The EGEE Project

David Foster
Communications Systems Group Leader

Bob Jones
EGEE Technical Director

CERN
Geneva, Switzerland

Healthgrid workshop, Oxford, UK

8th April 2005

Information Society

www.eu-egee.org

Enabling Grids for E-sciencE

- **The EGEE Project**
  - Overview and Structure
  - Grid Operations
  - Middleware
  - Networking Activities
  - Applications
    - Biomedical
    - How new users become involved

- **Summary and Conclusions**

- **Goal of EGEE: develop a service grid infrastructure which is available to scientists 24 hours-a-day**

- **The project concentrates on:**
  - building a consistent, robust and secure Grid network that will attract additional computing resources

  - continuously improve and maintain the middleware in order to deliver a reliable service to users

  - attracting new users from industry as well as science and ensure they receive the high standard of training and support they need

Enabling Grids for E-sciencE

EGEE is the largest Grid infrastructure project in Europe:

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

- Leveraging national and regional grid activities

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

- EU review, February 2005 successful

- Preparing 2nd phase of the project – proposal to EU Grid call September 2005

- 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 is a truly international under-taking**

- **Collaborations** with other existing European projects, in particular:
    - **GÉANT, DEISA, SEE-GRID**

- **Relations** to other projects/proposals:
    - **OSG: OpenScienceGrid (USA)**
    - **Asia: Korea, Taiwan, EU-ChinaGrid**
    - **BalticGrid: Lithuania, Latvia, Estonia**
    - **EELA: Latin America**
    - **EUMedGrid: Mediterranean Area**
    - **…**

- **Expansion of EGEE infrastructure in these regions is a key element for the future of the project and international science**

![EGEE logo — Enabling Grids for E-sciencE]

24% Joint Research

JRA1: **Middleware Engineering and Integration (5 partners, 16%)**

JRA2: Quality Assurance (2 partners, 2%)

JRA3: Security (5 partners, 3%)

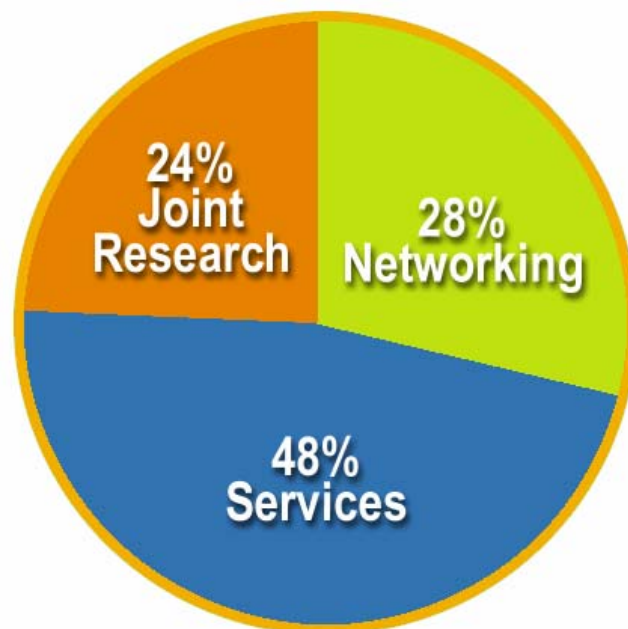**JRA4**: Network Services Development

28% Networking

NA1: Management

NA2: Dissemination and Outreach

NA3: User Training and Education
(22 partners, 4%)

NA4: **Application Identification and Support (20 partners, 12.5%)**

NA5: 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

48% Services

SA1: **Grid Operations (48 partners, 45%)**

SA2: Network Resource Provision

![egee - Enabling Grids for E-sciencE]

- **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

We're already exceeding deployment expectations!!!

