



Enabling Grids for
E-science in Europe

NA4 status V. Breton (CNRS)

All activity meeting
13/9/04

Credits: R. Barbera, F. Harris, M. Lamanna, J.
Montagnat



Deliverables and milestones

- DNA4.1 (PM3) sent to EU
- DNA4.2 (PM6) under edition
 - Skeleton approved by PEB
 - Contribution on virtuous cycle by Malcolm Atkinson
- MNA4.1: first implementation of NA4 test suite expected at PM6

Quality metrics

- **Quality metrics allow to quantify the growth of EGEE user communities and their experience of the infrastructure**
- **Metrics described in NA4 quality plan: target values from Technical Annex**
 - Extension and growth of user communities: number of disciplines, number of users, number of “scientific” VOs, geographic extension of the VOs
 - Experience of the infrastructure: quality of service relative to job success rate and execution time, VO specific usage of the infrastructure,...
 - Network usage, user support, test phases, ...
- **Information workflow is being set up**
 - JRA2 centralizes the information
 - VO managers to be contacted monthly by H el ene Ruelle for VO internal information
 - Collection of NA4/SA1 information to be discussed within NA4/SA1 joint group
- **Requirements added (M.Soberman) to data base regarding**
 - R0047: NA4 Metrics and QoS relative to jobs requests and execution
 - R0049: NA4 Metrics requirement - Abort codes classification and recording
 - R0050: NA4 Metrics: Network usage

Joint NA4/SA1 group

- *Mandate: It will deal with the top level problems of integrating new VOs into the LCG/EGEE infrastructure from the aspects of integrating new resources, negotiating the use of distributed resources and the human and technical interfaces necessary from the application VO(s) to the services provided by LCG/EGEE.*
- Membership: F. Harris, Y. Legré, A. Mills, N. Thackray, VO managers, ROC/CIC representatives, ...
- Venue
 - Monthly phone conference
 - first meeting Sept 15th at CERN is deliberately a compact one, involving relatively few people from NA4 and SA1 looking at just Biomedicine.
 - The goal is to move soon, in future meetings, onto the agreed new application areas

EGEE Generic Applications Advisory Panel



- EGAAP mandate: advise the project management on new application sectors in terms of
 - Scientific and technical relevance
 - level of commitment (how many persons, for how many months)
 - scope of the NA4 effort (ie create the VO, deploy F90 on so many sites,etc..)
- EGAAP is part of the “virtuous cycle”
 - Place where new applications can apply for integration on EGEE
 - Place where deployed applications can provide update and feedback
 - Following PEB/PMB approval, NA4 management
 - Identifies resources allocated
 - Writes mini-MoU with new applications
 - Follows up on application deployment
 - NA2/NA3 involvement welcome
- First meeting in June
 - Very satisfactory from the applicants point of view
 - Unsatisfactory as many EGAAP members did not attend
- Next meeting during EGEE second conference (November 25th)

Risk analysis

- Different perspectives from different user communities
- HEP top 5 risks (from highest to lowest risk)
 - Middleware related: service instability
 - Infrastructure related: available resources insufficient to attract users
 - User related: The integration of the data management infrastructure from the experiment framework could reduce the efficiency of the final system
 - User and Infrastructure related: How many interoperable infrastructures can experiment work with? Is there a minimal number?
 - User related: The systems do use custom components (from the experiment) which might be difficult to integrate obtaining final systems of sufficient quality (ease of use, etc...).
- Biomed top 5 risks (from highest to lowest risk):
 - security constraints prevent large scale deployment of applications
 - unable to attract biomedical community, especially far end-users (biologist, physicians...)
 - unable to demonstrate the relevance of grid for biomedical applications (no killer application demonstrated)
 - unable to port biomedical applications to the EGEE middleware (efficiency/complexity problems)
 - unable to deploy biomedical grid resources

Partner	FTE	MM	Assigned	To hire	FTE from TA	Deviation
CERN	8	96	8		8	0%
CESNET	1.6	19	0.8		1.6	50%
MTA	0.4	5	0		0.4	100%
UEDIN	1	12	1		1	0%
CNRS	13.85	165	12.85	1	10	0%
CSSI	2	24	1.6		2	20%
CRSA	2	24	1.7		2	15%
DKRZ	1	12	1		1	0%
FhG	1	12	1		1	0%
INFN	6	72	6		4	0%
FOM	2	24	1		2	50%
CSIC	2	24	1		2	50%
UPV	1.7	20	1.7		1.8	6%
IHEP	0.8	10	0.8		0.8	0%
IMPB RAS	0.6	8	0.6		0.6	0%
ITEP	1	12	1		1	0%
JINR	0.8	10	0.8		0.8	0%
PNPI	0.8	10	0.8		0.8	0%
RRC KI	1	12	1.		1	0%
SINP-MSU	1	12	1.		1	0%
TOTAL	>47	>564	>42		42,8	

