EGEE – applications and training

Vincent Breton, on behalf of NA4

Application identification and support manager

CNRS, France

Meeting with Microsoft, Geneva, October 4th 2005

www.eu-egee.org

Information Society

Content of the talk

- **Status of applications: what happened since Athens**
  - Deployment on the production infrastructure
  - Deployment on the PreProduction Service
  - Other achievements
- **Perspectives**
  - Short term: EU review in December
  - Mid term: User forum March 1-3 2005
  - Long term: EGEE-II
- **Conclusion**

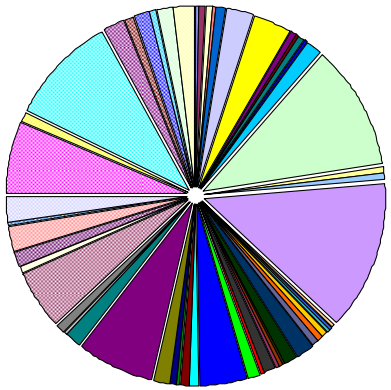
What happened since Athens ?

- **The number of users in VOs related to NA4 activity keept growing regularly**
  - from ~500 at PM9 to ~1000 at PM18
  - More than 20 applications are deployed on the production infrastructure
- **The usage of the grid by pilot applications has significantly evolved during the summer**
  - From data challenge to service challenge (HEP)
  - First biomedical data challenge (WISDOM)
- **Several existing applications have been migrated to the new middleware by the HEP, biomedical and generic teams**

Production

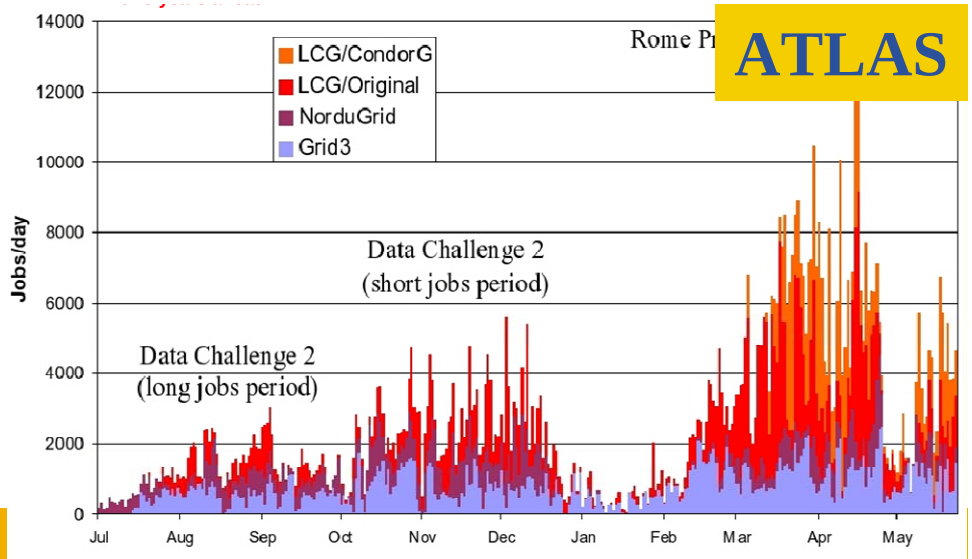
- **Fundamental activity in preparation of LHC start up**
  - Physics
  - Computing systems
- **Examples:**
  - LHCb: ~700 CPU/years in 2005 on the EGEE infrastructure
  - ATLAS: over 10,000 jobs per day
    - Comprehensive analysis: see S.Campana et al., “Analysis of the ATLAS Rome Production experience on the EGEE Computing Grid“, e-Science 2005, Melbourne, Australia
  - A lot of activity in all involved applications (including as usual a lot of activity within non-LHC experiments like BaBar, CDF and D0)
  - A lot more details in DNA4.3.1 (internal review)