○ Country providing resources
● Country anticipating joining EGEE/LCG

In EGEE-0 (LCG-2):
⇨ >120 sites
⇨ >12,000 CPUs
⇨ >5 PB storage

eGee

Enabling Grids for E-sciencE
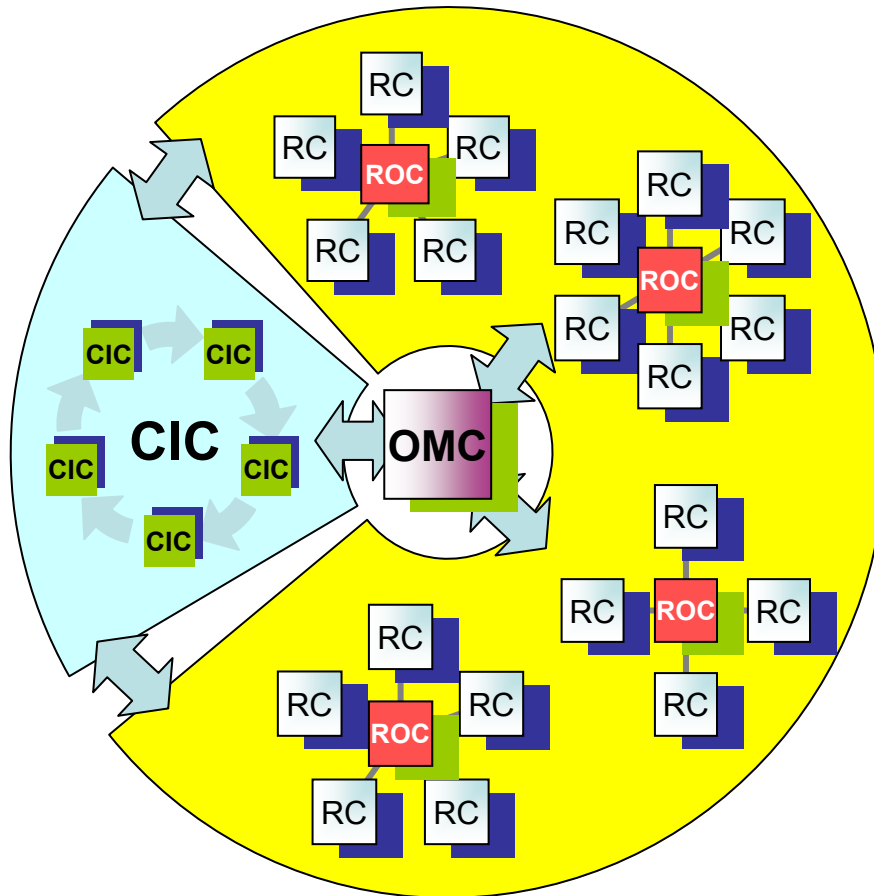
RC - Resource Centre
ROC - Regional Operations Centre
CIC – Core Infrastructure Centre

- **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**

- **Operation of Production Service: real-time display of grid operations**

- **Accounting information**

- **Selection of Monitoring tools:**

  – GIIS Monitor + Monitor Graphs

  – Sites Functional Tests

  – GOC Data Base

  – Scheduled Downtimes

  – Live Job Monitor

  – GridIce – VO + fabric view

  – Certificate Lifetime Monitor

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

- **Intended to replace present middleware (LCG-2)**
- **Developed mainly from existing components**
- **Aims to address present shortcomings and advanced needs from applications**
- **Regular, iterative updates for fast user feedback**
- **Makes use of web-services where currently feasible**

Application requirements http://egee-na4.ct.infn.it/requirements/

- The 1st release of gLite (v1.0) made end March'05
  - http://glite.web.cern.ch/glite/packages/R1.0/R20050331
  - http://glite.web.cern.ch/glite/documentation

- Lightweight services

- Interoperability & Co-existence with deployed infrastructure

- Performance & Fault Tolerance

- Portable

- Service oriented approach

- Site autonomy

- Open source license

gLite

Lightweight Middleware for Grid Computing

- **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)*

- **Information and Monitoring**
  - R-GMA *(EDG)*

- **Storage Element**
  - glite-I/O *(AliEn)*
  - Reliable File Transfer *(EGEE)*
  - GridFTP *(Globus)*
  - SRM: Castor *(CERN)*, dCache *(FNAL, DESY)*, other SRMs

- **Catalog**
  - File/Replica & Metadata Catalogs *(EGEE)*

- **Security**
  - GSI *(Globus)*
  - VOMS *(DataTAG/EDG)*
  - Authentication for C and Java based (web) services *(EDG)*

