egee

EGEE – application support and identification

Vincent Breton, NA4 activity manager, CNRS

EGEE 2nd EU Review

6-7/12/05

www.eu-egee.org

Information Society
and Media

Enabling Grids for E-sciencE

- **Status of application migration to gLite**

- **Capture of requirements and integration process**

- **Evaluation of user satisfaction**

- **Demonstration of the scientific benefit of the EGEE grid**

- **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, …)*
  - *Earth Science (Simulation of flooding)*

- **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………………

Examples of HEP gLite migration

- **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 (ARDA/ATLAS TF)**
- **Very nice and effective collaboration with JRA1 and experiments' people:**
  - x10 (Bulk submission)
  - x4 in matchmaking (improvements in the code)
- **Similar activities on other gLIte components (catalogues, R-GMA)**
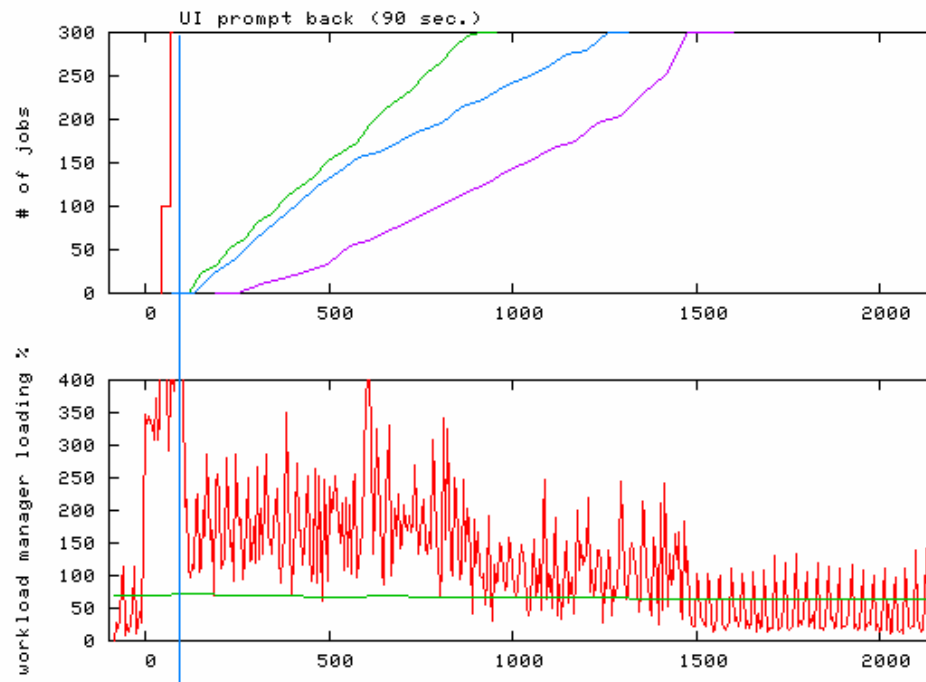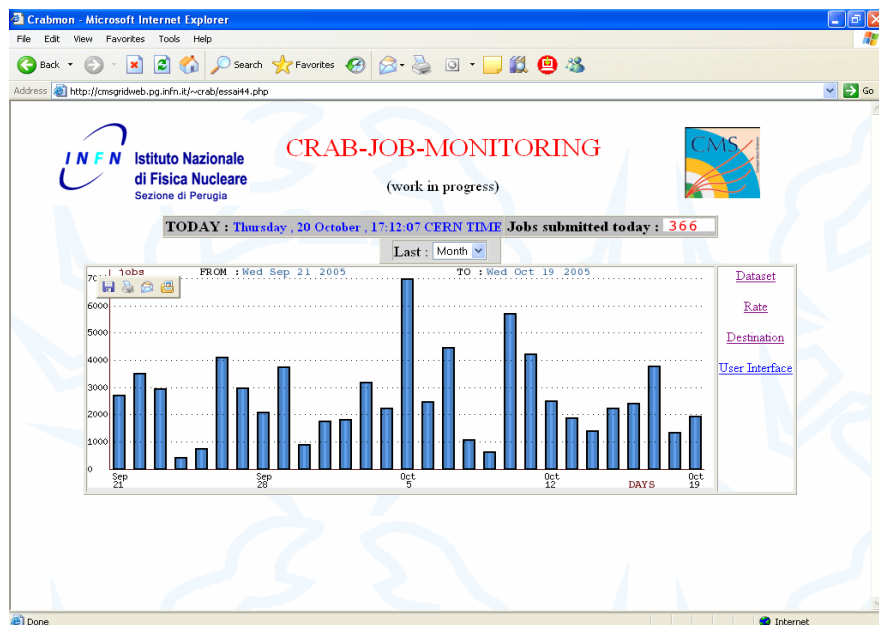
– **Efficient data access integration**: File catalogue ACL and fast data access via xrootd (ALICE prototype). See SuperComputing 2005
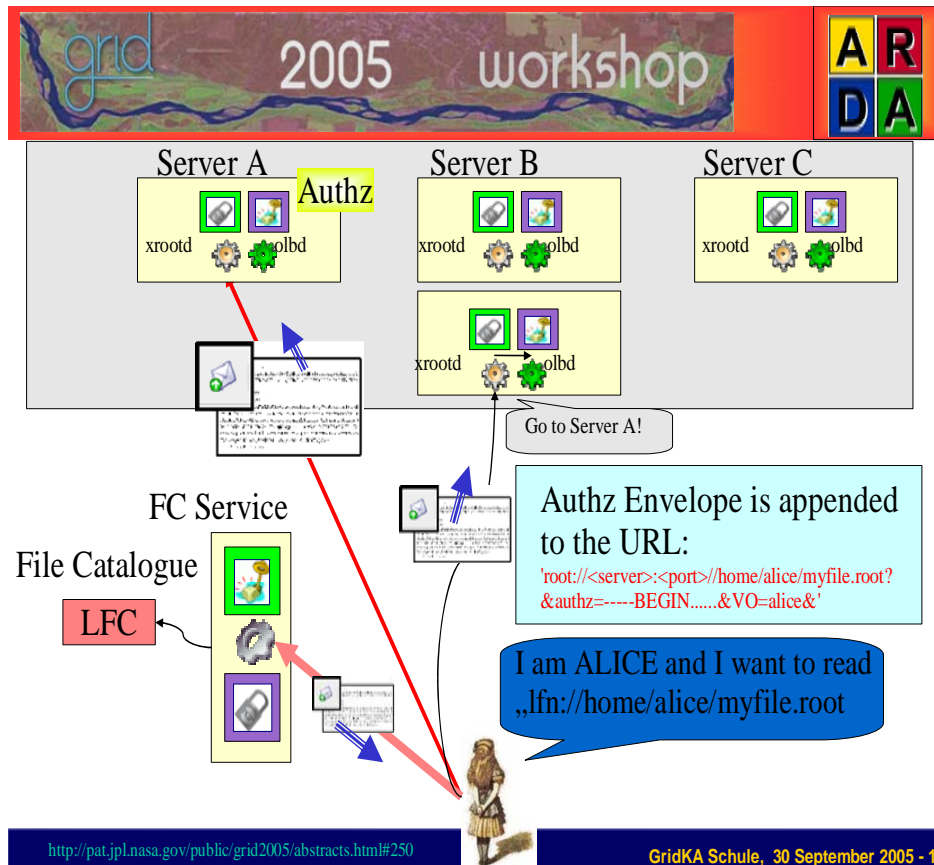
– **Metadata Services :** Integrated developments (AMGA) into gLite. Used also by non-HEP. Demos by Biomed and "Generic" at Pisa