Hiring status and manpower levels

- Hiring status
 - 2 of CNRS NA4 biomed “loose cannons” hired September 1st
- Resource management
 - HEP and Biomed on track
 - The process to allocate generic application resources to the new applications approved for deployment is under definition
 - Proposal (Roberto Barbera): to allocate resources from PEB/PMB approval to the following EGAAP meeting
- Issues
 - Some partners have sent technical reports directly to PO for QR1. Their content was not included in NA4 QR as the work was not known by NA4 management
 - feedback loop from PO to activity leaders on declared effort vs. true effort
 - Unfair situation: some partners do not fill time sheets

Industry forum

- The industry forum newsletter will be available next week on the industry forum web site
- A session at next IST2004 is under preparation
- Two work groups are being set up and are expected to report on their activity at next EGEE conference
- A new charter is under discussion

Status of non-HEP application deployment



Biomedical applications

- Applications description page
 - <http://egee-na4.ct.infn.it/biomed/applications.html>
- Three types of application
 - **Pilots**: LCG2 compliant applications at day 0
 - **Internal**: from project partners, to be deployed on EGEE
 - **External**: from other projects, to go through a selection procedure
- Applications available today
 - **Pilots**: GATE, GPS@, CDSS
 - **Internal**: SiMRI3D, PTM3D, xmipp_MLrefine
 - **External**: Mammogrid



GATE

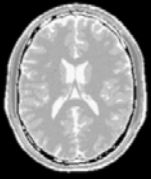
- **Contact**
 - Lydia Maigne, maigne@clermont.in2p3.fr
- **Description**
 - Monte Carlo simulation for radiotherapy planning
 - Demonstrated for the 3rd EDG review
- **Deployment and status**
 - Installed at CC-IN2P3
 - GENIUS interface
- **Users**
 - developers only
 - sets of 5 to 10 jobs, 4 minutes each, for testing
- **Plans**
 - to install on LPC (Clermont-Ferrand) and CNB (Madrid) clusters
 - larger scale tests when deployed on more nodes available
- **Problems**
 - too long waiting time (up to 30 minutes) for short jobs

- **Contact**
 - Christophe Blanchet, christophe.blanchet@ibcp.fr
- **Description**
 - Web portal for bioinformatics
- **Deployment and status**
 - NPSA is a production web portal hosting proteins databases and algorithms
 - GPS@ is the grid version under development deployed on LCG2
- **Users**
 - NPSA serves hundreds of bioinformaticians daily (about 3000 jobs/day) but limited resources (4 CPUs)
- **Plans**
 - to replace NPSA with GPS@ when showing similar robustness
- **Problems**
 - too long waiting time for short jobs
 - no service certificate for the portal
 - no RLS service for data management

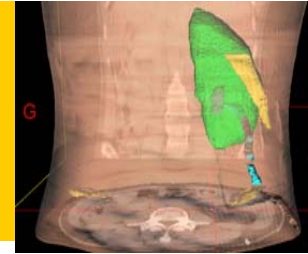


CDSS

- **Contact**
 - Ignacio Blanquer, iblanque@dsic.upv.es
- **Description**
 - Clinical Decision Support System: expert system for medicine
- **Deployment and status**
 - Original developed under serviced-based approach, now ported to LCG2
- **Users**
 - About 10 medical users from 5 organizations
 - About 10 runs per day (1 hour each)
- **Plans**
 - Migrate from current in-site LCG2 installation to EGEE infrastructure
 - Enlarge the user community when more resources become accesible
- **Problems**
 - None reported



Internal applications



- Name: SIMRI3D
- Contact: Fabrice Bellet, fabrice.bellet@creatis.insa-lyon.fr
- Description: Magnetic Resonance Images parallel simulator
- Deployment and status
 - MPI simulator implemented
 - Some performance study lead on local cluster
 - Tests on CINES supercomputers
- Users
 - Very large potential community
 - Today, only developers (5 users)
 - 1000 to 2000 jobs this year, minutes to weeks per job
- Plans: open the simulator as soon as gridification is achieved
- Problems: no MPI-enabled resources available on EGEE infrastructure