Now doing rigorous scalability and performance tests on pre-production service
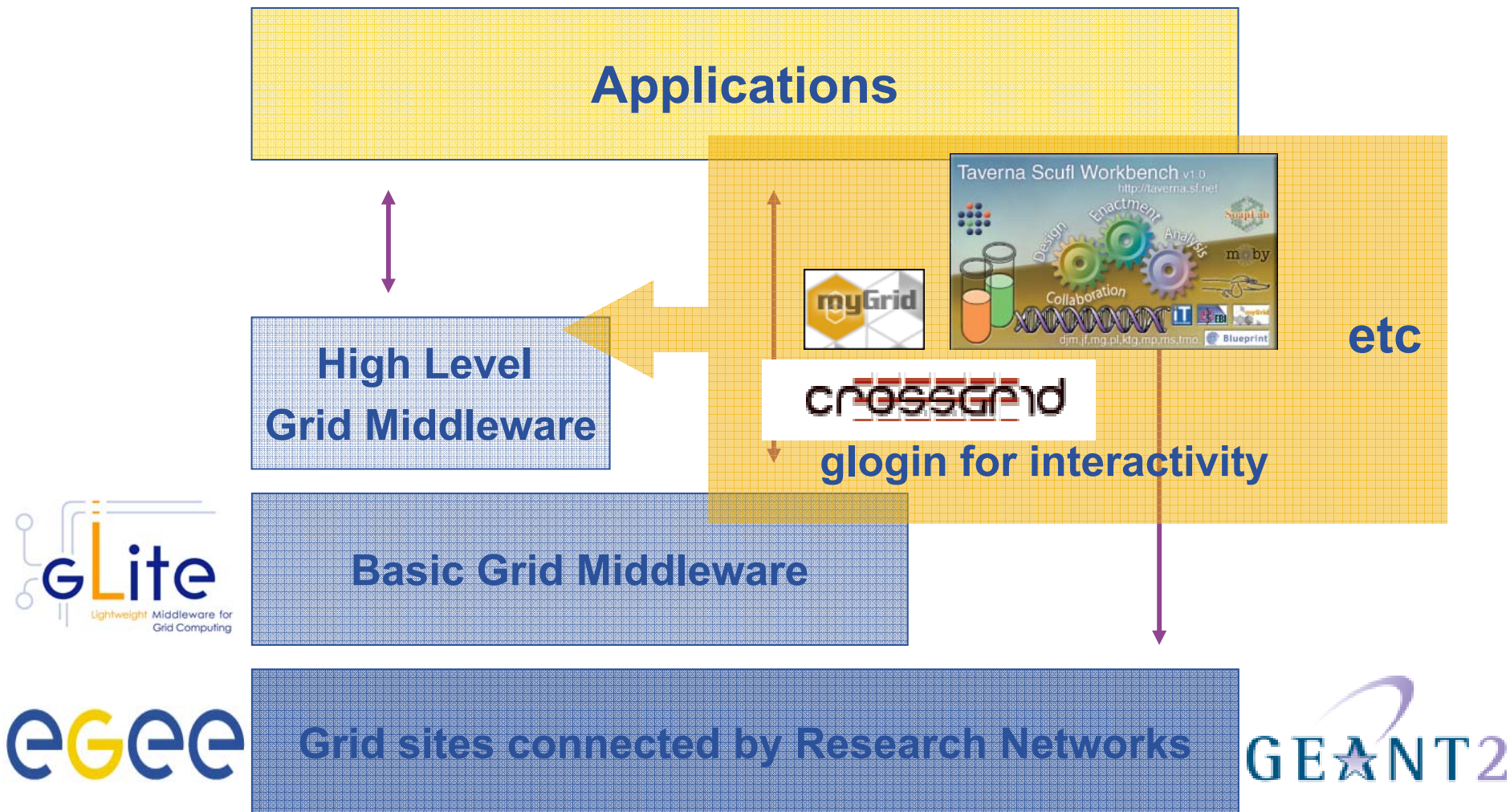
eGee

Enabling Grids for E-sciencE

Applications

High Level Grid Middleware

etc

glogin for interactivity

Basic Grid Middleware

Grid sites connected by Research Networks

eGee

Enabling Grids for E-sciencE

- 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)**

Emphasis in EGEE is on operating a production grid and supporting the end-users
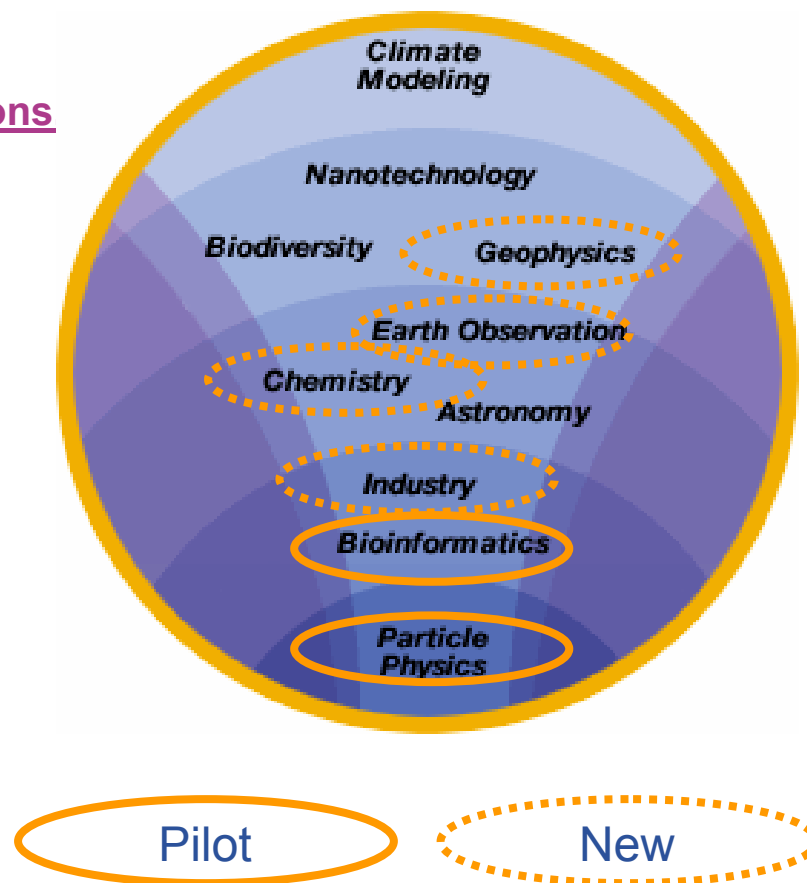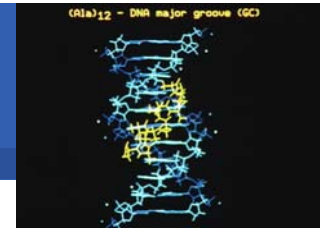
Enabling Grids for E-sciencE

- **Public and technical websites constantly evolving to expand information available and keep it up to date**