Metadata: ARDA Implementation
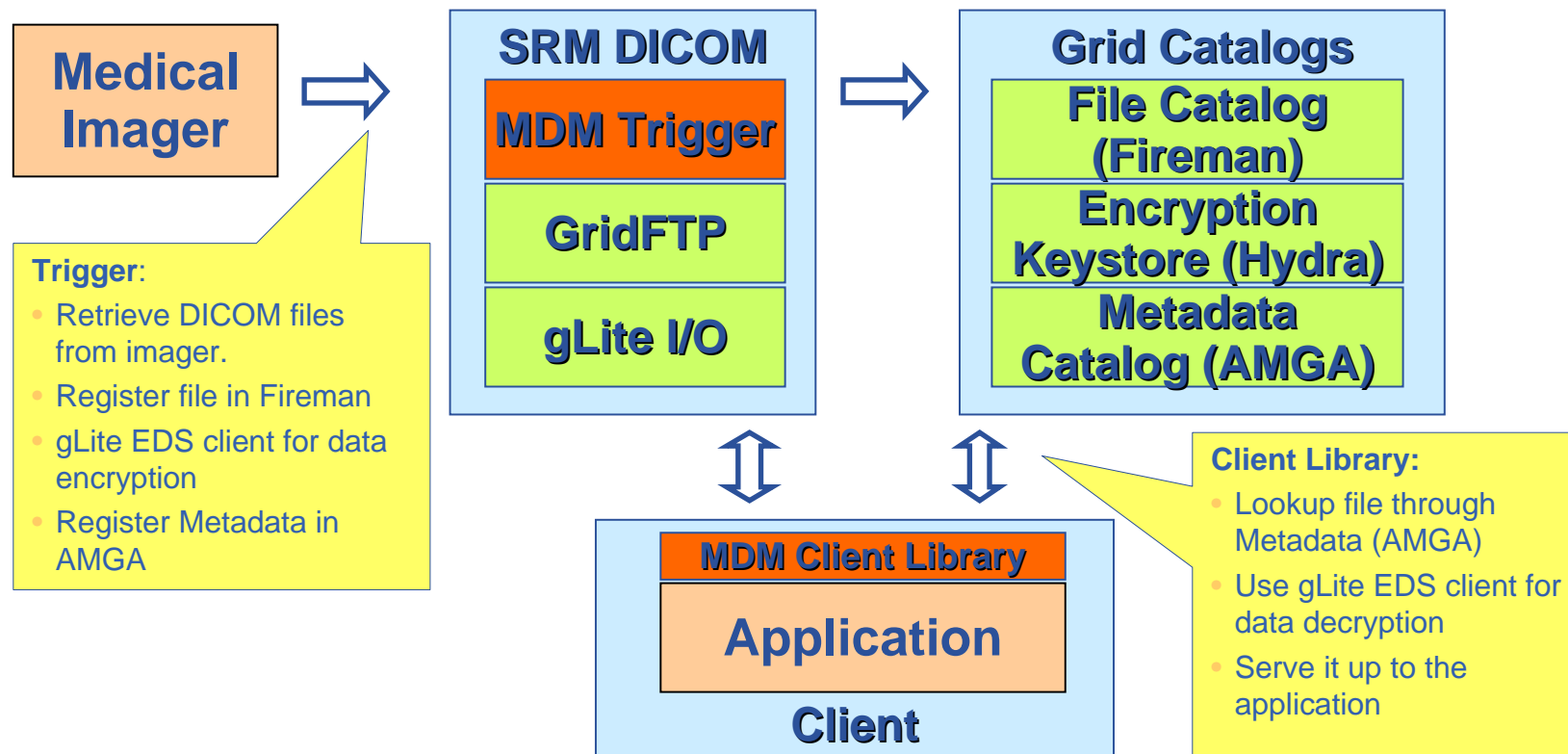
Enabling Grids for E-sciencE

- **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

Massimo Lamanna / CERN   15
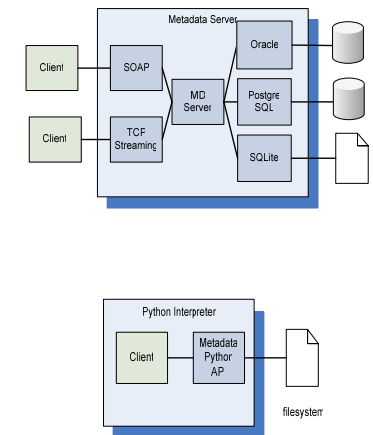
– **Improved user access:** GANGA (ATLAS and LHCb activities). Good feedback and demo in Pisa

eGee

- **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

Medical Imager

SRM DICOM
- **MDM Trigger**
- **GridFTP**
- **gLite I/O**

Grid Catalogs
- **File Catalog (Fireman)**
- **Encryption Keystore (Hydra)**
- **Metadata Catalog (AMGA)**

