egee

Enabling Grids for E-sciencE

Transition from EGEE-II to EGEE-III
High Energy Physics Cluster (TNA4.2.1)

Patricia Méndez Lorenzo (CERN, IT-GS/EIS)

NA4 Face-to-Face Steering Committee Meeting
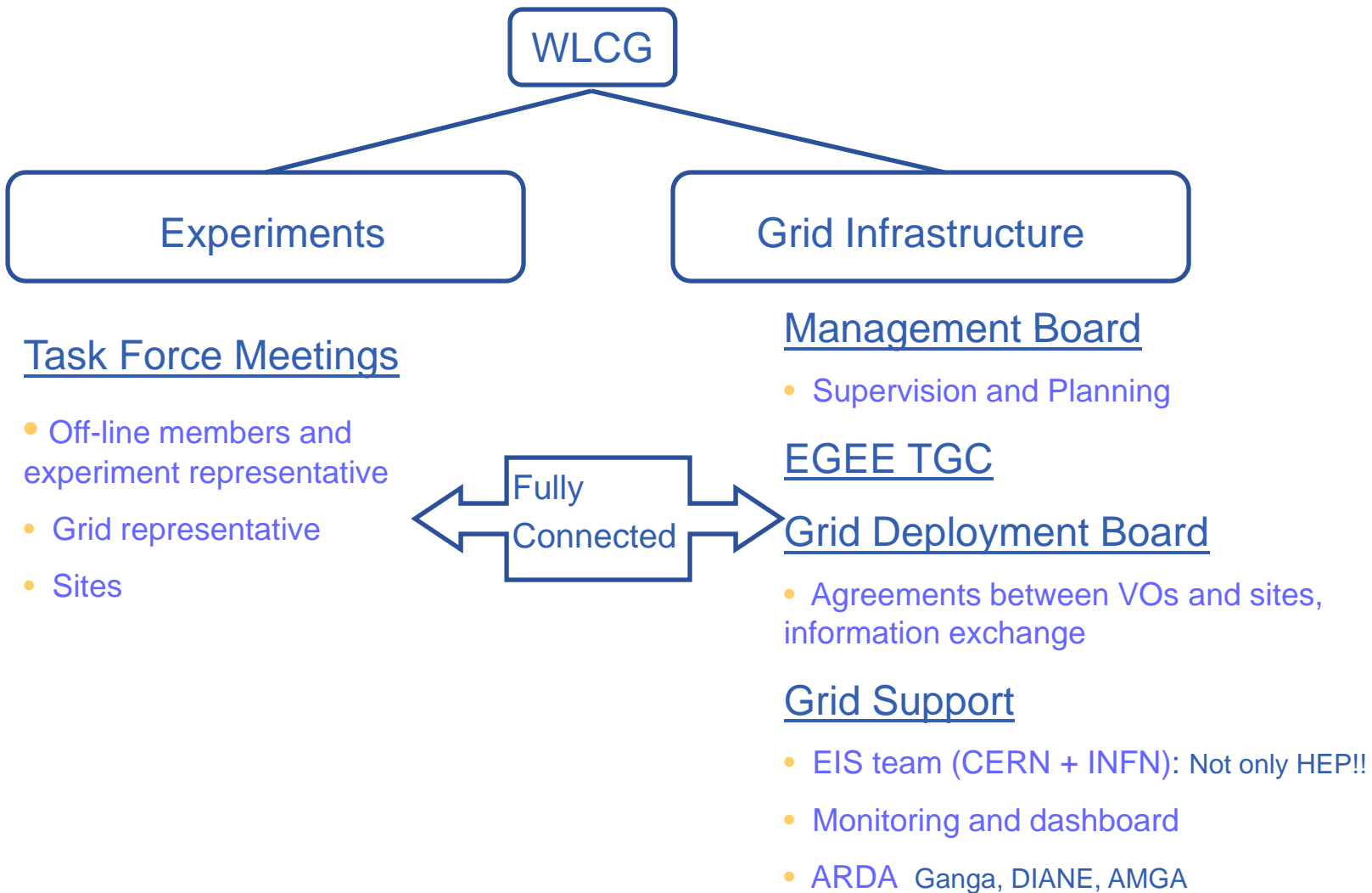*27-28 March 2008*
Université Paris VI

www.eu-egee.org

Information Society
and Media

EGEE and gLite are registered trademarks

- **Introduction: HEP Cluster**
- **WLCG infrastructure**
- **The CCRC`08**
- **Workgroup Plan**
- **Communication Plan and Tools**
- **Summary**