CPU used: 6,389,638 h
Data Output: 77 TB



DIRAC.Barcelona.es 0.214%
DIRAC.CERN.ch 0.571%
DIRAC.CracowAgu.pl 0.001%
DIRAC.LHCBONLINE.ch 0.779%
DIRAC.PNPI.ru 0.000%
DIRAC.ScotGrid.uk 3.068%
DIRAC.Zurich.ch 0.756%
LCG.BHAM-HEP.uk 0.705%
LCG.Bari.it 1.357%
LCG.CERN.ch 10.960%
LCG.CGG.fr 0.676%
LCG.CNAF.it 13.196%
LCG.CPPM.fr 0.242%
LCG.CY01.cy 0.103%
LCG.Cambridge.uk 0.010%
LCG.Durham.uk 0.476%
LCG.FZK.de 1.708%
LCG.Firenze.it 1.047%
LCG.GR-02.gr 0.226%
LCG.GR-04.gr 0.056%
LCG.HPC2N.se 0.001%
LCG.IFCA.es 0.022%
LCG.IN2P3.fr 4.143%
LCG.IPP.bg 0.033%
LCG.Imperial.uk 0.891%
LCG.JINR.ru 0.472%
LCG.Lancashire.uk 6.796%
LCG.Manchester.uk 0.285%
LCG.Montreal.ca 0.069%
LCG.NSC.se 0.465%
LCG.Oxford.uk 1.214%
LCG.PNPI.ru 0.278%
LCG.Pisa.it 0.121%
LCG.RAL-HEP.uk 0.938%
LCG.RHUL.uk 2.168%
LCG.Sheffield.uk 0.094%
LCG.Toronto.ca 0.343%
LCG.UCL-CCC.uk 1.455%
LCG.Zurich-spz.ch 0.003%
LCG.ACAD.bg 0.106%
LCG.Barcelona.es 0.281%
LCG.Bologna.it 0.032%
LCG.CESGA.es 0.528%
LCG.CNAF-GRIDIT.it 0.012%
LCG.CNB.es 0.385%
LCG.CSCS.ch 0.282%
LCG.Cagliari.it 0.515%
LCG.Catania.it 0.551%
LCG.Edinburgh.uk 0.031%
LCG.Ferrara.it 0.073%
LCG.GR-01.gr 0.349%
LCG.GR-03.gr 0.171%
LCG.GRNET.gr 1.170%
LCG.ICI.ro 0.088%
LCG.IHEP.su 1.245%
LCG.INTA.es 0.076%
LCG.ITEP.ru 0.792%
LCG.Iowa.us 0.287%
LCG.KFKI.hu 1.436%
LCG.Legnaro.it 1.569%
LCG.Milano.it 0.770%
LCG.NIKHEF.nl 5.140%
LCG.Napoli.it 0.175%
LCG.PIC.es 2.366%
LCG.Padova.it 2.041%
LCG.QMUL.uk 6.407%
LCG.RAL.uk 9.518%
LCG.SARA.nl 0.675%
LCG.Torino.it 1.455%
LCG.Triumf.ca 0.105%
LCG.USC.es 1.853%

LHCb



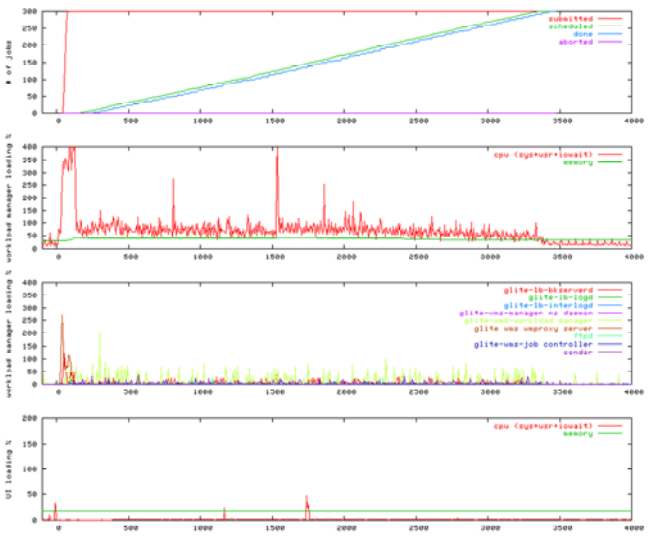
ATLAS

Integration and development

- **Task Forces**
  - EGEE + experiment people. Very close and positive collaboration. Emphasis on integration onto EGEE infrastructure
- **One example (direct contribution of NA4-HEP/ARDA within the ATLAS TF**
  - Detailed studies of advanced gLite feature: WMS bulk submission (ATLAS Task Force)
  - Other middleware tested in the framework of other ARDA protoypes and TFs
  - More under arda.cern.ch