Trigger:
- Retrieve DICOM files from imager.
- Register file in Fireman
- gLite EDS client for data encryption
- Register Metadata in AMGA

MDM Client Library

Application

Client

Client Library:
- Lookup file through Metadata (AMGA)
- Use gLite EDS client for data decryption
- Serve it up to the application

Enabling Grids for E-sciencE

- **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*

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
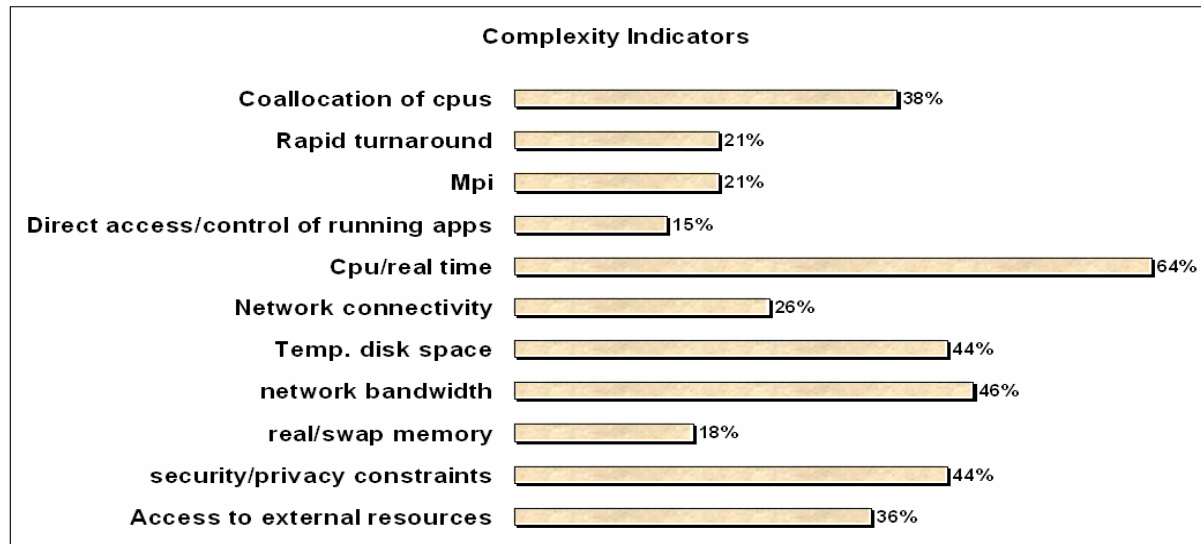
- **Evaluation of user satisfaction available in DNA4.3.2**
  - good for user support
  - intermediate on grid services
  - critical on documentation (with the exception of system manuals which are well appreciated)
- **Action: creation of the User information Group to address issues related to documentation in the project**
- **Next user survey at PM23**

- **EGEE is necessary for scientific communities**
  - Very large scale production of simulated data is essential for the commissioning of the LHC experiments
  - Distribution of LHC data on the grid is required to manage the output flow from the detectors

- **First grade scientific achievements impossible without EGEE**
  - In silico drug discovery for neglected diseases

- **Other measurements of EGEE growing scientific impact**
  - 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
  - More and more scientific communities are asking for collaboration with EGEE

