HealthGrid, status and perspectives
Yannick Legré
CNRS
legre@clermont.in2p3.fr
• A few generalities on biomedical grid applications from life science to healthcare

• Pioneering biomedical applications (DataGrid & Instruire experiences)

• Perspectives for biomedical grids in Europe: EGEE, EMBRACE, …

• Disclaimer: I will give the user perspective…
A recent example of the potential grid impact for healthcare

• Last summer heat wave killed more than 10,000 people in one of the European countries

• Mortality rate in excess of 10 to 50% in retirement homes and hospitals unnoticed for 2 weeks

• A monitoring system could have raised the alarm much earlier
  ▪ Requirements: collect information from hospitals and/or funeral services on the number of casualties

• From internet to grid
  ▪ Internet can do it, but it requires manpower to provide reliable information to a centralized web portal
  ▪ Grid added value: database federation (data left in hospitals, à la BIRN) + a grid service for mortality rate computation and monitoring

• Technology is available today…
My definition of a grid

- A grid is:
  - distributed on distant sites with large bandwidth network connection,
  - a large set of computing and storage resources
  - operated by a common or interoperable middleware(s)
  - providing a set of services allowing the deployment of large scale applications
  - by communities of end users
  - structured in virtual organizations
The challenges of a grid for life sciences

- Goal: empower biological analysis workflow on up-to-date exponentially growing data

- Technical challenges
  - data and tools integration: address data heterogeneity and legacy of tools and standards
  - provide the infrastructure and the services (database update and mirroring, grid portals, toolboxes)

- Human challenge: involve end users in the grid game
  - Grids are still very much in development and therefore user-unfriendly
  - Training and support to molecular biology tool and data providers, university hospitals, biology/medicine research centres, …
The challenges of an healthcare grid

• Goal: allow every physician to access a reliable grid for his daily practice
  ▪ New actors: hospitals, physicians, healthcare administrations

• Technical challenges
  ▪ Networking, User interface
  ▪ Grid quality of services (stability, scalability, security, privacy, …)
  ▪ Legal/ethical issues: obey the laws of International countries with respect to personal data ownership and data transfer

• Human challenge: new approach to healthcare delivery
  ▪ Change the way in which doctors/healthcare administrations conceive health
  ▪ Grids are still very much in development and therefore user-unfriendly
  ▪ Training and support to healthcare professionals
Early experience of grid deployment of biomedical applications (2001-2004)

• Several European projects within the 5th framework program
  ▪ Multidisciplinary projects: CrossGrid, DataGrid, EuroGrid, …
  ▪ Projects focused on healthcare issues: GEMSS, Mammogrid, …

• International/National projects
  ▪ UK e-science (Mygrid, …)
  ▪ French ACI grid (Medigrid, GRIPPS, GLOP, …)
  ▪ American projects (BIRN, NDMA, …)
  ▪ Japanese projects (Biogrid, …)
  ▪ Asia-Pacific (APBiogrid, …)
  ▪ Russian projects (RGrid, …)
  ▪ …
DataGrid at a glance

**People**
- 500 registered users
- 12 Virtual Organisations
- 21 Certificate Authorities
- >600 people trained
- 456 person-years of effort
- 170 years funded

**Software**
- > 65 use cases
- 7 major software releases (> 60 in total)
- > 1,000,000 lines of code

**Application Testbed**
- ~20 regular sites
- > 60,000 jobs submitted (since 09/03, release 2.0)
- Peak >1000 CPUs
- 6 Mass Storage Systems

**Scientific Applications**
- 5 Earth Obs institutes
- 10 bio-medical apps
- 6 HEP experiments

**Coordinator:**
F. Gagliardi, CERN

Credit: E. Laure, CERN
DataGrid applications

- DataGrid involved three research communities
  - High Energy Physics
  - Life sciences and medical imaging
  - Earth observation

- Obvious differences in communities organization and computing awareness had to be addressed
  - from the biology community which has no center of gravity and where most end-users are not skilled at using computers…
  - to High Energy Physics community which is extremely organized around CERN with skilled users
  - Setting-up pluridisciplinary coordination at application level turned out very beneficial

- Most user communities are not willing to play guinea pigs of grid technology
  - Added value to be demonstrated early on to biologists
Cultural and social lessons from application deployment in DataGrid

