WP10 Biomedical Applications

Vincent Breton (CNRS)
breton@clermont.in2p3.fr

Outline

- Overview of objectives and achievements

- Lessons learned

- Exploitation plan

- Concluding comments

- Questions and discussion

Objectives

- To demonstrate the relevance of grids for life science

- To test the EDG middleware and feedback requirements to the middleware developers

- To raise awareness on the impact of grids in the life science community

Demonstrate the relevance of grids for life science

- ◆ **When DataGrid started**, no clear view of where grids should apply to life sciences

- ◆ **First year** dedicated to identify potential applications of grids to life science
  - WP focus enlarged from biology to biomedical sciences (medical imaging)
  - Deliverables D10.1 on requirements and D10.2 on applications identified for deployment
  - First demo on distributed computing for a bioinformatics web portal

- ◆ **Second year** dedicated to the applications deployment on EDG testbed 1 and testbed2 (EDG1.4.x)
  - Limited deployment of 3 out of 10 applications by March 2003
  - Second demo on medical images handling in a grid environment
  - Deliverable D10.3 (March 2003)

Demonstrate the relevance of grids for life science (II)

- **Third year** dedicated to the applications deployment on EDG testbed 2 and testbed 3 (since October 20st, 2003)
  - Large scale deployment of biomedical applications was successfully achieved

- **Results** presented at conferences and submitted to several computer science journals
  - IST 2003 in Milano, GGF Life Science Research Group, ISGC 2003, PharmaGrid,…
  - 3 papers accepted for Healthgrid 2004 conference proceedings in Methods of Information in Medecine
  - 2 papers submitted to a special issue of Parallel Processing Letters
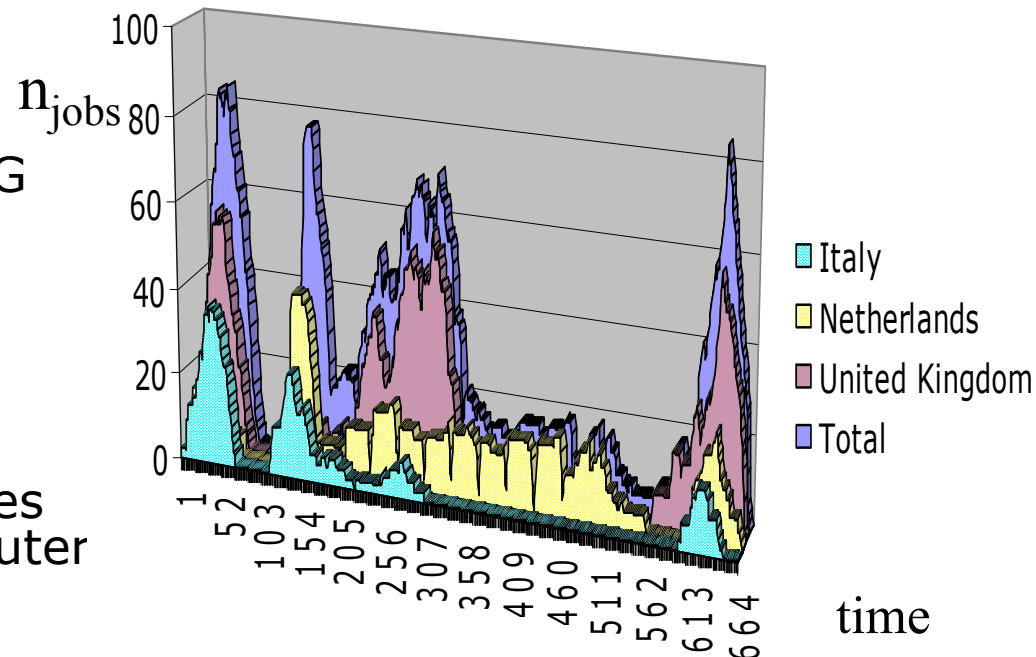
Successful deployment on EDG testbed2 of large scale phylogenetics analysis (450 jobs) (credit : T. Silvestre, CNRS)