- **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
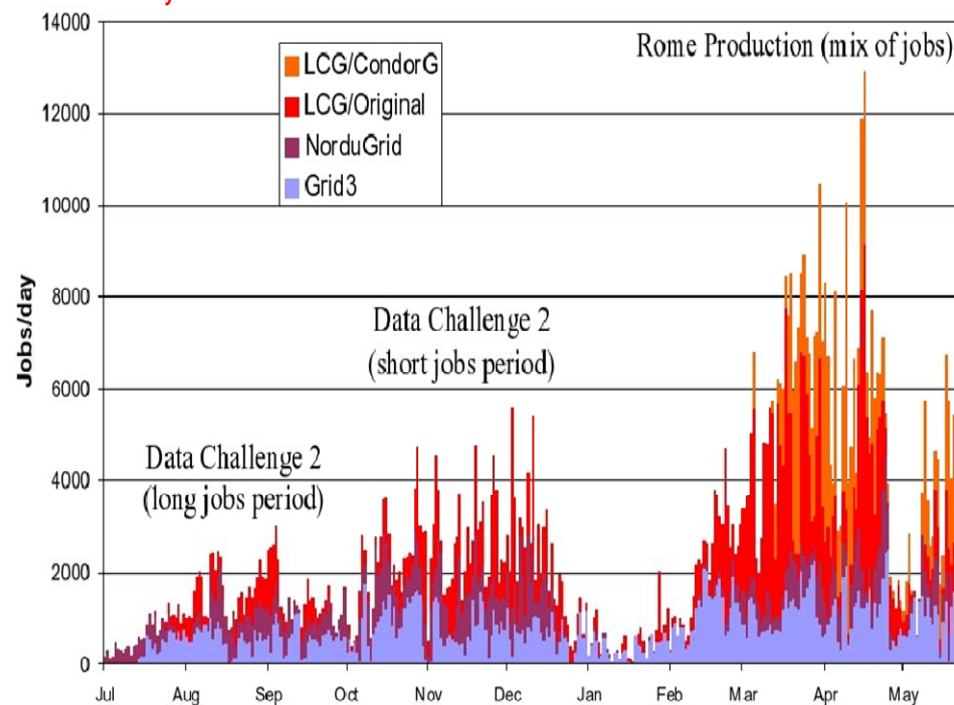   - Comprehensive analysis: see S.Campana et al., e-Science 2005, Melbourne, Australia

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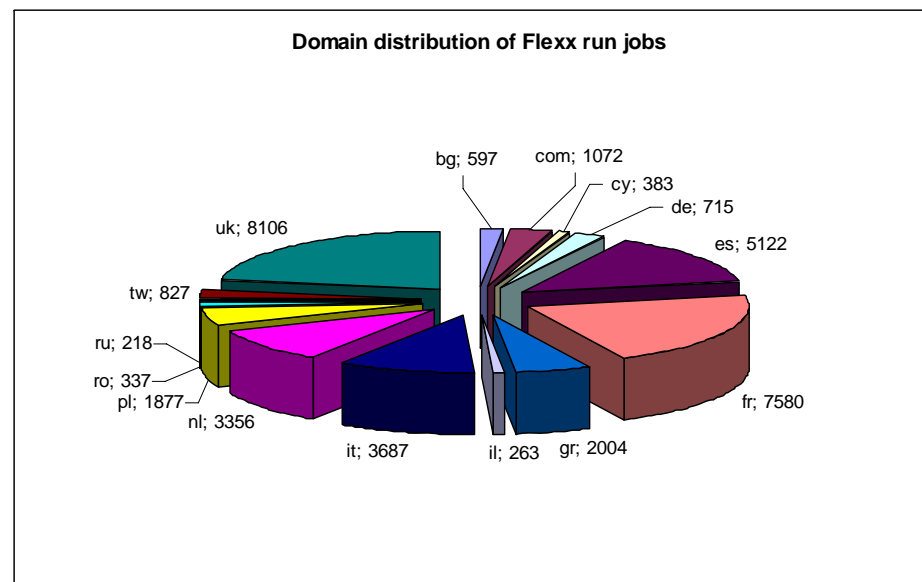   - 700 CPU/years in 2005 ; 77 TB of data