Use of special installation for detailed tests

Enabling Grids for E-sciencE



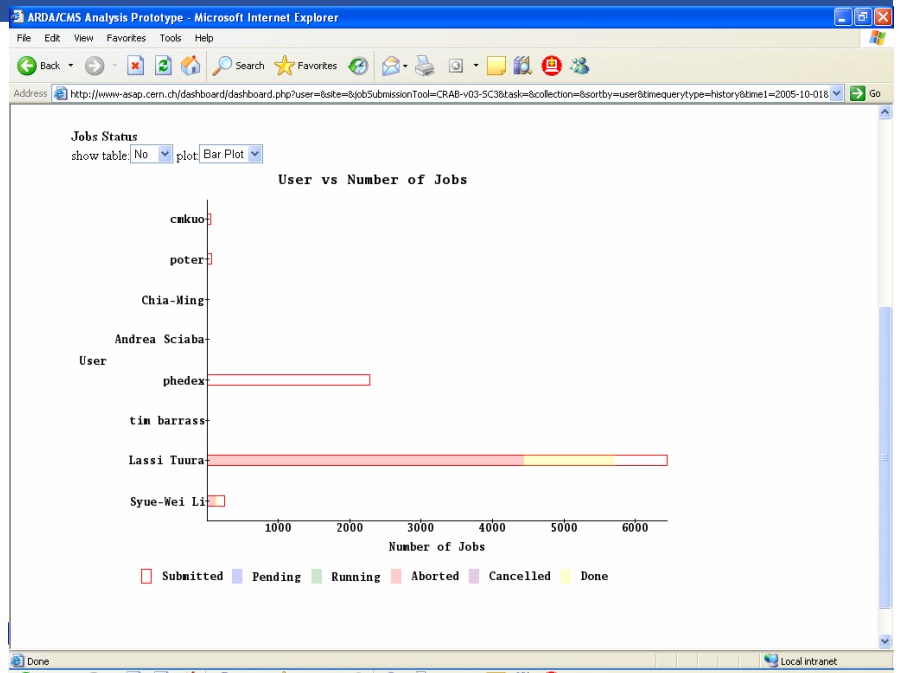
First gLite pre1.4 WMS performance measurements in Milano

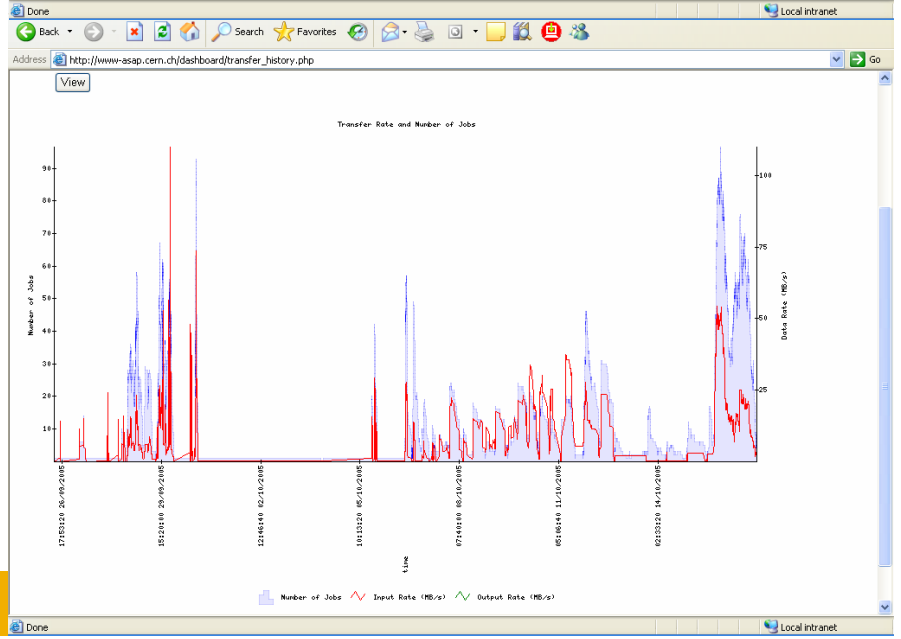
- gLite 1.4 WMS (4CPU, 3GB memory)
- 300 simple hello world jobs submitted by 3 parallel threads
- submission rate ~ 4.1 jobs/sec
- dispatching rate ~ 0.08 jobs/sec
- All jobs in bulk are submitted to the same CE: atlasce.lnf.infn.it

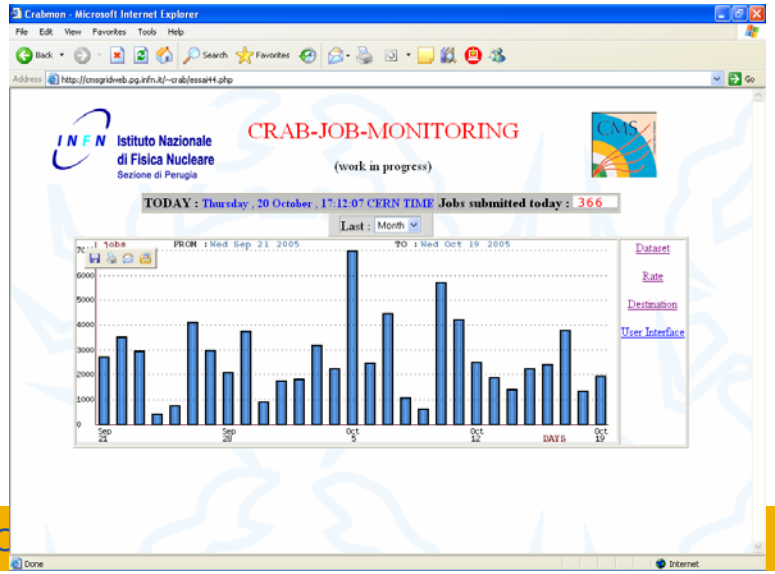
- Thanks to Elisabetta Molinari for setting up the gLite WMS in Milan

