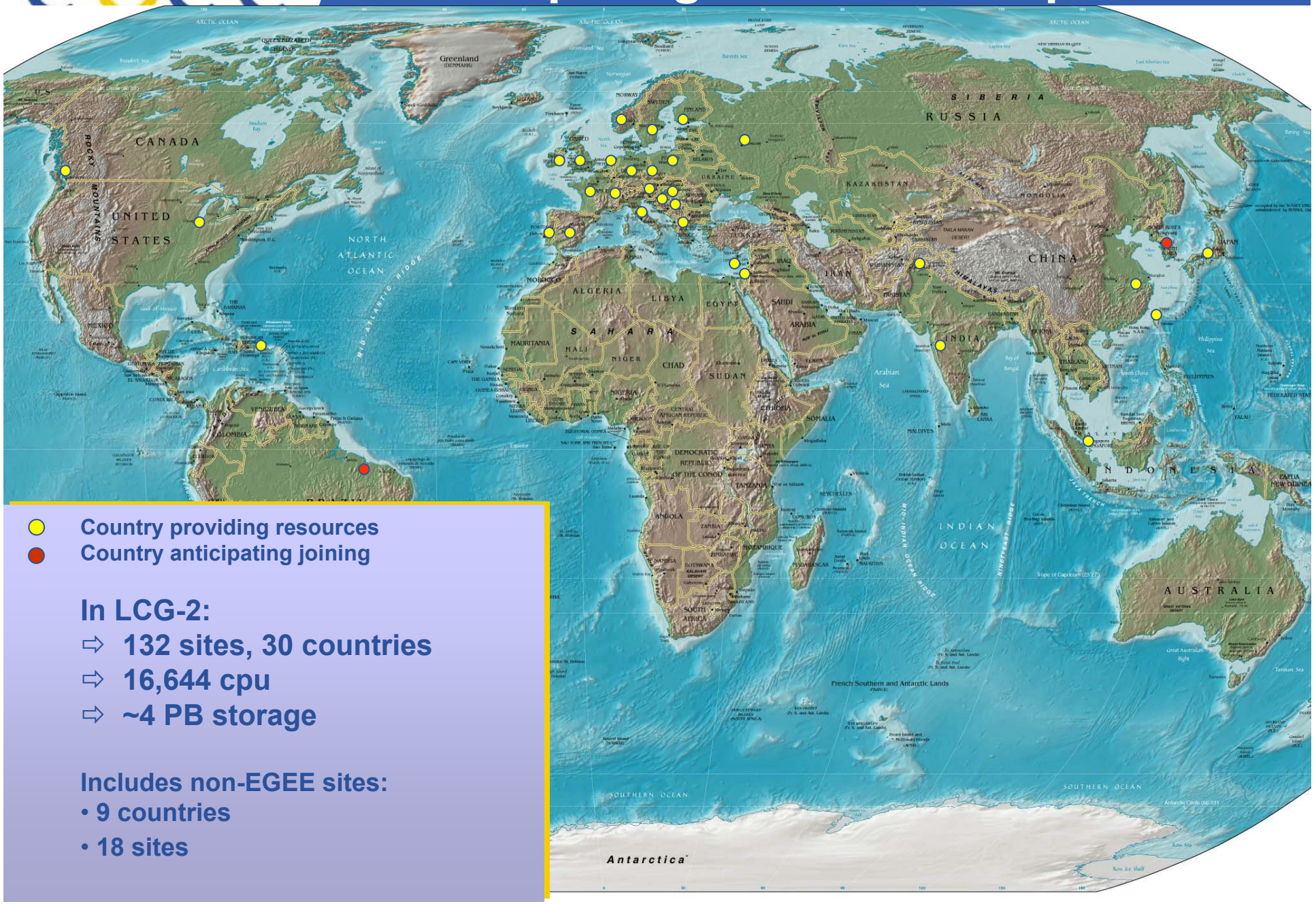
- Goals
- Organisation
- Activities and status
  - JR activities
  - **Operations**
  - (Human) Networking
- FAQ

# Operations - Introduction

- **Strategy has been to**
  - simplify as far as possible what is deployed, and to make that robust and useable.
  - In parallel construct the essential infrastructure needed to operate and maintain a grid infrastructure in a sustainable way.

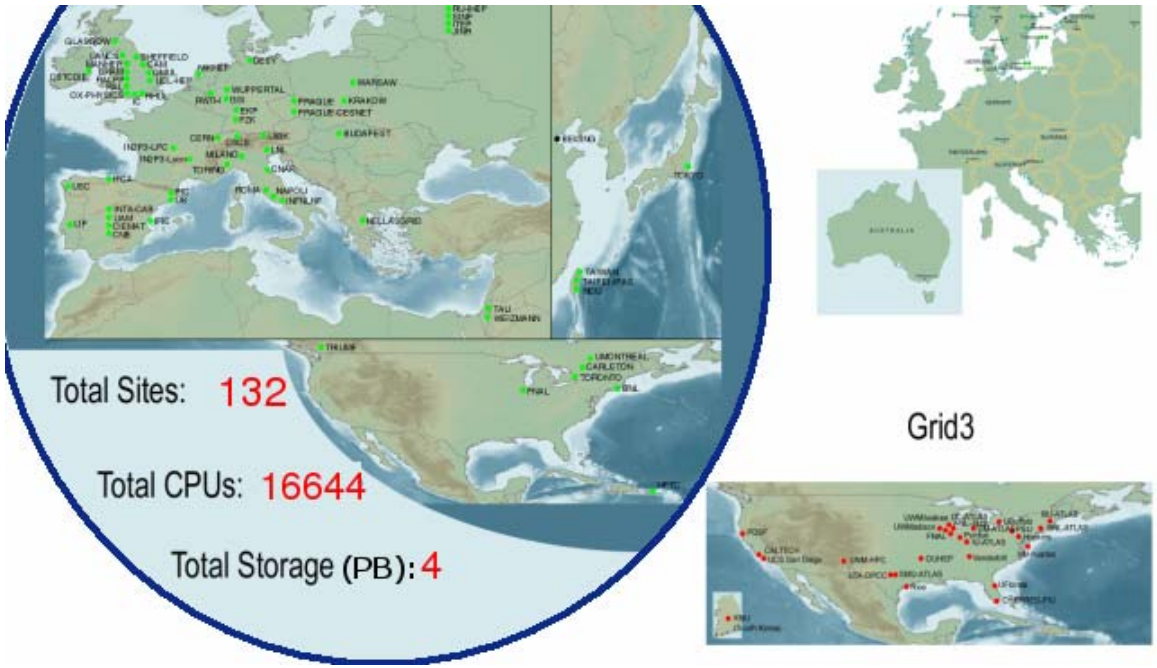
- **Current service based on work done in LCG – culminating in the current service ("LCG-2")**
  - Now at the point where in parallel we need to deploy and understand gLite – whilst maintaining a reliable production service.