- Name: gPTM3D
- Contact: Cécile Germain-Renaud, germain@lal.in2p3.fr
- Description: radiological data interactive segmentation and analysis
- Deployment and status
 - Application ported to LCG2 on top of the interactive job submission service
 - Deployed on Orsay resources
- Users
 - Developers
 - Potential medical users
- Problems: Interactivity made difficult due to bypass-based communication performance limitations

- Name: xmipp_MLrefine
- Contact: Angel Merino, AJ.Merino@cnb.uam.es
- Description: Macromolecular 3D structure analysis
- Deployment and status
 - Recently ported to LCG2 and tested both on Clermont and Madrid clusters
- Users
 - Developers
 - One experiment corresponds to about 500 jobs and one week of computations on Madrid site
- Plans: To shorten experiments time by using more resources
- Problems: Resource shortage

Show stoppers and planning

- Limitations induced by infrastructure
 - Limited VO acceptance (CNAF RB)
 - No RLS service for biomedical VO
 - CC-IN2P3 proposal to host the service, under study in September
 - CNAF proposal to temporarily set up a RLS service meanwhile
 - No MPI-enabled resources available for parallel applications
 - INFN reported that an MPI-enabled cluster exists on INFN-Grid
 - Lack of service redundancies
 - High sensitivity to CNAF RB health
- Future plans
 - to focus more on applications and less on infrastructure deployment
 - to set up a communication channels
 - to notify users of SA1 maintenance operations and problems
 - to provide feedback to SA1 on problems encountered
 - to set up a new community integration procedure with SA1 including:
 - VO creation
 - RB registration
 - RLS provision

Status of new Generic Applications

- **Earth Observation**: first cluster should be ready next week at IPSL. As soon as it comes alive we will incorporate it into GILDA. VO “esr” set-up at SARA. VO manager Wim Som de Cerff (sdecerff@knmi.nl).
- **Geophysics**: some people from CGG already did some tests and are installing some machines to join GILDA. RPM ready and installed on GILDA.
- **Chemistry**: the first cluster is going to be ready in Perugia next week and will be added to GILDA. VO “compchem” and RLS created at CNAF. VO manager Osvaldo Gervasi (osvaldo@unipg.it).
- **Hydrology**: first contacts with Philippe Renard (Neuchatel) and Giuditta Lecca at CRS4 in Sardinia.
- **ESA**: Salim Ansari after the tutorial of July in Catania has been able to run GAIA satellite simulations on GILDA. They are now investigating how to proceed with an ESA specific VO accepted in GILDA.
- **Astrophysics**: Italian National Institute of Astrophysics is driving the community and will be present in GILDA with a few sites starting from the second half of September. They will also send a proposal for the next EGAAP meeting.
- **Astroparticle-physics**: VO “magic” and RLS created at NIKHEF. VO manager Harald Kornmayer (harald.kornmayer@iwr.fzk.de). GILDA will be the porting testbed for the MAGIC Montecarlo code. RPMs of MAGIC Montecarlo code in preparation.
- **GRACE**: they are happily running on GILDA since may and did two demos for the EU. They will most probably send a proposal for the next EGAAP meeting.

The GILDA Testbed

- 5 academic sites
- 2 “commercial” sites within the collaboration with the EU Project GRACE

The screenshot shows the GILDA Testbed website interface. At the top, there's a navigation bar with the GILDA logo and the eGEE logo. Below this is a menu with items: HOME, TESTBED, GRID DEMONSTRATOR, 1) CERTIFICATION AUTHORITY, 2) VIRTUAL ORGANIZATION, 3) GRID TUTOR, MONITORING, and CONTACTS. A large map of Europe is centered on the page, with several sites marked: Catania, Rome (CNR), Bologna (CNAF), Padua, Turin (Telecom Italia Lab), Seregno (GL2006), and Vienna (University). A sidebar on the left contains a list of links: Grid tutorials, Video tutorials, Instructions for users, Instructions for sites, Useful links, Sponsors, and Usage Statistics. The browser window title is 'GILDA Testbed - Grid INFN Laboratory for Dissemination Activities - Mozilla' and the address bar shows 'http://gilda.ct.infn.it/'.