- **2 conferences organised**
    - **~ 300 @ Cork, ~ 400 @ Den Haag**
- **Athens 3rd project conference 18-22 April '05**
    - http://public.eu-egee.org/conferences/3rd/
- **Pisa 4th project conference 24-28 October '05**

- **More than 70 training events (including the GGF grid school) across many countries**
    - **~1000 people trained**
        - **induction; application developer; advanced; retreats**
    - **Material archive with more than 100 presentations**

- **Strong links with GILDA testbed and GENIUS portal developed in EU DataGrid**

- **Pilot applications**
  - **High Energy Physics**
  - **Biomed applications**
    - **http://egee-na4.ct.infn.it/biomed/applications**

- **Generic applications – Deployment under way**
  - **Computational Chemistry**
  - **Earth science research**
  - **EGEODE: first industrial application**
  - **Astrophysics**

- **With interest from**
  - **Hydrology**
  - **Seismology**
  - **Grid search engines**
  - **Stock market simulators**
  - **Digital video etc.**
  - **Industry (provider, user, supplier)**

Climate Modeling

Nanotechnology

Biodiversity   Geophysics

Earth Observation

Chemistry   Astronomy

Industry

Bioinformatics

Particle Physics

Pilot   New

- **GPS@: Grid Protein Sequence Analysis**

  – NPSA is a web portal offering proteins databases and sequence analysis algorithms to the bioinformaticians (3000 hits per day)

  – GPS@ is a gridified version with increased computing power

 – Need for large databases and big number of short jobs

- **xmipp_MLrefine**

  – 3D structure analysis of macromolecules from (very noisy) electron microscopy images

  – Maximum likelihood approach for finding the optimal model

  – Very compute intensive

- **Drug discovery**

  – Health related area with high performance computation need

  – An application currently being ported in Germany (Fraunhofer institute)
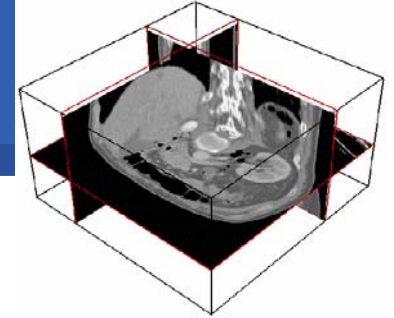
Medical imaging

- **GATE**

  – **Radiotherapy planning**

  – Improvement of precision by Monte Carlo simulation

  – Processing of DICOM medical images

  – **Objective**: very short computation time compatible with clinical practice

  – **Status**: development and performance testing

- **CDSS**

  – **Clinical Decision Support System**
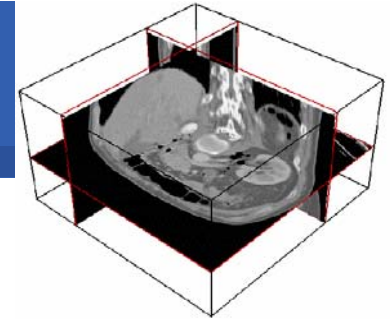
  – knowledge databases assembling

  – image classification engines widespreading

  – **Objective:** access to knowledge databases from hospitals

 – **Status**: from development to deployment, some medical end users

Talk at this workshop

Medical imaging

Enabling Grids for E-sciencE

- **SiMRI3D**
  - 3D Magnetic Resonance Image Simulator
  - MRI physics simulation, parallel implementation
  - Very compute intensive
  - **Objective**: offering an image simulator service to the research community
  - **Satus**: parallelized and now running on LCG2 resources
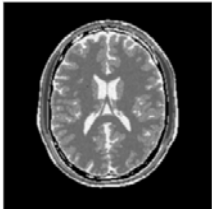
- **gPTM3D**
  - Interactive tool for medical images segmentation and analysis
  - A non gridified version is distributed in several hospitals
  - Need for very fast scheduling of interactive tasks
  - **Objectives**: shorten computation time using the grid
  - **Status**: development of the gridified version being finalized
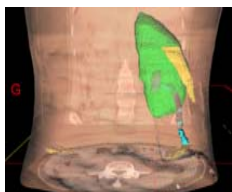
- **Goal: Enable PTM3D for the Grid**
  - PTM3D (Poste de Travail Médical 3D) is
    - an interactive radiological image visualization and processing tool
    - developed at LIMSI (CNRS)
    - with clinical usage
  - Step1: interactive response time for CPU-intensive volume reconstruction