# Computing Resources: April 2005

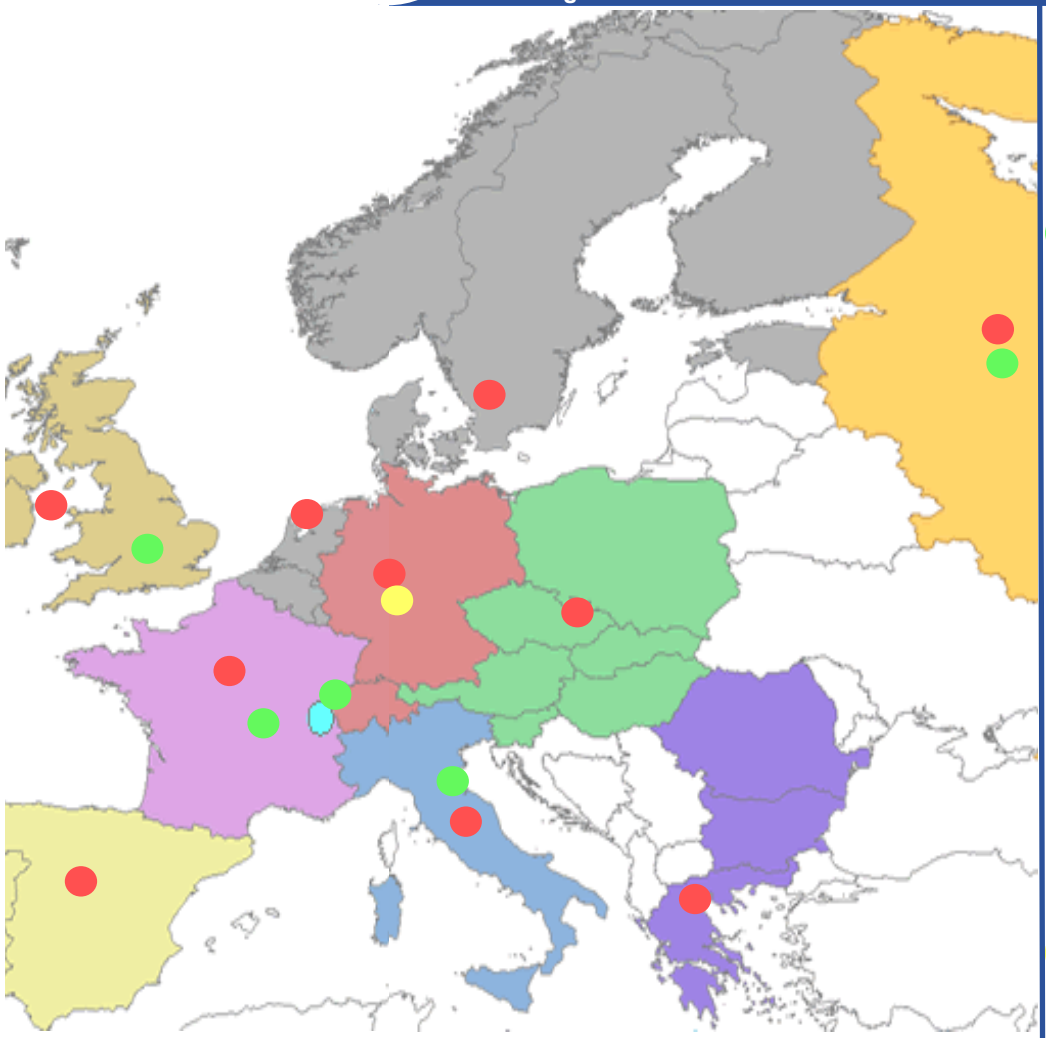


# LCG-2

- **Real-time monitor**
  - http://www.hep.ph.ic.ac.uk/e-science/projects/demo/index.html
- **Current status**
  - http://goc.grid-support.ac.uk/gridsite/monitoring/