Development and integration

- **From prototypes to coherent integration (CMS)**
  - NA4 HEP (ARDA/ASAP) prototype → converging on the CMS CRAB system
  - SC3 activity, analysis jobs, productions jobs → CMS dashboard
  - Clear signs of wide user activity!!!
    - ASAP and CRAB

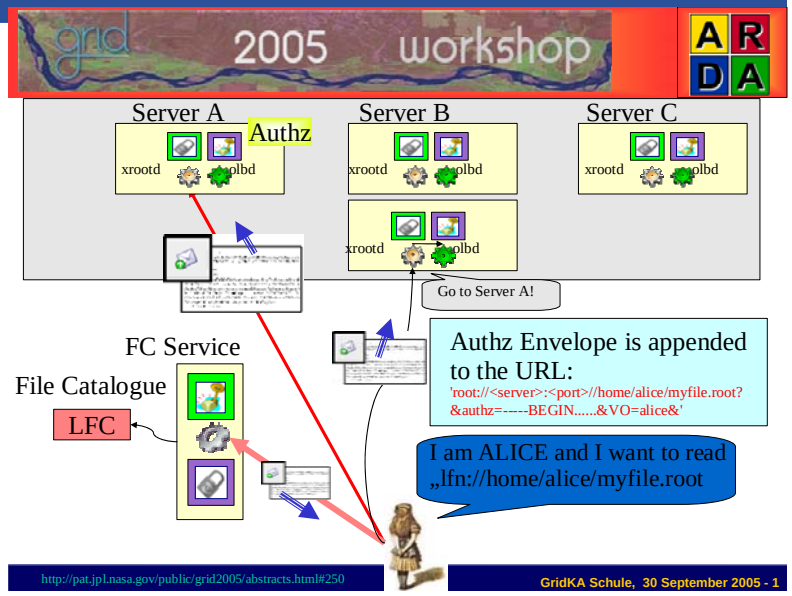


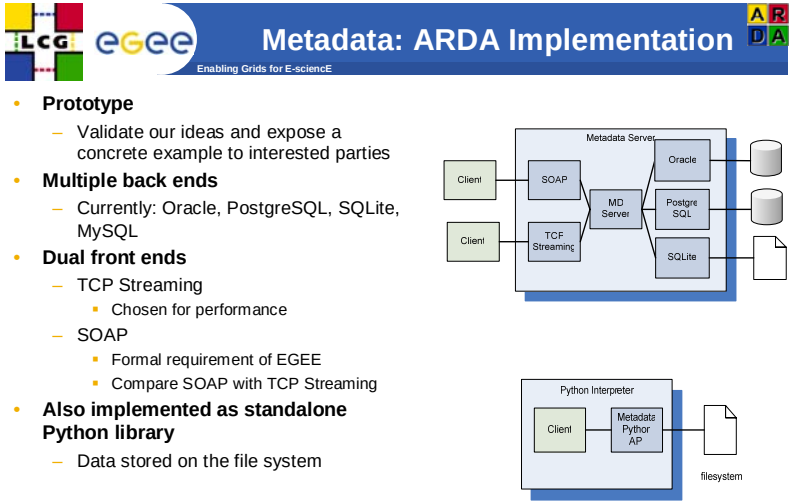




Development and integration

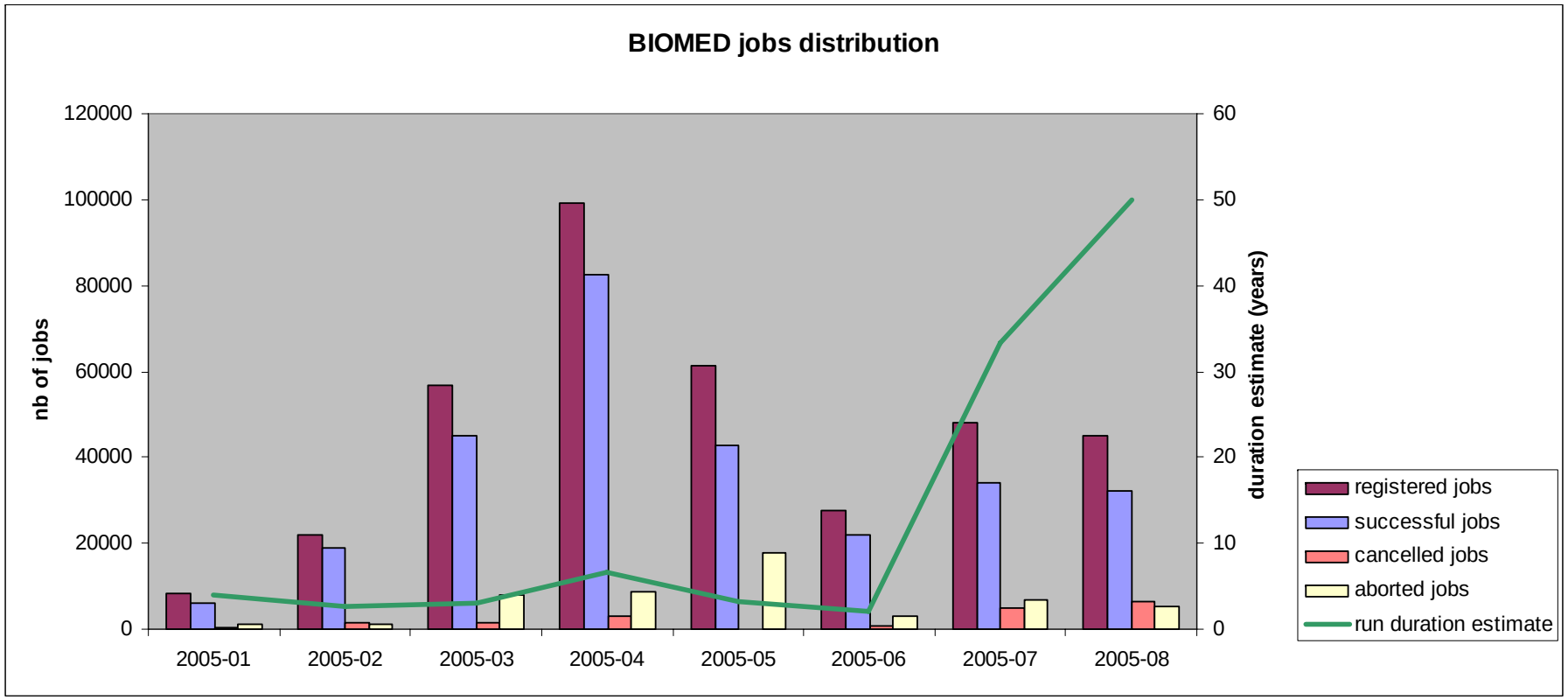
- **Other examples:**
  - Efficient data access integration: File catalogue ACL and fast data access via xrootd (ALICE prototype and Task Force)
