



Enabling Grids for E-science

EGEE-II NA4 Bioinformatics Applications Status - 26 Sept 2006

Christophe Blanchet (CNRS IBCP)

On behalf of EGEE bioinformatics users

www.eu-egee.org












Status of Bioinformatics in EGEE-2

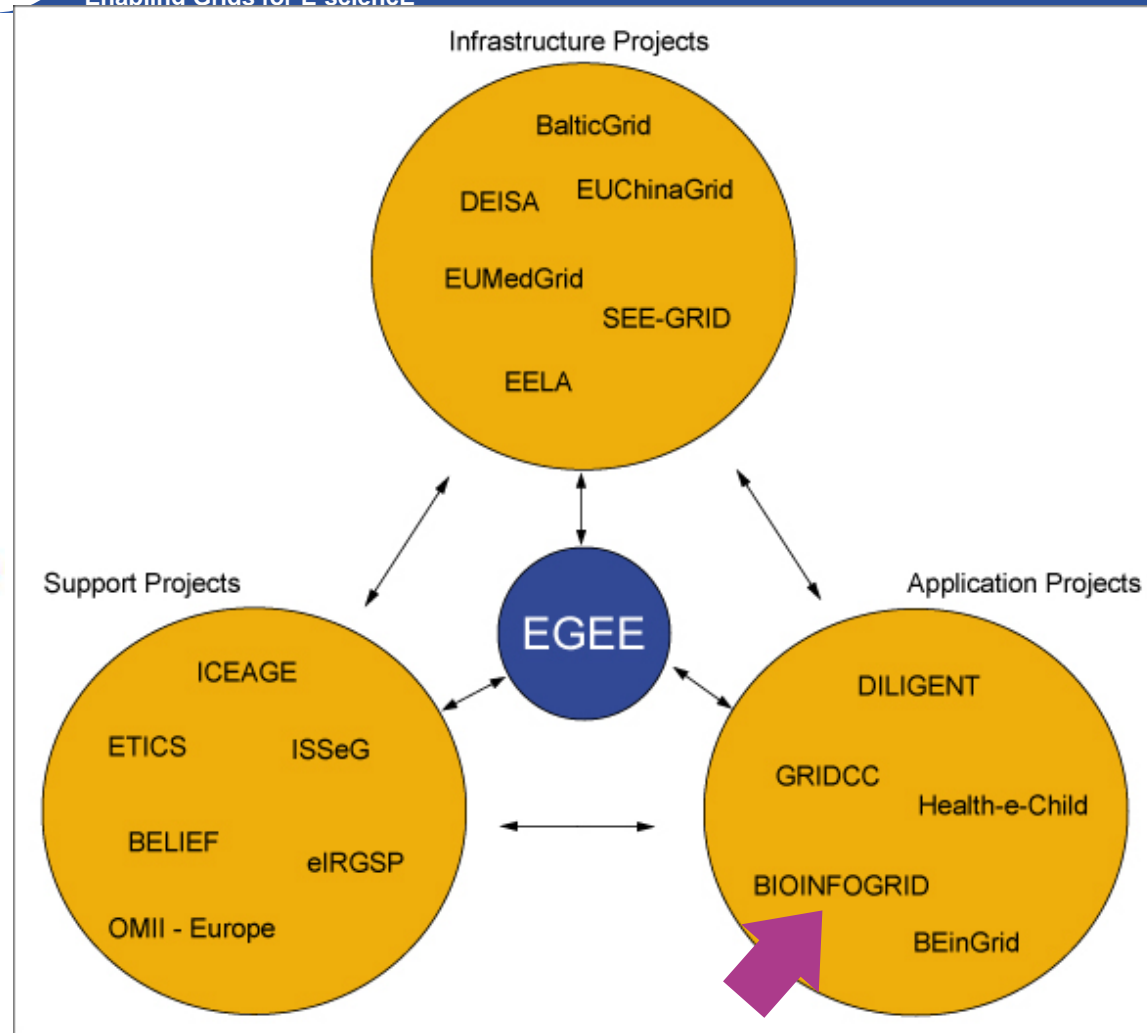
- **Work on requirements (participate to TCG)**
 - Define requirements and prioritize them
 - Give feedback about satisfaction with middleware developments
- **Deploy applications on the production platform**
 - 10 applications
 - Training, collaboration with ROC, Add new resources
 - Give strong FEEDBACK of GRID-ADDED value
- **Defined Priority:**
 1. Deploy updatable databases
 - > First EGEE-bioinformatics workshop on « Grid data replication, consistency and requirements », 26th May 2006, Pisa, Italy
 - Joined meeting between EGEE and EMBRACE European project
 2. Deploy legacy programs with special I/O
 3. Security of medical and industrial data
 4. Complex Workflow
 5. Portal for end-users: biologists, ...
- **Contact: Christophe.Blanchet@ibcp.fr**