Interaction

| Acquire | Explore | Analyse | Interpret | Render |
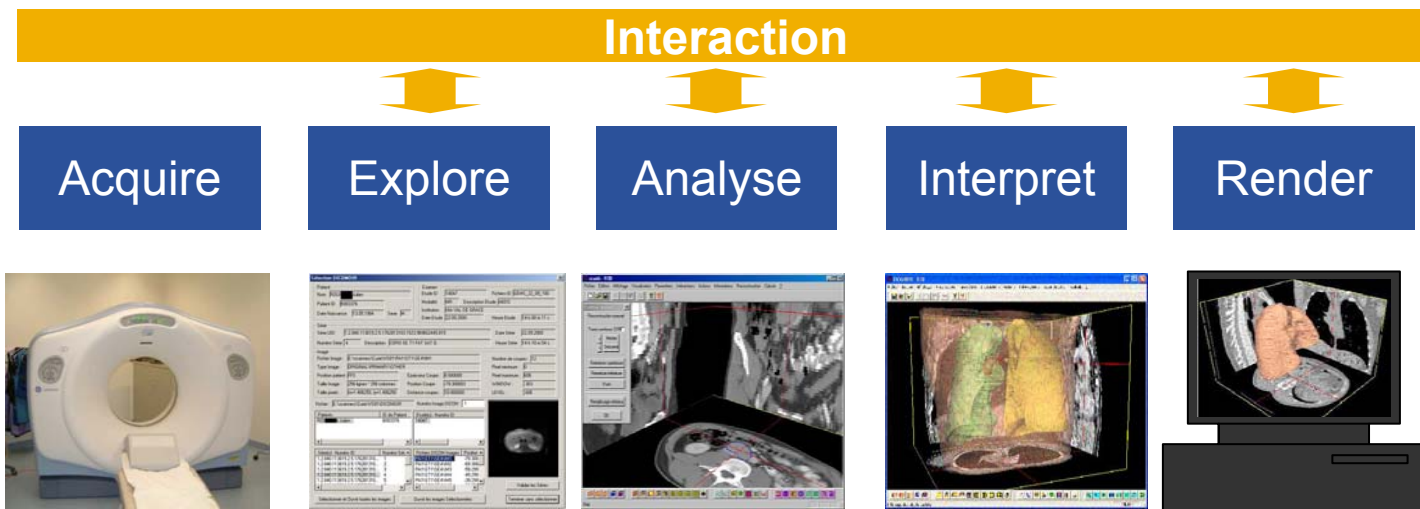
compare

	Dataset	Input data	Output data	Tasks	Standalone Execution	EGEE Execution 14 procs.
Small body	87MB	3MB 18KB/slice	6MB 106KB/slice	169	5min15s *1min54s*	37s *18s*
Medium body	210MB	9.6 MB 25KB/slice	57MB 151KB/slice	378	33min *11min5s*	2min30s *1min15s*
Large body	346MB	15MB 22KB/slice	86MB 131KB/slice	676	*18min*	*2min03*
Lungs	87MB	410KB 4KB/slice	2.3MB 24KB/slice	95	36s	24s

Status for Resource Broker IN2P3-LAL_grid09: Fri Jan 28 08:45:01 GMT 2005

RLS, VO LDAP Server: CC-IN2P3

4 RBs: CNAF, IFAE, LAPP, UPV

20 resource centres
- CEs: >2000 CPUs
- SEs: >20TB disk
15 RBs
2 RLS
1 LDAP Server

Biomed VO (jobs/month)