Enabling Grids for E-sciencE

- **HEP onto the EGEE-III Proposal:**
  - *"Largest community and the largest contributor of computing and storage resources to EGEE"*
  - *"Huge demand in terms of the size of stored data, number of jobs to manage, and number of active users"*
  - *"High scalability and reliability from the middleware and the availability of of application-level service are crucial"*
  - *"Four international collaborations comprising more than 4000 scientists all over the world"*
  - *"Close contact with other international facilities as DESY (Germany), FNAL (USA), SLAC (USA). The total number of users in HEP VOs exceeds 5000"*

- **Grid Users and HEP:**
  - <u>**Provisioned use**</u>:
    - Applications need quasi-dedicated resources
    - Offers an excellent long-term load ensuring the continuous use of the Grid
  - <u>**Scheduled use:**</u>
    - Many resources for a short time
    - Normally the resources are booked in advance for these communities
  - <u>**Opportunistic use:**</u>
    - Least demanding
    - Use of resources depending on their availability

- **HEP is a clear example of provisioned use community, providing already support and infrastructure (services and middleware) for the rest of the potential Grid users**

- **Much depending on a T0-T1-T2 hierarchical infrastructure**
  - Role of each site has been agreed by the experiments following their own Computing models. In general:
    - Raw data collection and 1st reconstruction at T0
    - Movement of raw data to T1 for custodial purposes and 2nd reconstruction
    - Simulation and analysis at T2 sites (LHCb foresees the analysis at T1 and larger T2 sites)
- **WLCG infrastructure**
  - Services
    - Level of service signed individually by each site through a MoU
      - *Contains the services provided by each site per experiment, the time schedule for recovery and the level of service*
    - Ranges from 99% for key services at T0 up to 95% for several services at T2 sites
    - A general procedure has been defined to ensure a good level of service: Checklist for new services, recommendations for middleware and DB developers, operational techniques and procedures
      - *Proposed to all T1 sites and will be extended to "crucial" T2 sites*
  - Operations
    - Daily operation meetings at CERN
    - Weekly operations meetings with the ROC responsibles
      - *Experiments represented at the meetings*
  - Application and User Support: EIS team
    - Each application has a dedicated support person (all of them high energy physicists)
    - Members of the experiment task force groups
    - Close collaboration with both service/operations and middleware (development and deployment)

WLCG

Experiments

Grid Infrastructure

Task Force Meetings

- Off-line members and experiment representative
- Grid representative
- Sites

Fully Connected

Management Board

- Supervision and Planning

EGEE TGC

Grid Deployment Board

- Agreements between VOs and sites, information exchange

Grid Support

- EIS team (CERN + INFN): Not only HEP!!
- Monitoring and dashboard
- ARDA Ganga, DIANE, AMGA

Enabling Grids for E-sciencE

Common Computing Readiness Challenge 08

- **The primary objective is to demonstrate that the 4 LHC experiments could run together at 2008 production scale**

- **It is a WLCG exercise and NA4-HEP support has participated together with the experiments**
  - This includes testing all functional blocks
  - We have established the CCRC simultaneously to the experiments FDR (Full Dress Rehearsals) which means the testing of the whole computing model using both real (cosmic) and simulated data
  - It is therefore a stress tests of the Grid infrastructure offered to the experiments under real data taking conditions
  - It is foreseen to have a yearly CCRC exercise during shutdowns of the accelerator

- **Two challenge phases have been foreseen**
  - February 08: Not all 2008 resources in place. Still adapting to new versions of some services
  - May 08: All 2008 resources in place. Full 2008 workload, all aspects of experiments` production chains
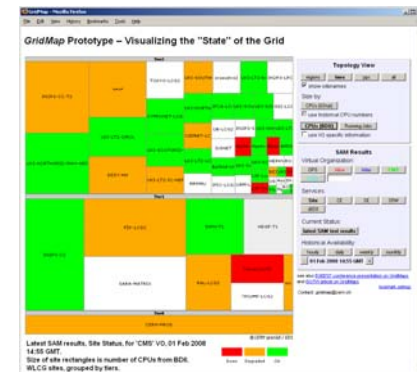
Enabling Grids for E-sciencE

