Health grid

Vincent Breton
ISGC 2003
Plan

• Health grid : concepts
• Health related grid projects in Europe
• Lessons from the Healthgrid conference
• The next step : welcome to the HEAlthgrid VENture
Healthgrid : concepts
A grid for health : concepts

• The goal is to meet the growing computing needs of health actors with the help of grid technology
  – Biologists (exponential growth of *omics databases)
  – Drug Designers
  – Health care professionals (Each hospital generates 1-7 TB of data per year including medical images, sounds, texts,...)
  – ...

• On a common infrastructure in order to proceed toward individualized healthcare
Individuated healthcare uses information at 5 levels...

Health grid

PublicHealth
Patient
Tissue, organ
Cell
Molecule

Patient related data

Association
Modeling
Computation

Databases

PublicHealth
Patient
Tissue, organ
Cell
Molecule

INDIVIDUALISED HEALTHCARE MOLECULAR MEDECINE

Computational recommandation
Definition

From this slide on, **health** will be used to cover all features related to healthcare from molecule to population.

The term **biomedical** will be used to address the whole area covered by both bio- and medical-informatics as well as the new disciplines arising from their synergy.
Grid technology is promising for both computing intensive applications and knowledge discovery.

- To connect databases of heterogeneous content (biology and medicine) enabling new knowledge discovery (research, drug design), better guidance and information (healthcare professionals).

- To increase computing power for imaging, simulation and modelling thus allowing these fields to take into account more data and therefore to provide more accurate results.

- To address security (integrity, confidentiality, authentication, authorization, non-repudiation, availability).
What are the challenges?

• The first challenge is to integrate grid technology into health practice
  – Deployment of pilot biomedical applications

• The second challenge is to integrate specific health requirements into grid technology
  – Standards, interfaces, protocols
  – Data heterogeneity

• The third challenge is to address the dispersion of the biomedical community
Integration of grid technology into health practice

Health-related grid projects in Europe
The “HealthGRID Cluster”
(http://www.healthgrid.org)

Overview
• Formation of the cluster in 2002

• 1st HealthGrid Conference, Lyon, Jan. 2003
  - intro to themes addressed and directions identified

*Particular thanks go to Guy Lonsdale, NEC, Sofie Nørager, European Commission
& Richard Mcclatchey, U. West England for their support via the input of information & material for the presentation*
Perform a trial for the introduction of the Grid approach in the biotechnology industry.

1st HealthGrid Cluster Workshop – Brussels, 20th Sept. 2002

Initial Aim: to unite the EU-funded projects using GRID technology in health areas.

Biology & Medical Apps:
- Medical imaging
- Computational Grid for “Virtual Arteries” Surgical support tool

BioGrid Task: operate a Grid for biomolecular simulations
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
A flood of data

• Currently one view taken of each breast → 3,000,000 images per year
• Increasing to 2 views per breast over the next 2-5 years
• Digitised at 50µ each mammogram yields 25-40Mbytes
• Total annual potential is 240x10^{12} bytes
• Compression must be lossless

… and there are 11,000,000 mammograms in the backlog!
Federated System Solution

- Knowledge is stored alongside data
- Active (meta-)objects manage various versions of data and algorithms
- Small network bandwidth required

Massively distributed data AND distributed analyses

- Shared meta-data
- Analysis-specific data
Grid challenges: database

• Large federated databases
  – Images and metadata
• Ontologies and metadata
  – Image formation parameters
  – Image features
  – Clinical information
  – Demographic data
• Effective data mining of a *rapidly growing* database
• Allow for complex queries involving executables
  – e.g. “give me breast densities for all women >45 with HRT”
  – e.g. “… and those with a suspect mass like this”
• Mirada and medical image analysis clients are
Grid challenges: communications

• Legal restrictions on access to data
  – Clinicians, researchers, developers, Govt, …

• Data resides in hospitals
  – Firewall protected

• Combining several databases

• Secure file transfer

• Large images to be transferred

• Develop API for black box third party applications
DataGrid, prototype of a biomedical grid

• DataGrid is a European funded project
• DataGrid has three ambitious goals:
  – Develop a middleware
  – Deploy a testbed
  – Have large scale applications running on this testbed
• The biomedical work package faces three challenges
  – Make the middleware meet biomedical specific requirements
  – Run biomedical applications
  – Deploy grid nodes in biomedical laboratories
Biomedical technical requirements

1. Large user community
   - anonymous/group login
2. Data management
   - data updates and data versioning
3. Security
   - disk / network encryption
4. Limited response time
   - fast queues
5. High priority jobs
   - privileged users
6. Interactivity
   - communication between user interface and CE’s
7. Parallelization
   - MPI site-wide / grid-wide
8. Pipeline processing
   - pipeline description language / scheduling

Critical for development
Mid-term requirement
Long-term requirement
## Analysis and design of biomedical data management grid services

