egee

Enabling Grids for E-sciencE

Status of the Project

Fabrizio Gagliardi, Project Director, CERN

EGEE 1st EU Review

9-11/02/2005

www.eu-egee.org

Information Society

EGEE EU Review – Agenda

Enabling Grids for E-sciencE

- **Wednesday, 09/02/05 PM:**   *Dieter Kranzlmüller*
  - Overview: Status of Project
  - Status of Production Service and applications

- **Thursday, 10/02/05 AM:**   *Mirco Mazzucato*
  - Networking, User Training and Induction
  - Application Demonstrations

- **Thursday, 10/02/05 PM:**   *Manuel Delfino*
  - Development/Re-Engineering (gLite)
  - Security and deployment

- **Friday, 11/02/05 AM:**   *Robin Middleton*
  - Dissemination, Outreach, and Policies
  - Plans for Next Period
  - Conclusions

- **Friday, 11/02/05 PM:**   *Fabrizio Gagliardi*
 - Feedback from Reviewers

- **Project goals and structure**

- **Financial status and resource consumption**

- **Accomplishments**

- **Overall issues and concerns**

- **Summary**

- **Several projects in FP5 (DataGrid, DataTAG, CrossGrid, etc…) and in other EGEE partner states (VDT, Globus, Condor, etc…) demonstrated the viability of Grid technology for data intensive science and produced a large amount of functional middleware**

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

- **Strong interest from dedicated user communities**
    – High Energy Physics
    – Biomedicine

EGEE committed to "hit the ground running" at the proposal submission time in 2003
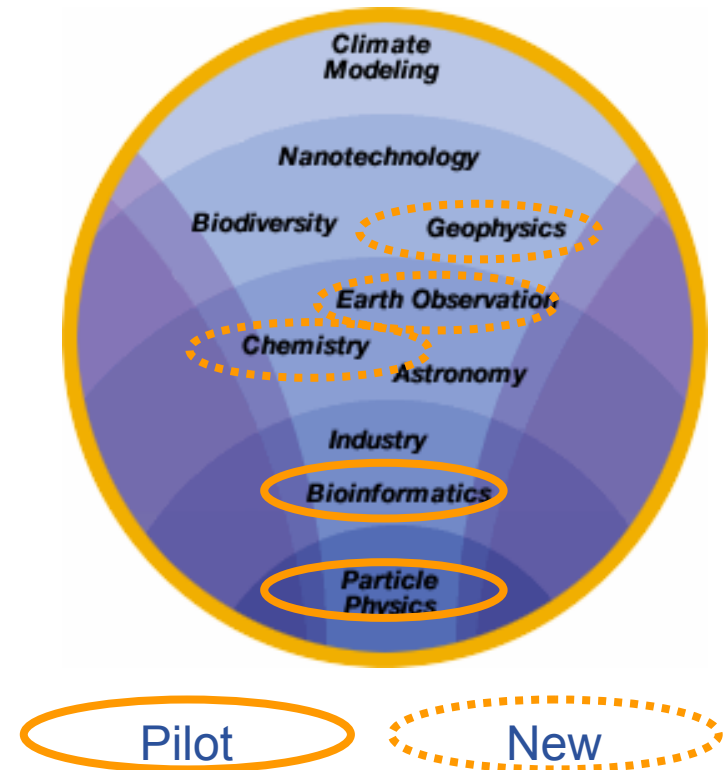
- Strong HEP community focused on the LCG project
- EGEE profits from the resources - no funded computing/data resources in EGEE
- EGEE profits from the HEP applications
- LCG obtains additional production and operation efforts
- LCG contributes specs to new production quality middleware and profit from EGEE S/W development
- LCG has strict deadlines and quality criteria that constantly push EGEE
- Shared management and technical infrastructure
- Many common partners and regional/national funding agencies