- New sites that will join after the upgrade to INFN Grid 2.2.0, fully compatible with LCG 2.2.0

- INAF (Catania, Trieste, ...) (Italy)
- ICI Bucharest (Romania)
- II-SAS Bratislava (Slovakia)
- IPSL Paris (France)
- University of Perugia (Italy)
- University of Merida (Venezuela)

The GILDA Tutorials

(<http://gilda.ct.infn.it/tutorials.html>)



- Edinburgh, 7 April 2004
- Tunis, 22-23 April 2004
- Edinburgh, 26-28 April 2004
- CERN, 17-19 May 2004
- Catania, 24-25 May 2004
- Dubna, 29 June - 2 July 2004
- Edinburgh, 6 July 2004
- Karlsruhe, 6 July 2004
- Catania, 14 July 2004 (NA Open Meeting)
- Vico Equense, 19 July 2004 (GGF Grid School)
- NeSC event, 31 August – 03 September 2004
- Vico Equense, 6-10 September 2004 (CERN School of Computing)
- Karlsruhe, 20-23 September 2004 (GridKa Grid School)
- Heidelberg, 11-14 October 2004 (Grid course)
- CERN, 16 October 2004 (CERN 50th Anniversary)
- Den Haag, 15-17 November 2004 (IST2004)
- Merida, 15-19 November 2004 (Latin America Grid School)
- Istanbul, 9-10 December 2004 (SEE-grid event)
- Tel Aviv, December 2004

ARDA in a nutshell



- ARDA is an LCG project whose main activity is to enable LHC analysis on the grid
- ARDA is coherently contributing to EGEE NA4 (using the entire CERN NA4-HEP resource)
- Use the grid software as it matures (EGEE project)
 - ARDA should be the key player in the evolution from LCG2 to the EGEE infrastructure
 - Provide early and continuous feedback (guarantee the software is what experiments expect/need)
- Use the last years experience/components both from Grid projects (LCG, VDT, EDG) and experiments middleware/tools (Alien, Dirac, GAE, Octopus, Ganga, Dial,...)
 - Help in adapting/interfacing (direct help within the experiments)
 - Every experiment has different implementations of the standard services, but:
 - Used mainly in production environments
 - Few expert users
 - Coordinated update and read actions
 - ARDA
 - Interface with the EGEE middleware
 - Verify (help to evolve to) such components to analysis environments
 - Many users (Robustness might be an issue)
 - Concurrent “read” actions (Performance will be more and more an issue)
- One prototype per experiment
 - A Common Application Layer might emerge in future
 - ARDA emphasis is to enable each of the experiment to do its job
 - About 2 FTEs per prototype
- Provide a forum for discussion
 - Comparison on results/experience/ideas
 - Interaction with other projects
 - ...

Experiment interfaces:
Piergiorgio Cerello (ALICE)
David Adams (ATLAS)
Lucia Silvestris (CMS)
Ulrik Egede (LHCb)

The experiment interfaces agree with the ARDA project leader the work plan and coordinate the activity on the experiment side (users)

Milestones (Level 1)



LCG ARDA End-To-End Prototype activity

Milestone	Date	Description
1.6.18	Dec 2004	E2E prototype for each experiments (4 prototypes), capable of analysis (or advanced production)
1.6.19	Dec 2005	E2E prototype for each experiments (4 prototypes), capable of analysis and production

Prototype



- Available for us since May 18th
- In the first month, many problems connected with the stability of the service and procedures
- A second site (Madison) available at the end of June
- At that point just a few worker nodes available
- Now the no. of CPU is increasing (50 as a target for CERN, hardware available) as well as the no. of sites
- CASTOR access to the actual data store being delivered recently (essential)
- Look for a larger installation to be able to attract users (few more sites, some ~100 CPUs)

Status



ALICE

- Grid activity
 - Use of the Glite testbed
 - Access system to gLite services being developed (Demo available in June 2004)
This is the key layer to allow ALICE software to be effectively use their prototype (evolution of their 2003 system, using gLite and PROOF)
- Other contributions
 - The access system is a generic piece of software (plans to be used in ATLAS and metadata access)
 - Tests of the metadata capabilities of the gLite file catalogue

ATLAS

- Grid Activity
 - Use of the Glite testbed
 - DIAL on gLite OK (Evolution of the DIAL demo)
 - ATHENA to gLite OK. Ready to expose to test users.
 - First skeleton of high level services
 - Detailed studies of the Don Quijote system (ATLAS data management system)
- Other contributions:
 - Detailed collaboration on AMI database (performance, support Oracle implementation,...)
 - Production and Combined Test Beam contributions