BioMed VO (% jobs status)
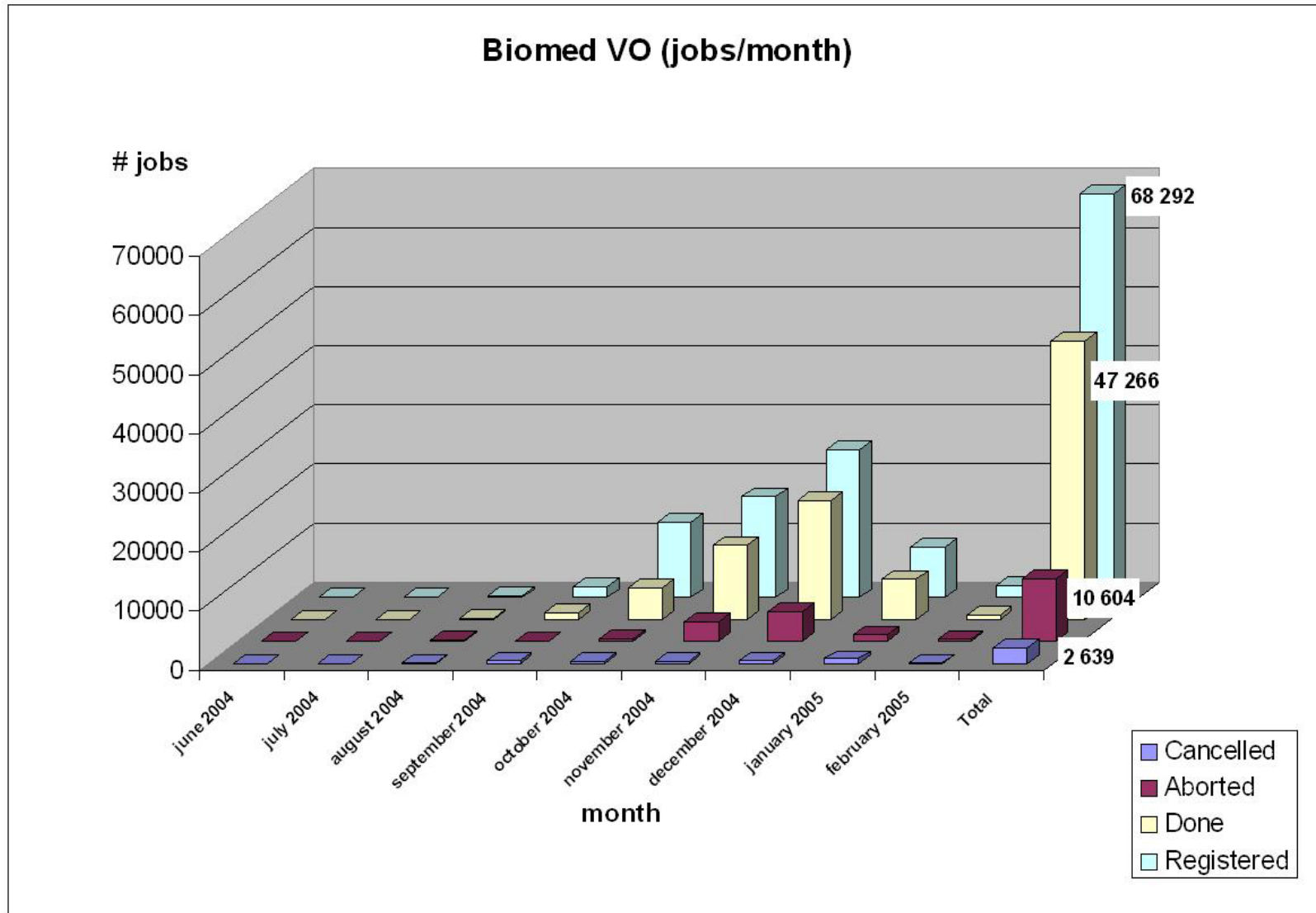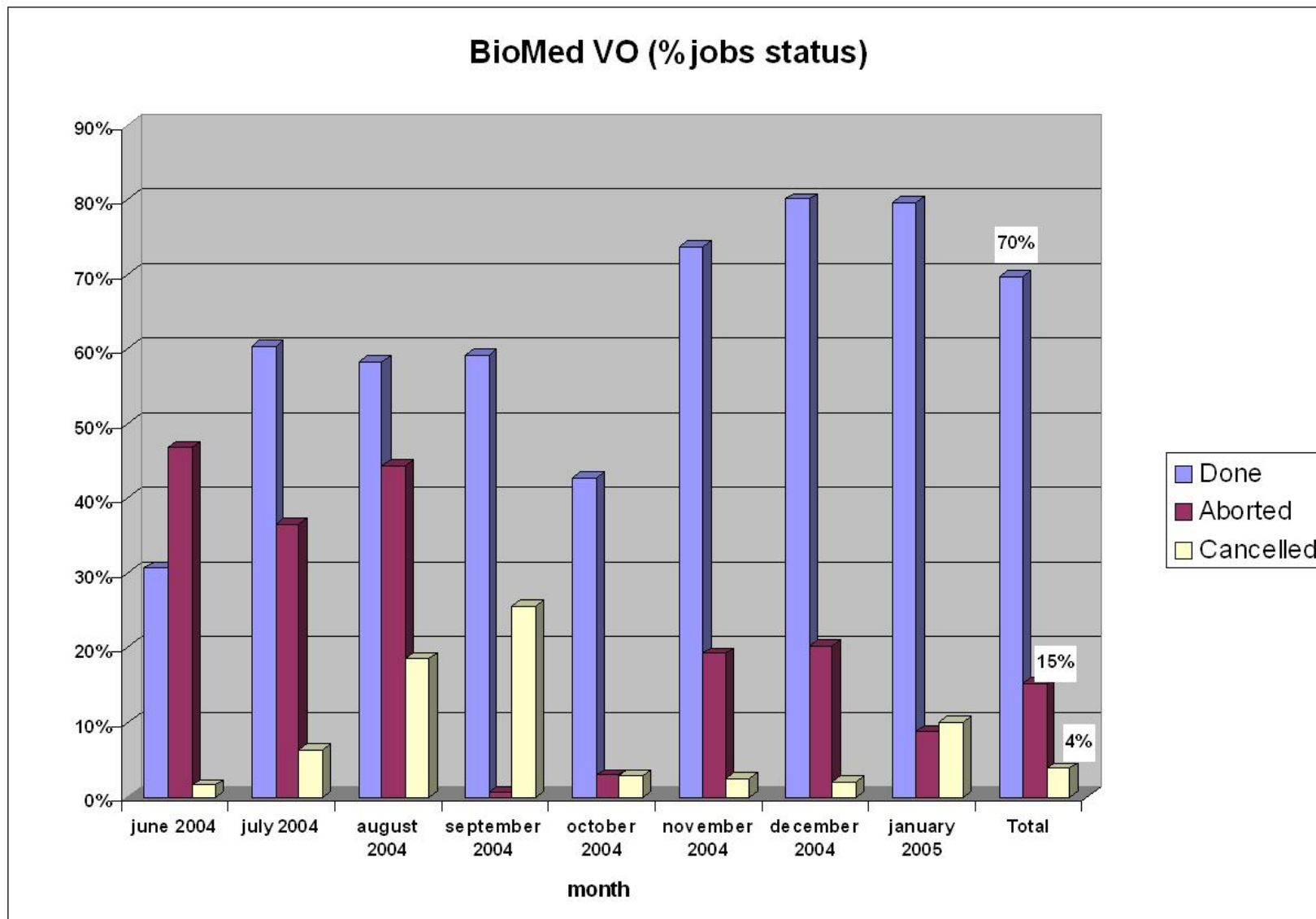
- **Growing interest of the biomedical community**
  - Partners involved proposing new applications
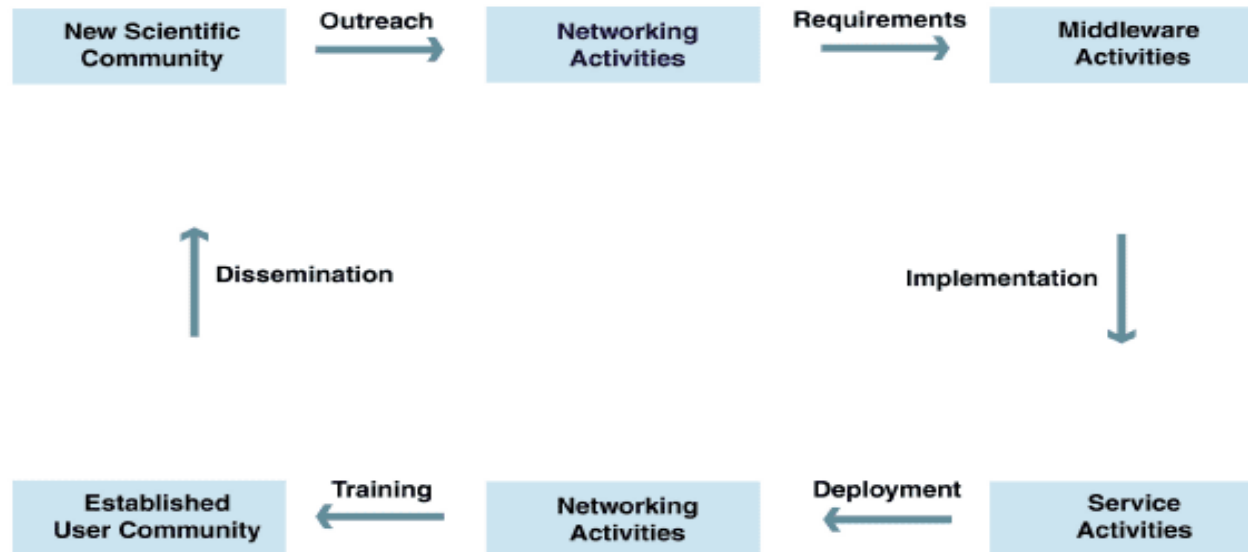 - New application proposals (in various health-related areas)
  - Enlargement of the biomedical community (drug discovery)