- **It is an important test to be undertaken by the whole support infrastructure which has taken into account the following aspects to define the successful rate:**
  - Scaling factors: set by the experiments and established with the sites to ensure the necessary resources
  - Critical services: Defined by the experiments depending on their computing models and including the impact of service degradation or interruption
  - WLCG MoU: level of service provided by each site, availability, interventions....
- **The key aspect of the exercise has been the stress of a general and multi-purpose Grid infrastructure applied and "individualized" for each community depending on their service requirements.**
  - This means production, this is what we will face with any community

Enabling Grids for E-sciencE

- **Principal Goals of the NA4-HEP cluster**
  - The support team has to be expert on the experiment computing models and on Grid infrastructures
  - Experiment Support and integration
    - Participation in CCRC`08 and experiment FDR
    - Acting as contacts between sites-services-experiments
    - The support tasks depend on the requirements of each experiment
  - Continuing gridification projects arriving at CERN also besides HEP
    - Ganga and DIANE are the key tools towards a successful gridification
    - Next project: Beam Collimation and simulation: Close infrastructure to Fusion initiatives (ITER)
  - Monitoring: SAM, Dashboard and Gridmap and expansion to other communities
    - Each EIS member is involved in the creation of the experiment specific tests for each service
    - The created infrastructure can be used by any other application

SAM implementation applied to ALICE

Gridmap view

Enabling Grids for E-sciencE

- Analysis tools and middleware development
  - Example: CMS (CRAB): INFN product. CLI oriented tool which handles the submission of the jobs supporting EGEE and OSG m/w infrastructures.
  - If required the support members are the perfect actors to provide experiments with missing functionalities
    - *Can be used by the rest of communities*
    - *Examples: VOBOXES used by all HEP experiments*
  - **Ganga**
    - *ATLAS and LHCb as HEP customers*
    - *Used by many other communities and established as the gridification tool by any new community*
    - *This is a tool which we would like to generalized to any community counting with your collaboration*
      - There are more Grid UIs, but where posible, Let`s unify tools within Ganga
  - **DIANE**
    - *Version 2.0 ready for production and will be used during the Geant4 production in May*
  - **AMGA**
    - *Good example of collaboration between different applications and institutes*
- Middleware testing: Performance and scalability
  - Following the experiment requirements, the m/w has to be put in production ensuring their needs
  - Examples: Testing of the gLite WMS following the ATLAS and the CMS requirements and the CREAM CE
    - *Collaboration with JRA1 and SA3 totally transparent to the end-users: experiments*
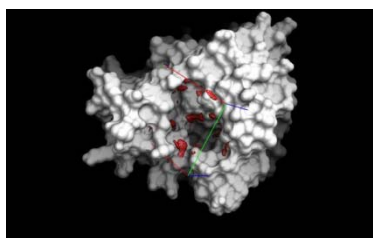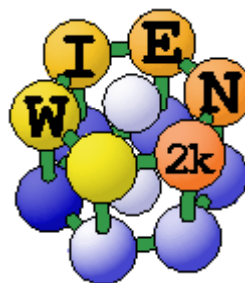
HARP

