Introduction to Grid and EGEE

Ladislav Hluchý
Institute of Informatics
Slovak Academy of Sciences
Slovakia

www.eu-egee.org

Information Society

Introduction to Grid and EGEE

Miroslav Dobrucký
Institute of Informatics
Slovak Academy of Sciences
Slovakia

www.eu-egee.org

Information Society

Acknowledgements

- **This talk is based on a module of the tutorials delivered by the EDG training team and slides from**
  - Andrew Grimshaw, University of Virginia
  - Bob Jones, EGEE Technical Director
  - Mark Parsons, EPCC
  - the EDG training team
  - Roberto Barbera, INFN
  - Ian Foster, Argonne National Laboratories
  - Jeffrey Grethe, SDSC
  - The National e-Science Centre
  - Dave Berry, NeSC

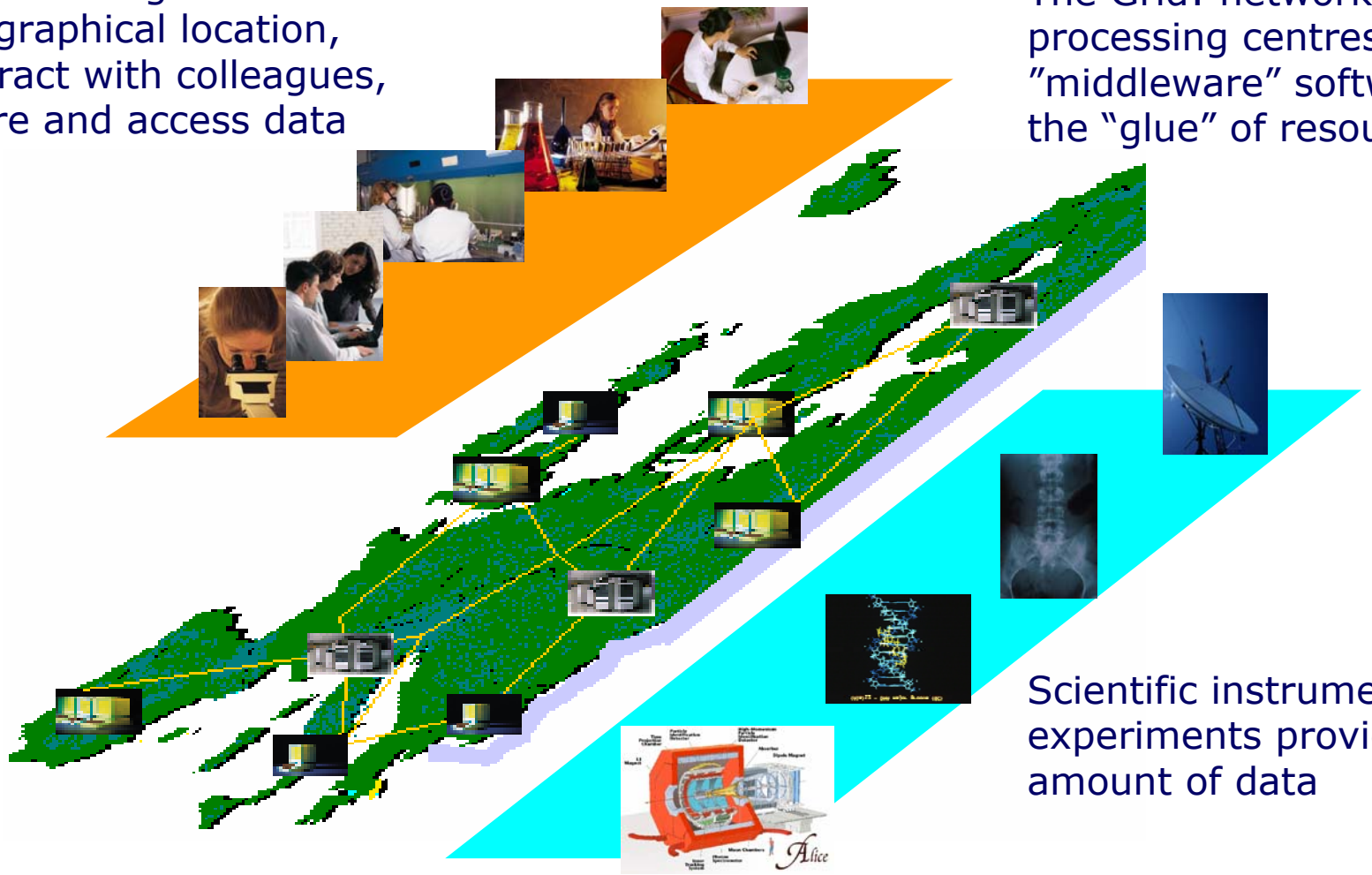
Overview

- What is Grid computing?
- Characteristics of a grid
- Applications (what's in it for the working scientist)
- European grids, and the world
- EGEE project
- K-WfGrid project

The (Science) Grid Vision

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

The Grid: networked data processing centres and ”middleware” software as the “glue” of resources.



Scientific instruments and experiments provide huge amount of data

What is Grid Computing?