Today: Biomed meeting in EGEE'06 Conference, Geneva, Sept 25-29, 2006, <http://www.eu-egee.org>

Presented
Today

	GPS@	CNRS IBCP	Christophe Blanchet (IBCP) Christophe.Blanchet@ibcp.fr	Prototype	http://gpsa-pbil.ibcp.fr/
	SPLATCHE	External	Dr. Nicolas Ray nicolas.ray@zoo.unibe.ch	Production	http://cmpg.unibe.ch/software/splatche/
	Large-scale Pathway Analysis	CNRS IBCP	Ralf Herwig herwig@molgen.mpg.de	Porting	
	bioDCV	INFN, ICTP (E-GRID)	Cesare Furlanello furlan@itc.it	Prototype	http://biodev.itc.it/
	Phylojava	CNRS	Manolo Gouy mgouy@biomserv.univ-lyon1.fr	Porting	http://pbil.univ-lyon1.fr/software/phylojava/phylojava.html
	BiG	UPV	Alexandre Dehne Garcia dehneg@prabi.fr	Porting	
	Superlink-online	TAU	Ignacio Blanquer iblanque@dsic.upv.es	Porting	
	3DEM	TAU	Prof. David Horn horn@post.tau.ac.il	Feasibility	http://bioinfo.cs.technion.ac.il/superlink-online/
	CAST	TAU	Mark Silberstein marks@techunix.technion.ac.il	Feasibility	
	3DEM	CNB/CSIC	Jose-Maria Carazo carazo@cnb.uam.es	Porting	http://3dem.ucsd.edu/
	CAST	UCY	George Tsouloupas (UCY) georget@ucy.ac.cy	Feasibility	
	Dengue Docking Project	UCY	Maria Poveda (UCY) mpoveda@cs.ucy.ac.cy	Feasibility	
	Dengue Docking Project	CSCS	Michael Podvinec (Biozentrum Basel, CH)	Prototype	

- **« A European Model for Bioinformatics Research and Community Education »**
- **Goals:**
 - simplify and standardize the way in which biological information is served to the researchers who use it.
 - Integrating biological data and bioinformatics tools in grid
- **Network of Excellence (2005-2010)**
 - From Feb 1st, 2005
 - partners: EBI (PI), EMBL, SIB, CNRS, MPI_MG, INRA, ITB CNR, CNB, ...
- **Funded by the European Union (EU-FP6, LHSG-CT-2004-512092)**
 - EMBRACE uses a test problem driven development method. The services will be developed through a set of test problems, which will use tasks from real biological research, designed to stretch the system in critical ways



- **Details agenda online.**

<http://indico.cern.ch/sessionDisplay.py?sessionId=130&slotId=0&confId=1504#2006-09-26>