DataGrid : status of biomedical applications

Legend:
- 🟩 Deployed
- 🟧 Under test
- 🟥 Not deployed

◆ **Bio-informatics**

- **Phylogenetics : BBE Lyon (T. Sylvestre)**
- **Search for primers : Centrale Paris (K. Kurata)**
- **Bio-informatics web portal : IBCP (C. Blanchet)** *NEW*
- **Parasitology : LBP Clermont, Univ B. Pascal (N. Jacq)**
- **DNA chips analysis portal : Karolinska (R. Martinez)** *NEW*
- **Geometrical protein comparison : Univ. Padova (C. Ferrari)**

◆ **Medical imaging**

- **MR image simulation : CREATIS (H. Benoit-Cattin)**
- **Secure access to medical data : CREATIS (J. Montagnat), to be interfaced with SE**
- **Mammographies analysis ERIC/Lyon 2 (S. Miguet, T. Tweed)** *NEW*
- **Simulation platform for PET/SPECT based on Geant4 : GATE collaboration (L. Maigne)** *NEW*

# Applications on testbed3 and beyond...

◆ GPSA
Grid Protein Sequence Analysis

   ▪ Web portal devoted to protein sequence analysis on the GRID

   ▪ Oriented to high-sized genomic data analysis

   ▪ Providing biologists with a simple and powerful interface for the facilities of the DataGrid testbed

◆ Karolinska DNA chips analysis portal

◆ GATE platform for nuclear medical imaging simulation (3rd year demo)
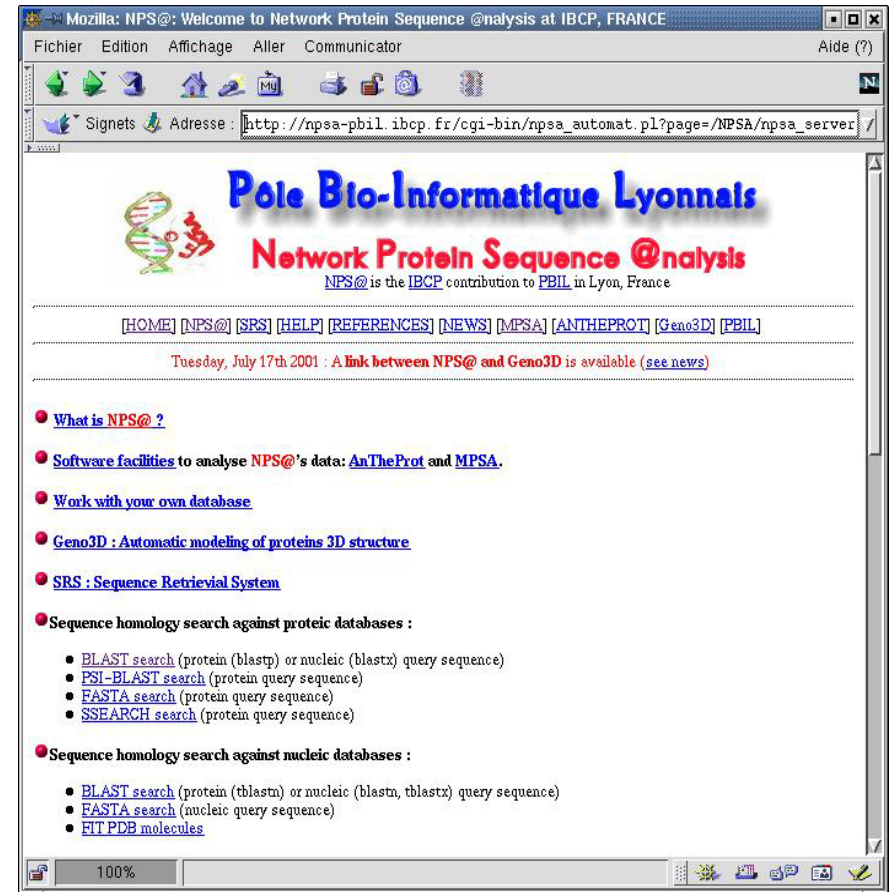
Achievements : requirements feedback

