



Enabling Grids for E-scienceE

# NA4: Application Identification & Support

*C. Loomis (LAL-Orsay)*

*All Activity Meeting (CERN)  
24-25 August 2006*

[www.eu-egee.org](http://www.eu-egee.org)



- **Management**
- **Communication Channels**
- **Status & Plans**
  - DNA4.1: Application Deployment Plan
  - Status of each scientific discipline.
- **Cross-Activity Groups**
- **VO Managers' Group**
- **EU Recommendations**
- **EGEE'06 Conference**
- **Execution Plan**
- **Conclusions & Issues**

- **Activity Manager: C. Loomis**
- **Scientific Discipline Representatives:**
  - HEP: Massimo Lamanna, Andrea Sciabà
  - Biomed: Vincent Breton, Johan Montagnat, Christophe Blanchet
  - Earth Science: Monique Petitdidier
  - Fusion: Francisco Castéjon
  - Comp. Chemistry: TBN
  - Astrophysics: TBN
- **GILDA: R. Barbera**
- **VO Managers' Group: Frederic Schaer**
- **Emeritus: Frank Harris**

- **Internal Management:**
  - Weekly NA4 Steering Committee meetings.
- **NA4 Partners**
  - Mailing list: [project-eu-egEE-na4-partner-contacts@cern.ch](mailto:project-eu-egEE-na4-partner-contacts@cern.ch)
- **NA4 Participants**
  - Mailing list: [project-eu-egEE-na4-member-contacts@cern.ch](mailto:project-eu-egEE-na4-member-contacts@cern.ch)
- **EGEE Users**
  - NA4 website: <http://egee-na4.lal.in2p3.fr/>
- **Virtual Organizations**
  - VO Managers' Group (Frederic Schaer, chair)
  - Mailing list: [project-egEE-vo-managers-group@cern.ch](mailto:project-egEE-vo-managers-group@cern.ch)
- **Users Forums & EGEE Conferences**

- **Application Deployment Plan (DNA4.1)**
  - Covered all scientific disciplines defined in TA.
  - Initial census of deployed and soon-to-be-deployed apps.
  - Information from ~20 VOs concerning around 70 apps.
  - Start to see grid “patterns” emerging from the applications.
- **Status:**
  - Reviewed and approved by PEB.
  - EDMS (<https://edms.cern.ch/document/722131/2>)
- **Future:**
  - Automate collection of information from VOs.
  - Primary forum for communication will be VO Managers’ Group.

- **Job Reliability**
  - Spot weak sites using the actual user job information
  - Careful study of middleware failures to pin down middleware issues
  - Presentations in various forums (notably EGEE TCG)
- **Experiment Dashboard**
  - Now with ATLAS and CMS
  - Single point of aggregation of all grid activity of all users (analysis and production).
  - Milestone demonstration for EGEE'06.
- **Ganga**
  - Continuing its evolution; used inside and outside HEP.
- **Other activities**
  - CDF, D0, BaBar, H1, Zeus, ...
  - UNOSAT, ITU, Geant4

- **Experiment Integration and Support (EIS)**
  - Key activity to support operation of large production and analysis activities on the grid
  - Dedicated tests of the gLite 3.0.2 middleware (close contacts with SA1, SA3 and TCG).
  - Validation used CMS analysis and ATLAS production jobs with encouraging results.
- **Experiments' activities**
  - ALICE: DC6 run; 2k-3k concurrent jobs.
  - ATLAS: Extended prod. on 3 infrastructures (LCG dominant).
  - CMS: 30k jobs/week, analysis jobs and 70+ users & 85+ sites
  - LHCb: Preparing DC06; over 6k concurrent jobs.

- **Drug Discovery**

- Successful runs for malaria and avian flu virus.
- Similar work to be done for neglected diseases in EGEE-II.
- WISDOM: 1 October - 1 December
  - ⌘ Requirements circulated to SA1 through OAG.
  - ⌘ 1 October - 1 December, 500 CPU-years, 5 TB storage (x2 backup)
  - ⌘ Discussions underway for finalizing docking targets.
  - ⌘ Other grid projects participating through resources or development
  - ⌘ Collaboration with DIANE/GANGA for better data mgt.



- **Medical Imaging**

- Kickoff meeting on July 12 in Sophia Antipolis
- Three application services offered from partners
  - ⌘ MDM installation/configuration being finalized
  - ⌘ MOTEUR workflow manager
  - ⌘ P-GRADE portal (license to be clarified)
- 6 applications from EGEE, 5 new in EGEE-II

- **Bioinformatics**

- Kickoff meeting on May 26 in Pisa
  - ⌘ Joined meeting between EU-EGEE, EU-EMBRACE and INFN-Grid
  - ⌘ About Grid data replication and consistency in Bioinformatics
- ~10 different bioinformatics applications
  - ⌘ Couple production systems being integrated.
  - ⌘ 3 applications from EGEE, 6 new in EGEE-II
- Defining Bioinformatics priorities in EGEE
  - ⌘ Deploying updatable databases (1st workshop), deploying legacy programs with special I/O, security of medical and industrial data, workflow, portal for biologist end-users, ...