- **Growing scale of the applications**
  - Progressive migration from prototypes to pre-production services for some applications
  - Increase in scale (volume of data and number of CPU hours)

1. *Outreach events* inform people about the grid / EGEE
2. Application experts discuss *specific characteristics* with the users
3. *Migrate application* to EGEE infrastructure with the help of EGEE experts
4. *Initial deployment* for testing purposes (using GILDA testbed)
5. Production usage - user community contributes computing resources for heavy production demands - "*Canadian dinner party*"

How new communities join EGEE

1. **New user community establishes contact with EGEE application group for initial discussions**
   - http://public.eu-egee.org/join/

2. **Clarifies needs and characteristics of application via a questionnaire**

3. **Prepares submission to EGAAP (EGEE Generic Applications Advisory Panel) that makes recommendations taking into account**
   - Scientific interest of the proposed work and the grid added-value
   - Coordination and grid-awareness of the community
   - Agreement to the various EGEE policies and especially the security and resources allocation policies

4. **Community and EGEE plan in greater detail the work to be performed**
   - Establishes a Memorandum of Understanding (MoU) signed by the community representatives and EGEE management formalising the engagements of each party

5. **Progress of work is monitored regularly by the project**
   - Training
   - Porting of application (to GILDA, private infrastructure or production infrastructure)
   - Support for creation of virtual organisation and access to resources
   - Results achieved

Enabling Grids for E-sciencE

- **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

  - Same approach for new middleware (gLite)

