



Enabling Grids for  
E-science in Europe

*NA4 Biomed requirements - PTF meeting  
June 17, 2004*

# Biomedical Application Requirements

Johan Montagnat



- **Heterogeneity**
  - developers, biologists, physicians, chemists, etc.
  - biology, medical imaging, health...
- **Not necessarily grid aware**
  - we have to demonstrate grid benefit
- **Confidentiality requirements**
  - sensitivity of data
  - defiance toward out-of-the-hospital data management
- **Time constraints**
  - clinical use
- **Not necessarily convinced by grid technologies**
  - other priorities
  - hospitals have existed long before grid

# Requirements server

- <http://egee-na4.ct.infn.it/requirements>

## Application requirements

### Browse and update requirements

<a href="#">All requirements</a>	<b>By middleware:</b> <ul style="list-style-type: none"><li>● <a href="#">User interface</a></li><li>● <a href="#">Job submission</a></li><li>● <a href="#">Data management</a></li><li>● <a href="#">Information system</a></li><li>● <a href="#">Storage</a></li><li>● <a href="#">Network</a></li><li>● <a href="#">Operation</a></li></ul>	<b>By type:</b> <ul style="list-style-type: none"><li><input checked="" type="checkbox"/> Function</li><li><input type="checkbox"/> Interoperability</li><li><input type="checkbox"/> Performance</li><li><input type="checkbox"/> Reliability</li><li><input type="checkbox"/> Security</li><li><input type="checkbox"/> Portability</li><li><input type="checkbox"/> Accessibility</li></ul> <input type="button" value="Search"/>	<b>By area:</b> <ul style="list-style-type: none"><li><input checked="" type="checkbox"/> HEP</li><li><input type="checkbox"/> Biomed</li><li><input type="checkbox"/> Generic</li><li><input type="checkbox"/> Geophysics</li><li><input type="checkbox"/> Earth Observation</li><li><input type="checkbox"/> Astronomy</li></ul> <input type="button" value="Search"/>

### Create a new requirement

Related to:

- [User interface](#)
- [Job submission](#)
- [Data management](#)

## R0005 - Parallel jobs execution

<b>Version</b>	2
<b>Creation date</b>	2004-06-02
<b>Last update</b>	2004-06-03
<b>By</b>	Johan Montagnat
<b>Contact</b>	iblanque@dsic.upv.es
<b>Description</b>	Parallel execution is mandatory for some applications. The middleware is expected to enable parallel job submission (by specifying a number of hosts to allocate at submission time, at least at a site scale) and provide a message passing interface (preferably MPI)
<b>Area</b>	HEP Biomed Generic
<b>Priority</b>	Very high
<b>Type</b>	Function Performance
<b>Requirement date</b>	PM1
<b>Implementation hint</b>	MPICH-G2 is satisfying: it has already been experimented. The job submission system need to take into account parallel jobs.
<b>Status</b>	Required

[Update requirement](#)

### Applicable usecases

AWG AWG UC#PJ  
HEPCAL UC#paralleljobsubmit

### Track record

Date	User	Comment
2004-06-02	iblanque	PM0: Since late 2003, EDG middleware 2.1 provides access to MPICH-G2
2004-06-02	iblanque	PM0: LCG2 middleware does not support parallel jobs

# Requirements in a nutshell

**Critical for development**

mid-term requirement

long-term requirement

- **1. Large user community**
  - **anonymous/group login**
  - **multiple VOs/subgroups**
- **2. Data management**
  - **application metadata**
  - **data updates** and **data versioning**
- **3. Security**
  - **application data filtering**
  - **disk / network encryption**
- **4. Limited response time**
  - **fast queues for short jobs**
  - **high priority jobs/privileged users**
- **6. Interactivity**
  - **communication between user interface and WNs**
- **7. Parallelization**
  - **MPI site-wide**
- **8. Pipeline processing**
  - **pipeline description language / scheduling**
- **9. Network**
  - **outbound connectivity**
- **10. Development**
  - **C++/Java APIs**

- User Interface
  - User interface installation and portability
  - User Interface configuration.
  - Application Programming Interface.
  - Group/anonymous login (portals)
- Job requirements
  - Short jobs execution
  - Parallel jobs execution
  - Prioritised jobs execution
  - Multiple data jobs
  - Compound jobs execution (pipelining)
  - Interactive jobs
  - Job access to data
  - Simple job submission
  - Job execution control
  - Job killing
  - Resources reservation

- Data management requirements
  - File names translation
  - File access interface
  - Fine grain control of access rights
  - Group of files
  - Metadata associated to files
  - Access rights delegation
  - Data updates and versioning
  - File name changes
  - Data replication control
  - Data registration, retrieval, and deletion
  - Data access cost estimation
  - Partial file access
  - File browsing
  - Scalability

- Information system requirements
  - Jobs information and status notification
  - Top level information system index
  - Resource brokers index
  - Grid resource browsing
- Storage requirements
  - On-disk encryption
  - Interface to new storage systems
  - Hook on data privacy manager
- Network requirements
  - Communications encryption
  - Outbound connectivity
  - Guaranteed bandwidth

- Security requirements
  - Data privacy
  - User privacy
  - Computations privacy
  - Data encryption and encryption keys
- Operation requirements
  - VO creation
  - User control
  - User login
  - Software package publication
  - Robustness
  - VO management and subgroups