- **A *Virtual Organisation* is:**
  - People from different institutions working to solve a common goal
  - Sharing distributed processing and data resources

- **Grid infrastructure enables virtual organisations**

  "Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations"
  (I.Foster)

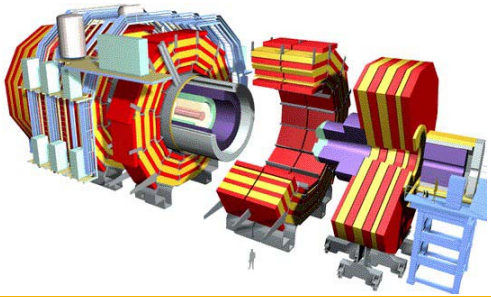
Grids vs. Distributed Computing

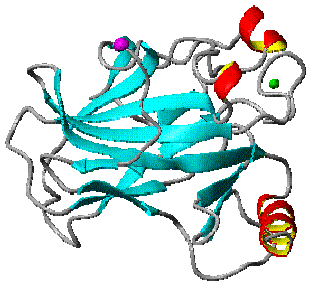
- **Existing distributed applications:**
  - tend to be *specialised systems*
  - intended for a single purpose or user group

- **Grids go further and take into account:**
  - Different kinds of *resources*
    - Not always the same hardware, data and applications
  - Different kinds of *interactions*
    - User groups or applications want to interact with Grids in different ways
  - *Dynamic* nature
    - Resources and users added/removed/changed frequently

The main drivers behind Grid

- **The relentless increase in microprocessor performance**
  - you can buy multi-gigaflop systems for less than €800
- **The availability of reliable high performance networking**
  - in Europe the GEANT network links 32 countries at speeds of up to 10Gbps (and beyond)
  - in the UK we have gone from 100Mbps -> 10Gbps academic backbone since 2000
  - 1Gbps is commonly available to the desktop
- **The desire to push the boundaries of scientific discovery by computational analysis and simulation – e-Science**

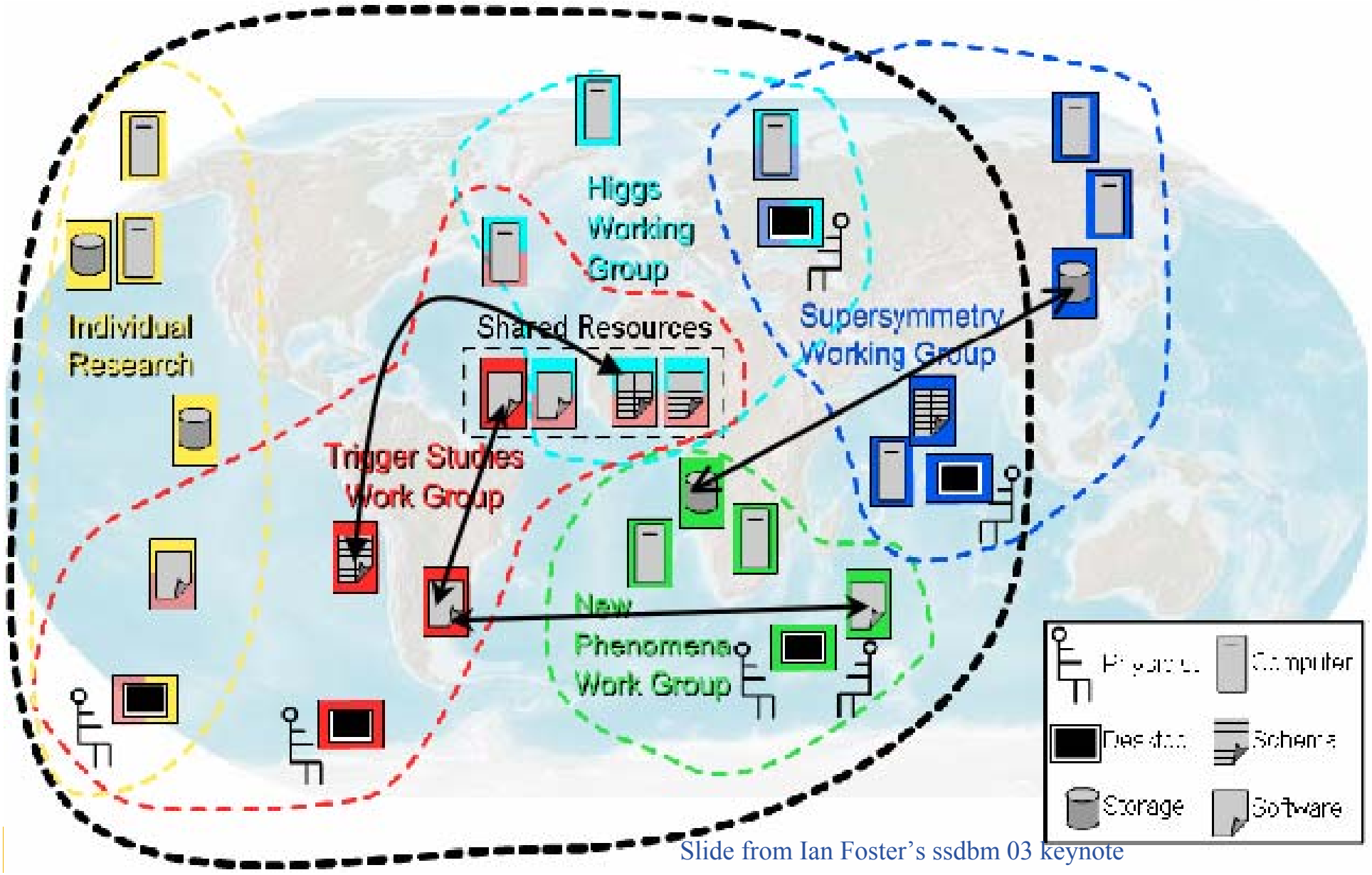




Challenges

- Must share data between thousands of scientists with multiple interests
- Must ensure that all data is accessible anywhere, anytime
- Must be scalable and remain reliable for more than a decade
- Must cope with different access policies
- Must ensure data security

The Emergence of Global Knowledge Communities



Slide from Ian Foster's ssdbm 03 keynote

- What is Grid computing?
- **Characteristics of a grid**
- Applications (what's in it for the working scientist)
- European grids, and the world
- EGEE project
- K-WfGrid project