CMS

- The CMS system within ARDA not 100% defined
- Grid and other contributions:
 - Use of the Glite testbed
 - Successful ORCA job submission to Glite ☺.
 - Investigating with the package manager
 - Access to files directly from CASTOR
 - Glite file catalog tests
 - RefDB studies and evolution

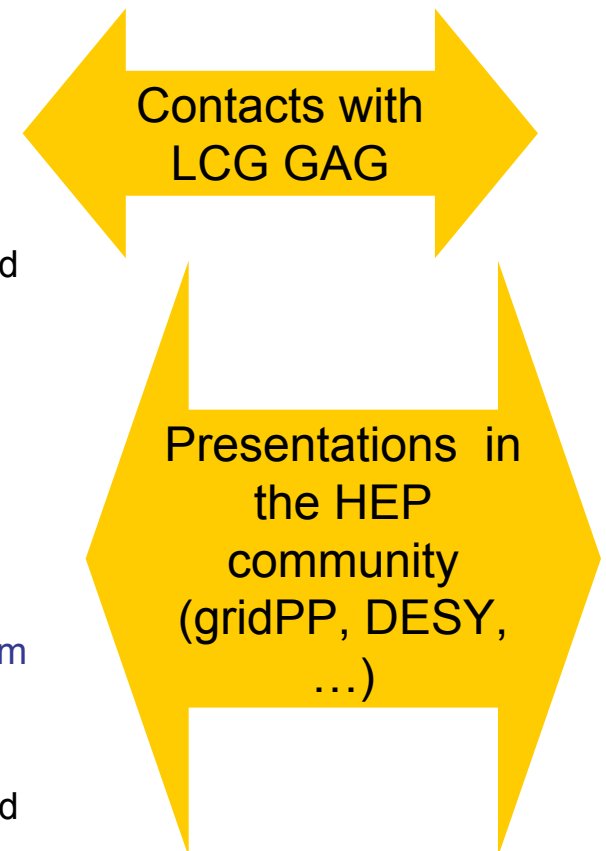
LHCb

- Grid activity:
 - Use of the Glite testbed
 - “Regular” DaVinci jobs onto Glite exposed to test users (outside the ARDA team)
 - DaVinci jobs from Ganga to Glite
- Other contributions:
 - GANGA release mgt and software process (CVS,...)
 - Contributions to DIRAC
 - Metadata catalogue (Perform. test also in Taiwan)

ARDA workshops and liaison activity



- **1st ARDA workshop (January 2004 at CERN; open)**
- **2nd ARDA workshop (June 21-23 at CERN; by invitation)**
 - “The first 30 days of EGEE middleware”
 - **Main focus on LHC experiments and EGEE JRA1 (Glite)**
- EGEE NA4 (Applications) meeting mid July
 - NA4/JRA1 and NA4/SA1 sessions organised by M. Lamanna and F. Harris
 - EGEE/LCG operations new ingredient!
- **3rd ARDA workshop (October 20-22 2004; → registration open ←)**
 - “The LCG ARDA prototypes”
 - →
http://lcg.web.cern.ch/LCG/peb/arda/LCG_ARDA_Workshops.htm
- EGEE Conference meeting mid November
 - NA4/JRA1 and NA4/SA1 sessions organised by M. Lamanna and F. Harris



Update on relationship with other activities

- New contacts points between NA4 and other project activities
 - G. Romier (UREC, CNRS) will act as the interface between NA4 and networking activities
 - Joint NA4/SA1 working group coordinated by F. Harris
 - M. Soberman acts as the interface with JRA2
- PTF to meet tomorrow
- Internal NA4 issues
 - AWG meeting weekly after August break
 - C. Loomis in charge of supervising NA4 test team

Next steps before Den Haag

- Application deployment on LCG-2
 - HEP: data challenges under way
 - Biomed: pilot applications (GATE, CDSS, GPS@) on LCG2 to be demonstrated in Den Haag
 - Non-HEP: further integration/deployment of non-HEP VOs on LCG2 to be discussed by joint NA4/SA1 group
- NA4 specific issues
 - Biomed technical team must get up to speed to test Glite
 - Integration (MoU's, resource allocation) and follow-up must be settled for 1st round of new approved applications before next EGAAP
 - First implementation of NA4 test suite
- Implementation of virtuous cycle on external projects
 - NA4 cannot do it alone for non HEP external projects
 - Identify a few high profile external projects for a coordinate NA1/NA2/NA3/NA4/SA1 work