Enabling Grids for E-sciencE
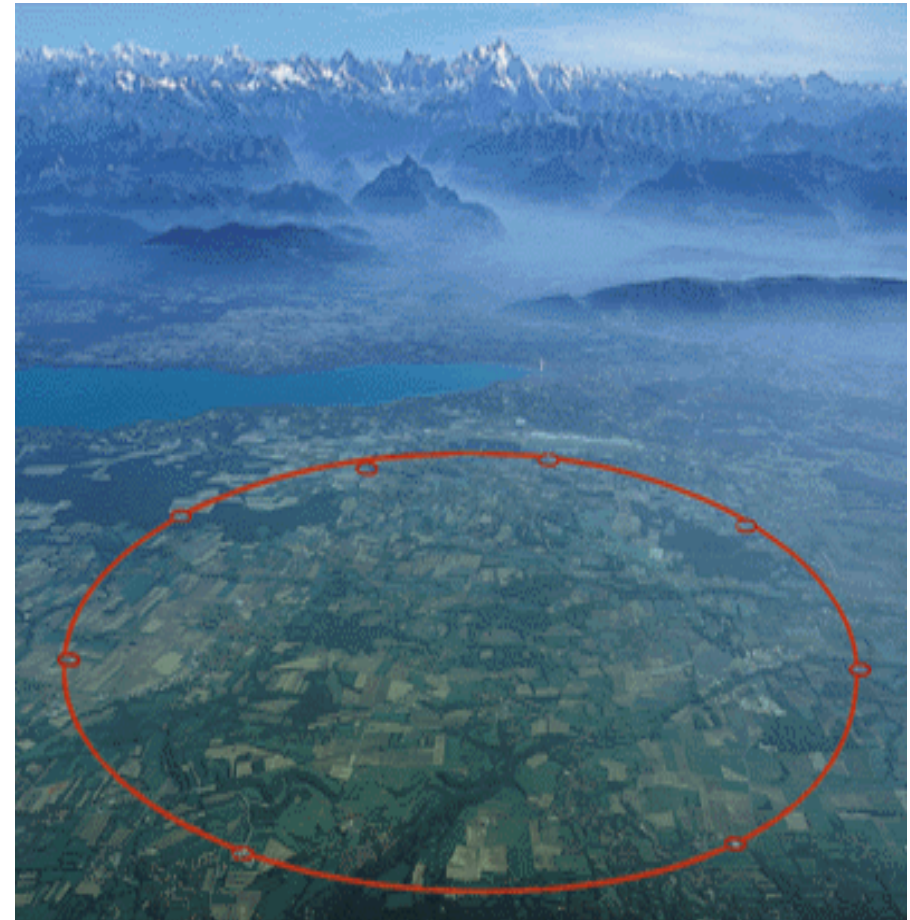
Within a four year programme:

- **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**

eGee
Enabling Grids for E-sciencE

- **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**

Climate Modeling
Nanotechnology
Biodiversity
Geophysics
Earth Observation
Chemistry
Astronomy
Industry
Bioinformatics
Particle Physics

Pilot   New

EGEE

Computing Resources – Dec. 2004

Enabling Grids for E-sciencE

We're already exceeding deployment expectations!!!

○ Country providing resources
● Country anticipating joining EGEE/LCG

In EGEE-0 (LCG-2):
⇨ > 100 sites
⇨ > 10,000 CPUs
⇨ > 5 PB storage

- **70 leading institutions in 27 countries, federated in regional Grids**

- **~32 M Euros EU funding for first 2 years starting 1st April 2004**

- **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

eGee

Enabling Grids for E-sciencE

- **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**

24% Joint Research

28% Networking

48% Services

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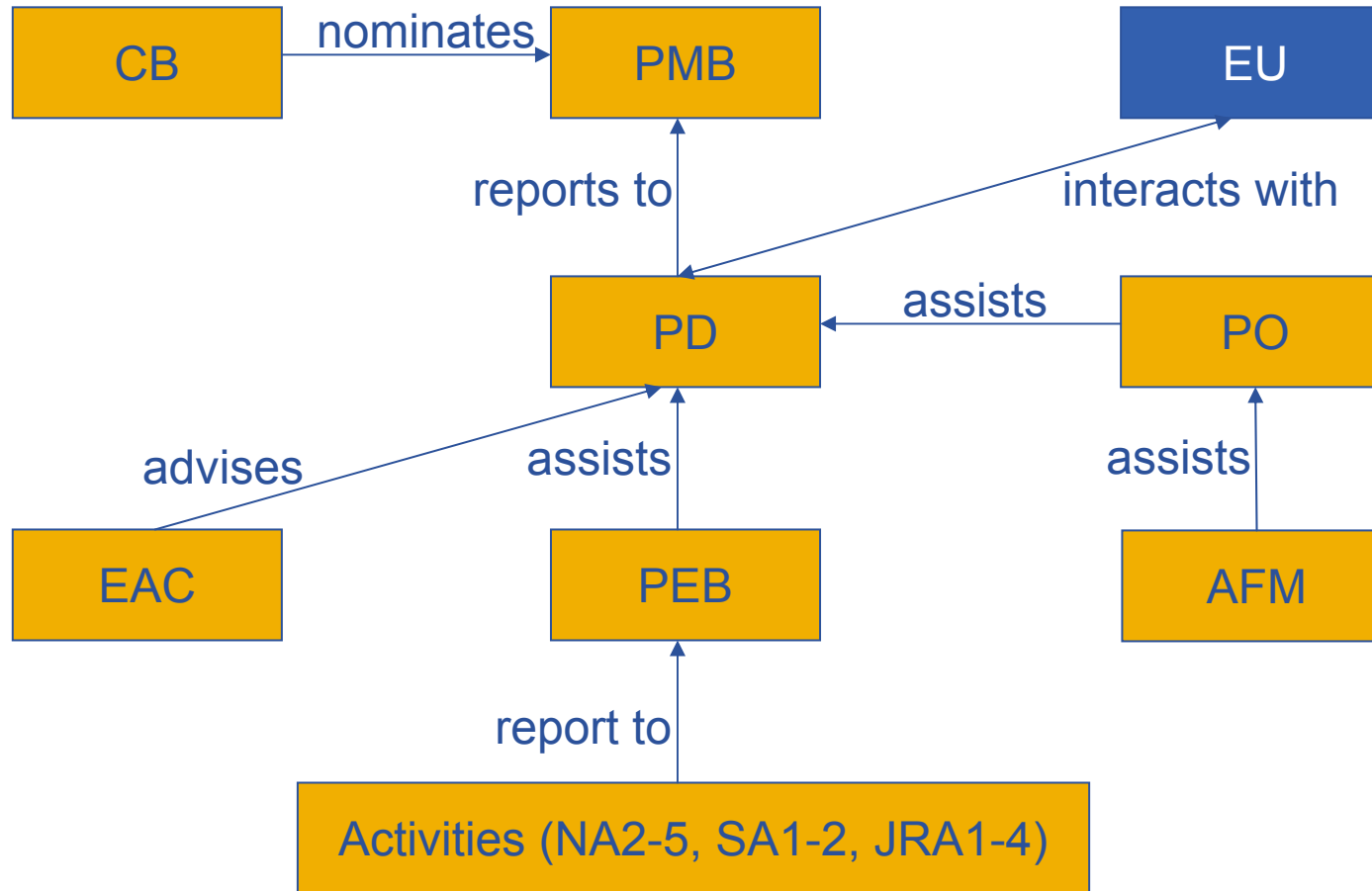
egee

