



Enabling Grids for
E-science in Europe

*NA4 open meeting
Catania, July 15, 2004*

Overview of biomedical work

Johan Montagnat



- Quick-off meeting in Cork, April 18th
 - set up executive and technical structures
 - pilot applications identification
 - external application integration procedure
- Web
 - <http://egee-na4.ct.infn.it/biomed/>
 - project-eu-egee-na4-biomed-applications@cern.ch
- Human resources
 - 4FTE became 5 (LPC + CREATIS = LPC + CREATIS + IBCP)
 - CNB, Madrid: Angel Merino
 - UPV, Valencia: 4 * 0.25 FTE coordinated by Ignacio Blanquer
 - LPC, Clermont-Ferrand: Yannick Legré
 - CREATIS, Lyon: one candidate, starting September 1st
 - IBCP, Lyon: one candidate to hire very soon

- VO management
 - internal VO management (Y. Legré)
 - biomed VO hosted at CC-IN2P3
- Technical team
 - middleware expertise
 - testing internal to biomed applications
- Participation to other groups
 - Application Working Group (I. Blanquer, J. Montagnat)
 - Project Technical Forum (C. Blanchet, J. Montagnat)
 - Middleware Security Group (C. Blanchet)
 - Data management (J. Montagnat)
- Web
 - Webmaster: Y. Legré
 - Detailed meetings minutes

- **Role**
 - Group of technical people (acquiring and) providing expertise, at the interface between middleware developers and users
- **Contact**
 - project-eu-egge-na4-biomed-twg@cern.ch
- **Composition**
 - 5 funded engineers
 - 2 test team representatives
 - 8 unfunded participants
- **First objectives**
 - Test middleware
 - Acquire expertise
 - Provide feedback

- Meetings
 - biomed plenary: Cork (April), Lyon (June), Catania (July), Madrid (October)
 - AWG (phone, weekly)
 - MWSG: CERN (May, June), Stockholm (August)
 - PTF: CERN (June)
 - JRA1: CERN (June)
 - ARDA: CERN (June)
 - JRA2: phone (June)
 - BIRN: CERN (June)
- Participation to NA4 tasks and deliverable
 - TNA4.1: biomedical requirements
 - DNA4.1: applications interface
- Testing and training
 - Access to LCG2
 - LCG2 deployment
 - Pilot applications deployment
 - Access to glite

TNA4.1: Requirements

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

Critical for development

mid-term requirement

long-term requirement

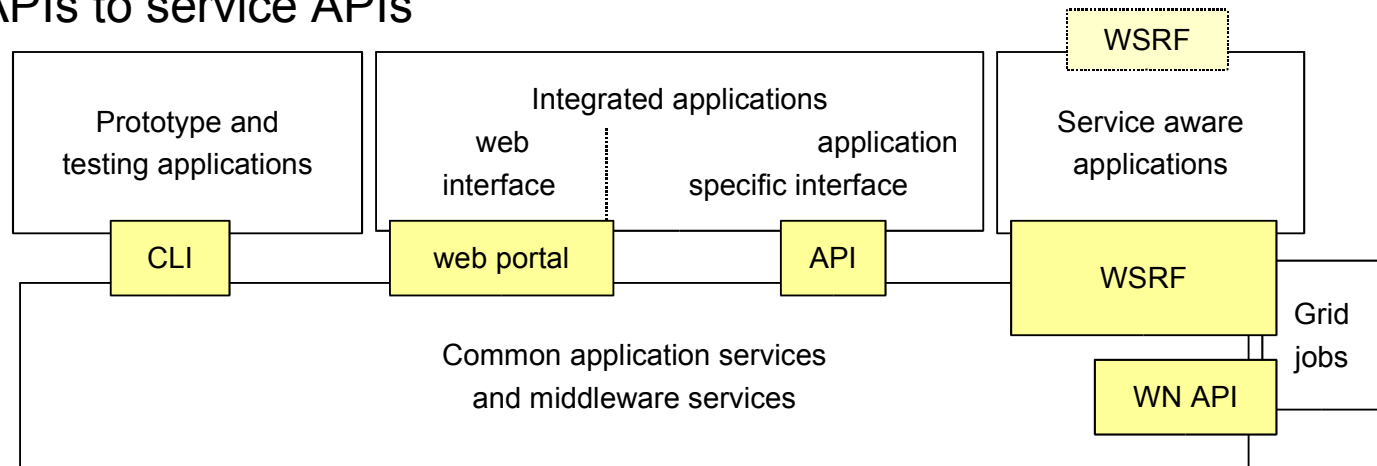
- **1. Large user community**
 - **anonymous/group login**
 - **multiple VOs/subgroups**
- **2. Data management**
 - **application metadata**
 - **data updates** and **data versioning**
- **3. Security**
 - **application data filtering**
 - **disk / network encryption**
- **4. Limited response time**
 - **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**

