



Enabling Grids for E-science

EGEE – application support and identification

Vincent Breton, on behalf of NA4

Application identification and support manager

CNRS, France

Second EGEE review

December 6th 2005



www.eu-egee.org

- **The talk is organised around the recommendations made for NA4 at the last review :**
 - 1. *Have all current applications migrated to gLite*
 - 2. *Building on the experience of previous FP5 projects, capture the full requirements of future user groups*
 - 3. *Clarify the true motivation of users from new application areas from the beginning of their relationship with EGEE*
- **Other issues addressed**
 - 4. *The evaluation of user satisfaction*
 - 5. *The demonstration of the scientific benefit of the EGEE grid*

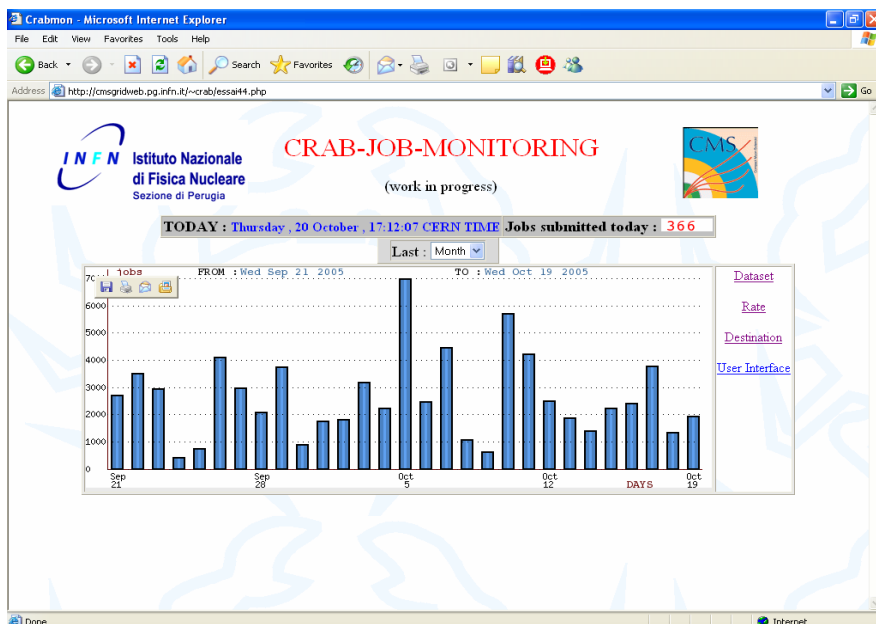
1. The status of the applications migration to gLite

- **gLite offers new functionalities for user communities**
- **The migration of several existing applications to gLite has been achieved (see demos)**
 - *HEP Data Analysis prototypes (experiments + NA4/ARDA)*
 - *Several biomed applications (Pharmacokinetics, CDSS, ...)*
- **gLite was available for early deployment on the GILDA infrastructure**
- **Using GILDA, gLite tutorials have been provided for several hundred users**