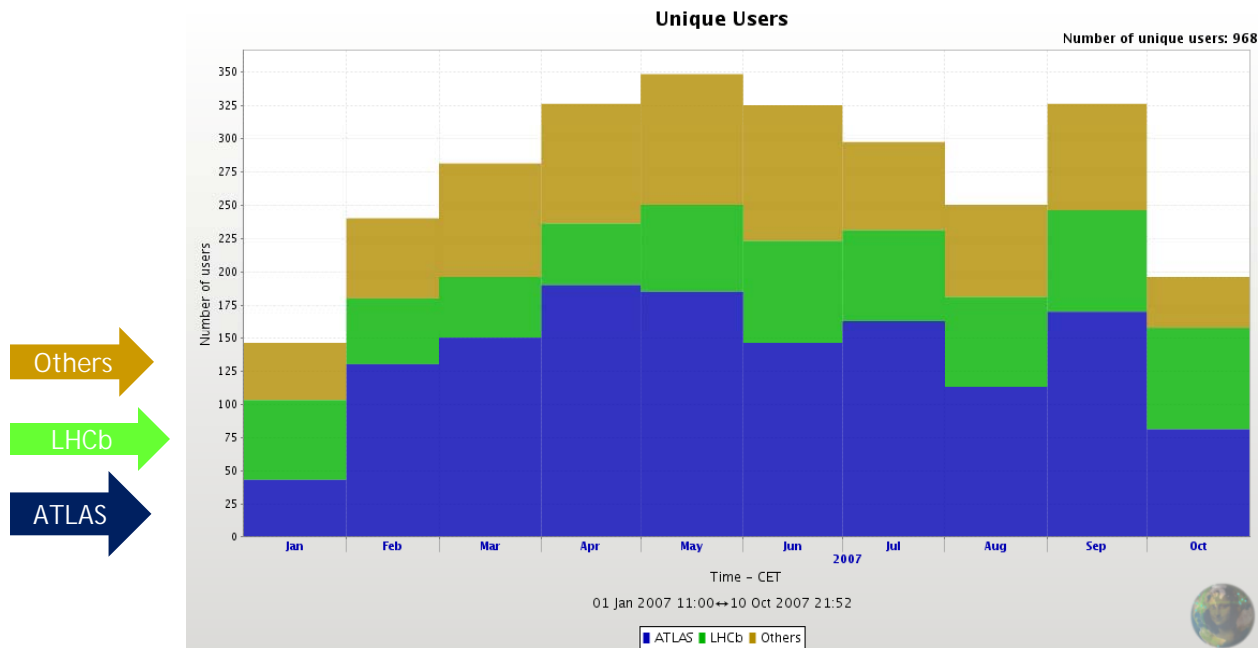Garfield

Enabling Grids for E-sciencE
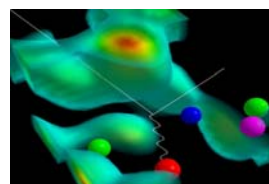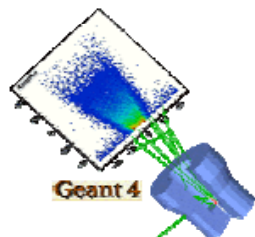
- **User friendly job submission tool**
  - – Able to hide all grid and m/w specifications
  - – Applicable to many backends
- **Analysis tool used by ATLAS and LHCb and defined as the gridification toolkit (combined with Diane if needed) for new applications arriving at CERN**

eGee

Enabling Grids for E-sciencE

- **Services**
  - Software: testing and creation are key purposes (WLCG)
  - Documentation: maintenance of the gLite User Guide
  - m/w and services expertise
  - Tutorials: We consider that new applications and users should be trained on upper layer tools (Ganga, Diane)
    - FTS, LFC, WMS... cannot be considered user tools but must be implemented with the support team assistance into the computing models but transparently to the end users
- **Subtasks**
  - VO Support will be fundamental towards new VO creation
  - The collaboration between VO and Application support is mandatory at this point
- **Clients**
  - HEP experiments (mostly off-line groups), new applications (application developers)
- **Metrics**
  - Difficult to established but this could be an initial point:
    - Number of users (not number of certificates), site efficiency (SAM and dashboard included), new applications (matter of quality not only number)
- **Policies**
  - Each application will be considered individually depending on the available manpower and resources at that moment

Enabling Grids for E-sciencE

- **The support team is fully engaged with each experiment**
  - The communication is performed on a daily based collaboration
  - Active participation into the TF meetings (weekly in most of the cases)
  - Daily representation during the CCRC operations meetings
  - Experiment contacts with operations and services
    - Represented in the GDB, TMB, operations meetings, etc
- **Communication with EGEE through the NA4 contact person**
- **Active participation on EGEE conferences and User Forums**
  - HEP participation will be always ensured
  - Fundamental to share information with the rest of communities
- **GGUS has been established as the individual user support toolkit into HEP applications**
  - Improvement in this sense is needed if we want to avoid users contacting directly with sites or application support
  - It does not provide individual user support

- **The NA4-HEP Cluster will face a crucial year due to the real data taking beginning in 2008**

- **The support team is embedded also into the WLCG structure collaborating with the experiments, services and operations**

- **The CCRC`08 is the key exercise to test the Grid infrastructure that will be provided to the HEP experiments during the real production**
  - Each experiment has its own computing model, own requirements and needs in terms of support and services
  - CCRC`08 has to observed each individual application onto a general Grid infrastructure

- **Support, software testing and software creation are the key tasks of the team at CERN**
  - Ganga, Diane and AMGA will continue the development and their establishment as the application tools for HEP application and others
  - Collaboration with other communities are oriented towards the general establishment of these tools

- **The gridification of other communities will be one of our goals but always depending on the man-power available at each moment**