egee

Enabling Grids for E-sciencE

Status of the Project

Fabrizio Gagliardi, Project Director, CERN

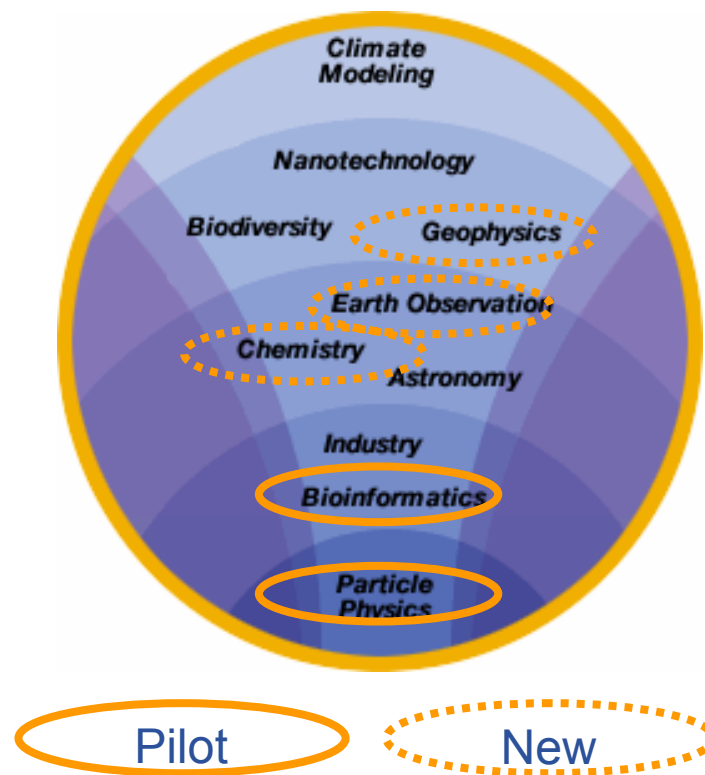*EGEE 1st EU Review*

9-11/02/2005

www.eu-egee.org

Information Society

- **Project goals and structure**

- **Financial status**

- **Manpower levels**

- **Activities main accomplishments and issues**

- **Overall issues and concerns**

- **Summary**

- 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

Enabling Grids for E-sciencE

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

- **32 M Euros EU funding**

- **O(100 M Euros) total budget for first 2 years starting 1st April 2004**

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**
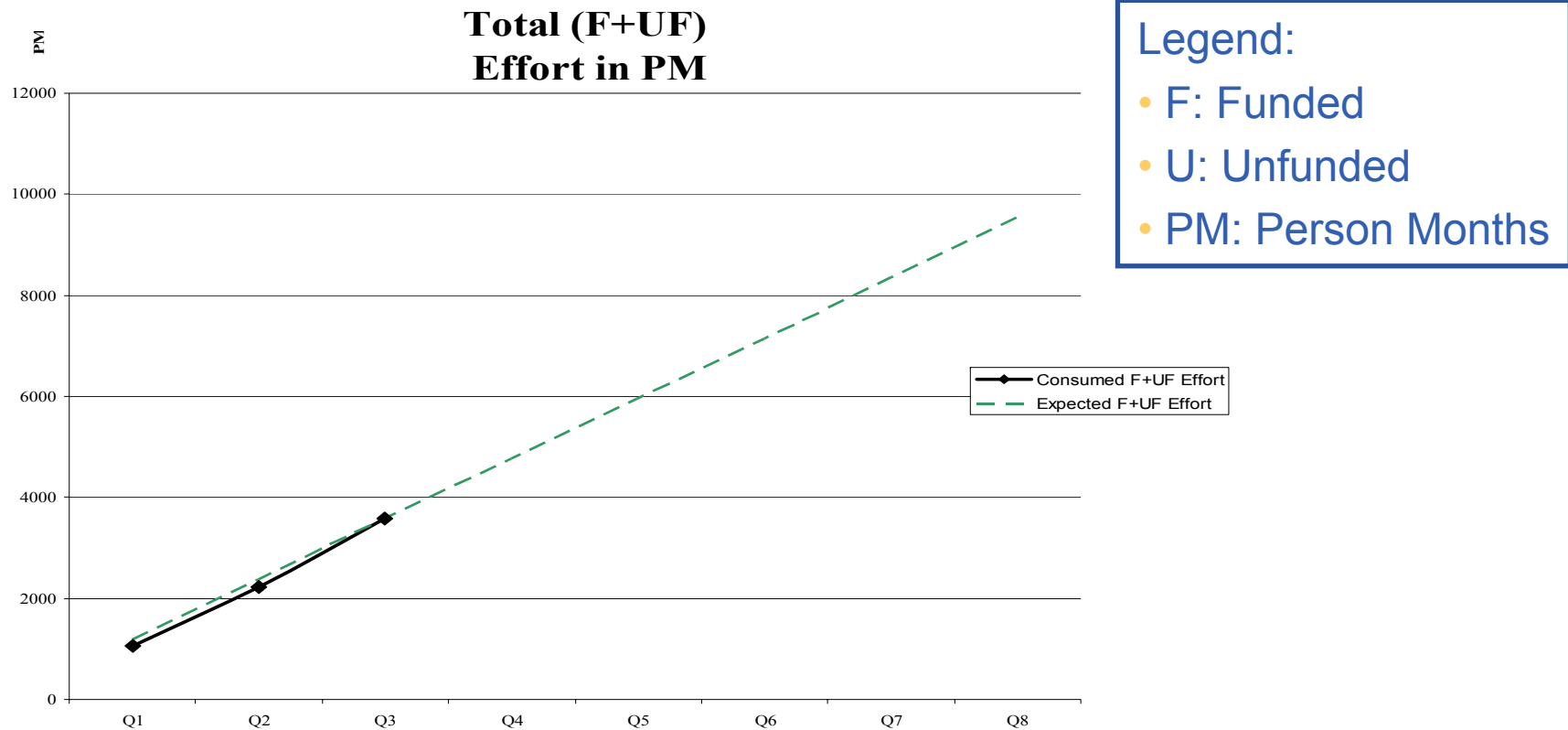
- **Joint Research Activities**
    - JRA1: **Middleware Reengineering + Integration**
    - JRA2: **Quality Assurance**
    - JRA3: **Security**
    - JRA4: **Network Services Development**

24% Joint Research

28% Networking

48% Services

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

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 compliment

- **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: help to implement PPT tool to monitor project effort consumption (timesheets)**
  - 850 people registered around Europe and Russia

Total (F+UF) Effort in PM

Legend:
- F: Funded
- U: Unfunded
- PM: Person Months

Consumed F+UF Effort
Expected F+UF Effort

Manpower Distribution

eGee — Enabling Grids for E-sciencE

Federation	Funded	UnFunded	Grand Total
CENTRAL EUROPE	203	56	259
CERN	326	308	634
FRANCE	426	34	460
GERMANY	107	12	119
ITALY	371	220	592
NORTHERN	156	84	240
NRENS	31	21	52
RUSSIA	435	30	465
SOUTH EAST	126	44	171
SOUTH WEST	136	48	184
UK/IRELAND	326	67	393
(USA & Korea)	0	7	7
Grand Total	2,644	930	3,573

Figures in person months

- **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)
      - provide recommendation document (PM15) using the results from the MJRA3.6 "Security operational procedures – first revision" (PM12)
    - 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)

Accomplishments

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

- **Limited resources available in the NA1 project office**

- **New FP6 rules and guidelines for reporting**

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

EGEE builds on the work of LCG to establish a grid operations service

- CERN and partners are building and operating the LHC Grid (LCG) to handle data processing for the experiments at the Large Hadron Collider (LHC)

- **A collaboration between:**
  - The physicists and computing specialists from the LHC experiment
  - The projects in Europe and the US that have been developing Grid middleware
  - The regional and national computing centres that provide resources for LHC
  - The research networks