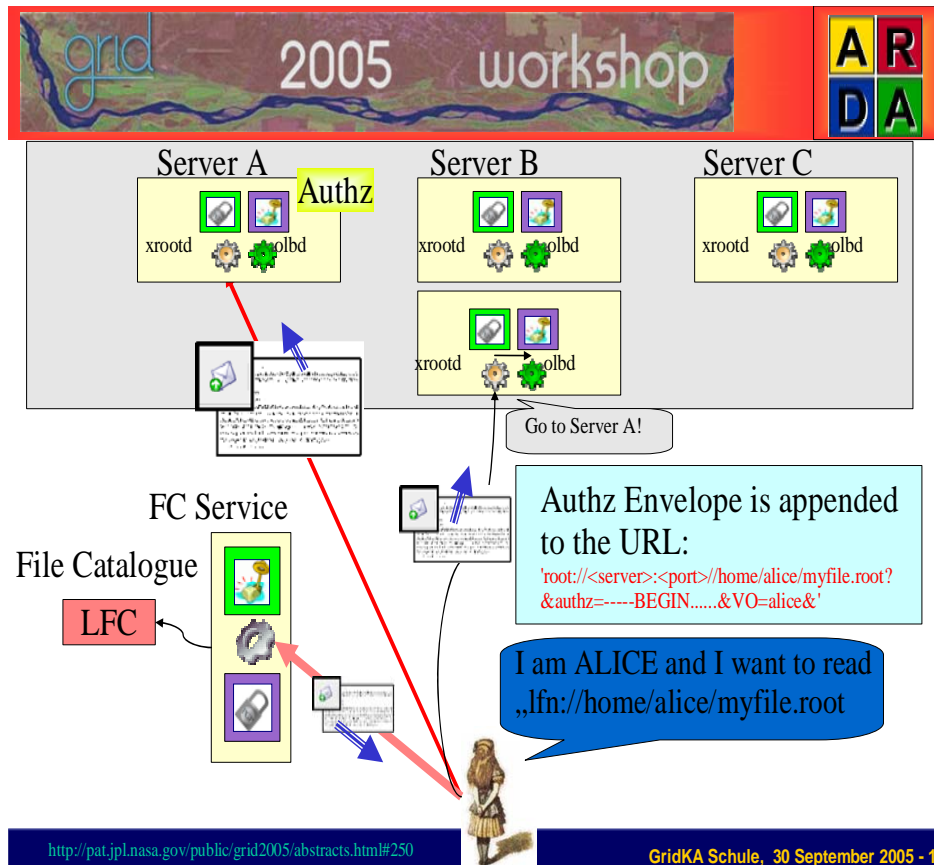
- **Experiment Task Forces**

- Developments have been based on task forces composed of NA4/ARDA and experiment people. Each experiment task force has concentrated on migrating key physics analysis system components into gLite
- There has been synergy between the 4 experiment task forces, and non-HEP applications have also benefited from these focussed efforts (e.g. see work on metadata processing)
- We show now some examples of key testing and migration accomplishments.....

- NA4/ARDA prototype → converging on the CMS CRAB data analysis system
- Service challenge activity, analysis jobs, productions jobs → CMS dashboard
- Substantial user activity
 - ARDA and CRAB
 - ~150k jobs/month
- Evaluations with ATLAS of the performance of bulk submission with gLite WMS
- Studied submission and dispatching rates with parallel threads



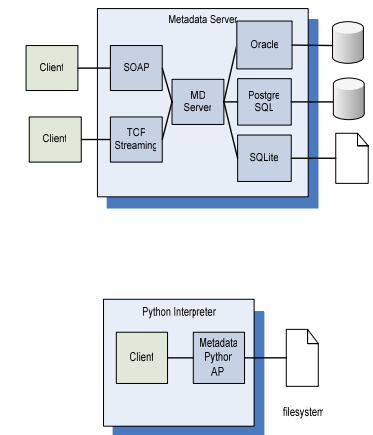
- **Efficient data access integration:** File catalogue ACL and fast data access via xrootd (ALICE prototype). See SuperComputing 2005



- **Metadata Services :** Integrated developments (AMGA) into gLite. **Used also by non-HEP.** Demos by Biomed and “Generic” at Pisa