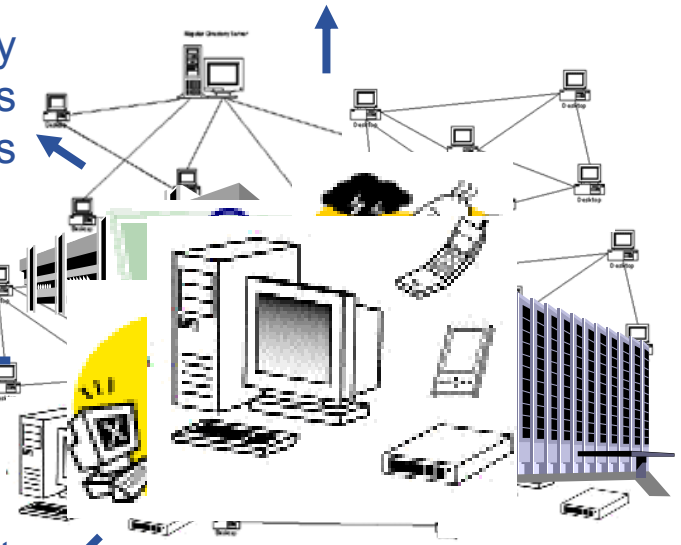
What are the characteristics of a Grid system?

Numerous Resources

Ownership by Mutually Distrustful Organizations & Individuals

Different Security Requirements & Policies Required

Potentially Faulty Resources



Resources are Heterogeneous

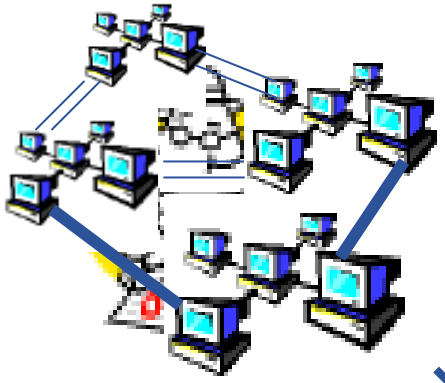
What are the characteristics of a Grid system?

Standards

Numerous Resources

Ownership by Mutually Distrustful Organizations & Individuals

Connected by Heterogeneous, Multi-Level Networks



Different Security Requirements & Policies Required

Different Resource Management Policies

Potentially Faulty Resources

Geographically Separated

Resources are Heterogeneous

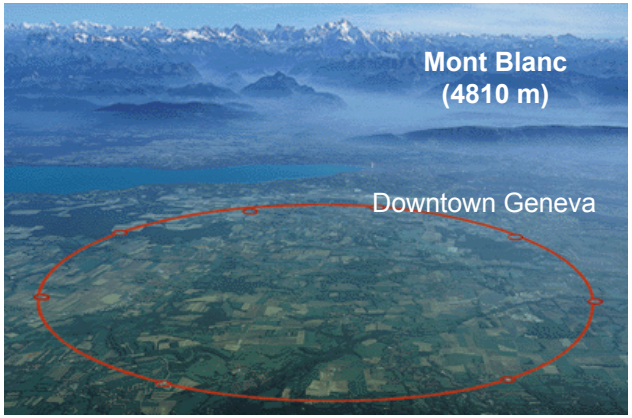
- What is Grid computing?
- Characteristics of a grid
- **Applications (what's in it for the working scientist)**
- European grids, and the world
- EGEE project
- K-WfGrid project

Grid 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)***

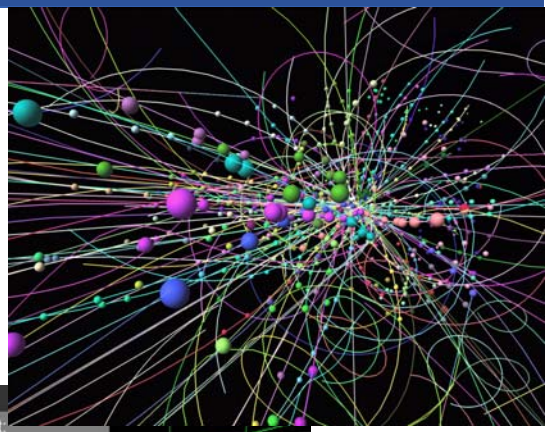
CERN: Data intensive science in a large international facility

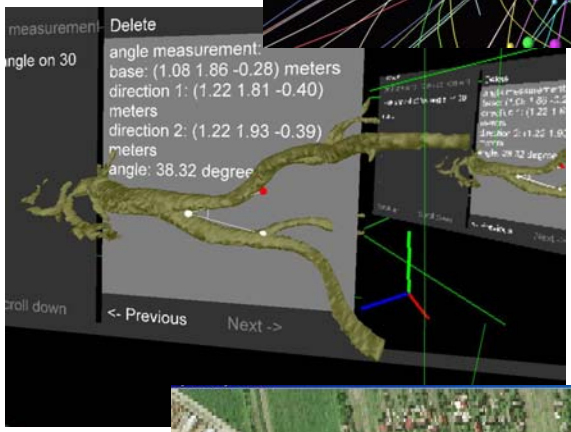
- **The Large Hadron Collider (LHC)**
  - The most powerful instrument ever built to investigate elementary particles physics
- **Data Challenge:**
  - 10 Petabytes/year of data !!!
  - 20 million CDs each year!
- **Simulation, reconstruction, analysis:**
  - LHC data handling requires computing power equivalent to ~100,000 of today's fastest PC processors!