EGEE and related infrastructure
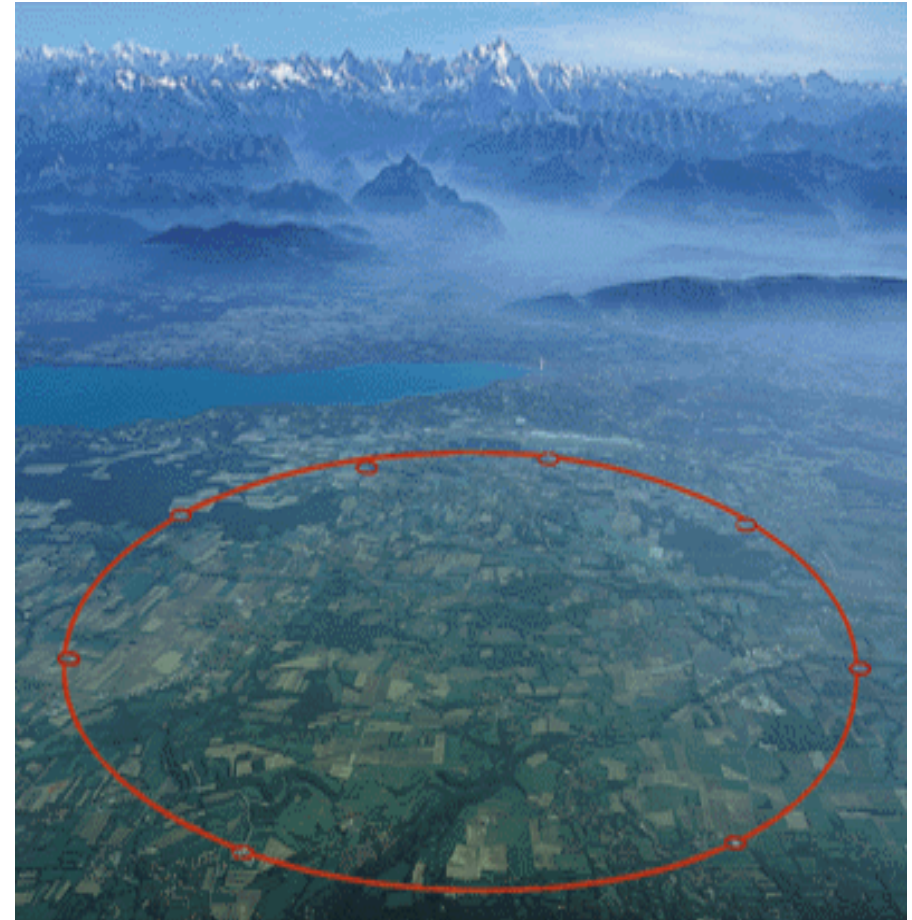
- **GEANT** **– Network provision for EGEE**

- **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**

- **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 fro the EDG license and takes into account feedback from the World Intellectual Property Office (WIPO)

Accomplishments

- **Public Website with over 4000 visitors a month**

- **A wide-range of publicity material (information sheets, fact sheet, folders, templates for presentations, posters, word documents, etc.)**

- **Organised 2 successful EGEE conferences**

- **Over 90 media cuttings, 2 radio interviews and 2 TV interviews worldwide**

- **EGEE present at over 100 different events across the world**

- **Regular newsletters keep all members informed of latest developments**

Issues

- **Activities concentrate principally on technical activities, thus not aware of newsworthy themes**

- **Monthly communication of metrics by NA2 partners to TERENA is not automatic**

- **NA2 partners have to ensure that their dissemination activities fit in to the overall strategy for NA2**

NA3 Major Accomplishments

eGee
Enabling Grids for E-sciencE

Accomplishments

- **Establishment of an effective and federated collaboration for training across whole EGEE geographic area**

- **Significantly exceeding first year goals in training and induction**

- **Identifying the requirements for t-Infrastructure and pioneering its provision with GILDA and the shared training material repository**

Issues

- **Sustaining high-quality training throughout the region**

- **Experts are needed to inform planning, preparation and delivery of new courses**

- **The demand for training is growing rapidly and the breadth of requirements expands with each new community and operational advance**

eGee

Enabling Grids for E-sciencE

Accomplishment

- **The successful deployment of several biomedical and "generic" applications**

- **Successful outreach to new generic communities**
    - Well established process, providing education and application migration
    - Using GILDA and GENIUS as tools
    - For new application areas selected by EGAAP

- **Demonstration of prototype analysis system using gLite for all 4 LHC experiments**

Issues

- **Integration and support of multiple user communities,**

- **Foresee difficulty in supporting significant increase in the number of EGEE active users**

- **Migration of application to gLite (currently deployed on LCG2)**

- **Availability of security-enhanced services for data manipulation and job execution**