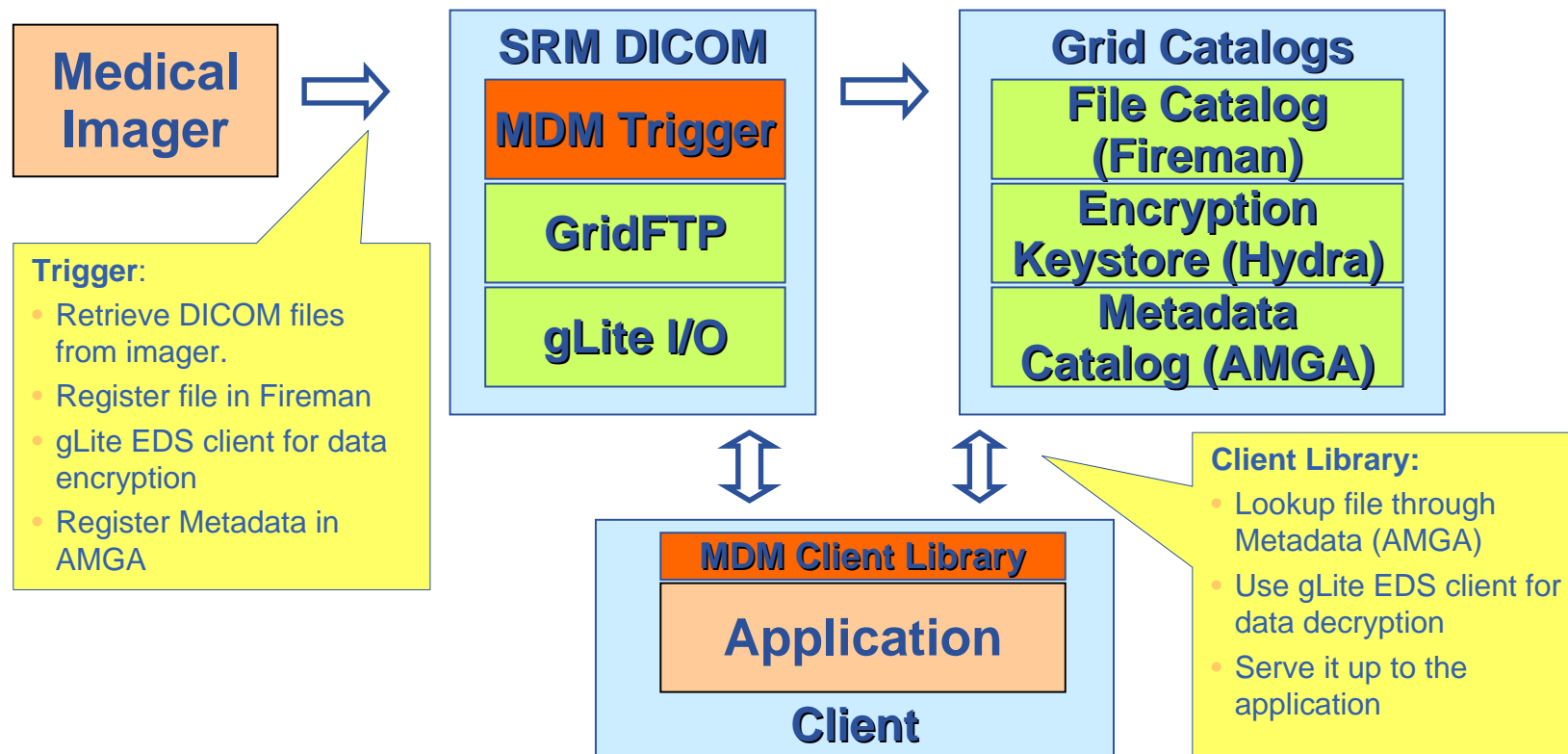


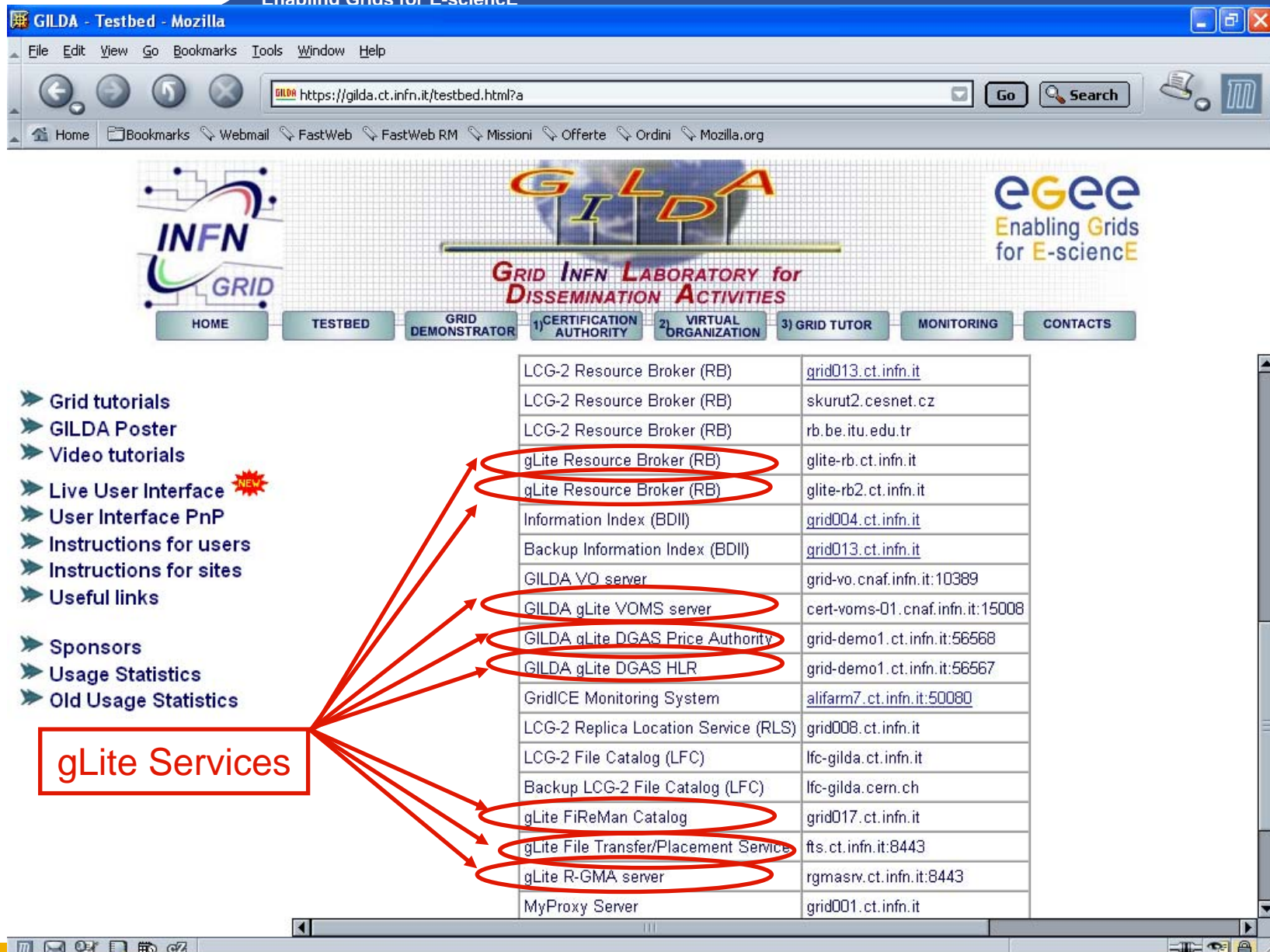
- **Prototype**
 - Validate our ideas and expose a concrete example to interested parties
- **Multiple back ends**
 - Currently: Oracle, PostgreSQL, SQLite, MySQL
- **Dual front ends**
 - TCP Streaming
 - Chosen for performance
 - SOAP
 - Formal requirement of EGEE
 - Compare SOAP with TCP Streaming
- **Also implemented as standalone Python library**
 - Data stored on the file system



- **Improved user access:** GANGA (ATLAS and LHCb activities). Good feedback and demo in Pisa

- **PPS components tested:** Weekly testing of WMS, FiReMan, security related components (ACLs for file access), R-GMA
- **Other gLite testing**
 - Site installation (v1.0, 1.1, 1.2 and 1.3)
 - Advanced data management on prototype testbed





gLite Services

LCG-2 Resource Broker (RB)	grid013.ct.infn.it
LCG-2 Resource Broker (RB)	skurut2.cesnet.cz
LCG-2 Resource Broker (RB)	rb.be.itu.edu.tr
gLite Resource Broker (RB)	glite-rb.ct.infn.it
gLite Resource Broker (RB)	glite-rb2.ct.infn.it