- **Application software maintains its own licensing scheme**

  - Sites must obtain appropriate licenses before installation

  - EGEE will investigate policies for managing commercially licensed software

For applications that must operate in a closed environment, EGEE middleware can be downloaded and installed on closed infrastructures
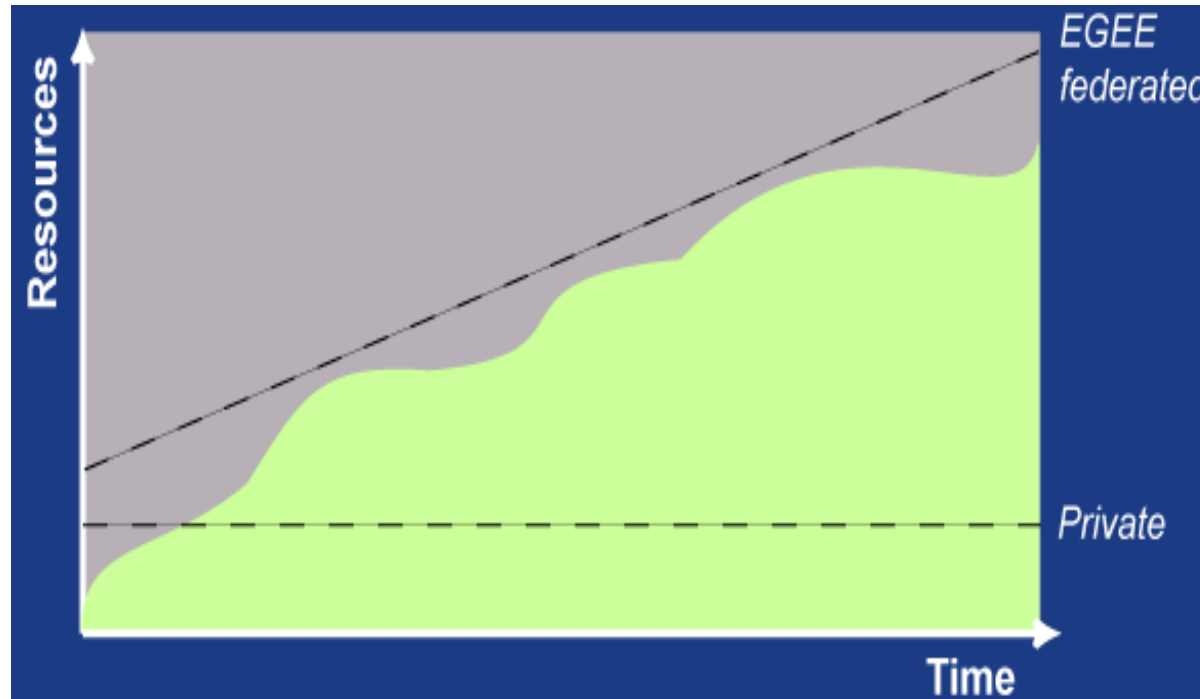
EGEE sites are administered/owned by different organisations

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

Enabling Grids for E-sciencE

- **e-Infrastructures deployment creates a powerful new tool for science – as well as applications from other fields**

- **Investments in grid projects and e-Infrastructure are growing world-wide**

- **Applications are already benefiting from Grid technologies**

- **Open Source is the right approach for publicly funded projects and necessary for fast and wide adoption**

- **Europe is strong in the development of e-Infrastructure also thanks to the initial success of EGEE**

- **Collaboration across national and international programmes is very important**

Enabling Grids for E-sciencE

- EGEE **is the first attempt to build a worldwide Grid infrastructure for data intensive applications from** many scientific domains

- **A** large-scale production grid service **is already deployed and being used for HEP and BioMed applications with new applications being ported**

- **Resources & user groups are** expanding

- **A process is in place for** migrating new applications **to the EGEE infrastructure**

- **A** training programme **has started with many events already held**

- "*next generation*" **middleware is being tested (**gLite**)**

- **First project review by the EU successfully passed in Feb'05**

- **Plans for a** follow-on project **are being prepared**

Enabling Grids for E-sciencE

- Identifier: **FP6-2004-Infrastructures-6**

- Research Infrastructures "**eInfrastructure: Grid Initiatives**"

- Publication date: **May 11, 2005**

- Closing date: **Sept. 7, 2005**

- Indicative Budget: **€ 55 Million**

European Commission

- **Task force appointed by EGEE Project Management Board (PMB)**
  - Fabrizio Gagliardi (Project Director), Dieter Kranzlmueller (TF Meeting Chair), Anna Cook (Project Secretary), Fotis Karayannis, Christian Saguez, Neil Geddes, Olof Barring, Giorgio Maggi, Klaus Ullmann, Anders Ynnerman
  - **The task force has dealt with generalities and did not produce a written description of the proposal**

- Defined the boundary conditions for EGEE 2
  - **Consortium:** very similar to EGEE (approx 70 partners)
  - **Finance:** similar budget (€ 32M from EU + contributions from partners)
  - **Duration:** 2 years starting 1st April 2006
    - No extension to EGEE phase 1
    - Personnel contracts must be extended/renewed beforehand to ensure staff retention

From the 1st EU Review:

- Additional activities could be spun off to other related projects

- EGEE should support other projects which use the EGEE infrastructure ➔ EGEE as an **incubator**

- The actual scope of the proposal will be influenced by approval of the recently submitted related proposals (EU Calls March: 16 proposals related to EGEE) and possible future proposals (EU Calls September)

Enabling Grids for E-sciencE

- A continuation of the project's first phase with mostly the same partners and necessary minor adjustments on the basis of the input from the activities and especially taking into account the recommendations and comments from the 1st EU review (February 2005)

- Emphasis on providing an infrastructure for eScience

- Increased support for applications
  - **Maintain role of pilot applications (HEP, Biomedical)**
  - **Address needs of a wide user base**

Enabling Grids for E-sciencE

- **EGEE Website**

  http://www.eu-egee.org

- **How to join**

  http://public.eu-egee.org/join/

- **EGEE Project Office**

  project-eu-egee-po@cern.ch