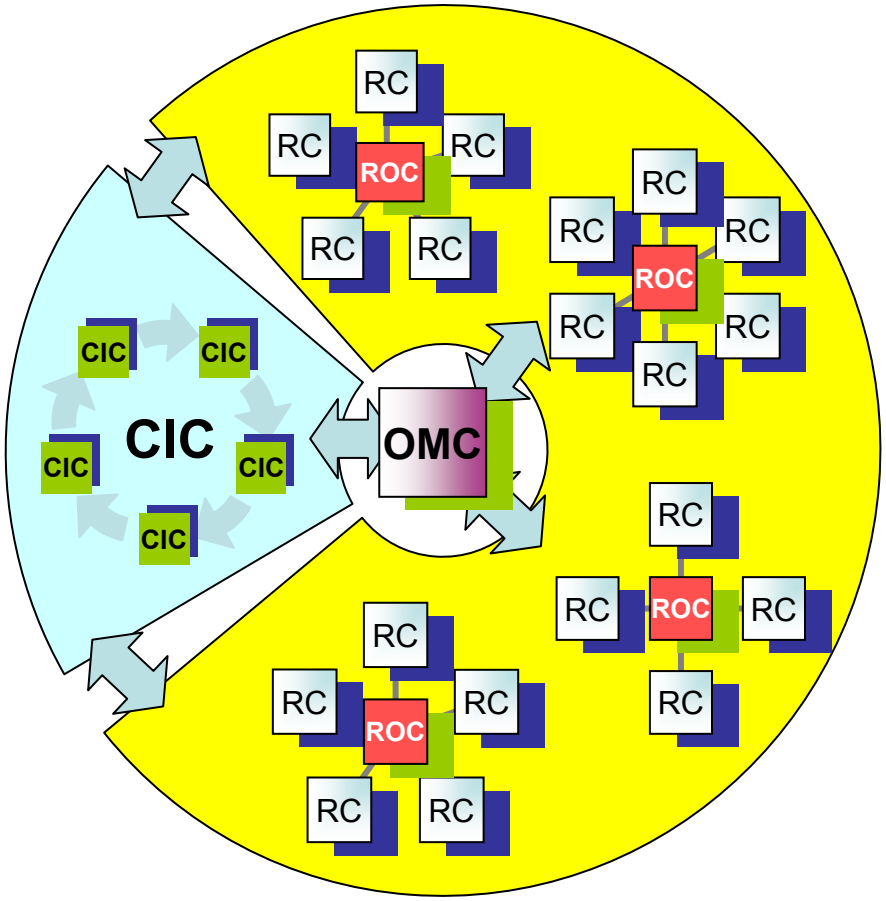


# SA1 – Operations Structure



- **Operations Management Centre (OMC):**
  - At CERN – coordination etc
- **Core Infrastructure Centres (CIC)**
  - Manage daily grid operations – oversight, troubleshooting
  - Run essential infrastructure services
  - Provide 2nd level support to ROCs
  - UK/I, Fr, It, CERN, + Russia (M12)
  - Taipei also run a CIC
- **Regional Operations Centres (ROC)**
  - Act as front-line support for user and operations issues
  - Provide local knowledge and adaptations
  - One in each region – many distributed
- **User Support Centre (GGUS)**
  - In FZK – manage PTS – provide single point of contact (service desk)
  - Not foreseen as such in TA, but need is clear

# Grid Operations




RC = Resource 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**

# Operations: Overall status

- **The EGEE production grid service (LCG-2) is quite stable**
  - The services are quite reliable
- **The biggest problem is stability of sites**
  - Configuration problems due to complexity of the middleware
  - Fabric management at less experienced sites
  - In large tests, selecting stable sites, achieve >>90% efficiency
- **Operations workshop last November to address this**
  - Fabric management working group – write fabric management cookbook
  - Tighten operations control of the grid – escalation procedures, removing bad sites

☞ **Complexity is in the number of sites – not number of cpu**

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - JR activities
  - Operations
  - **(Human) Networking**
- FAQ

# (Human) Networking Activities

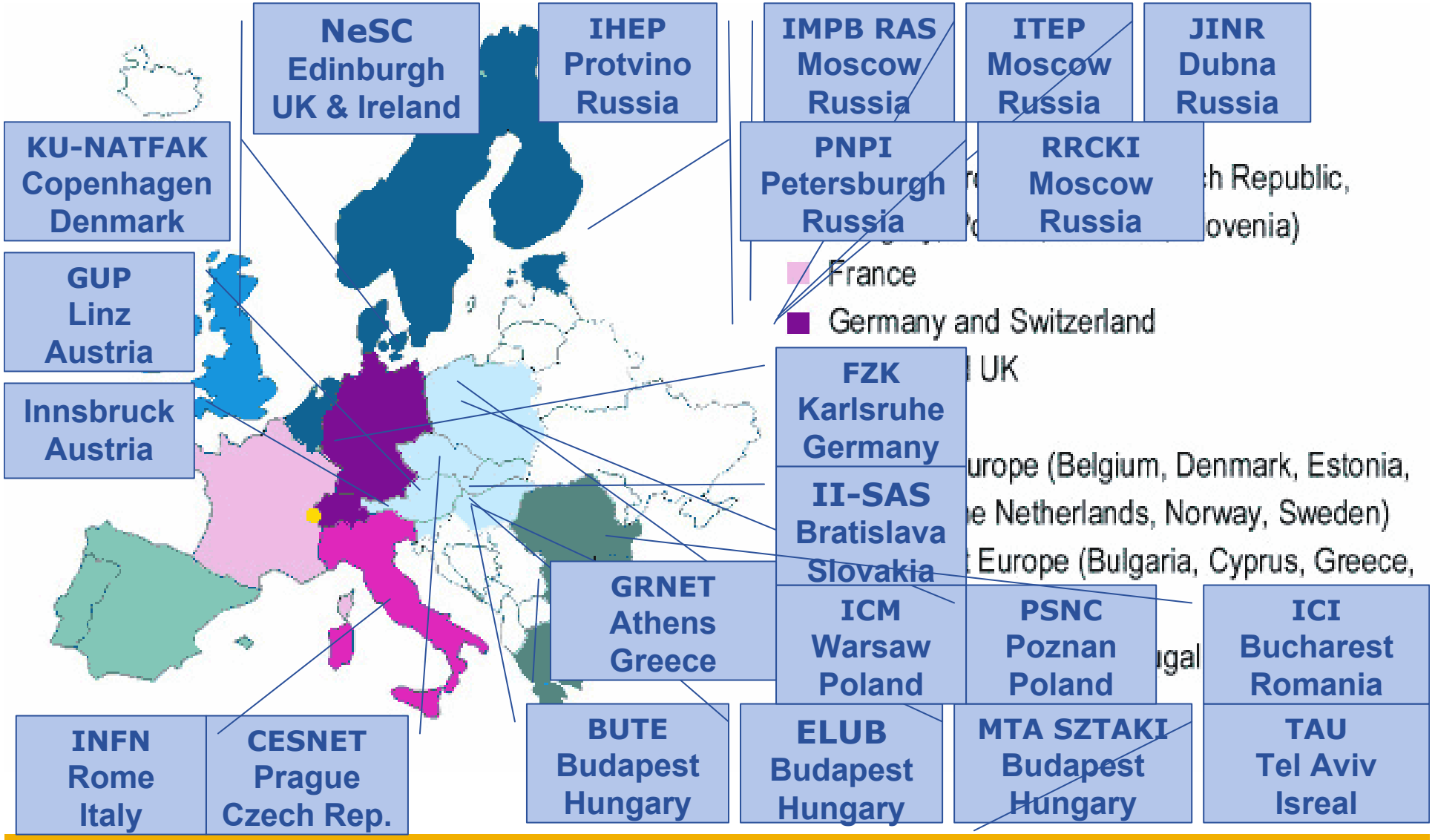
- **Dissemination and Outreach: 5% of EGEE budget**
  - Dissemination – to actively promote and raise awareness of the EGEE project
  - Outreach – to identify and contact potential new user communities
- **Training and Induction: 4% of EGEE budget**
  - Induction – to introduce and orient - users and members
  - Training – to create, collate, make available and deliver material and courses
- **Application Identification and Support**
  - Process for selecting new application areas
  - Supports selected VO's in porting applications
- **International cooperation**