Information Index (BDII)	grid004.ct.infn.it
Backup Information Index (BDII)	grid013.ct.infn.it
GILDA VO server	grid-vo.cnaf.infn.it:10389
GILDA gLite VOMS server	cert-voms-01.cnaf.infn.it:15008
GILDA gLite DGAS Price Authority	grid-demo1.ct.infn.it:56568
GILDA gLite DGAS HLR	grid-demo1.ct.infn.it:56567
GridICE Monitoring System	alifarm7.ct.infn.it:50080
LCG-2 Replica Location Service (RLS)	grid008.ct.infn.it
LCG-2 File Catalog (LFC)	lfc-gilda.ct.infn.it
Backup LCG-2 File Catalog (LFC)	lfc-gilda.cern.ch
gLite FiReMan Catalog	grid017.ct.infn.it
gLite File Transfer/Placement Service	fts.ct.infn.it:8443
gLite R-GMA server	rgmasrv.ct.infn.it:8443
MyProxy Server	grid001.ct.infn.it

2. The capture of requirements, and the assessment of middleware

- **The status at PM18**
 - A database of requirements has been constructed by the Project Technical Forum(PTF)
 - This database already contains >400 requirements
 - This database is populated by the scientific communities deploying applications on EGEE and by the FP6 projects (eg Diligent, Seegrid,...)