DNA4.1: application interfaces

- GENIUS: generic portal
- Interface to middleware services
 - exposition to middleware
 - required services

R/S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1				X											
2				X											
3									X						
4									X						

- Interface level
 - Application **Programing** Interfaces!
 - from APIs to service APIs



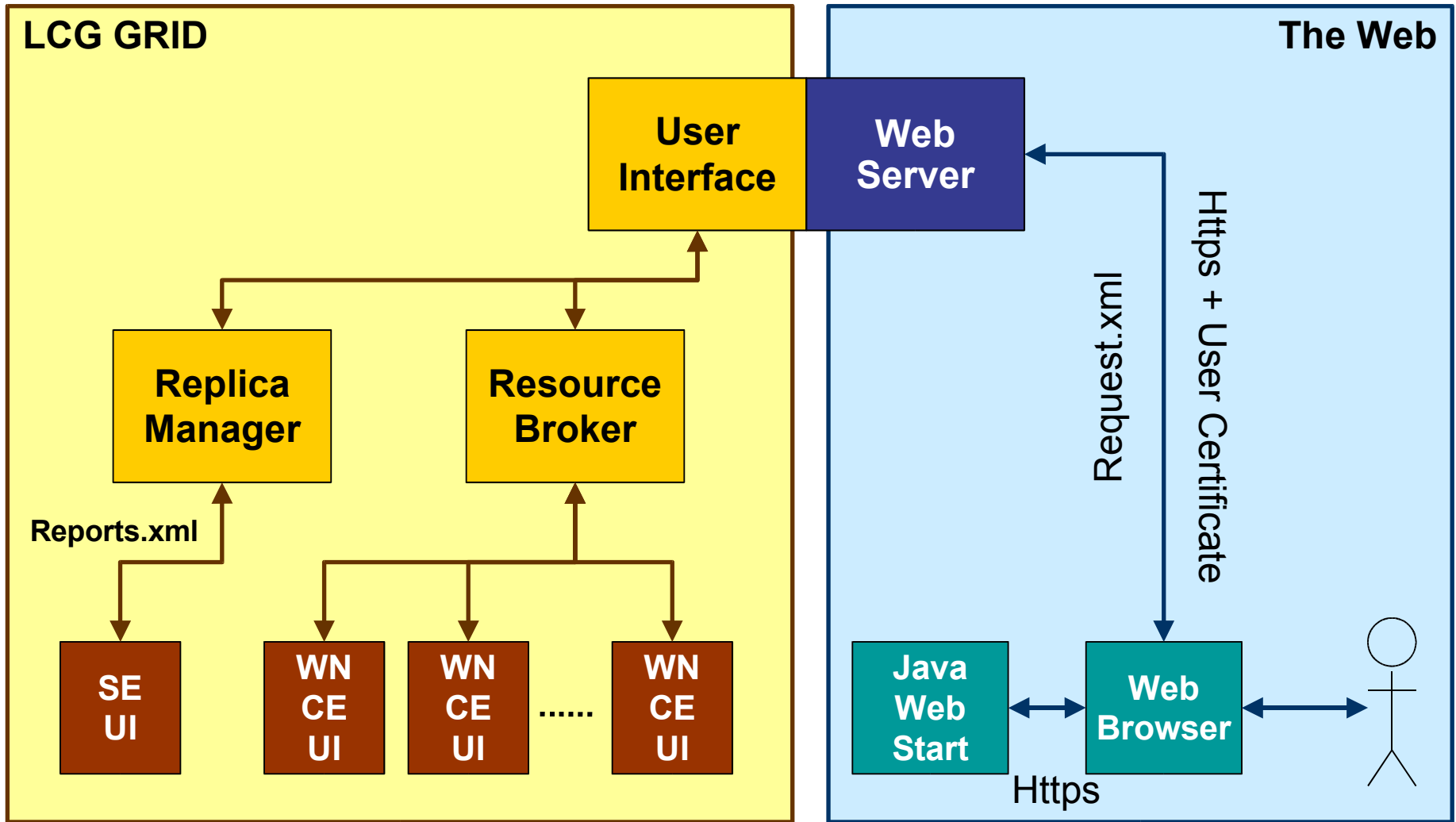
- Deploy applications and testing
 - Access to infrastructure
 - Pilot applications for early testing
 - Internal applications
- Identify health-related applications
 - External application integration procedure
 - Mammogrid
 - Other contacts (Geneva hospitals, Phylogeny BBE, Gene expression MPBA)
- Deploy biomed resources
 - Hardware resources
 - LCG2 deployment