# Dissemination

- 1st project conference, Cork, April
- 2nd conference in The Hague
  - 22-26 November
  - http://public.eu-egee.org/conferences/2nd
  - Over 300 delegates
- 3rd conference… you know where…

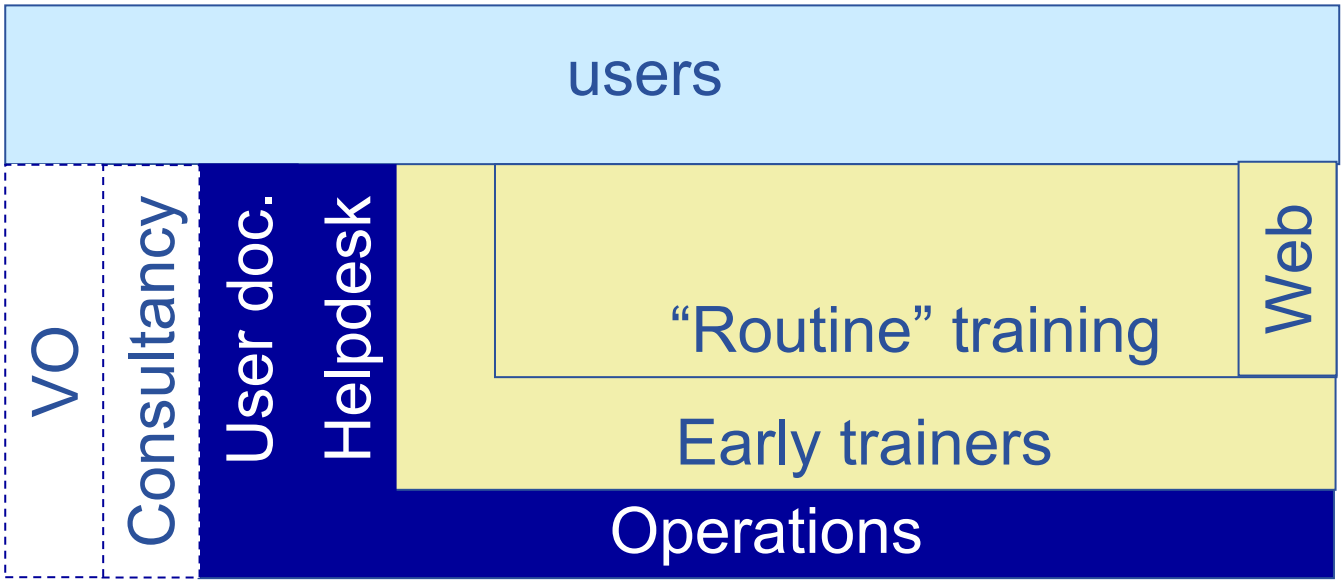
- Websites, Brochures and press releases
  - For project and general public **www.eu-egee.org**
  - Information packs for the general public, press and industry

# Training and Induction



# What are trainers for?

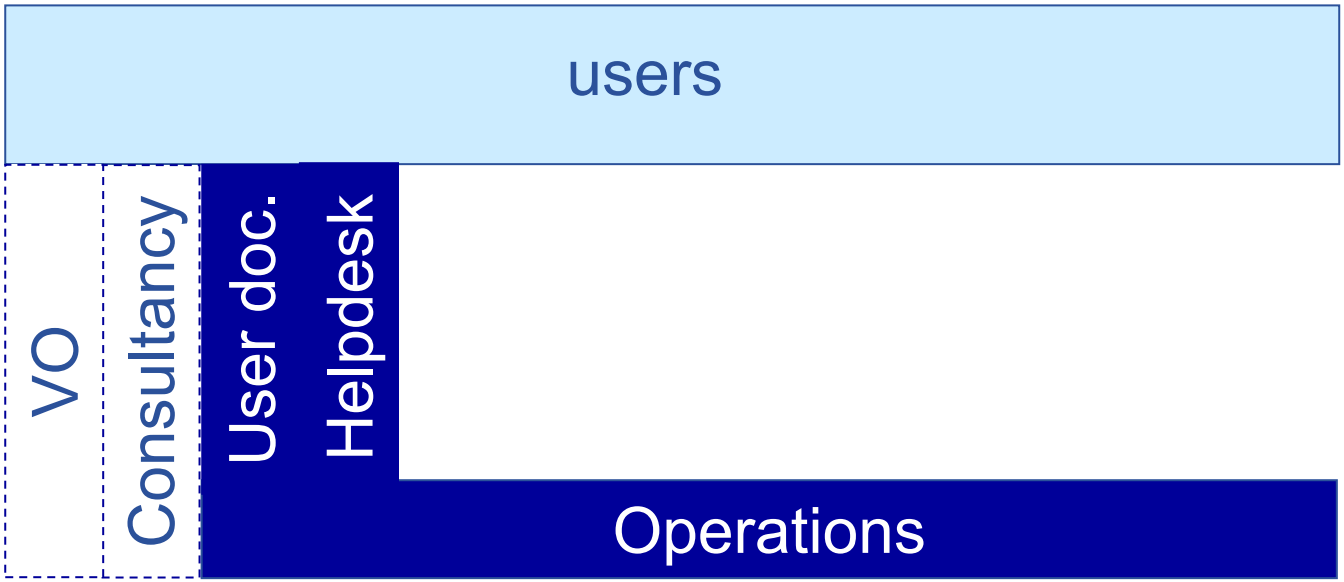
- **EGEE Goal**
  - many effective and active users in many locations
- **Technology is supported by a small number of people**
- **Training helps to make this feasible and scalable**
  - EDG demonstrated the necessity of training