Enabling Grids for E-sciencE

- **Provisional Cost Claim to December 2004: 33% spending of the overall maximum EC contribution after 9 months of operation**
  - We are 10% under-spent due to hiring process but are now at full complement

- **Expenses are as follows**
  - Personnel: 91%
  - Travel and Subsistence: 7%
  - Other Costs: 2%
  - Audit Costs are not accounted for as audits have not yet been performed

- **CERN/IT: customized the PPT tool to monitor project effort consumption (timesheets)**
  - 850 people registered in the EGEE consortium
  - Total equivalent of ~400 FTEs

Total (F+UF) Effort in PM
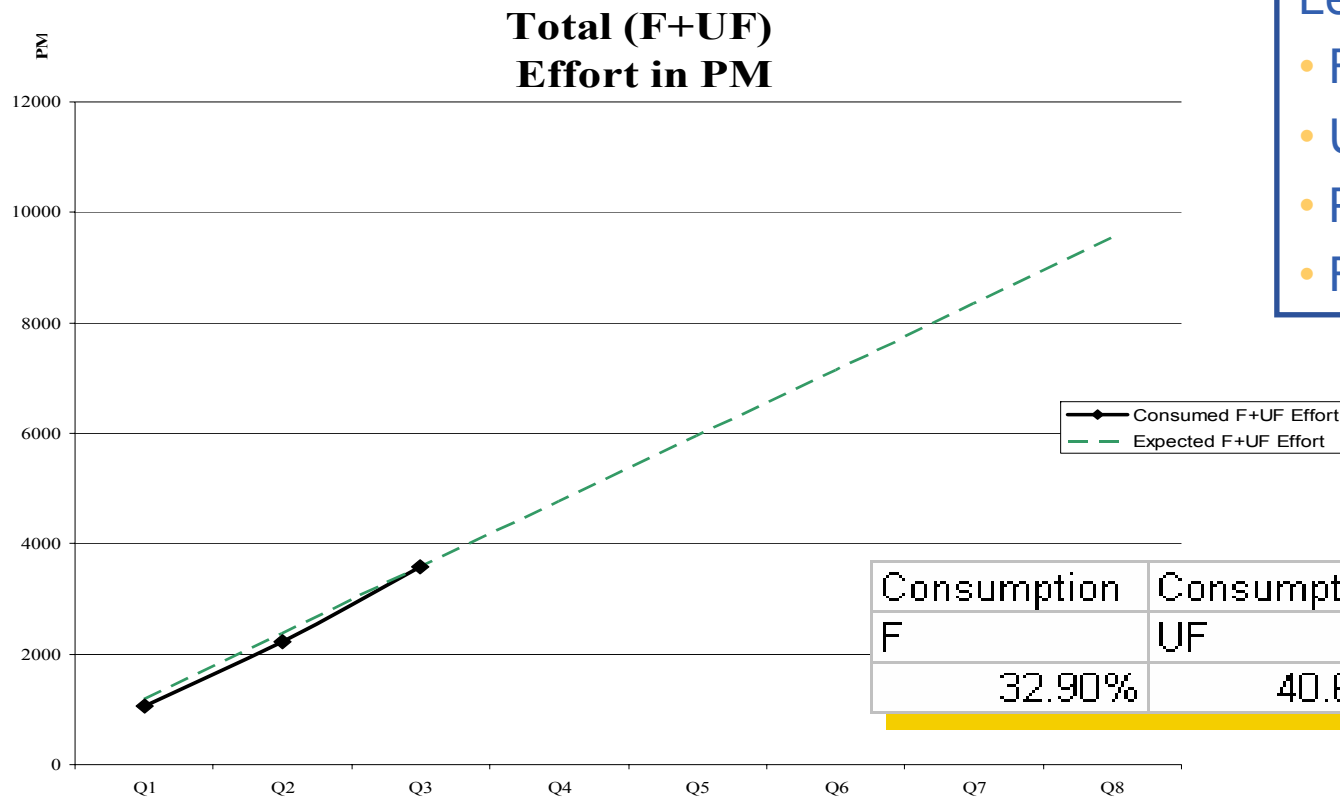
Legend:
- F: Funded
- U: Unfunded
- PM: Person Months
- FTE: Full time equiv.