• Cultural: difficulty to establish a common language between the project partners

• Cultural: Need for intermediate levels between middleware developers and “end”-users
  ▪ Application developers aware of middleware issues needed to act as an interface with end-users (biologists, physicists, physicians)

• Social: a grid deployment project brings a real sense of identity
  ▪ DataGrid ended up as a true collaboration with a team spirit, to the benefit of EGEE
DataGrid: successful deployment of biomedical applications

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

- **Medical imaging**
  - MR image simulation: CREATIS (H. Benoit-Cattin)
  - Medical data and metadata management: CREATIS (J. Montagnat)
  - Mammographies analysis ERIC/Lyon 2 (S. Miguet, T. Tweed)
  - Simulation platform for PET/SPECT based on Geant4: GATE collaboration (L. Maigne)
Radiotherapy treatment planning in a grid environment

1°) Obtain scanner slices images

The head is imaged using a MRI and/or CT scanner

2°) Treatment planning

Calculation of deposit dose on the tumor (~1mn):
A treatment plan is developed using the images

3°) Radiotherapy treatment

Irradiation of the brain tumor with a linear accelerator

Credit: L. Maigne, CNRS-IN2P3
Better treatment requires better planning

• Today: analytic calculation to compute dose distributions in the tumor
  ▪ For new Intensity Modulated Radiotherapy treatments, analytic calculations off by 10 to 20% near heterogeneities

• Alternative: Monte Carlo (MC) simulations in medical applications

• The GRID impact: reduce MC computing time to a few minutes

Credit: L. Maigne, CNRS-IN2P3
Conclusion and future prospects

• The parallelization of GATE on the DataGrid testbed has shown significant gain in computing time (factor 10)

• It is not sufficient for clinical routine
• Necessary improvements
  ▪ Job prioritization
  ▪ Submission from the PC farms grid to a supercomputer

Credit: L. Maigne, CNRS-IN2P3
INSTRUIRE: Regional Auvergne Infrastructure

- Operate at a regional level computing and storage resources to meet growing needs of:
  - Research labs
  - Universities
  - Public administrations and services (hospitals, …)
  - Small and Medium Enterprises
- Favors regional-international collaborations

**e.g.** Federation of patient databases for clinical trials and epidemiology in developing countries (Relational DB, SRB)

Preparation and follow-up of medical missions in developing countries of the French NPO “Chain of Hope”

Support to local medical centres in terms of second diagnosis, patient follow-up and e-learning

Current actions:
- China: paediatric neurosurgery
- Burkina Faso: ophthalmology
A look at the future: the HealthGrid vision

In this context "Health" does not involve only clinical practice but covers the whole range of information from molecular level (genetic and proteomic information) over cells and tissues, to the individual and finally the population level (social healthcare).
Key issues for successful adoption of grids in the health area

• Large scale collaboration
  ▪ Not one single project can make it all
  ▪ HealthGrid is a long term vision involving several generations of European and International projects

• Feed middleware developers with specific requirements from life science and healthcare

• Build bridges towards the HealthGrid vision
  ▪ Deploy large scale applications to drive development and demonstrate relevance (see examples)
The HealthGrid initiative

- To provide a place of dialog and exchange between European and international projects
  - Web site: www.healthgrid.org

- To produce collaborative documents
  - HealthGrid White Paper (download)

- To organize conferences and workshops on Health grids

Third European HealthGrid Conference

Location: Oxford (UK)

Dates: April 7th – 9th 2005

Information and registration (available soon): http://oxford2005.healthgrid.org
Ongoing work on biomedical requirements in Europe…

• Life sciences
  - Within the framework of EGEE: please visit our requirements database on http://egee-na4.ct.infn.it/requirements/
  - Within the framework of Embrace, EC funded network of excellence dedicated to a grid for bioinformatics

• Health
  - HealthGrid white paper publicly available on HealthGrid web site since June 15th, 2004 (http://www.healthgrid.org)
  - Technical work to be done to extract requirements from it

• Drug discovery
  - Use case: grid to address diseases of developing countries (Dengue, Malaria)
  - Discussed at last PharmaGrid conference (www.pharmagrid.org)
# HealthGrid White Paper: Contributors