- **The capture of requirements from FP5 projects**
 - Several FP5 projects were granted access to the database (*Grace, Mammogrid*)
 - But it has proved to be difficult to collect their requirements
 - *No formalization of requirements*
 - *Projects finished*
 - Requirements were collected mainly from the scientific communities (Crossgrid, Datagrid)

- **The assessment of middleware w.r.t requirements**
 - There is on-going work with JRA1 within the framework of the PTF and the TCG

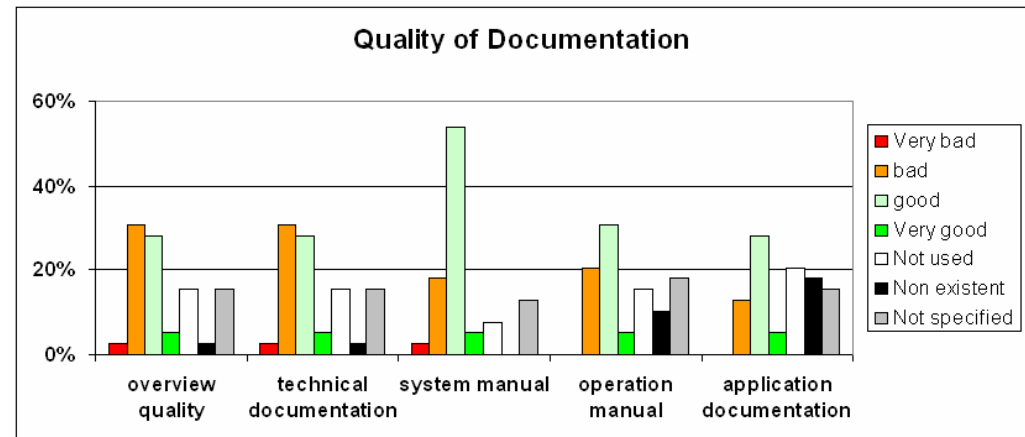
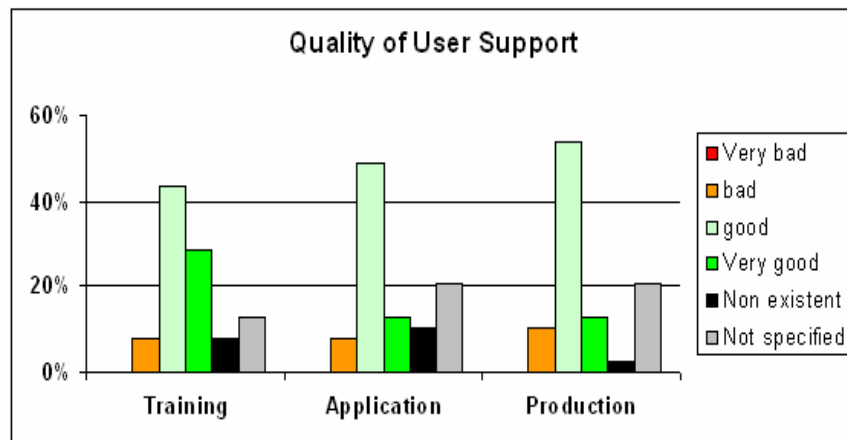
- An improved Memorandum of Understanding(MOU) between EGEE and the scientific communities has been designed and implemented
 - *Detailed SA1 questionnaire to evaluate the required resources*
 - *Regular phone conferences with the Operation Advisory Group*
- An improved application integration process is under design
 - *Creation of lightweight ‘recognised’ VOs vs ‘supported’ Vos*
 - *Only ‘supported’ VOs will be asked to write an MOU*
 - *Decentralised integration vs centralised EGAAP-like approach*
 - *Initial deployment will be at regional level*

The status of the MoUs with the generic supported applications

Actions	Comp Chem	Magic	Planck	Drug Discovery	Egeode	ESR	Egrid	Fusion
MoU filled by applications	OK	OK	OK	OK	Waiting feedback	OK	Just started	Just started
SA1 questionnaire	OK	OK	OK	OK	OK	OK		
SA1	OK	OK	OK	OK	Pending	OK		
SA2	OK	OK	OK	OK	Pending	OK		
NA3	OK	OK	OK	OK	Pending	OK		