  - Improved user access: GANGA (ATLAS and LHCb activities). Public beta is out. Good feedback and demo in Pisa
  - Contribution to services used also outside HEP and contributed to gLite (AMGA)





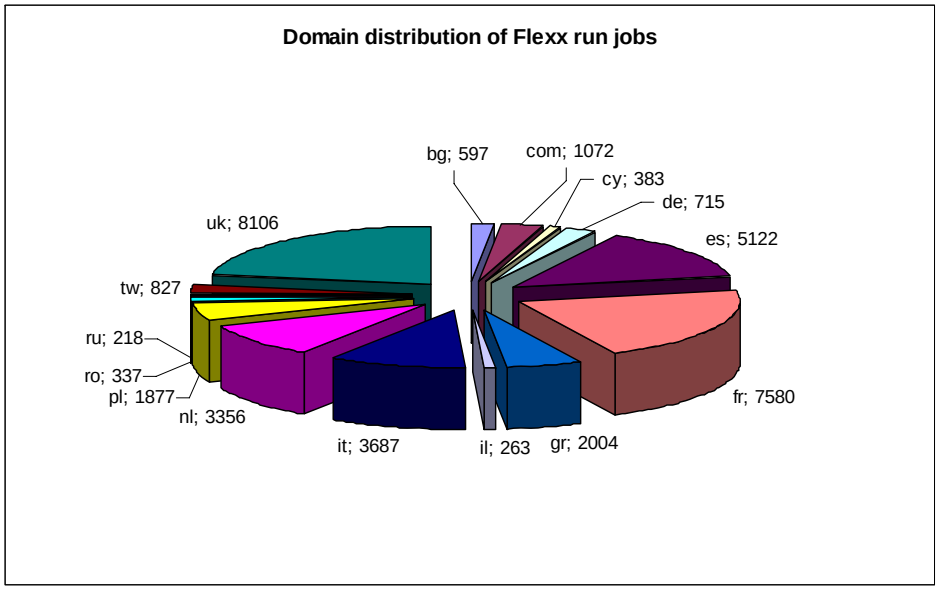
Biomed Virtual Organisation

- ~ 70 users, 9 countries
- > 12 Applications (medical image processing, bioinformatics)
- ~3000 CPUs, ~12 TB disk space
- ~100 CPU years, ~ 500K jobs last 6 months



First biomedical data challenge: World-wide In Silico Docking On Malaria (WISDOM)

- **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)

