



Enabling Grids for E-scienceE

# NA4: Application Identification & Support

*C. Loomis for NA4  
EGEE EU Review (CERN)  
May 23, 2006*

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



- **Recommendation 11:**
  - For each application, produce a matrix of requirements versus the gLite components that will fulfill these requirements. Use a color coding scheme to identify priorities and importance.
  
- **Recommendation 12:**
  - Create an inventory of all application users. For each user, compile a list of all the applications she is using, together with the relative importance of each application to the user and applications characteristics (for instance, compute bound, I/O bound, etc.). Establish a process to keep this inventory up to date.

- **LCG Baseline Working Group** <http://lcg.web.cern.ch/LCG/peb/bs/>
  - Lead by Ian Bird
  - Spring 2005: HEP starts working group to prioritize the grid services.
  - June 2005: Final report ( <http://lcg.web.cern.ch/LCG/peb/bs/BSReport-v1.0.pdf> )

- **These findings are an important input for the EGEE TCG activity.**

<i>Service</i>	<i>ALICE</i>	<i>ATLAS</i>	<i>CMS</i>	<i>LHCb</i>
<i>Storage Element</i>	A	A	A	A
<i>Basic transfer tools</i>	A	A	A	A
<i>Reliable file transfer service</i>	A	A	A/B	A
<i>Catalogue services</i>	B	B	B	B
<i>Catalogue and data management tools</i>	C	C	C	C
<i>Compute Element</i>	A	A	A	A
<i>Workload Management</i>	B/C	A	A	C
<i>VO agents</i>	A	A	A	A
<i>VOMS</i>	A	A	A	A
<i>Database services</i>	A	A	A	A
<i>Posix-I/O</i>	C	C	C	C
<i>Application software installation</i>	C	C	C	C
<i>Job monitoring tools</i>	C	C	C	C
<i>Reliable messaging service</i>	C	C	C	C
<i>Information system</i>	A	A	A	A

A = High priority and mandatory

B = Standard solutions are required but experiments could select different implementations

C = Common solution desirable but not essential

- Biomed community has created matrices for 8 different applications.
- Critical functionality provided by:
  - Fireman
  - gLite I/O
  - AMGA
  - Hydra

CDSS\_Pharmacokinetics

DU = Direct Usage  
IU = Indirect Usage  
PU = Planned Usage

S = Satisfied  
PS = Partly satisfied  
NS = Not satisfied

Services usage Requirements	PU	VOIMS	ROAMA	IU	DU	PU	DU	IU	Accounting	SRM	FTS	gLite	GFAL	Fireman	LFC	Hydra	AMGA	Others	Piroquet	EnvFile
100733 Critical services need to be deployable in a redundant way	NS	PS	NS	PS	NS	S								NS						
100477 Performance of short jobs submission				NS																
100479 Middleware overhead for short jobs				NS																
100542 On-disk encryption of data																				
100474 Accessibility to MPI functionalities				NS			NS													
100553 User authorization	S			S			S													
100552 VO authorization	S																			
100498 Automatic translation of LFNs into physical file names												S	S							
100500 Automatic output files registration												S	S							
100510 Data registration on grid							S					S	S							
100511 Grid data retrieval							S					S	S							
100512 Grid data removal							S					S	S							
100515 Group of files							NS					S								
100516 Submission of multiple data jobs using group of files							NS					S								
100517 Replication of group of files												S								
100521 ACLs on files and metadata	NS											S	NS	NS	NS					
100524 Data updates							NS					S								
100526 Data replication							S					S								
100528 Data access cost estimation							NS					NS								
100535 Jobs listing												S								
100762 Encryption during transmission												S								
100544 Hook on data privacy manager	NS						NS					S								
100545 Communications encryption												S								
100560 Files listing												S								
100561 Files collection												S								
100563 Jobs listing control							PS					S								
100564 Jobs co-allocation on one site							PS					S								
100600 Jobs control							PS					S								
100762 Encryption during transmission												S							S	S
100731 Middleware has to allow for complete job traceability								S				NS			NS					
100706 Store data in encrypted form												NS	NS	S			NS			
100685 Resource access								S		NS		NS		S			NS			
100606 Job killing							S					S								

- **Database contains requirements from both internal and external applications.**
  - 382 requirements
  - <https://savannah.cern.ch/support/?group=egeeptf>
- **MoU process requires new applications to review requirements and add missing ones.**
- **No longer a formal part of the prioritization and evaluation process, but still valuable resource.**

- **Excellent venue to see:**
  - How real users are using the system.
  - What problems they are having.
  - What common services would help them.
- **Collected this information through:**
  - Formal users' survey.
  - Preparation of parallel summary talks.
  - Submitted abstracts.
- **Important for users to see beyond their own app.:**
  - Buy into (and perhaps help develop) new common services.
  - Reuse or be inspired by usage patterns in other domains.
  - Results: WISDOM, avian flu, EGEE/ITU, TCG working groups

- **Each application has different priorities:**
  - Stage of application development (porting, testing, production,...)
  - Functional requirements (e.g. privacy constraints for biomed.)
  - Application's infrastructure (e.g. use of existing databases)
- **NA4 tries to balance all of the different priorities to keep a practical, realistic overall prioritization.**
- **Technical Coordination Group (TCG) decides overall project priorities.**
  - EGEE NA4
  - LHC Experiments
  - Biomed Task Force

- **Users' surveys completed (MNA4.3, MNA4.4):**
  - Online version limited to those with grid certificates.
  - Very limited response to questionnaire. □
- **Reaffirmed known issues (e.g. documentation).**
- **However, not really effective for obtaining the information we want:**
  - Functionality required by application.
  - Persistent problems encountered by the application.
  - Contribution to and utilization of the infrastructure.
- **Gathering this information requires continuous dialog with (representatives of) each community.**



- **‘VO interviews’ in MoU process give a detailed view of:**
  - Resource requirements,
  - Application behavior, and
  - Interactions with grid services.
- **However,**
  - Extremely time consuming and not scalable.
  - Doesn’t track the evolution of the application.
  - Too early lifecycle to provide operational, middleware feedback.

- **Propose:**
  - Group consisting of all VO Managers (or community contacts).
  - Chaired and managed by NA4.
- **Allows:**
  - More detailed view of the VOs using the grid.
  - Better and clearer mechanism for two-way communication.
  - More integrated view of VO's needs and usage.
- **Prerequisites:**
  - Inventory of VOs (with contacts, parameters, etc.)
  - Extremely lightweight registration procedure (and follow-up).

- **Preliminary inventory in DNA4.4 (sec. 5.2):**
  - Total 9 VOs (HEP=4, Biomed=1, Generic=4)
  - Identified operational and functional needs.
    - § Big VOs well represented, e.g. in TCG.
    - § Easy to miss important requirements from smaller VOs with great potential.
- **Need to expand to all VOs using the infrastructure.**
  - E.g. 41 VOs have used > 1 CPU\*day/week in 2006.
  - Dialog often difficult because parameters are not known.

- **Rec. 11: Collection and prioritization of requirements**
  - Application Matricies
  - Requirements Database
  - User Forums
  - Technical Coordination Group (TCG)
  
- **Rec. 12: Expansion of and communication with VO's**
  - User Surveys
  - Memoranda of Understanding
  - VO Managers' Group
  - Lightweight VO registration