- **Applications accepted by the Project Executive Board(PEB)**
 - Earth Science Research (Earth Observation, Hydrology, Climate)
 - Geophysics (Industry)
 - Computational Chemistry
 - Astrophysics (MAGIC and Planck collaborations)
 - Finance (EGRID)
- **New Applications recommended to the PEB by the EGEE Generic Applications Advisory Panel (EGAAP)**
 - Fusion (ITER)
 - Archaeology
 - EC projects (EELA, EUMEDGRID, EUCHINAGRID, BIOINFOGRID)

- **First user survey at PM15**
 - 34 users filled a questionnaire available online
- **Results available in DNA4.3.2**
 - User satisfaction is good for user support
 - User satisfaction on grid services is intermediate
 - Users would like improvements in documentation (with the exception of system manuals which are well appreciated)
- **Action: creation of the User information Group to address issues related to documentation in the project**
- **Next user survey at PM21**



- **The number of users in the VOs related to NA4 activity has continued to grow**
 - *from ~500 at PM9 to ~1000 at PM18*
 - *More than 20 applications are deployed on the production infrastructure*

WE HAVE DEMONSTRATED

- **added value in terms of deployment scale**
 - The very large scale production of simulated data for HEP experiments
 - In particular this has been essential for the commissioning of the LHC experiments
- **added value in terms of response time**
 - *The computation of the seismic epicentre by ESR*
 - *The search for new drugs: large scale in silico docking (WISDOM) - access to large scale resources gives much improved turnaround*
- **added value in terms of data distribution**
 - The distribution of LHC data on the grid, and the large users activity (~150k jobs/month in the CMS case)

- **Fundamental activity in the preparation of the LHC experiments start up**
 - Physics and detector studies
 - Computing systems commissioning

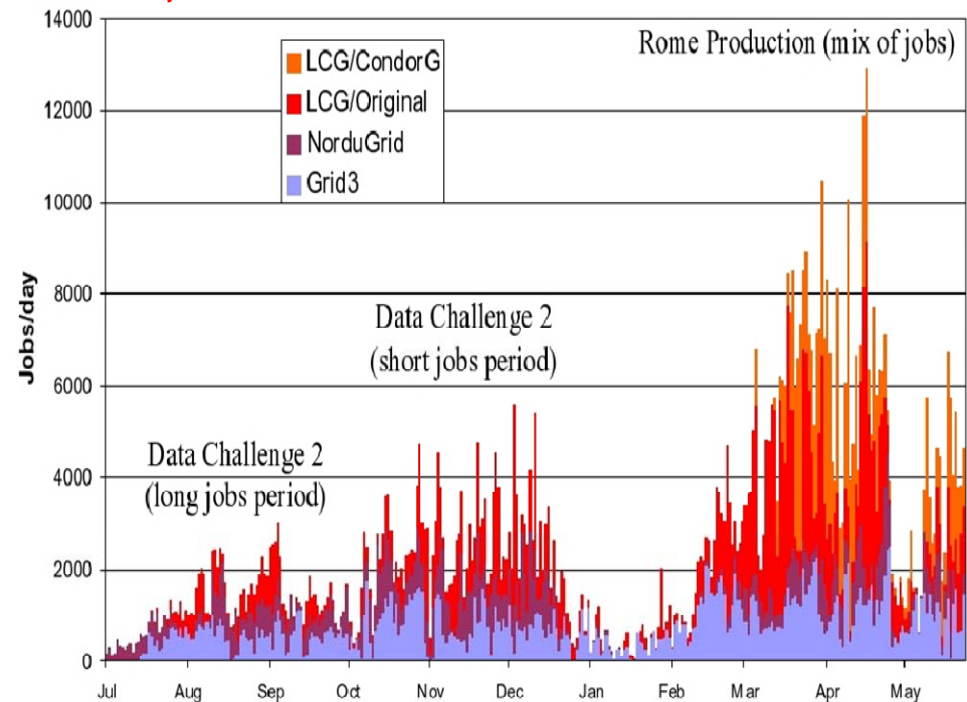
- **Examples:**
 - **ATLAS:**
 - Ran on 3 grids (2/3 of work on EGEE)
 - Peaks of 10,000 jobs per day on EGEE
 - 200 CPU years in 2005 ; 45 TB data
 - Comprehensive analysis: see S.Campana et al., e-Science 2005, Melbourne, Australia