  - 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
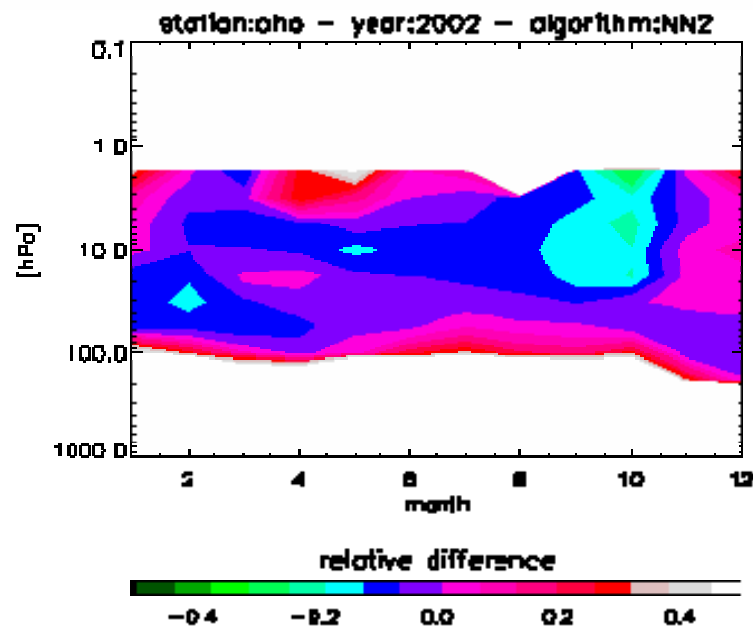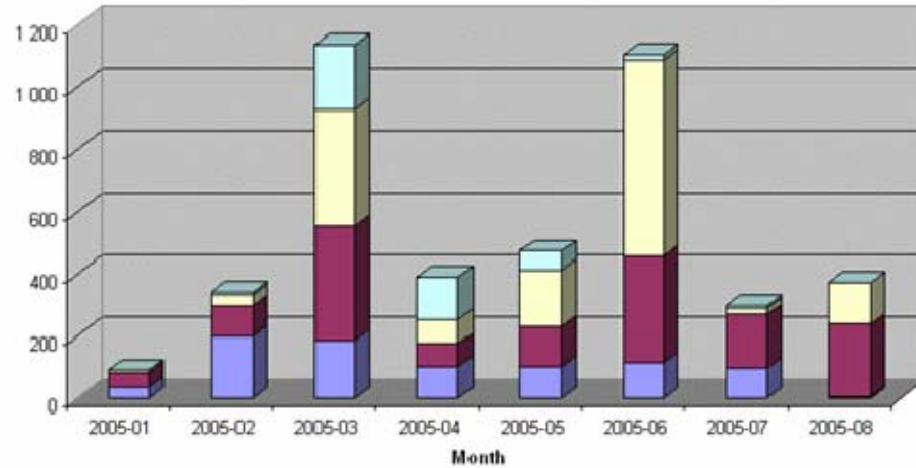 - Average crunching factor ~600

Domain distribution of Flexx run jobs

bg; 597 com; 1072 cy; 383 de; 715 es; 5122 uk; 8106 tw; 827 ru; 218 ro; 337 pl; 1877 nl; 3356 it; 3687 il; 263 gr; 2004 fr; 7580
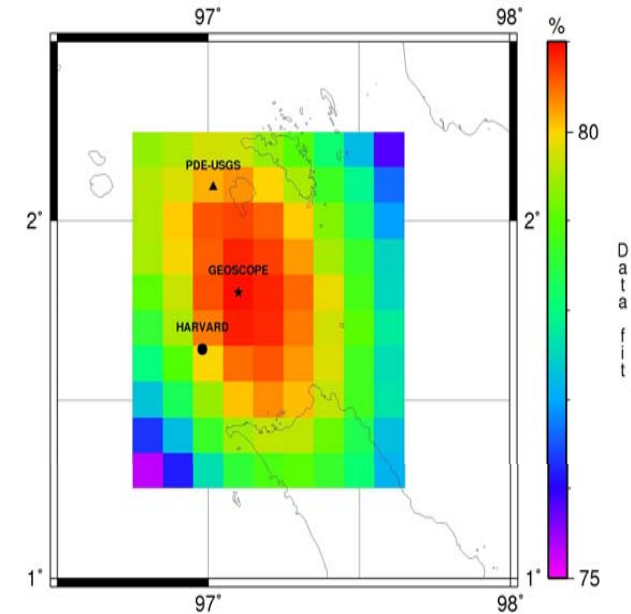
WISDOM open day

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

eGee

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*