# Without training….

**Poor training ⇒**

- Users uninformed
  - Unproductive and dissatisfied
- Support staff overloaded
  - With mundane questions

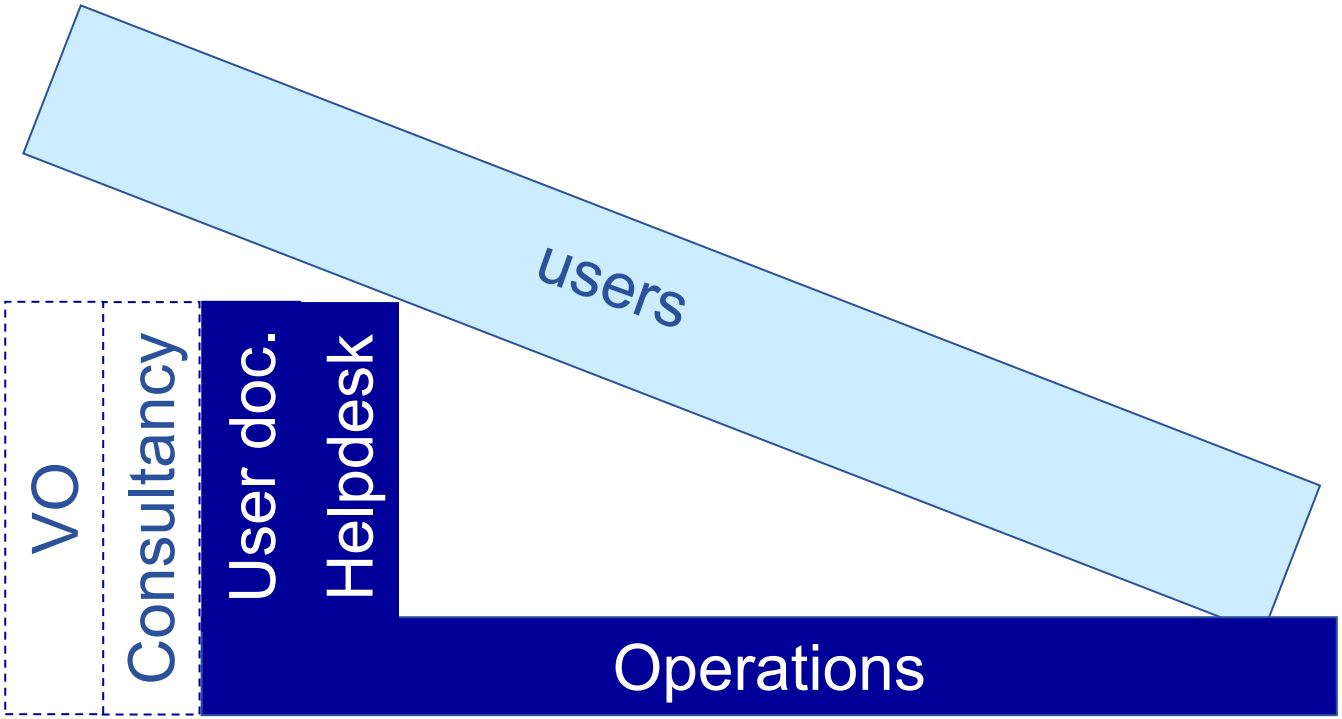


# Without training….

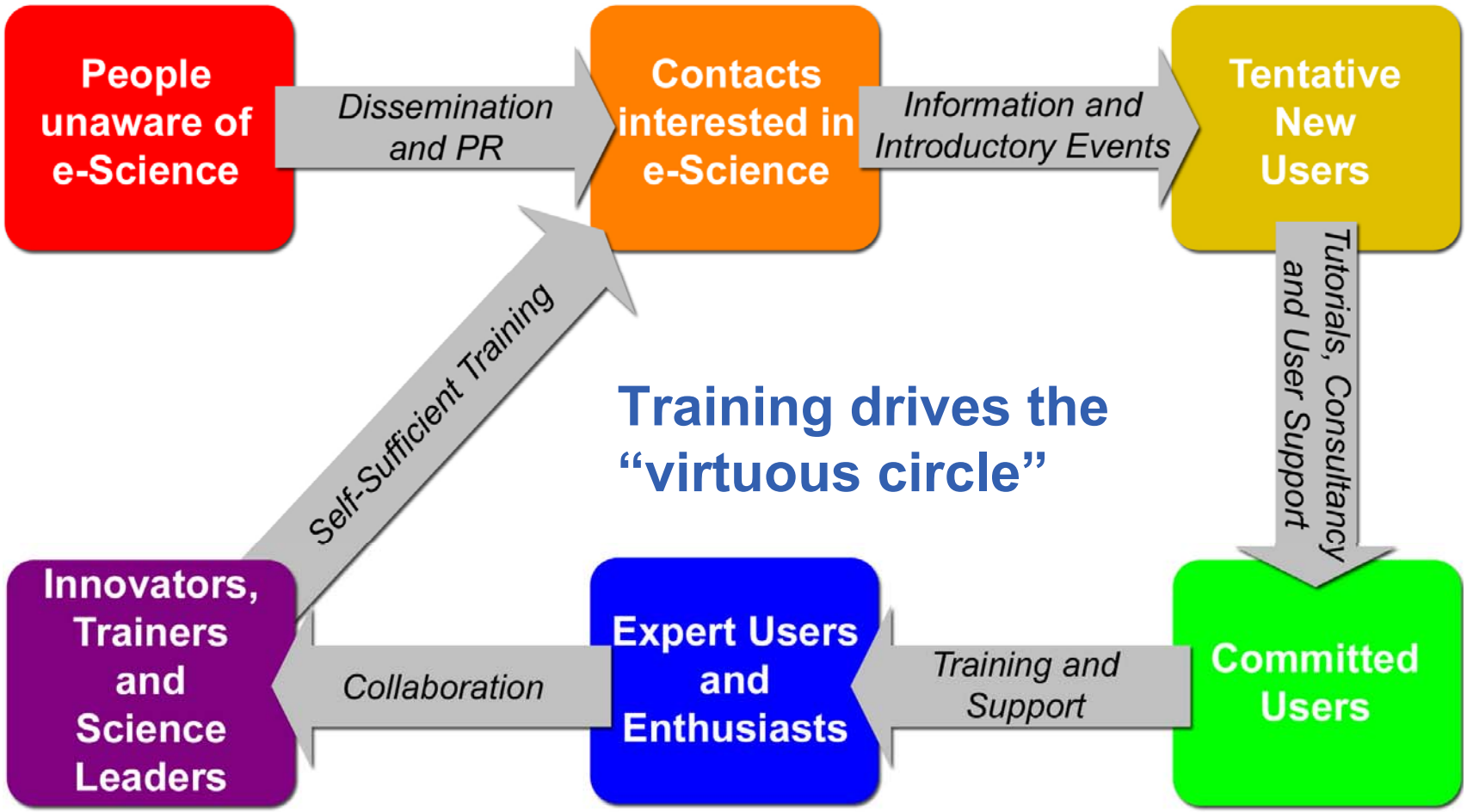
**Poor training ⇒**

- Users uninformed
  - Unproductive and dissatisfied
- Support staff overloaded
  - With mundane questions

If you think training is expensive, try ignorance
Roy Crock, Founder of McDonalds



# Gaining new and effective users



# User training and induction

- Training material and courses from introductory to advanced level
- Train a wide variety of users both internal to the EGEE consortium and external groups from across Europe
- Experience with GENIUS portal and GILDA testbed
- Courses inline with the needs of the projects and applications
- See http://egee.nesc.ac.uk/

# Planned courses and repository

- The EGEE training material is being made available to the whole community.

- EGEE repository allows flexible searches of the available material... Take a look!!!

- Courses and material via…

http://egee.nesc.ac.uk

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - Middleware Re-engineering
  - Operations
  - (Human) Networking
    - *Supporting application communities*
- FAQ

# Application communities and EGEE

- **LCG and Bio-informatics from day 1**
- **New application communities are selected by the EGEE Generic Applications Advisory Panel**
  - See: EGEE web site (NA4 activity) and also http://agenda.cern.ch/age?a042351
- **Selected are:**
  - Computational chemistry
  - Earth sciences
  - Earth observation