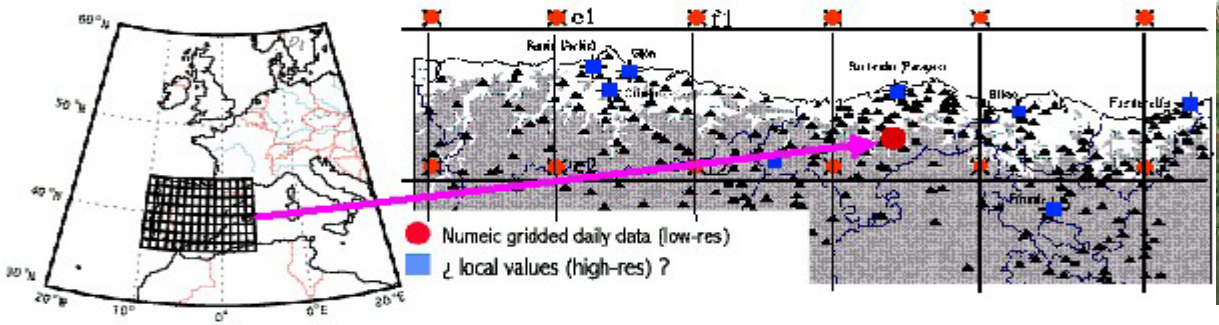


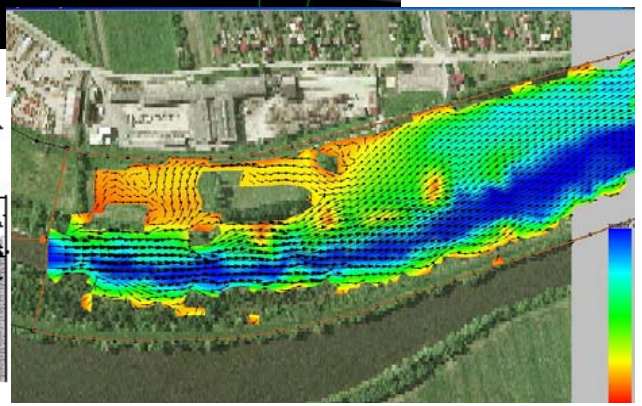
CrossGrid

- 1. Interactive biomedical simulation and visualization
- 2. Flooding crisis team support
- 3. HEP distributed data analysis
- 4. Weather forecasting and air pollution modelling







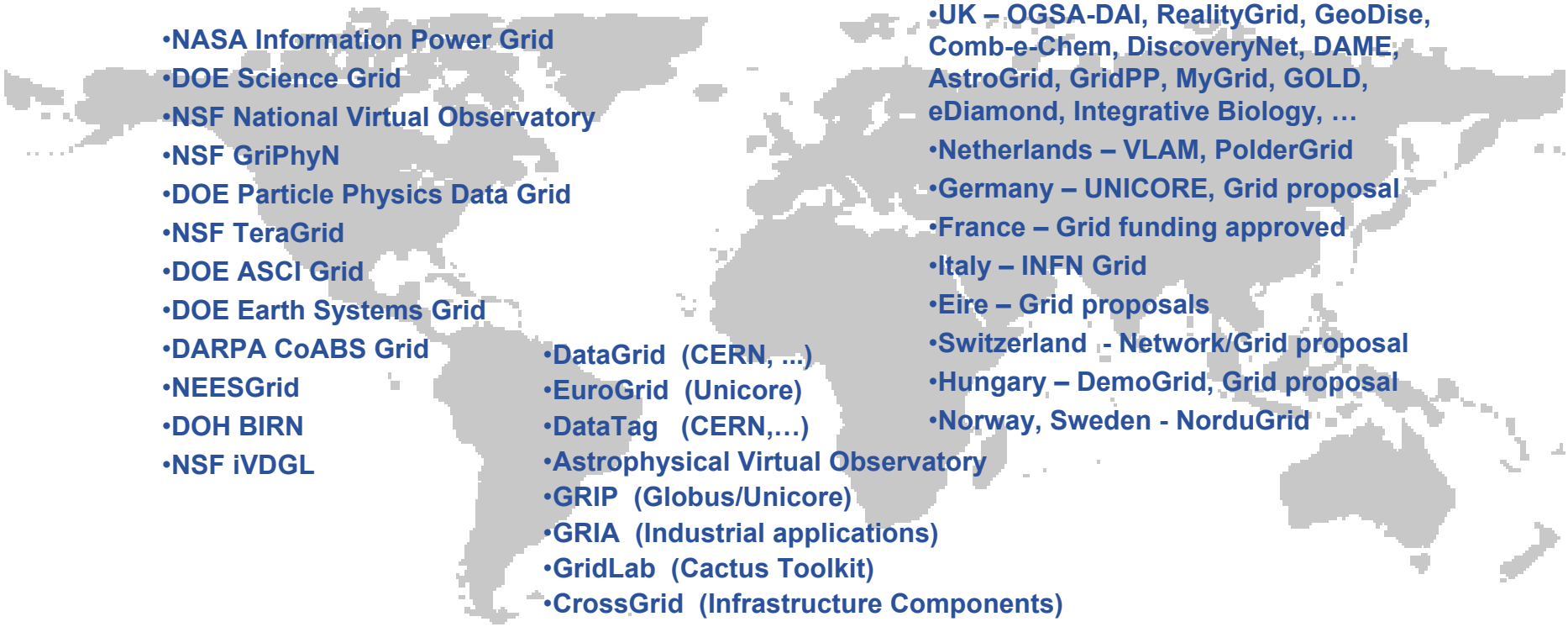


Connecting *People*: Access Grid

- What is Grid computing?
- Characteristics of a grid
- Applications (what's in it for the working scientist)
- **European grids, and the world**
- EGEE project
- K-WfGrid project

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 – OGSA-DAI, RealityGrid, GeoDise, Comb-e-Chem, DiscoveryNet, DAME, AstroGrid, GridPP, MyGrid, GOLD, eDiamond, Integrative Biology, …
- 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

Major EU GRID projects

European DataGrid (EDG)	www.edg.org
LHC Computing GRID (LCG)	cern.ch/lcg
CrossGRID	www.crossgrid.org
DataTAG	www.datatag.org
GridLab	www.gridlab.org
EUROGRID	www.eurogrid.org