- **Tight collaboration with DEGREE related project.**
  - Update of earth science requirements
  - Based on existing list from EDG, EGEE, and EGAAP report
  
- **ESR (Earth Science Research) VO**
  - Earth observation.
  - Solid earth physics.
    - § Seismology collaboration with EELA.
  - Hydrology
    - § U. Neuchâtel porting code for coastal aquifer analysis.
    - § Installing a cluster to integrate into EGEE.
  - Climate.
  - Pollution.

- **EGEODE VO**
  - Industrial, geophysics application.
  - Plan to organize a meeting in Paris for Geocluster users.
  
- **Two milestone demos planned for EGEE'06.**
  - Seismology
  - Climate

- **Fusion**
  - Porting and initial use of EGEE infrastructure.
- **Astrophysics**
  - PLANCK: active and growing
  - MAGIC: need to clarify future plans
- **Computational Chemistry**
  - Perugia group still active and using grid.
  - Partners have identified other comp. chemistry applications.
  - Need to forge collaboration between the various groups.
- **Others**
  - Many other applications through related projects.
  - Tentative contacts:
    - § Nanotechnology, Molecular dynamics, Computer science

- **Operations Advisory Group (OAG)**
  - High-level contacts between NA4 and SA1 activities.
  - Manages VO procedures (registration & MoUs).
  - Brokers resource requests.
- **VO Registration Procedure**
  - Procedure: <https://edms.cern.ch/document/503245/6>
  - Lightweight, web-based form for collecting minimal information.
  - CIC portal: DB to keep contact and configuration information.
  - Principal tool to keep track of VOs.
- **DILIGENT MoU**
  - Very advanced—need some input from SA3.
  - EDMS: <https://edms.cern.ch/document/748998/1>

- **User Information Group (UIG)**
  - Cross-activity group with strong NA4 participation.
  - Detailed (working) use cases as means to unify docs.
  - Working to define format and delivery.
  - NA4 will start in earnest with use cases once format is decided.

Naïve	Normal	Skilled
Get certificate	Resource/service discovery	Software installation
Run a job	Jobs with data requirements	Large-scale data transfer
Copy/register/access files	Environment setup	Monitoring status (R-GMA)
Recovering results	Monitoring status	Data encryption
Monitoring job status	Software installation	AMGA metadata
Preparing a job	Short-deadline job submission	MPI
		Workflow examples
		VO deployed services
		Biomed app. kernel
		Geo. app. kernel

- **TCG**
  - LHC experiments
  - Biomedical apps.
  - NA4 (C. Loomis, M. Lamanna)
  
- **QAG**
  - Haven't completed NA4 quality plan yet.
  - Have replaced and updated the NA4 web site.

- **Mandate**

- Coordinate information flow between EGEE NA4 and VOs.
- Provide forum for information exchange between VOs.

- **Status**

- Registering the VOs currently supported by infrastructure.
- Planning for first meeting at EGEE'06 conference.
- Contact: [project-egEE-vo-managers-group@cern.ch](mailto:project-egEE-vo-managers-group@cern.ch)

- **VO Registration Status**

- 165 VOs visible on the production infrastructure.
- 99 have entry in CIC portal.
- 34 have complete registration.



- **Rec. 11 (2nd): Matrix of requirements**
- **Rec. 12 (2nd): Inventory of all application users.**
- **Rec. 10 (3rd): Externalization of VOs**
  - Keep track of which ones exist (CIC portal)
  - Contact them routinely (VO Mgrs.' Group)
  - Identify major institutional constituents (Deliverables)
  - Identify their objectives (CIC portal, Deliverables)
  - Identify key achievements (Deliverables)
  - Document VO's impact on its community (Deliverables)
  - Identify VO's available resources (CIC Portal)
  - Document how to join them (CIC Portal, NA4 Web)
  - “Brand” their public website (won't do; inappropriate)

- **Sessions**

- NA4 General Session
- NA4/SA1/JRA1 Joint Meeting
- Application Session
- Life Sciences Session
- VO Managers' Group
  
- Demo session
- Mini-Users' Forum
  
- HealthGrid Session

- **NA4 Execution Plan**

- EDMS: <https://edms.cern.ch/document/759410/1>
- Tasks and effort as defined in TA (with INFN changes).
- List of people named by partners (along with declared effort).
- Distribution of deliverable reviews between partners.

- **Status**

- NA4 has 40 partners and 228 named people.
- 27 partners missing some effort  $\Rightarrow$  significant total
- Few unregistered people: VR, ELETTRA
- Other issues:
  - § INFN: people, task assignment, mgt. effort
  - § UEDIN: people
  - § ICI, RUG: confirmation of information

- **Conclusions**

- Broad production use, large number of existing VOs:
  - § Good sign for the uptake of grid technology.
  - § Makes organization and support more difficult.
- Natural expansion, but...
  - § Must ensure NA4 is aware of new applications & needs.
  - § Procedures and support must scale to accommodate all VOs.

- **Issues**

- Finalizing administrative details.
- Getting VO Managers' Group ramped-up ASAP.
  - § Cooperation with VO registration.
- Resource allocation procedure for new VOs.
  - § Information, procedures for new participants
  - § Supplying and integrating resources quickly
- Additional hurdles for people outside of CA coverage.