  - Astrophysics
- **Also working with DILIGENT:**
  - Virtual digital data libraries
- **GILDA grid for new applications and testing before migrate to production grid**

# EGEE pilot application: BioMedical

- **BioMedical**
  - Bioinformatics (gene/proteome databases distributions)
  - Interactive application (human supervision or simulation)
  - Security/privacy constraints
    - Heterogeneous data formats - Frequent data updates - Complex data sets - Long term archiving
- **BioMed applications deployed**
  - **GATE -** Geant4 Application for Tomographic Emission
  - **GPS@ -** genomic web portal
  - **CDSS -** Clinical Decision Support System

**http://egee-na4.ct.infn.it/biomed/applications.html**

# Geophysics Applications

## Seismic processing Generic Platform:

- Based on Geocluster, an industrial application – to be a starter of the core member VO.

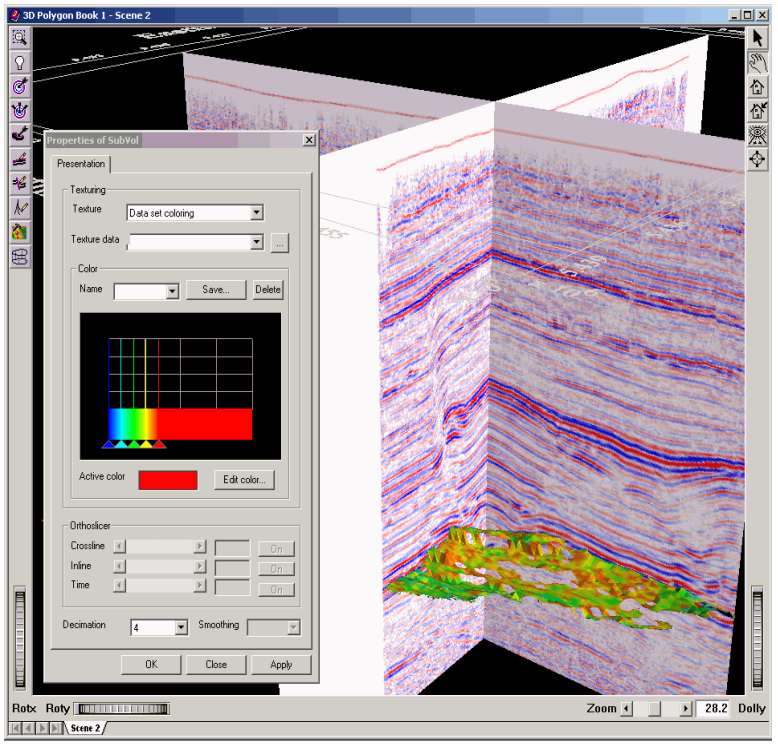
- Include several standard tools for signal processing, simulation and inversion.

- Opened: any user can write new algorithms in new modules (shared or not)

- Free for academic research

-Controlled by license keys (opportunity to explore license issue at a grid level)

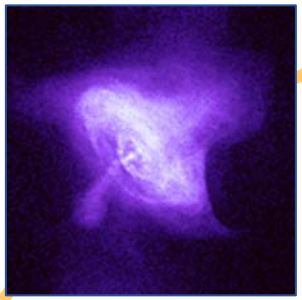
- initial partners F, CH, UK, Russia, Norway



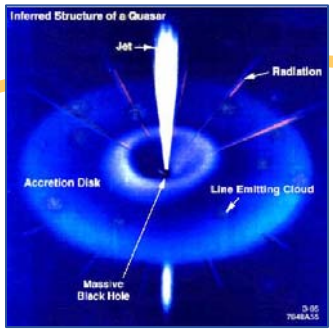
# The MAGIC telescope

- Largest Imaging Air Cherenkov Telescope
  **(17 m mirror dish)**
- **Located on Canary Island**
  La Palma **(@ 2200 m asl)**
- **Lowest** energy threshold **ever obtained with a Cherenkov telescope**
- **Aim: detect** $\gamma$**-ray sources in the unexplored energy range:**
  30 *(10)*-> 300 GeV