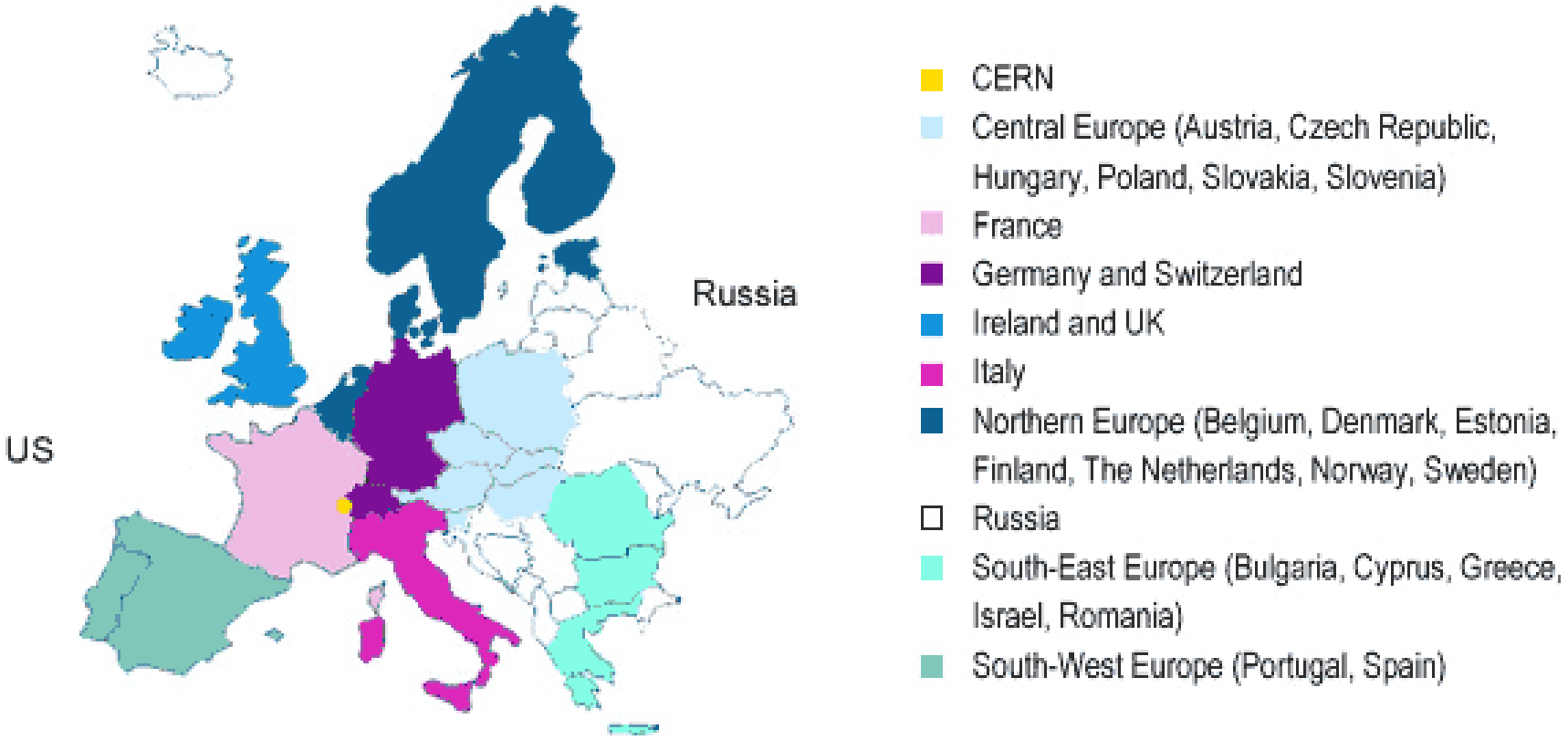
European National Projects:

- INFNGRID,
- UK e-Science Programme,
- NorduGrid

- What is Grid computing?
- Characteristics of a grid
- Applications (what's in it for the working scientist)
- European grids, and the world
- **EGEE project**
- K-WfGrid project

The EGEE Project

- **Leverage national resources for broader European benefit**
- **70 institutions in 27 countries, federated in regional Grids**



Why EGEE?

- **A lot of investment from previous projects both at national and international level**
- **For once Europe is not lagging behind (yet) more advanced IT regions (US and Japan)**
  - NYT article on 11/11/03 gives EU a 12-18 lead to Europe on Grid deployment
- **Important to keep momentum and preserve the human asset and resource investment so far O(100 MEuros) in FP5**
- **100 M Euros already invested in first FP6 phase, another 160 M foreseen in second phase**
- **More investment possible in FP7 (if success in FP6 continues)**
- **Project Director and senior partners already working on this**