- LCG2 infrastructure
 - First “Hello biomed world” job executed on June 23
 - VO hosting and User Interface at CC-IN2P3
 - VO acceptance at CNAF
 - Lot of difficulties related to VO:
 - ★ Adding new VOs in LCG2 (tedious manual configuration, LCFG overwriting new VOs every night...)
 - ★ VO acceptance (political level)
 - ★ VO acceptance (technical level: RB, resources...)
 - Jobs submitted from a biomed UI at IBCP
 - RC to be hosted at CC-IN2P3
- glite prototype
 - First glite job executed on June 17
 - Few testing done today, focus on LCG2
- Low expertise level

- **CDSS: Clinical Decision Support System (UPV)**
 - Application that Extracts Medically Relevant Knowledge from a Large Set of Information with the Objective of Guiding the Practitioners in their Clinical Practice.
 - Trained Databases Available
 - ★ Classification of Tumours of Soft Tissues
 - ★ Diagnostic and Classification of Thalassemia and Other Anaemia
 - ★ Bioinformatics: Study of the Human Genome
- **Migration to LCG2**
 - CDSS ported on LCG2 (UPV installation)
 - To high job pay-off
 - Service oriented architecture
 - Redesign needed for taking into account the batch-oriented nature of LCG2



Pilot applications report



Pilot applications report

- GPS@: genomics web portal (IBCP)
 - Already ported on EDG through an application C++ API
 - <http://gpsa.ibcp.fr/>

ibcp Institut de Biologie et Chimie des Protéines

Grid Protein Sequence @analysis **Data GRID WP10**

GPSA intends to provide Biologist with a **GRID web portal** dedicated to protein sequence analysis.

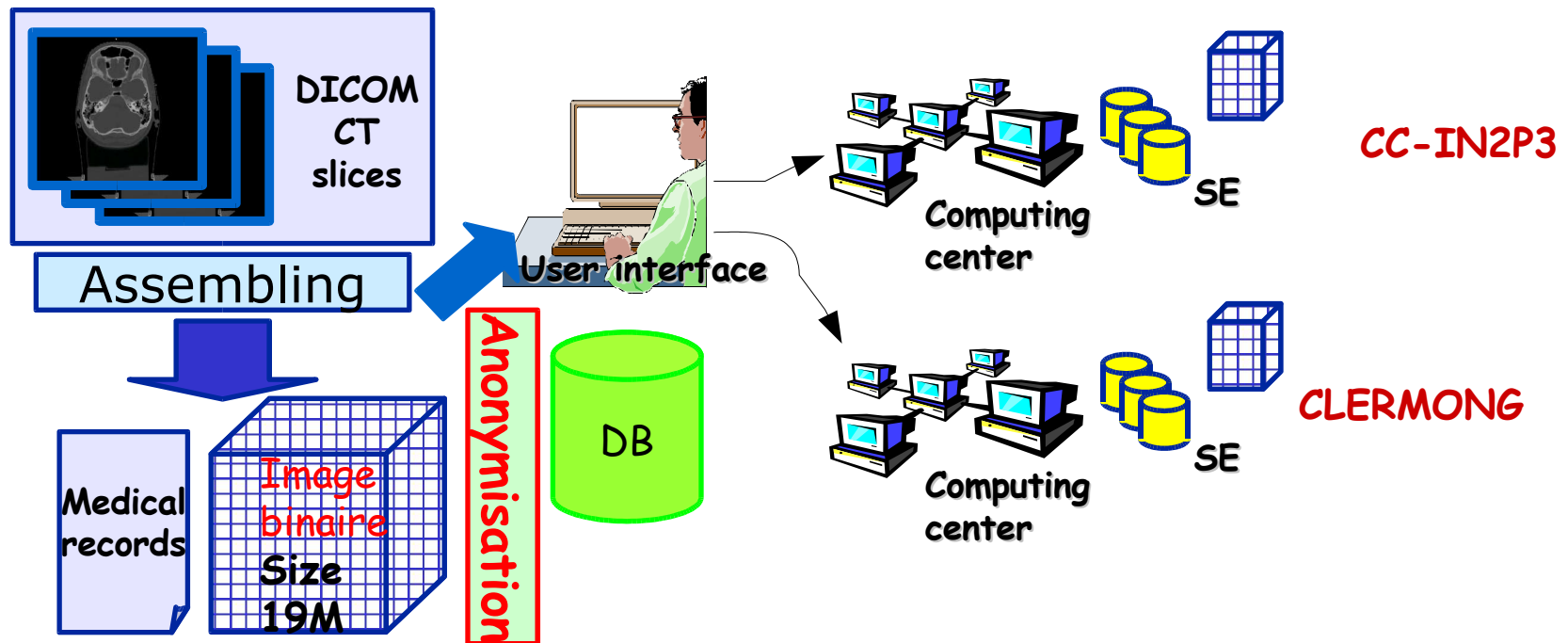
[[GPSA](#)] [[mySEQ](#)] [[HELP](#)] [[REFERENCES](#)] [[NPS@](#)] [[PBIL-Gerland](#)] [[PBIL](#)]