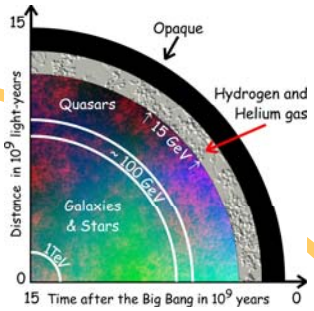
# The MAGIC Physics Program



- Pulsars



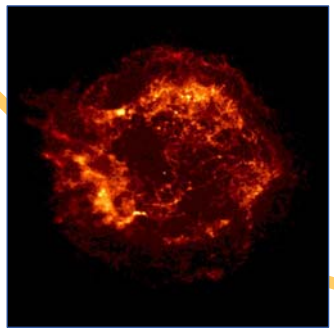

- AGNs



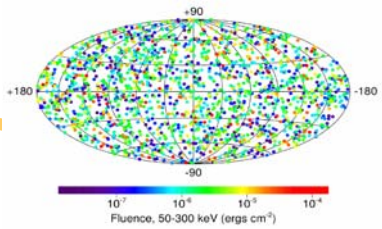

- Cosmological γ-Ray Horizon

- Origin of Cosmic Rays

- Tests of Quantum Gravity effects



- SNRs




- GRBs


- Cold Dark Matter

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - Middleware Re-engineering
  - Operations
  - (Human) Networking
    - ***Policy and international relations***
- FAQ

# Policy and international relations

- **Further releases of the eIRG white paper on grid policy**
  - Closer relations with the European Strategy Forum on Research Infrastructures (ESFRI)
- **Participation in EU concertation meetings**
  - Contribute to organisation
  - Leadership of key working groups (e.g. security via JRA3)
- **EU synergy roadmap revision**
  - With SEE-GRID and DEISA
- **Continue work with OASIS, GGF etc**

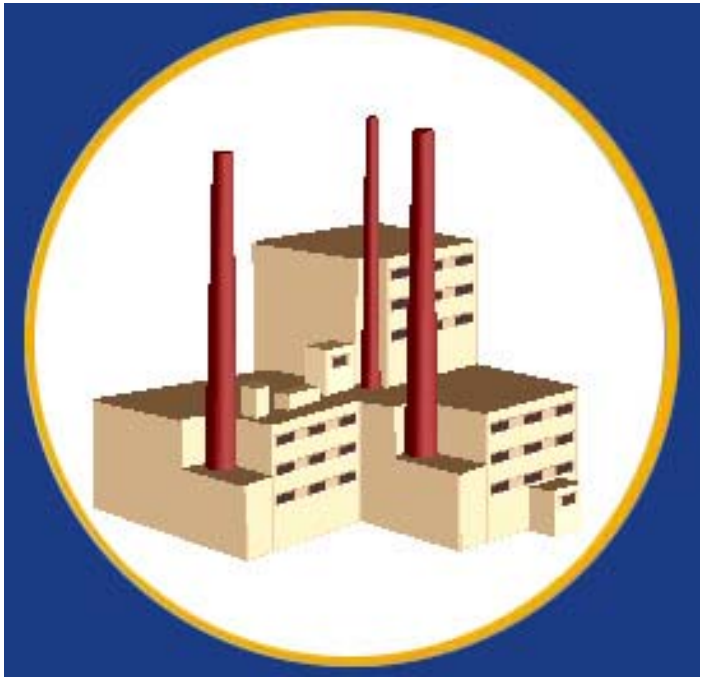
# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - Middleware Re-engineering
  - Operations
  - (Human) Networking
    - *Supporting application communities*
- **FAQ**

# Links to industry?

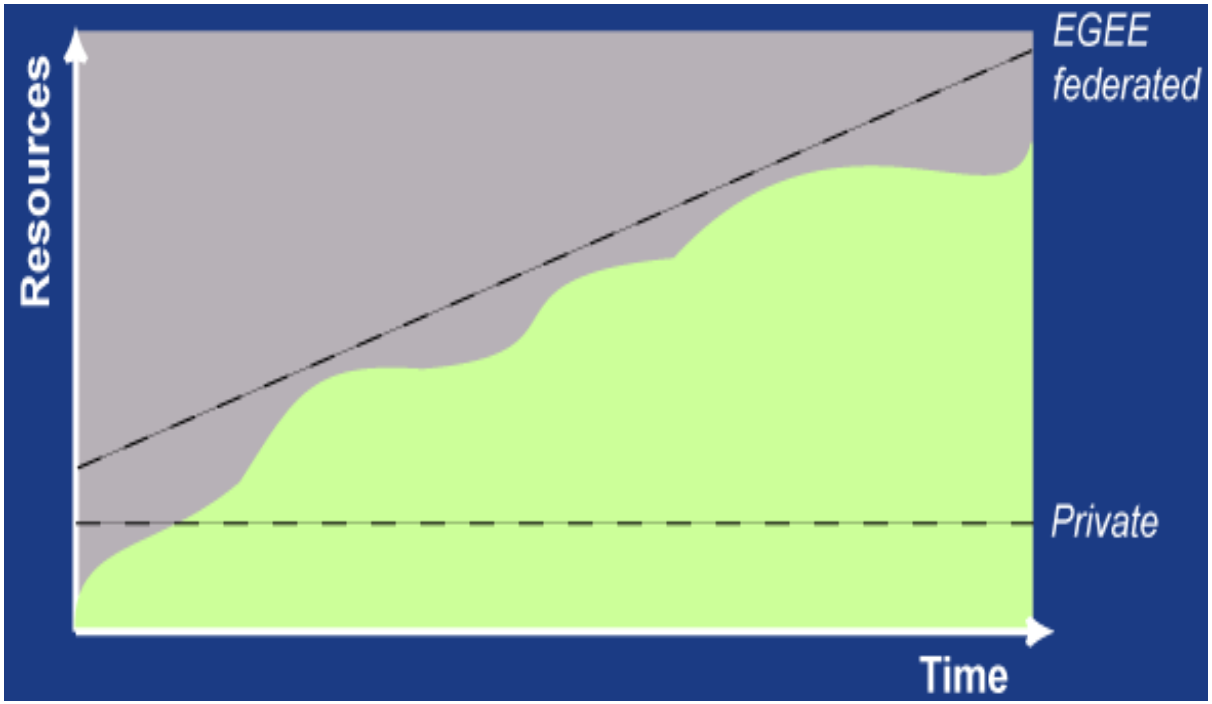
- **EGEE Industry Forum**
  - raise awareness of the project in industry to encourage industrial participation in the project
  - foster direct contact of the project partners with industry
  - ensure that the project can benefit from practical experience of industrial applications