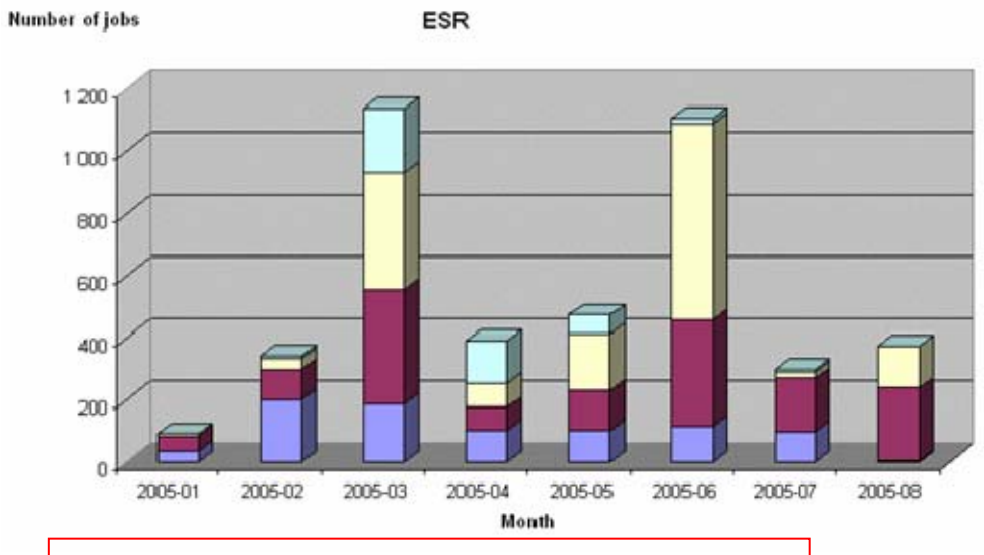
Discuss Data Challenge results

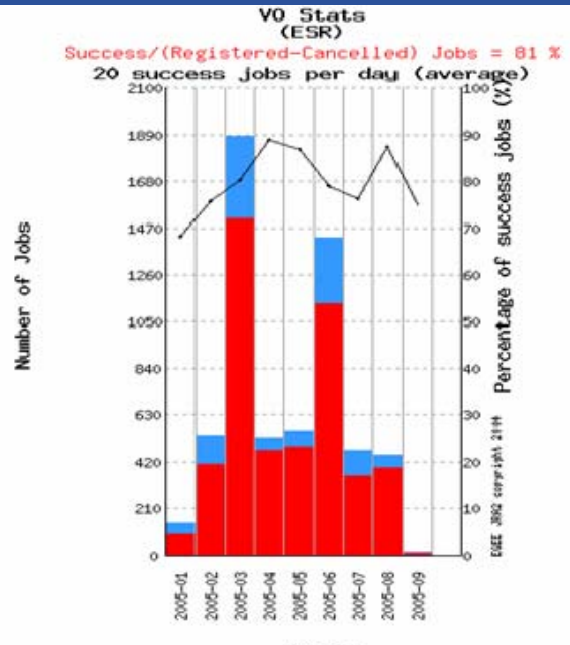
Prepare next steps towards a malaria Grid (EGEE-II, Embrace, Bioinfogrid)