Enabling Grids for E-sciencE

Accomplishment

- Development of the eIRG White Paper process

- Providing continuous effort to maintain momentum behind the Den Haag eIRG White Paper work

- "Concertation" event planning and follow-up (together with NA2)

- Synergy roadmap development in collaboration with SEE-GRID and DEISA

Issues

- Late availability of dedicated resources

- Scope and goals of the activity adjusted after project conception

- Deliverables depend on EU events, outside EGEE control

SA1 Major Accomplishments

Accomplishment

- **Functioning production grid infrastructure in place**
  - 110 sites in 31 countries (10 countries and 18 sites outside Europe)
  - >10,000 CPU + ~5 PB of storage
  - **Exceeding all of the project milestones in this respect**
- **Support extensive and intense LHC experiments' data challenge (2004)**
  - 1 Million SI2k-years of CPU
  - Peaks of 4000 jobs in parallel for a single VO
  - Other VO's now being deployed successfully on the Production Services
- **Operations support infrastructure in place, including procedures for**
  - Problem management
  - Escalation
  - Incident response

Issues

- **Improving the quality, reliability and efficiency of the operations**
- **Understanding how to approach "24x7" global operations**
- **Continuing to develop the user support aspects in order to build reliable user support infrastructure**
- **Lighten new VO deployment (still much too heavy)**

Grid operations screens are visible during the review

eGee
Enabling Grids for E-sciencE

Accomplishment

- **Progress in the work from the "Application requirements" to the Network SLAs**

- **Proposed models well received by the networking community (EGEE/JRA4, GN2, TNLC)**

- **Initiated QoS experimentation for an EGEE application to get a real use-case**

Issues

- **Network support at the middleware level**

- **Diversity of the site connectivity and the associated network services can make the network user support more complex**

- **GN2 began six months after EGEE adds difficulties for EGEE networking activities**

Accomplishment

- **Produced gLite Architecture and Design documents**
    - Already referenced outside EGEE
- **Software development processes in place**
    - Prototyping
    - Design
    - Re-engineering
    - Integration
    - Testing
    - Release
- **Software available to pilot application representatives**
- **Actively following emerging standards (GGF, WSRF, etc.), but not early adopters**

Issues

- **Disagreement on what software is made available**

- **Discontinuity between the early prototypes and contents of the first release**

- **Staffing in Testing/Integration under pressure at this stage**

- **Communication with software clusters not optimal (difficult to follow their work)**

Accomplishment

- **QA organisation is in place**

- **Main procedures, metrics and tools are in place**

- **Follow-up and feedback will continue to verify the project does deliver according to the agreed quality levels**

Issues

- **Metrics collection not automated (labour intensive)**

Accomplishment

- **Producing key security deliverables (well received in the community)**
  - Global Security Architecture
  - Site Access Control Architecture

- **Delivered a number of security modules, of which four will be part of gLite v1**

- **Driving community level agreements for middleware and policy**
  - EUGridPMA

Issues

- **Geographically distributed teams**

  Mitigation
  - Cross activity groups, more F2F meetings, especially in handing over security modules

Issues

- **Conflicting/challenging security requirements from applications**

  Mitigation
  - Proposed solutions meeting both sets of requirements as much as possible

Enabling Grids for E-sciencE

Accomplishment	Issues
• **NPM prototype demonstrated that**	• **Deployment of NPM and BAR services on the EGEE fabric**
– Collect and publish Network Performance data from heterogeneous sources	– JRA4 is currently short-listing available NPM tools to be deployed on the fabric
• **Activity management changed during first period**	– Currently no provision in the EGEE contract for the maintenance or deployment of these tools
	• **Mitigation**
	– Collaborate with JRA1 to substantiate the need and location for the deployment of these tools and
	– Same with SA1 for their deployment

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

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

- **The 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

  – Same applies for Technical Coordination

- **Not enough resources to support many more user communities**

  – Already reached our PM24 targets

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**
- **Review process is heavy to meet quality goals (~ 1 month)**
- **Problems encountered by all activities**
  - Effort exceeds expected levels at proposal writing
  - All activities find it hard to provide required effort to support
    - Deliverable and milestone preparation
    - Review process

Enabling Grids for E-sciencE

- **We have exceeded contractual commitments in many areas**
  - Work load more linked to number of partners than budget (but 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**