- **For more info:**
  *http://public.eu-egee.org/industry/*

# Private or Federated Resources?

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




**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

# Open Source Software License

- **The existing EGEE grid middleware (LCG-2) is distributed under an Open Source License developed by EU DataGrid project**
  - Derived from modified BSD - no restriction on usage (academic or commercial) beyond acknowledgement
  - Approved by Open Source Initiative (OSI)


- **Same approach for new middleware (gLite)**
  - New license agreed by partners is derived from the EDG license and takes into account feedback from the World Intellectual Property Office (WIPO)

# Contents

## EGEE: Enabling Grids for E-sciencE

- Goals
- Organisation
- Activities and status
  - Middleware Re-engineering
  - Operations
  - (Human) Networking
    - *Supporting application communities*
- FAQ
- **Summary**

# Deliverables/Milestones Timeline

Enabling Grids for E-sciencE

| Month | Ref | Description |
|---|---|---|
| 10 | MJRA1.5 | Integrated Release Candidate 1 enters testing and validation period (Release 1) |
| 12 | DJRA1.3 | Software and associated documentation (Release 1) |
| 12 | DJRA2.2 | Annual Report on EGEE Quality Status, including software and Grid operations and plan for second year |
| 12 | MJRA3.6 | Security operational procedures (first revision) |
| 12 | MJRA3.7 | Framework for policy evaluation accepted in GridPMA policies and CA service authorities for EGEE |
| 12 | MJRA4.3 | Prototype tool to access network performance metrics from a limited set of measurement points |
| 12 | DNA1.1.4 | Quarterly periodic report |
| 12 | DNA2.6.2 | Dissemination Progress Reports revision |
| 12 | MNA3.3 | First external review of User Training and Induction with feedback |
| 12 | MNA4.3 | First external review of Applications Identification and Support with feedback |
| 12 | DNA5.3 | Progress report on International Cooperation Activities |
| 12 | DSA1.4 | Assessment of initial infrastructure operation and plan for next 12 months |
| 12 | MSA2.3 | Operational interface between EGEE and GEANT/NRENs. |
| 12 | DSA2.2 | Institution of SLAs and appropriate policies |
| 14 | DJRA1.4 | Architecture and Planning (Release 2) |
| 14 | DSA1.5 | First release of EGEE Infrastructure Planning Guide (“cook-book”) |
| 14 | DSA1.6 | Release notes corresponding to the full production Grid infrastructure operational |
| 14 | MSA1.3 | Full production Grid infrastructure (20 Resource Centres) operational. |
| 15 | DJRA1.5 | Design of grid services (Release 2) |
| 15 | DNA1.1.5 | Quarterly periodic report |
| 15 | MJRA4.4 | Prototype bandwidth reservation within static network configuration |
| 15 | MJRA4.5 | Specification of end-to-end bandwidth reservation system. |
| 15 | DNA2.4.3 | Dissemination Plan revisions including a formal planning for using and disseminating knowledge |
| 15 | DNA3.1.3 | Training Plan revision |
| 15 | DNA3.3.2 | Training Progress Report update |
| 15 | DNA4.3.2 | EGEE Application Migration Progress report |
| 15 | DNA5.1.3 | eInfrastructure Forum White Papers in conjunction with the EGEE Project Conferences |
| 16 | DJRA3.3 | Global security architecture (first revision) |
| 18 | MJRA1.6 | Test plan for core Grid components and overall Integration (Release 2) |
| 18 | MJRA3.8 | Security operational procedures (second revision) |
| 18 | MJRA3.9 | Set-up of accounting techniques and distributed budgets |
| 18 | MJRA4.6 | Specification of high-level monitoring and diagnostic tools. Revision of metrics. |
| 18 | MJRA4.7 | Dynamic re-configuration of key ingress points in response to reservations. |
| 18 | DJRA4.3 | Report on implications of Ipv6 usage for the EGEE Grid |
| 18 | DNA1.1.6 | Quarterly periodic report |
| 18 | DNA1.3.2 | Periodic report |
| 18 | DNA2.6.3 | Dissemination Progress Reports revision |
| 18 | DNA5.4 | European Grid project synergy report (in collaboration with DEISA, SEE-GRID) |
| 18 | MNA1.2 | Successful completion of second review |

**Feb 2005**

**35 deliverables/milestones**

**(29 more in final 6 months)**

# Project Metrics

- The project has already met many of the targets set for the first 2 year phase
- Current metrics are defined by the simplest means of measurement – these will evolve as we understand better the true effect of EGEE

| Metric | Current status | Target end Year 2 |
|---|---|---|
| Number of Users | ~ 500 | ≥ 3000 |
| Number of sites | >110 | 50 |
| App. breadth (disciplines) | 6 | ≥ 5 |
| Multinational (countries) | 30 | ≥ 15 |
| Reviewed applications | 23% | ≥ 15% |

# Summary

- **EGEE runs the largest and probably the only multi-disciplinary production grid infrastructure**
- **A process is in place for** migrating new applications **to the EGEE infrastructure**
- **A** training programme **is established**
- **gLite "*next generation*" middleware released**
- **Plans for a** follow-on phase **are being developed**

# Further Information

**EGEE** www.eu-egee.org
**gLite** www.glite.org

**LCG** lcg.web.cern.ch/LCG/

**The Grid Cafe** www.gridcafe.org

- **More EU sites**:
  - http://www.cordis.lu/ist/grids/fp6_grid_projects.htm
  - http://www.gridstart.org/concertation_mtg.shtml
  - "e-Infrastructures Reflection Group http://www.e-irg.org
- **NeSC** www.nesc.ac.uk