The EGEE challenges

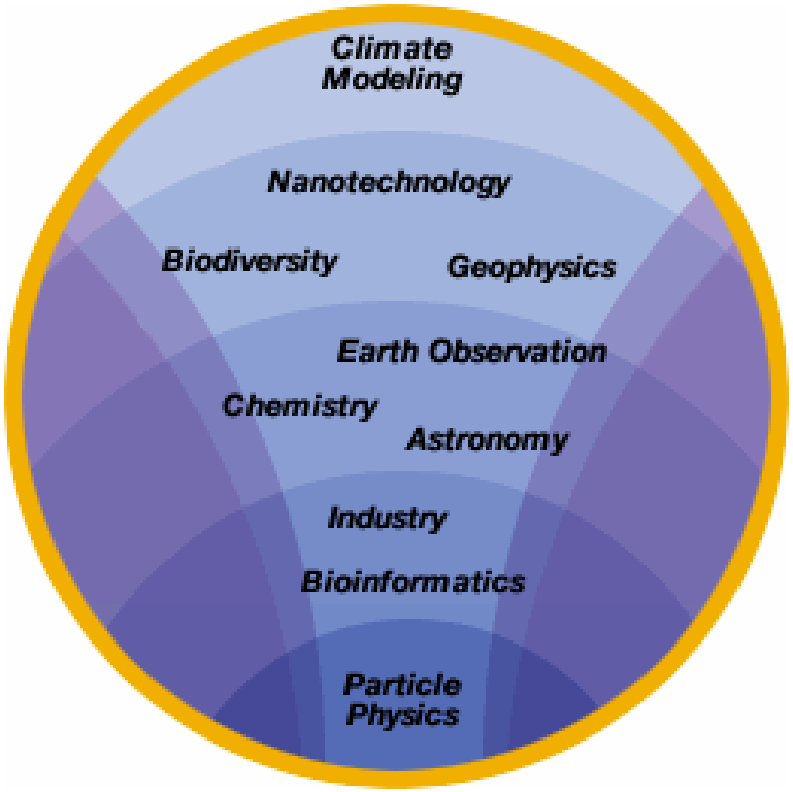
- **A large investment in a short time (32 M Euros/ 24 months):**
  - The rationale is to mobilize the wider Grid community in Europe and elsewhere and be all inclusive
  - Demonstrate production quality sustained Grid services for a few relevant scientific communities (at least HEP and Bio-Medical)
  - 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
- **Move from R&D Middleware and testbeds to industrial quality software and sustained production Grid infrastructure performance**
- **Implement a highly distributed software engineering process while maintaining efficiency and a fast release cycle (development clusters)**
- **Harmonize EGEE activities with national and international activities**
- **Cope with new FP6 rules and different and often conflicting EU Grid plans and activities**

EGEE Partners

- 70 leading institutions in 28 countries, federated in regional Grids
- Leverage national resources in a more effective way for broader European benefit

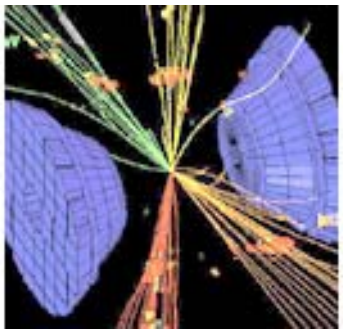
From the EGEE proposal: Applications

- **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**
- **Two pilot applications selected to guide the implementation and certify the performance and functionality of the evolving infrastructure:** Physics & Bioinformatics

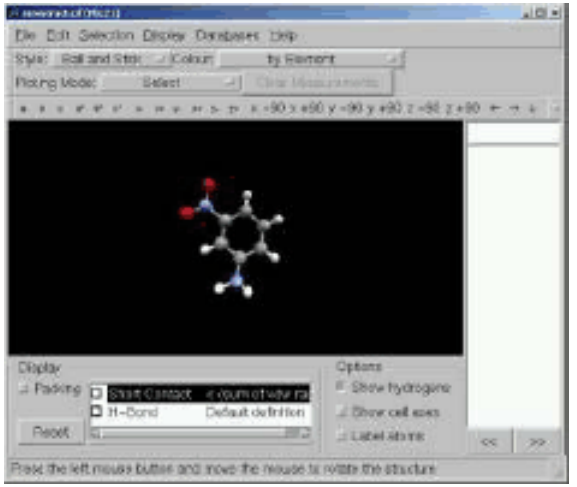


The pilot applications

- **High Energy Physics** with LHC Computing Grid (www.cern.ch/lcg) relies on a Grid infrastructure to store and analyse Petabytes (10^{15} bytes) of real and simulated data. LCG is a major source of resources, requirements and hard deadlines with no conventional solution available



- In **Biomedics** several communities are facing equally daunting challenges to cope with the flood of bioinformatics and healthcare data. Need to access large and distributed non-homogeneous data and important on-demand computing requirements



EGEE Related projects

- **From the EGEE mandate, be open and play an infrastructure role:**
  - SEE-GRID, South Eastern European Grid-enabled eInfrastructure development: extends EGEE to South East Europe http://www.see-grid.org/
  - DEISA , Distributed European Infrastructure for Supercomputing Applications: Supercomputing grid http://www.deisa.org/
  - Diligent: A Testbed Digital Library Infrastructure on Grid Enabled Technology: (in advanced negotiation) starts in September or October 2004
  - GRID-CC (in advanced negotiation): Real-time Grid applications
  - US projects (Trillium, GRID3, OSG etc.)
  - BioMedical and other EU projects from the current round of EU negotiation (will be known by June)
  - Other countries have expressed strong interest in the project: Korea, Taiwan, Egypt, Pakistan, India, Cuba, Chile, Iran…