Information: http://wisdom.eu-egee.fr

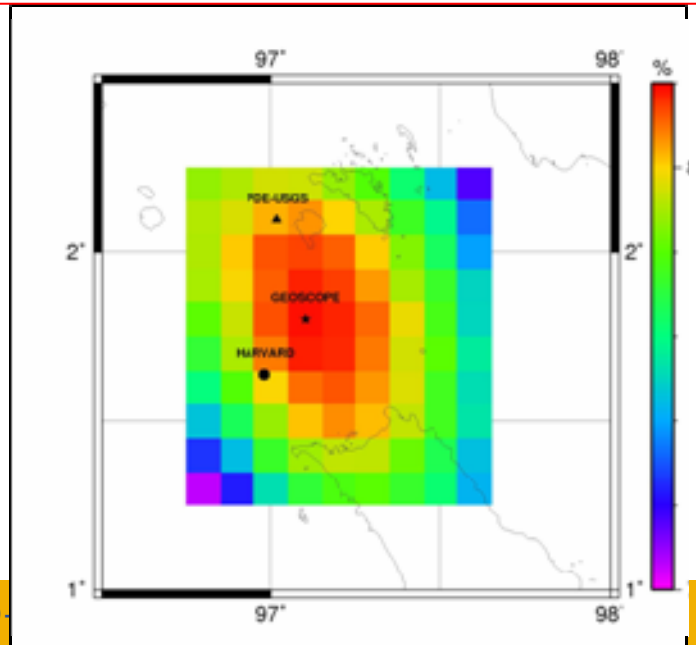
Generic Applications' use of EGEE

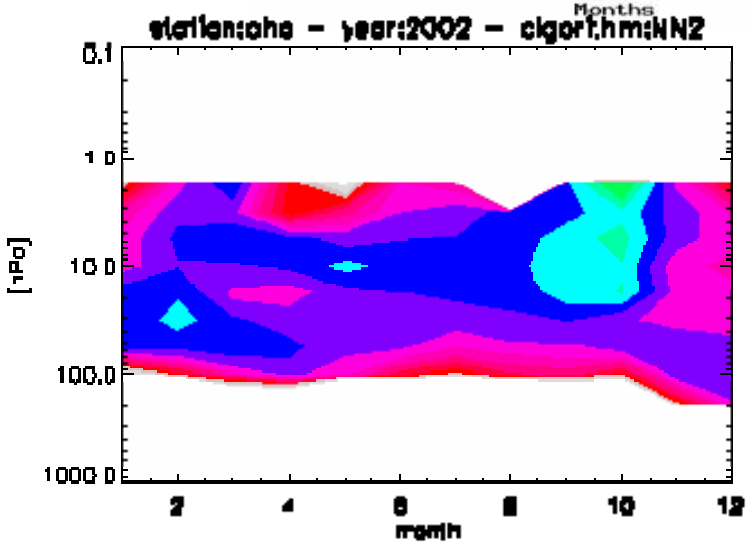
Enabling Grids for E-sciencE





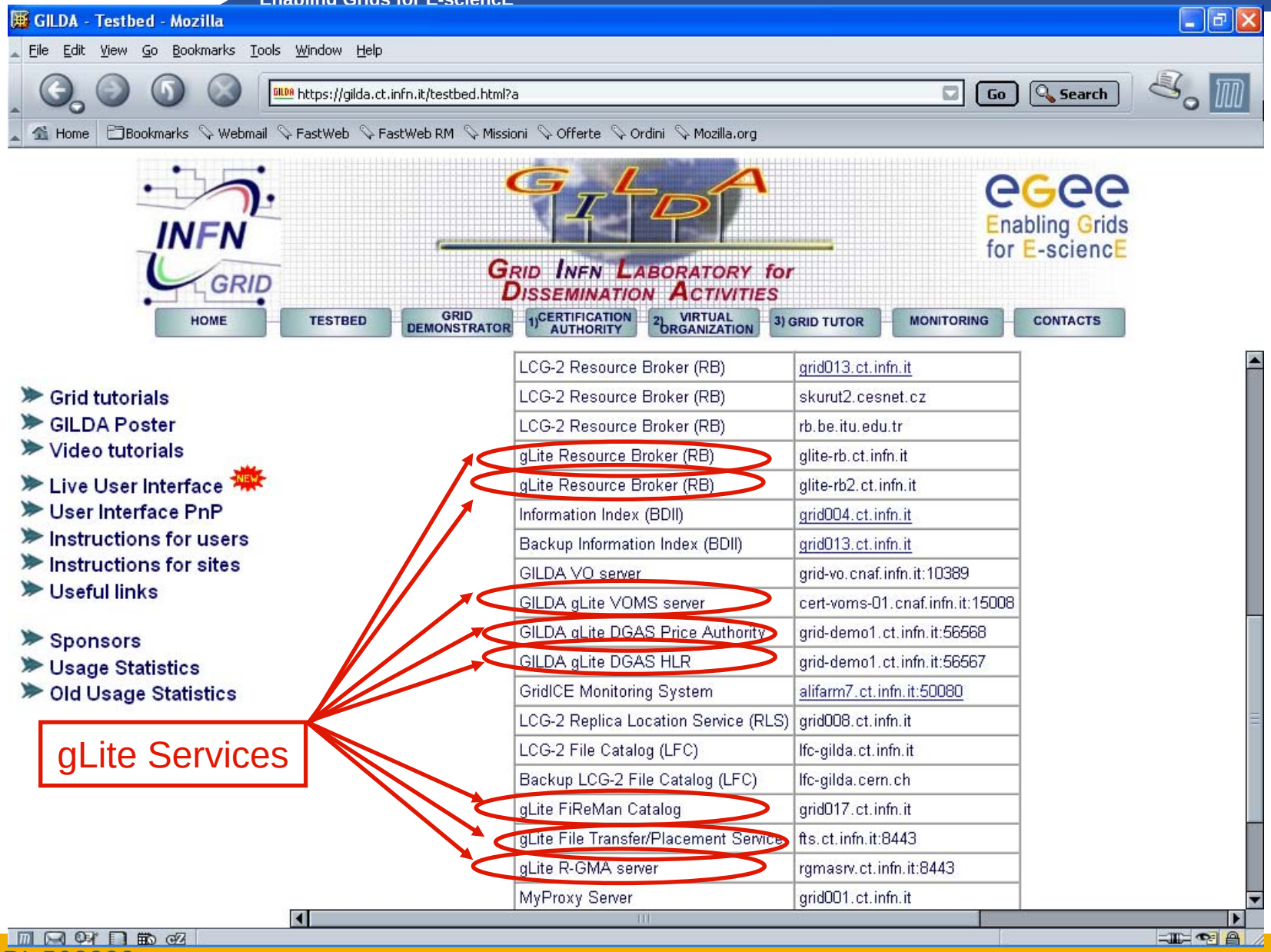
Earthquakes' epicenter determination





Ozone maps

All gLite Services available and tested on GILDA



gLite brought to 100s during EGEE tutorials with GILDA

- First gLite tutorial on GILDA, Catania, 13-15 June 2005
- ESR retreat to move to gLite, Bratislava, 27-30 June 2005
- GGF Grid School, Vico Equense, 10-22 July 2005
- EGEE Summer School, Budapest, 11-16 July 2005
- Healthgrid Workshop, Clermont-Ferrand, 25-27 July 2005
- UK grid event, Swansea, 6 August 2005
- EGEE Tutorial, Taipei, 22-23 August 2005
- EGEE tutorial for summer students, CERN, 24 August 2005
- EGEE Tutorial, Tokyo, 25-26 August 2005
- EGEE Tutorial, Seoul, 29-30 August 2005
- EGEE Tutorial for the Crossgrid Project, Lausanne, 5-9 September 2005
- Cern School of Computing, Saint-Malo, 12-15 September 2005
- GridKa School for Grid Application Developers, Karlsruhe, 26-30 September 2005
- MAGIC Coll. Meeting, Tenerife, 16 October 2005
- EGEE-4 Training day, Pisa, 23 October 2005

Other achievements:MoUs almost finalized

Actions	Comp Chem	Magic	Planck	Drug Discovery	Egeode	ESR
MoU filled by applications	OK	OK	OK	OK	Waiting feedback	OK
SA1 questionnaire	OK	OK	OK	Waiting feedback	OK	OK
SA1	OK	OK	OK	OK – Biomed VO resources	-	OK
SA2	OK	OK	OK	OK	-	OK
NA3	OK	OK	OK	OK	-	Waiting feedback

Special thanks to Rolf Rumler (SA1-OAG)

Short term perspectives

- **NA4 related sessions during this conference**
  - Demo session Wednesday (16 demos) and posters all week long
  - EGEE Generic Applications Advisory Panel meeting Thursday afternoon
  - NA2-NA3-NA4-NA5 joint session Tuesday afternoon to discuss relationship between EGEE and other EU-funded projects
- **Approval of deliverable DNA4.3.2 due at PM18**
  - Very positive feedback from reviewers
- **EU-review beginning of December**

EGEE first User Forum

- **Dates: March 1-3 2006**
- **Location: CERN, Switzerland**
- **Target attendance: 150 participants**
- **Goals**