Consumption	Consumption	Consumption
F	UF	F+UF
32.90%	40.60%	36.10%

All **64 deliverables and milestones** met on time,

with exception of:

- **MJRA3.5 "Secure Storage Credential procedures":**
  - Split into 2 parts:
    - *Collect relevant information (PM10), used in the work on MJRA3.6 "Security operational procedures - first revision" (PM12) - Provide recommendation document (PM15) using the results from the MJRA3.6*

  - **No effect on overall schedule**

- **MJRA1.4 "Software for the Release Candidate 1"**
  - One month late
  - Management review committee established. Conclusions and updated plan published
  - **No effect on overall schedule** – release 1 software will be delivered at the end of March 2005 as foreseen (DJRA1.3)

- **Design team set-up with representatives from all the contributing groups**
    - Resulted in architecture & design taking best features of the middleware toolkits

- **Prototypes built**
    - Using various code bases to explore different possibilities

- **Management review committee created**
    - Detailed plan established leading to release v1.0 of gLite at end of March 2005

- **Distributed software development has intrinsic overhead compared with a centralised model, but this is the cost of acceptance by the large & diverse communities (users, sites, etc.)**

- **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)

Accomplishments

- **Management structures set up and running (PMB, PEB, AFM, EAC, CB)**
  - AFM created to address EGEE's administrative needs

- **Contract, Consortium Agreement signature coordination**
  - 70 partners, and approximately
  - 35 non-contracting participants

- **First Contract Amendment**
  - Minor contractor changes
  - 3 new non-contracting partners

- **Dissemination activity in the PO**
  - Presentations worldwide by the Project Management

Issues

- **New FP6 rules and guidelines for reporting (not known at proposal preparation)**

- **Work load of PO is more linked to number of partners than budget**

- **Effort required to manage a project of this scale is underestimated, mitigation for PO**
  - Deputy PD appointed (Dieter K.)
  - 0.5 FTE moved from CERN/NA2 to NA1 working on Technical Coordination, compensated by 0.5 FTE moved from NA5 to NA2

- **Partner issues (catch all for the rest of the project)**

- **GEANT** **–** **Network provision for EGEE**

- **LCG** **–** **shares the same grid operations and infrastructure**

- **DEISA** **–** **Investigating interoperability with super-computers**

- **SEE-GRID** **–** **Extension of EGEE to Balkans**

- **Grid3/OSG (USA)** **–** **common middleware base and limited inter-operability (with LCG)**

- **Further extensions of EGEE are foreseen for Baltic states, Mediterranean, Latin America and Asia**

Enabling Grids for E-sciencE

- **EGEE includes delivery of over 130 deliverables + milestones**
    - Over 60 in this first reporting period

- **Deliverable review process requires effort from all activities including technical experts**

- **Review process is heavy to meet quality goals (>1 month) because of the large distributed nature of the project and the large number of participants**

Enabling Grids for E-sciencE

- **We have exceeded contractual commitments in many areas**

- **Management and administrative work load linked to number of partners but resources limited by budget**

- **We are the largest and probably the only multi-disciplinary production grid infrastructure**

- **We are exposed to dedicated and demanding communities which are both a strength and a weakness**

- **Plans for long term grid infrastructure will be developed during the second year**

- **This review will help us to assess our progress and plan for the future**