EGEE Project Structure

32 Million Euros EU funding over 2 years starting 1st April 2004

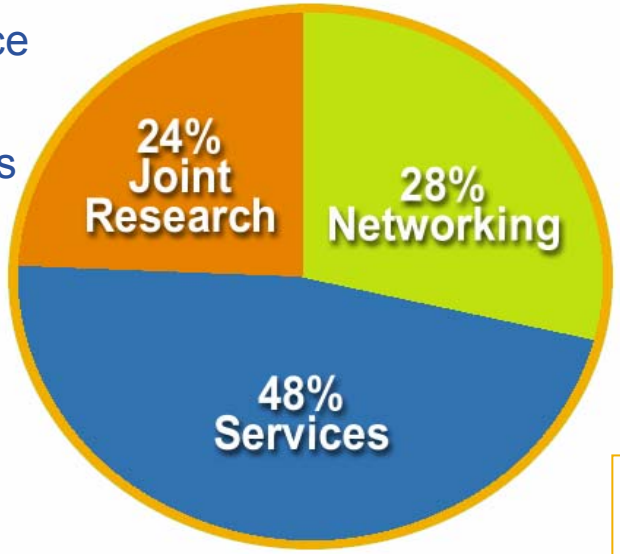
24% Joint Research

JRA1: Middleware Engineering and Integration

JRA2: Quality Assurance

JRA3: Security

JRA4: Network Services Development



28% Networking

NA1: Management

NA2: Dissemination and Outreach

NA3: User Training and Education

NA4: Application Identification and Support

NA5: Policy and International Cooperation

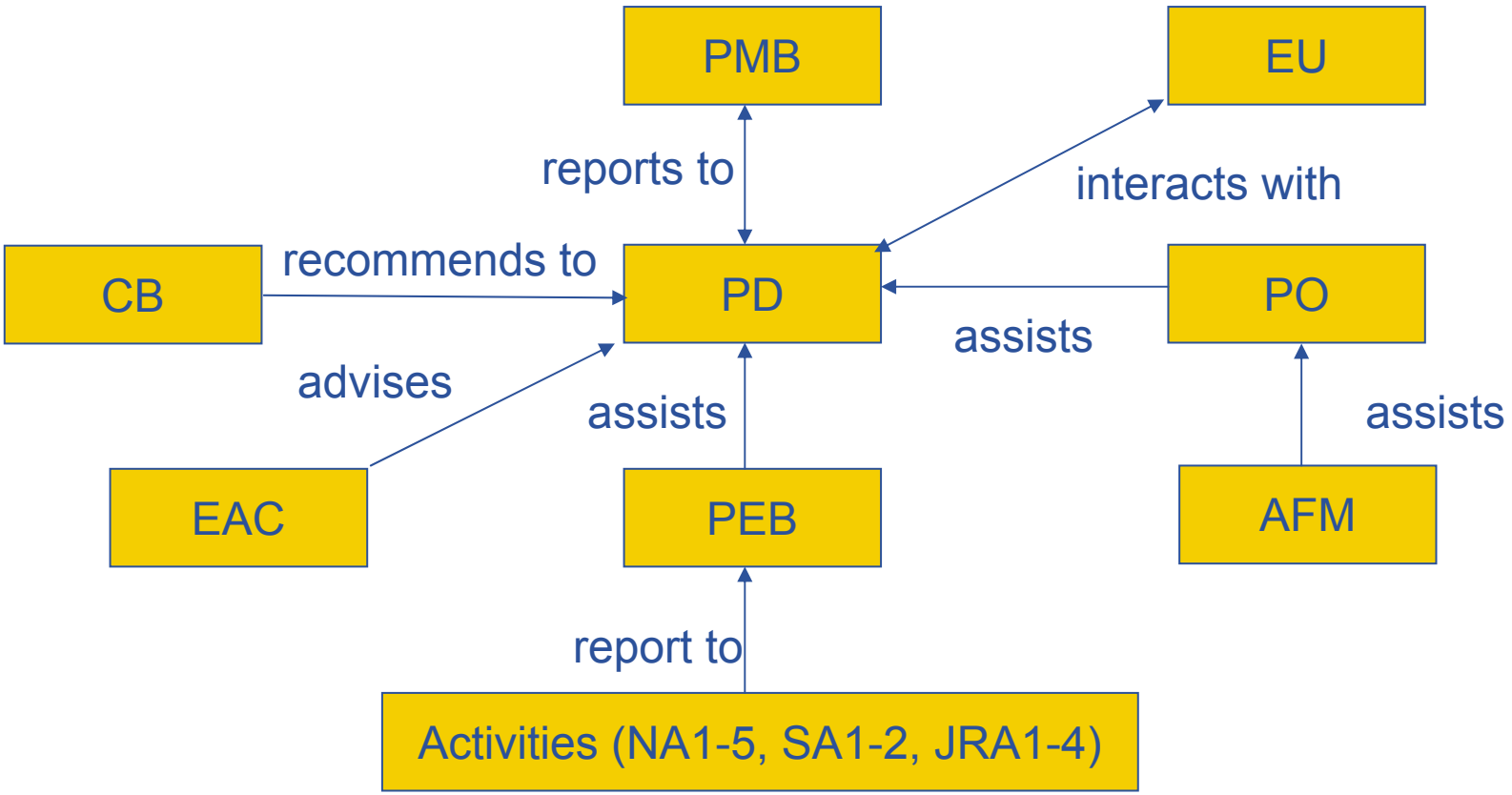
48% Services

SA1: Grid Operations, Support and Management

SA2: Network Resource Provision

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

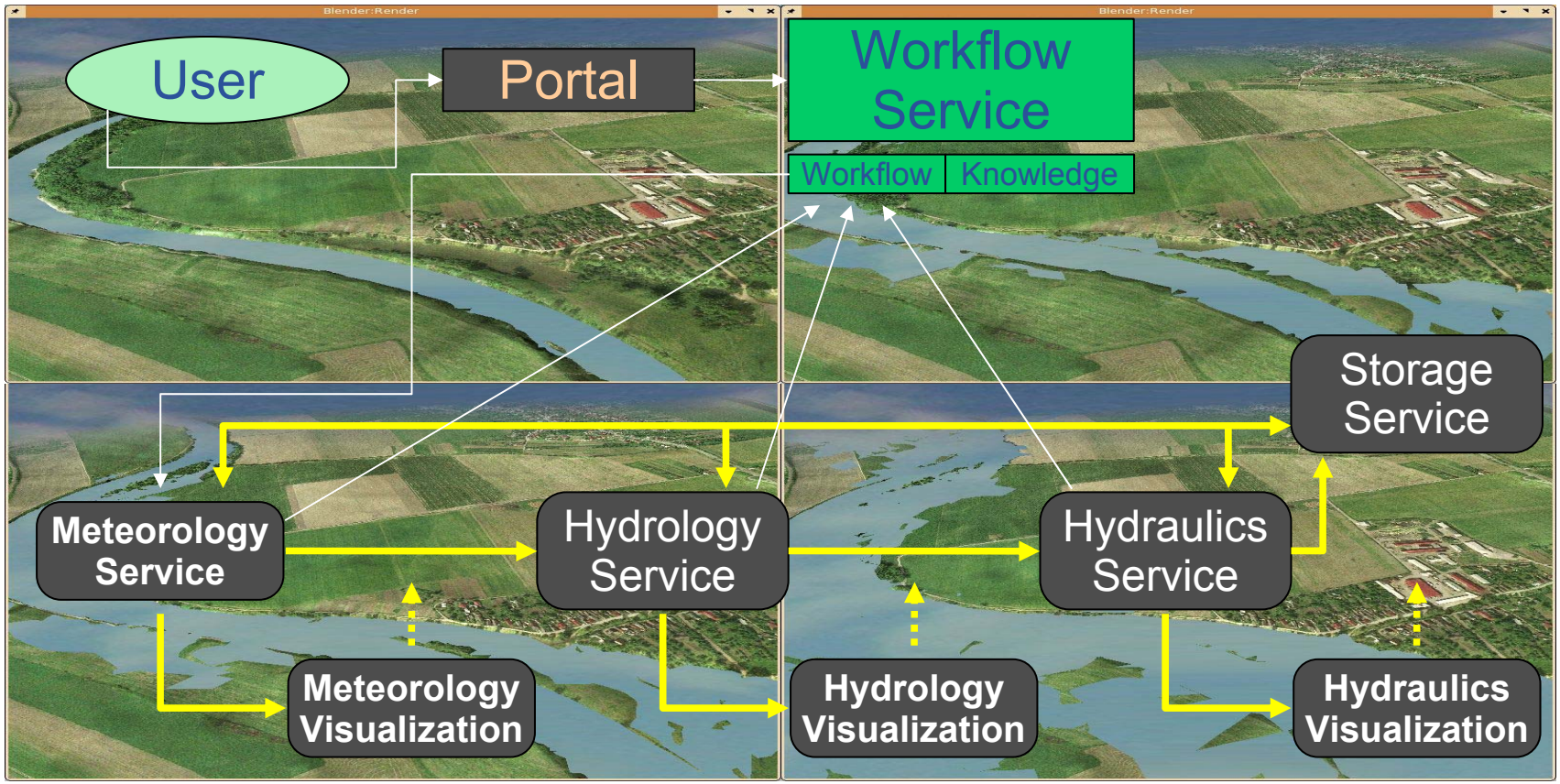
EGEE Management structure



CB	**Collaboration Board**
EAC	**External Advisory Committee**
EU	**European Union**
PD	**Project Director**

PEB	**Project Executive Board**
