



Enabling Grids for E-scienceE

Joint JRA1/JRA3/NA4 session

M.Lamanna NA4-HEP and LCG ARDA

www.eu-egee.org



- **Introduction**
- **Presentations:**
 - “Generic Application”
 - Biomed
 - ARDA
- **Panel discussion**

- **Objectives**
 - Where are we within NA4 in using the new middleware?
 - Focus on technical key points
 - Way forward?
- **Presentations**
 - “Generic” + Biomed + HEP (ARDA)
- **Panel discussion**
 - Why:
 - Technical discussions ≠ presentations
 - Who:
 - Panelists + YOU
 - What:
 - Preparation work (list of hot issues)
 - Discuss issues emerging from the presentations/discussions

- **Panelists:**
 - JRA1
 - Frederic Hemmer
 - Erwin Laure
 - JRA3
 - Olle Mulmo
 - David Groop
 - NA4
 - Birker Koblitz, Massimo Lamanna, Dietrich Liko (HEP)
 - Johan Montagnat, Ignacio Blanquer (Biomed)
 - Roberto Barbera, Giuseppe Andronico (“Generic” applications)

- **And all the audience!**

- **Discovery services:**
 - More or less missing
- **Security related**
 - Encryption in the storage of data (avoid people with local administrator rights to compromise privacy) and in the communication layer
 - ACL granularity
 - Groups / Users
 - Files / Groups of files (“directories”)
 - Hook on privacy management through applications
 - Stub to provide anonimisation by 3rd party applications
- **Grid access via shell:**
 - Value and limitation of the current approaches. Possible extensions
- **Job Definition Language:**
 - Motivation: ease of use and interoperability
- **Job submission services:**
 - Missing functionality: access to stdout/stderr and worker node details during execution
 - Motivation: debugging

- **Compute Element:**
 - Compare the experience of DIRAC, gPTM3D, others... with the current plans
 - Motivation: best exploit successful experience on LCG2 and other systems
 - Goals: Will this work on gLite? Should gLite explicitly account for it?
- **Metadata:**
 - Present the current status of the joint ARDA/gLite effort for feedback and discussion
- **Software installation services:**
 - Compare the use cases (experience) not only in HEP
 - Motivation: it is a critical component. More applications should be involved (almost only HEP so far)

- **As a follow-up of the 3rd ARDA workshop, we have agreed to have a persistent link with OSG**
- **Preparation for the next OSG workshop going on**
 - Mid December
- **Agreed to select issues to discuss**
- **Current brainstorming ideas from OSG**
 - Data Management
 - What is the definition, role and scope of LFNs
 - What are the responsibilities of the Replica Catalog - what are the precise means of the requirements and constraints of its capabilities
 - What are the responsibilities of the File Catalog - what are the requirements
 - JDL
 - Interesting subject
- **Sign in the arda@cern.ch (from arda.cern.ch) to join our meeting (the idea is to have one meeting every fortnight (phone + VRVS))**



Enabling Grids for E-science

1. Position in the queue;
2. Splitting information (if applicable);
3. Estimated time before running;
4. Estimated cost (arbitrary units);
5. Actual cost (arbitrary units);
6. Time and date of submission;
7. Time of start of execution;
8. Time of completion;
9. Priority;
10. Completion status;
11. CPU time used;
12. Real time elapsed;
13. Input I/O (amount and rate);
14. Output I/O (amount and rate);
15. CPU time left;
16. Executable running;
17. Dataset accessed;
18. CE, WN and SE used;
19. **Current stdout, stderr;**
20. Job status;
21. Job environment variables;
22. User who has submitted the job;
23. User attributes for the job specified in the job catalogue;
24. Queue used;

USE CASE: JOB MONITORING

Identifier	UC#jobmon
Goals in Context	Monitor a single running job
Actors	User
Triggers	Curiosity; need to know the status of a job;
Included Use Cases	Grid login;
Specialised Use Cases	
Pre-conditions	A job has been submitted; User knows the job identifier;
Post-conditions	
Basic Flow	<ol style="list-style-type: none"> 1. User submits a query using the job identifier as key and information about the job. 2. The amount and type of information retrieved can be sp options in the query.

From the HepCAL document

- **Compare the experience of DIRAC, gPTM3D, others... with the current plans**
 - Agent-based “mobile systems: in DIRAC these agents are distributed by the LCG2 WMS and then they collect tasks from an experiment-specific service
 - Robustness
 - Handcrafted control channels: in DIRAC, the communication layer is done via instant-message technologies
 - Duplication of effort on the application side?
 - Provisions for interactivity control
 - gPTM3D and ALICE (and maybe others)
 - Response time
 - Quality-of-service and (low) latency requirements

- **Data encryption**
 - on disk AND during network transmission
 - raise the problem of encryption keys control
- **Fine grain access control**
 - access control needed at individual AND group level
- **Both data and metadata concerned**
 - metadata may be even more confidential than data
 - same level of protection/access pattern expected
- **Hook on privacy manager**
 - sometimes need to control data privacy at the application level
(need middleware hook to filter data being set out)
- **Need to protect privacy of grid users**
 - hiding user names and programs being executed