<p>[300] Status of Bioinformatics in EGEE-2 by Dr. Christophe BLANCHET (CNRS IBCP) (Conf. Room 2: 14:00 - 14:05)</p>
<p>[310] Service to Encrypt Biological Data on Grid by Mr. Remi MOLLON (CNRS IBCP) (Conf. Room 2: 14:05 - 14:15)</p>
<p>[218] Enabling bioinformatics applications to access files over the grid via a GFAL plugin to Parrot by Dr. Giacinto DONVITO (INFN-BARI) (Conf. Room 2: 14:15 - 14:25)</p>
<p>[303] GPS@, Web interface for Protein Sequence Analysis on Grid by Dr. Christophe BLANCHET (CNRS IBCP) (Conf. Room 2: 14:25 - 14:35)</p>
<p>[345] Blast In Grid (BIG) by Dr. Ignacio BLANQUER ESPERT (DSIC) (Conf. Room 2: 14:35 - 14:45)</p>
<p>[134] Bringing 3D-EM to the Grid by Mrs. Germán CARRERA (Centro Nacional de Biotecnología CSIC) (Conf. Room 2: 14:45 - 14:55)</p>
<p>[309] Parallelised Monte-Carlo simulation of large biological networks using the EGEE grid by Dr. Christoph WIERLING (Max-Planck-Institute for Molecular Genetics) (Conf. Room 2: 14:55 - 15:05)</p>
<p>[327] Status of the Biomedical Applications in EELA by Vicente HERNANDEZ-GARCIA (Polytechnical University of Valencia) (Conf. Room 2: 15:05 - 15:15)</p>
<p>[174] A GRID PLATFORM FOR ITALIAN BIOINFORMATICS by Dr. Giacinto DONVITO (INFN-BARI) (Conf. Room 2: 15:15 - 15:25)</p>
<p>[308] Perspectives in Bioinformatics area by Dr. Christophe BLANCHET (CNRS IBCP) (Conf. Room 2: 15:25 - 15:30)</p>

PERSPECTIVES

In Bioinformatics Area

- **Very different applications ...**
 - Different requirements and priorities
 - Different resources involved
 - Hardware
 - Human
 - Different scientific communities adressed
 - But all are biologists
 - Don't care of the « infra »-structure
- **... but some common requirements**
 - Data
 - Deploying updatable databases
 - Security of biological data (medical or industrial)
 - Tools
 - Integrating numerous, complex programs
 - Legacy application
 - Portal and user interfaces
 - MPI parallel applications

- **Short Jobs (<5 min): SDJ workgroup**
 - SDJ WG has defined some CE setup rules to decrease grid middleware overhead to ~2 min
 - But only one site (LAL) is enabled (at least publishing it!)
- ⇒ **deploying SDJ recommendations on other biomed sites, with adequate publication (CE named with « sdj » tag)**

- **Data management: still some security issues in gLite data management system.**
 - DLI interface cannot be activated on biomed LFC
 - No scheduling according to replica (SFN, « closeSE »)
 - Why ?
 - Activating DLI means publishing all the SFN to all (VO?) users
 - With current SE, knowing a SFN means having access to.
- ⇒ **RB could use user delegation to access DLI interface on LFC**

- **Others: software deployment tools, efficient tools to manage jobs, lightweight UI, ...**

- **Agenda**

- Opening and status of EGEE project
- Applications status and feed-back
- Key themes:
 - Biological data on EGEE, access and security.
 - *Data Virtualization: Enabling bioinformatics applications with grid data access*
 - SE DPM ? GFAL ? Parrot/Perroquet ? Fuse ?
 - *Security: Working on security issues about biological data*
 - MDM ? EncFile ?
 - Computation management
 - *Portal and user interfaces*
 - *Workload mgmt: short job (SDJ), prioritized job, pilot job*
 - *Complex job: parallel job (MPI) , application built with several programs*
- AOB
- Conclusions

- **Location and date**

- Institute of Biology and Chemistry of Proteins, Lyon, France
- Friday October 20, 2006

***** Partners are invited to host some of the next meetings *****