PMB	**Project management Board**
PO	**Project Office**
AFM	**Administrative Federation Meeting**

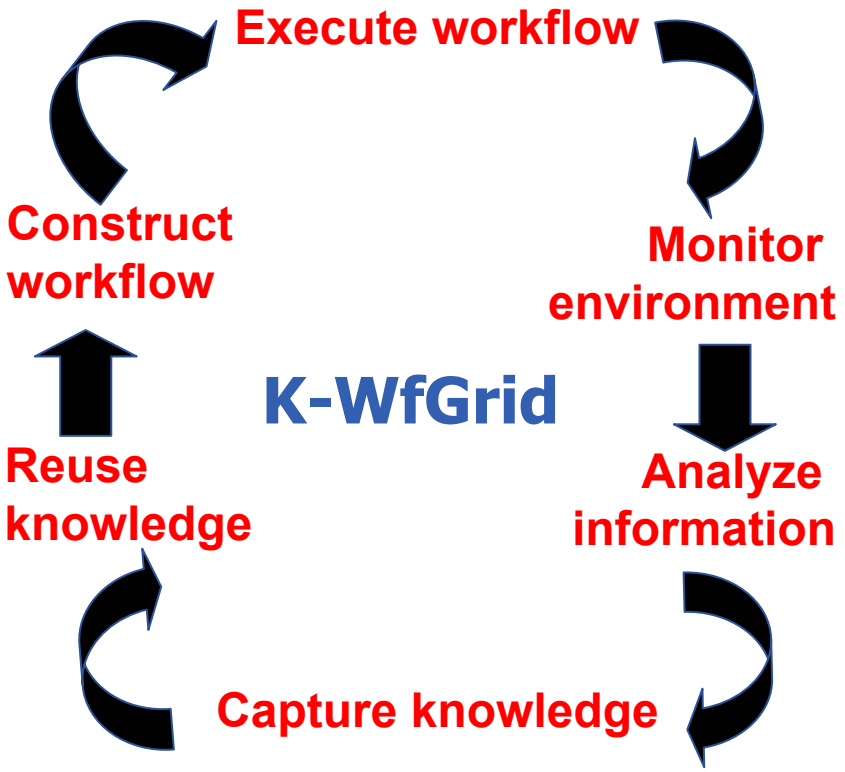
- What is Grid computing?
- Characteristics of a grid
- Applications (what's in it for the working scientist)
- European grids, and the world
- EGEE project
- **K-WfGrid project**



Applications are complex and dynamically constructed from services.
Current solutions rely on a human as a source of knowledge.

Objectives of K-WfGrid

- Integrating Grid services into coherent application scenarios
- Enabling automatic construction and reuse of workflows with knowledge gathered during operation
- Involving Grid monitoring and knowledge acquisition services in order to provide added value for end users



Technologies: service-oriented Grid architecture, software agents, ontologies, dynamic instrumentation

Thank you

www.eu-egee.org