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section Authors</th>
<th>Coordinators</th>
<th>Reviewers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapt.1</td>
<td>I. Blanquer, V. Hernandez, E. Medico, N. Maglaveras, S. Benkner, G. Lonsdale</td>
<td>Ignacio Blanquer Espert, Vicente Hernandez, Guy Lonsdale</td>
<td>Nikolay Tverdokhlebov</td>
</tr>
<tr>
<td>Chapt.2</td>
<td>K. Dean, S. Lloyd</td>
<td>Kevin Dean, Sharon Lloyd</td>
<td>Siegfried Benkner</td>
</tr>
<tr>
<td>Chapt.3</td>
<td>K. Hassan, R. Mc Clatchey, S. Miguet, J. Montagnat, X. Pennec</td>
<td>Richard McClatchey, Johan Montagnat, Mike Brady</td>
<td>Siegfried Benkner</td>
</tr>
<tr>
<td>Chapt.5</td>
<td>H. Bilofsky, R. Ziegler, M. Hofmann, V. Breton, C. Jones</td>
<td>Howard Bilofsky, Chris Jones</td>
<td>Irina Strizh</td>
</tr>
<tr>
<td>Chapt.6</td>
<td>M. Hofmann, T. Solomonides, N. Maglaveras</td>
<td>Martin Hofmann, Tony Solomonides</td>
<td>Clive Tristram</td>
</tr>
<tr>
<td>Chapt.7</td>
<td>M. Cannataro, P. Veltri, G. Aloisio, S. Fiore, M. Mirto, N. Maglaveras, I. Chouvarda, V. Koutkias, A. Malousi, V. Lopez, I. Oliveira, J. P. Sanchez, F. Martin-Sanchez</td>
<td>Ilidio C. Oliveira, Juan Pedro Sanchez, Victoria López</td>
<td>Irina Strizh</td>
</tr>
<tr>
<td>Chapt.8</td>
<td>G. De Moor, B. Claerhout</td>
<td>Georges De Moor, Brecht Claerhout</td>
<td></td>
</tr>
<tr>
<td>Chapt.9</td>
<td>J. A.M. Herveg</td>
<td>Jean A.M. Herveg</td>
<td>Yves Poullet</td>
</tr>
</tbody>
</table>
White Paper recommended areas for early grid deployment

• Medical imaging and medical image processing
  ▪ Use case: breast cancer early diagnosis
  ▪ Existing projects: Mammogrid, eDiamond, …

• Computational models for therapy planning and computer assisted interventions
  ▪ Use cases: radiotherapy, face surgery
  ▪ Existing projects: GEMSS, GATE, …

• Grid-enabled pharmaceutical R&D: pharmagrids
  ▪ Use case: grid for a rare disease

• Epidemiological studies
  ▪ Use case: genetic epidemiology

• Genomic medicine and biomedical informatics
MammoGrid

European federated mammogram database implemented on a GRID infrastructure

- Epidemiology of breast cancer from a European perspective
  - Open source architecture

- Use of Grid in developing quality control techniques for breast cancer screening

- Development of some CADe techniques

- Participants:
  - CERN & University of West of England - Grid infrastructure.
  - Mirada Solutions - Image analysis workstation and key technologies.
  - Universities of Pisa and Sassari - Image analysis algorithms
  - University Hospitals of Cambridge, Udine & Ospedale Valdese Torino - Clinical validation
  - University of Oxford - Grid-based quality control
A flood of data

• A flood of data in Europe: $120 \times 10^{12}$ bytes
  - Currently 3,000,000 images per year, each mammogram yielding 25-40Mbytes
  - Compression must be lossless

• The grid challenge on databases
  - Large federated databases
  - Ontologies and metadata
  - Effective data mining of a rapidly growing database
  - Allow for complex queries involving executables

• The grid challenge on communications
  - Legal restrictions on access to data
  - Data resides in hospitals
  - Secure file transfer
  - Large images to be transferred
  - Develop API for black box third party applications
**GEMSS: Grid-Enabled Medical Simulation Services**

Project Duration: 30 months, Start: 1.9.2002

Grid Software /solutions | Simulation /Imaging Software | Bio-numeric modelling | Medical Expertise | Legal Aspects

Medical simulation service + networked compute resources

User-site (SW installed)

Pre- & Post-processing

Could also be moved to the services portal

Simulation Service System
GRID SW (interface)
Applications SW
GRID SW (service use)