  - Get a consistent understanding across the EGEE related projects of expectation, present status and possible evolution
  - Promote cross-application fertilisation
  - Prepare EGEE2
- **Participation open to external projects and EGEE members**
- **Format: 3-days workshop**
  - Presentations by thematic areas selected by invitation and through a call for contributions
  - EGEE presentations
  - With a lot of time for discussion

NA4 in EGEE-II

- Ongoing support to 2 pilot areas (High Energy Physics, Biomedicine) with established user communities and large scale applications
- On-going support to GILDA, the virtual grid laboratory
- Increased support to 4 scientific disciplines (Astrophysics, Computational Chemistry, Earth Sciences, Fusion), 3 of them coming out of EGEE generic applications
- Increased effort to foster a cohesive application community at NA4 management level
- Increased complexity from the wider regional distribution and the associated projects

Task	Goal	Resource allocation in PM	Fraction
TNA4.1	Support teams to help with consulting and porting of applications	1005	51%
TNA4.2	Evolution of pilot applications with evolving middleware services	215	11%
TNA4.3	Validation of the efficient production and continuous availability of data for the scientific communities	282	14%
TNA4.4	Grid laboratory to attract and expose new applications to grid computing	192	10%
TNA4.5	Management of the large and diverse set of application and institutes involved in NA4	269	14%

NA4 achievements and issues

- **Achievements**
  - The number of users in VOs related to NA4 activity has doubled
  - The usage of the grid by pilot applications has significantly evolved
  - The migration of several existing applications to gLite is achieved

- **Issues**
  - The migration to gLite
  - The availability of gLite software on the pre-production and production services;
  - The amount of effort required by the VOs in order to obtain useful results from the use of the grid when compared to alternatives such as clusters
  - The availability, accessibility and quality of user documentation

- **Perspectives: EC review, User Forum, EGEE-II**