- ◆ Requirements described in deliverable D10.1 (June 2001)

- ◆ October 2003 : contribution to the joint list of use cases (WP8, WP9, WP10) edited by the Application Working Group

- ◆ Requirements beyond HEP needs can be roughly summarized as follows :

  - ▪ Fine grain Access Control List (Storage Elements, Job submission)

  - ▪ Ability to submit parallel jobs (MPI on grid clusters)

  - ▪ "Grid in the box": easy installation and configuration of grid nodes

Achievements : test of EDG middleware

◆ Testbeds evaluation described in several documents :

- March 2003 : deliverable D10.3 on our experience of testbed 2 (EDG1.4)

- June 2003 : contribution to the joint list of recommendations (WP8, WP9, WP10) edited by the Application Working Group

- November 2003 : deliverable D10.4 on our experience of testbed 3 (EDG2.0, 2.1)

◆ Final version of EDG middleware (2.1) opens new perspectives

- Largely improved stability

- New functionalities: MPI, C++/Java API for job submission

- Beyond EDG2.1, proposed solution to handle confidentiality (see David Kelsey's talk)

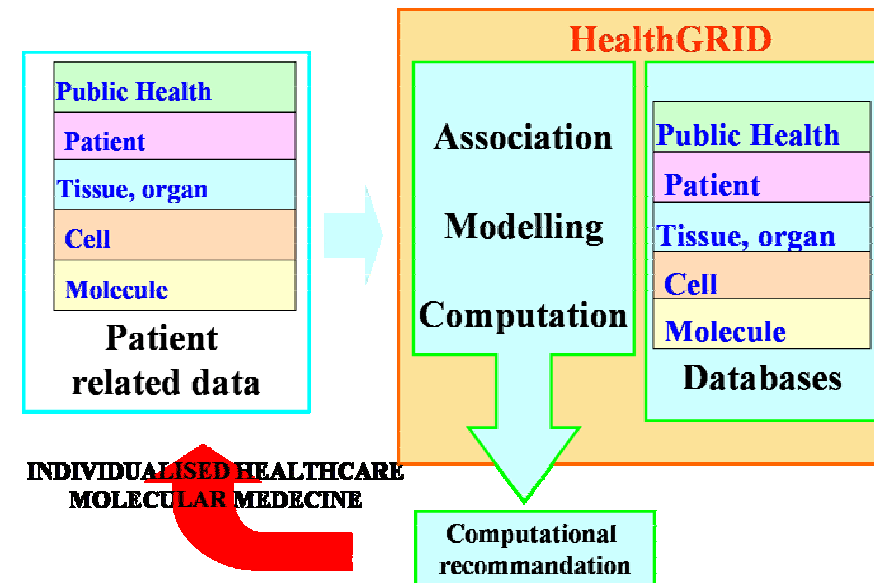# Achievements: tests of EDG middleware 2.0 and 2.1 (from D10.4)

Test title	EDG 2.0	EDG 2.1
Grid login	Passed	Passed
Replicas listing	Passed	Passed
edg-job-list-match	Passed	Passed
Browse resources using the LRC	Passed	Passed
Browse resources using the RMC	Passed	Passed
Remove an alias	Passed	Passed
Upload and register a file	Passed	Passed
Replicate files	Mostly passed	Passed
Retrieve files	Partly passed	Passed
Delete files	Passed	Passed
List replicas	Mostly passed	Passed
Register hundreds of files	Partly passed	Passed
Getting file access cost	Passed	Passed
Getting closest replica	Passed	Passed
Job submission	Partly passed	Mostly passed
Job output retrieval	Passed	Passed
Job control	Passed	Passed
MPI jobs	Mostly failed	Under test
C++ API	Mostly passed	Mostly passed
Interactive jobs	Failed	Partly passed