 - **LHCb:**
 - Peaks of 5,500 jobs/day on EGEE
 - 700 CPU/years in 2005 ; 77 TB of data

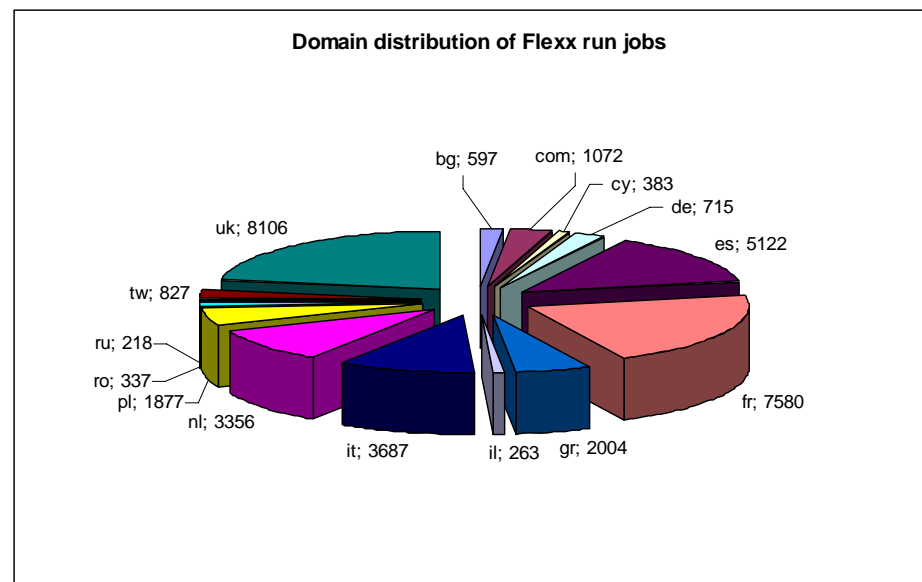
 - A lot of activity within non-LHC experiments (**BaBar,CDF, D, ZEUS,H1**)

 - More details in DNA4.3.2

ATLAS production evolution 2004-2005



- **Significant biological parameters**
 - two different molecular docking applications (Autodock and FlexX)
 - about one million virtual ligands selected
 - target proteins from the parasite responsible for malaria
- **Significant numbers**
 - Total of about 46 million ligands docked in 6 weeks
 - 1TB of data produced
 - Up 1000 computers in 15 countries used simultaneously corresponding to about 80 CPU years
 - Average crunching factor ~600



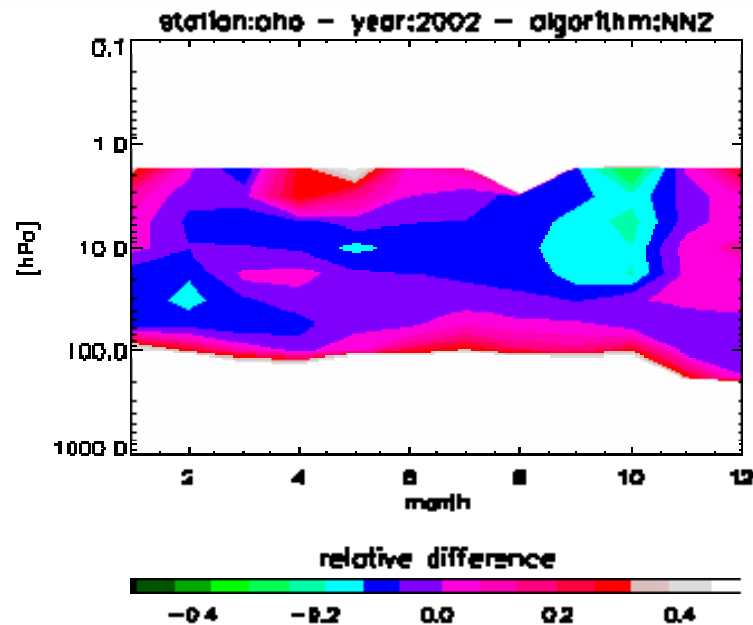
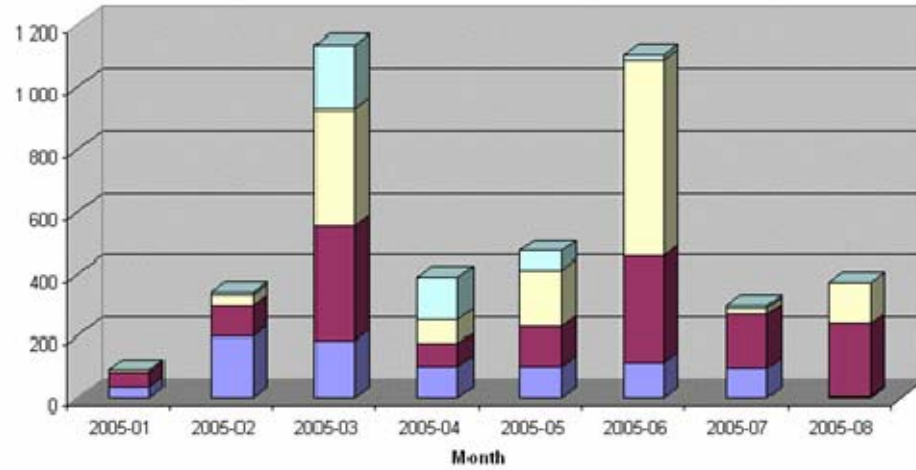
WISDOM open day