Internet or Intranet

**Courtesy of GEMSS Project - HealthGrid Workshop, Brussels; HealthGrid Conference, Lyon**

http://www.gemss.de
Main GEMSS Goals:

- Secure and lawful Grid provision of medical simulation services,
- Build 6 Grid-enabled medical prototype applications,
- Build suitable middleware on top of common standards,
- Install and evaluate a GEMSS test-bed,
- Anticipate privacy, security and other legal concerns related to providing medical services over the Internet.
Grids for rare diseases and diseases of the developing world

In silico drug discovery process
(EGEE, Swissgrid, ...)

Support to local centres in plagued areas (genomics research, clinical trials and vector control)

The grid impact:
- Computing and storage resources for genomics research and in silico drug discovery
- Cross-organizational collaboration space to progress research work
- Federation of patient databases for clinical trials and epidemiology in developing countries

interested to join?

Contact V. Breton (breton@clermont.in2p3.fr)
or M. Hofmann (martin.hofmann@scai.fhg.de)
EGEE biomedical sector

• **Goals**
  - To deploy biomedical applications on EGEE infrastructure
  - To participate to the early testing of EGEE middleware
  - To set up virtual organizations and to integrate new resources into the EGEE infrastructure
  - To establish collaborations with (national, European, Worldwide) biomedical projects for deployment on EGEE

• **Partners**
  - CNRS (Lyon, Clermont-Ferrand): J. Montagnat (CREATIS)
  - Universitat Politecnicad de Valencia: V. Hernandez
  - CSIC (CNB Madrid): J.-M. Carazo

• **EC funding:** 760k€ for 2 years
What EGEE offers to biomedical applications

• Access to large-scale infrastructure
  Thousands of processors and 1/3 petabyte online data storage

• Production ready grid middleware
  More than 3 years of large-scale testing/deployment experience

• Grid expertise
  Small team of technically competent people ready to help applications get up and running
  Training
The success of EGEE is measured by the impact it has on collaborative European science

The goal is to support communities of users

Therefore induction and training have a high priority from the outset

**Invitation**

- Declare your requirements for induction and training
  - The sooner the better – so that you influence planning
- First point of contact:
  - John Murison, Training Manager, NeSC : john@nesc.ac.uk
Security & Intellectual Property

- The existing EGEE grid middleware (LCG-2) is distributed under an Open Source License
  - No restriction on usage (scientific or commercial) beyond acknowledgement
  - Same approach for new middleware

- Application software maintains its own licensing scheme
  - Sites must obtain appropriate licenses before installation

- For applications that must operate in a closed environment
  - EGEE middleware can be download and installed on closed infrastructures
How to work with EGEE

0 Review information provided on the EGEE website (www.eu-egee.org)

1 Establish contact with the EGEE applications group
   Overall Coordinator: Vincent Breton (breton@clermont.in2p3.fr)
   Biomedical Coordinator: Johan Montagnat (Johan.Montagnat@creatis.insa-lyon.fr)

2 Provide information by completing a questionnaire describing your application
   (http://egee-na4.ct.infn.it/biomed/na4-biomed-questionnaire.html)

3 Applications are selected for direct support based on scientific criteria, Grid added value,
   effort involved in deployment, resources consumed/contributed etc.

4 Follow a training session

5 Migrate application to EGEE infrastructure with the support of EGEE BMI technical experts

6 Initial deployment for testing purposes

7 Production usage
   Contribute computing resources for heavy production demands
• We have a dream …
  - of a world where access to quality healthcare could be extended to the poorest: eHealth for all

• Grids are the infrastructures to empower the vision
  - Because it provides the mean to share resources, information and benefits at the minimal cost
  - Provided an on-going public effort keeps developing open source technology

• After three years of pioneering work in Europe, biomedical applications of grids enter into a new era where societal issues can be addressed
  - Breast cancer early diagnosis
  - Rare disease (malaria, dengue, …)

• There are more human challenges than technical challenges

• The key factor to successes are open source and worldwide collaboration: you are welcome to join!
Thank you for your attention!

The Auvergne Region

pariou & puy de dôme – © pierre soissons
puy de dôme – © jérome chabanne
saint-ilpize village – © christophe camus
chaîne des puys – © christophe camus