Raise awareness on the impact of grids in the life science community

- ◆ When project started, life science community showed scepticism and mistrust
  - ▪ "physicist" project

- ◆ WP10 meetings were widely open right from the beginning
  - ▪ Invited talks from European Bioinformatics Institute, European Molecular Biology network representatives

- ◆ A real momentum was gained during year 2002
  - ▪ Participation to the project of a grid for bioinformatics (European Bioinformatics Institute)
  - ▪ Up to 34 users coming from 15 laboratories in 4 different countries (April 2003)

- ◆ Year 2003 : birth of the Healthgrid initiative

The Healthgrid initiative

◆ Healthgrid: eInfrastructure for health (life sciences, drug discovery, healthcare,…)

- Long term vision : no single project can make it happen

◆ History

- First invitation to present DataGrid biomedical activities at a conference on the synergy between bio- and medical informatics in December 2001

- Creation of the Healthgrid cluster of projects in September 2002

- First and second Healthgrid conferences in Lyon (January 2003), Clermont-Ferrand (January 2004), next in Oxford (2005)

HealthGRID

Public Health
Patient
Tissue, organ
Cell
Molecule

Patient related data

Association
Modelling
Computation

Public Health
Patient
Tissue, organ
Cell
Molecule

Databases

INDIVIDUALISED HEALTHCARE
MOLECULAR MEDECINE

Computational recommandation

The Healthgrid initiative (II)

- The Healthgrid initiative provides a glue between the projects
  - To foster exchanges between projects, end users and technology developers
    - To avoid reinventing the wheel
    - To improve the take-up of grid technology

  - To disseminate information on grids for health
    - Summaries and links to health related grid projects
    - Available tools (software platforms, middleware,…)
    - Tutorials
    - Conferences

  - To promote standards
    - Involvement in GGF Life Science Research group

- Following Healthgrid conference in 2004, drafting of a White paper under way

Limitations

- Limited resources/time to test thoroughly testbed3

  - New functionalities of great interest for WP10 were made available (parallel job submission, data management)

- No task force ever built between WP10 and middleware work packages

- Installation of a grid node in a biomedical laboratory was not achieved during the project lifetime

  - Node installation and configuration

Lessons learned

- The importance of dialog among applications

  - Application Working Group experience was extremely positive
  - Joint list of use cases of direct use for the definition of EGEE middleware

- The importance of dialog with middleware

  - Usage of EDG middleware requires direct interaction with middleware groups (integration team)

- The importance of providing a stable environment to attract the biomedical community

  - Some external WP10 users were lost during the third year of the project because of testbed2 instability

- The importance of resource allocation within an application work package

  - WP10 resources allocated too close to "end"-users

Exploitation plan

- **EGEE**

  - Biomedical activity within NA4 will take advantage of the experience acquired in DataGrid

  - 1 or 2 applications are selected for early deployment on EGEE infrastructure (GATE medical imaging simulation platform)

- **Other FP6 projects**

  - Euromedim2 (NOE, LifeSciHea, DG Research) : simulation for molecular imaging in a grid environment

  - Embrace (NOE, LifeSciHea, DG Research) : grid for bioinformatics

- **Healthgrid, an eInfrastructure for health**

  - Healthgrid white paper including WP10/AWG work on requirements and use cases

Concluding comments

- ◆ Initial objectives were reached
  - To demonstrate the relevance of grids for life science
  - To test the EDG middleware and feedback requirements to the middleware developers
  - To raise awareness on the impact of grids in the life science community (for instance the Healthgrid initiative)

- ◆ Large scale deployment of grid biomedical applications was achieved for the first time in Europe
  - Bio-informatics web portals have been partially grid-enabled
  - Large scale simulations for medical imaging and radiotherapy are routinely achieved on DataGrid testbed3

- ◆ WP10 has been a pluridisciplinary melting pot
  - Cultural gap between middleware developers and WP10 users
  - However, resources allocated too close to "end"-users

- ◆ A pioneering work has been done on which future projects can build ...