Protein analysis methods available on GPSA

- Adaptation to LCG2 middleware in progress
 - Application API migration
 - XML description of bioinformatic jobs
 - Need last RedHat distribution for the web portal
 - service certificate needed
- Tested on LCG2 for 4 algorithms (PattInProt, secondary structures prediction)

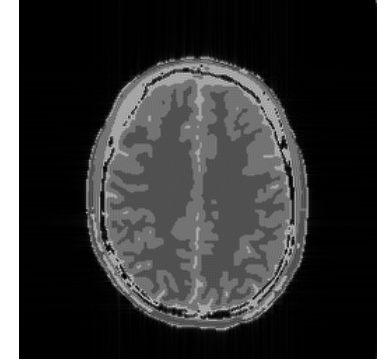
Pilot applications report

- GATE: radiotherapy planning (LPC)
 - Demonstrated at the last EDG review
 - Ported to LCG2



- Deployment
 - On CC-IN2P3
 - On Clermont-Ferrand site

- **SIMRI: parallel MRI simulator (CREATIS)**
 - migration to MPICH-G2
 - on-going tests at CC-IN2P3
 - no parallel jobs on LCG2
 - on-going test at CINES (supercomputing center)
- **g-PTM3D: interactive radiological images processing (LAL)**
 - radiological images manipulation
 - interaction and jobs execution
 - LCG2 pay off compensated by internal scheduler
 - interaction through bypass unsatisfying (std input/output used for communication, far too slow)
- **Mammogrid**
 - developed on AliEN
 - to be ported on glite



- **Material resources deployed**
 - Clermont-Ferrand: > 100 nodes, waiting (?) for LCG2 certification
 - UPV: 3 AMD/Pentium PCs for LCG2 services, 20 processor cluster running RH7.1
 - IBCP: 10 PCs dedicated to LCG2
 - CREATIS: up to 8 PCs available
 - CNB: 20 processors
- **Software deployed**
 - Clermont-Ferrand: UI, CE, SE, WNs, GENIUS web portal
 - UPV: RB, MDS, BDII, UI, CE, SE, WNs
 - IBCP: UI, CE, SE, WNs, 1 web portal
 - CREATIS: problems encountered installing multiple services on PCs
 - CNB: UI, CE, SE, WNs, hardware compatibility problems with RH7.3
- **Storage Resource Broker (SRB) testing**
 - Production stable at CC-IN2P3
 - Evaluation at Clermont-Ferrand and I3S

- Work done
 - Ready to go on LCG2
 - Pilots ported to LCG2
 - Applications emerging
 - Contributions to NA4 general activities
- Main problem encountered and weaknesses
 - Late access to LCG2, low level of expertise
 - Weight of LCG2 deployment
 - Weak response from e-health EU projects
 - Loose integration of biomedical community – external evaluation mentioned it as one of the few risk factor
- Future and roadmap
 - Infrastructure testing on-going
 - Applications porting on-going
 - Mammogrid integration on glite
 - Demonstration for the first project review (GPS@?)