December 16th, 2005, Bonn (Germany)

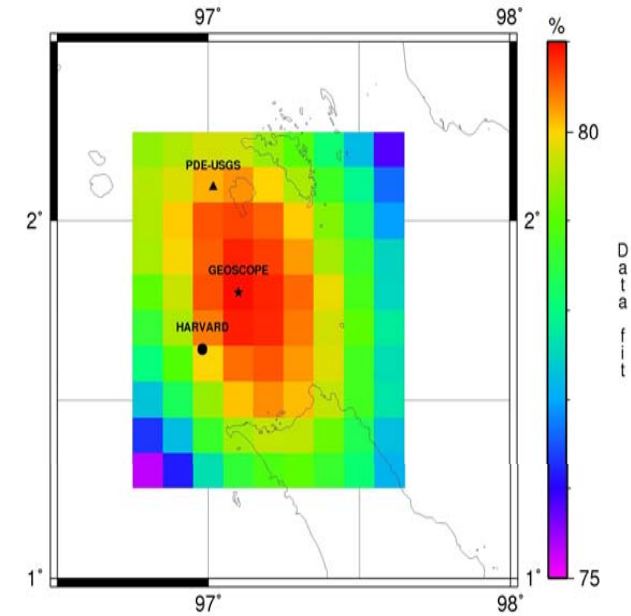
Discuss Data Challenge results
Prepare next steps towards a malaria Grid (EGEE-II, Embrace, Bioinfogrid)
Information: <http://wisdom.eu-egge.fr>

Number of jobs

ESR



Ozone maps



Earthquakes epicentre determination

- **Achievements**

- *The number of users in the VOs related to NA4 activity has doubled*
- *The usage of the grid by the pilot applications has significantly evolved into large scale production*
- *EGAAP applications have also achieved production status on EGEE (ESR,MAGIC)*
- *The migration of several existing applications to gLite has been achieved*
- *A detailed MOU procedure has been defined and executed for several selected applications*
- *The first User Survey has been performed*

- **Issues being actively addressed**

- *Reduction of the amount of skilled effort required by the VOs for production use of the grid (EGEE is improving middleware, monitoring, user interfaces and user support)*
- *The availability, accessibility and quality of user documentation (an inter-activity group is working on this)*
-

- **Looking forwards :** **User Forum, EGEE-II**

- **Dates: March 1-3 2006**
- **Location: CERN, Switzerland**
- **Target attendance: 150 participants**
- **Goals**
 - *Obtain a consistent understanding across the EGEE related projects of their expectations, present status and possible evolution*
 - *Learn how to improve EGEE quality of service*
 - *Promote cross-application fertilisation*
 - *Prepare EGEE-II*
- **Participation open to external projects and EGEE members**
- **Format: 3-day workshop**
 - *Presentations by thematic areas selected by invitation and through a call for contributions*
 - *EGEE presentations (integration of new applications, access to resources, status of middleware,...)*
 - *With a lot of time for discussion*
- **URL:**
 - *<http://egee-intranet.web.cern.ch/egee-intranet/User-Forum/index.html>*