<table>
<thead>
<tr>
<th>Use cases</th>
<th>EDG Middleware Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remote processing of a data set</td>
<td>Available</td>
</tr>
<tr>
<td>Image registration</td>
<td>Available / Next release</td>
</tr>
<tr>
<td>A patient retrieves one of its images</td>
<td>Next release</td>
</tr>
<tr>
<td>A physician retrieves all medical files of a patient</td>
<td>Next release</td>
</tr>
<tr>
<td>Image search on public metadata</td>
<td>Available</td>
</tr>
<tr>
<td>Automatic indexation of new images</td>
<td>Next release</td>
</tr>
<tr>
<td>Algorithm registration</td>
<td>Not planned</td>
</tr>
<tr>
<td>Automatic data parallelization</td>
<td>Not planned</td>
</tr>
<tr>
<td>Pipeline processing</td>
<td>Next release</td>
</tr>
<tr>
<td>Interactive modeling</td>
<td>Next release</td>
</tr>
</tbody>
</table>

J. Montagnat  CREATIS
Status of biomedical applications

- **Bio-informatics**
  - Phylogenetics: BBE Lyon (T. Sylvestre)
  - Search for primers: Centrale Paris (K. Kurata)
  - Statistical genetics: CNG Evry (N. Margetic)
  - Bio-informatics web portal: IBCP (C. Blanchet)
  - Parasitology: LBP Clermont, Univ B. Pascal (N. Jacq)
  - Data-mining on DNA chips: Karolinska (R. Médina, 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: GA, collaboration (L. Maigne)
Laboratories involved in DataGrid biomedical activities

- Dublin
- Hinxton (EBI)
- Padova
- Saclay (CEA)
- Lyon
- Clermont
- Montpellier
- Madrid (CNB)
- Uppsala
- Stockholm (Karolinska)
- Moscow
- Kosice
- Evry (CNG)
- Lausanne (EPFL, SIB)

Laboratories testing EDG Middleware

Other labs.
Integration of specific health requirements into grid technology
Objectives:
bring together technology developers and end-users
(a first step towards establishing a HealthGrid
Community)

“End-users” = healthcare professionals / providers
+ academic & industrial researchers and developers
from bio-informatics and medical-informatics.
A non-exhaustive list of issues/themes addressed (1/2):

- **Knowledge & Info. Discovery across distributed/federated databases**
  - Many areas: bio-informatics, pharmaceutical apps, neuroscience research, medical images (→ large data volumes)
  - **Very** rich semantics (→ investigations on mediation for integration)

- **Computational Grids**
  - Medical simulation services
  - Systems biology (eg for identification of drug targets)
  - Surgery support tools (requirements to move to real-time)

- **Integration**
  - Need for workflow systems
  - Self-learning, data-mining, modelling (e.g. HIV drug therapy support system)
  - Multi-scale Integration (eg connecting molecular biology to biology of cells, organs, organisms, populations)
A non-exhaustive list of issues/themes addressed (2/2):

- **Virtual Organisations**
  (e.g. Regional Health “Economies” – [virtual] electronic health records shared between all levels of healthcare providers)

- **Privacy & Security**
  (patients rights on use of data, data protection acts; commercial security e.g. Pharma industries)

- **Healthcare reluctance to introduce new IT**
  (e.g. doctors not always happy to use “computer systems” → PDA access )
Current Status:

- First projects are starting to apply GRID technologies both at national and international levels in fields such as: medical imaging, sequence analysis and biological database combination, modelling of medical data, database knowledge extraction for providing guidelines for health professionals.
- The success of these pioneers will be the first real showcases needed to convince the end-user community of the possible impact on healthcare.
- Regional health networks have been working on solutions for some of the data-access, data-sharing issues and can be seen as a good basis for the introduction of GRID technologies in the Healthcare sector.
Looking Ahead:

- Grid-based applications should provide quick & easy access to non-IT experts
- Healthcare needs not only storage and transfer of large amounts of data but also access, update and intelligent combination of heterogeneous data to create new knowledge.
- Homogenised access to remote medical data will help to face their natural fragmentation over different sites involved in the healthcare of a single individual.
- Introduction to Healthcare will be a difficult task: healthcare professionals are often reluctant to change their ways of working, policy makers are doubtful about taking risks and changing systems, citizens must trust the system to let it handle personal data.
- Potential for open standards: in the health area there is a differentiated end-user community - room for both open source tools and commercial products.
The next step: HEAlthgrid VENture
HEAVEN, a Network of Excellence to

• Form the core of a community pioneering the use of GRID-enabled medical applications and, through that early use, contributing to the development, usability and wide applicability of such applications.

• Create a wide awareness across the medical community, including both users and suppliers, of the potential and capabilities of GRID-enabled applications.

To prepare the health info-structure of the European Research Area
HEAVEN players

- End users (health professionals, biologists, …)
- Structures involved in administrating, providing and monitoring health care
- Computer scientists involved in
  - medical informatics
  - Bio-informatics
  - Grid technology development
  - …
- In interaction with
  - Associations of physicians
  - Other networks and structures (OMG, GGF, PRISM, …)
  - Lawyers
  - Economists
  - …
HEAVEN partners

- Healthcare centers (German Cancer Research Institute, University hospital Geneva, …)
- Institutes involved in national grid initiatives acting as an interface between the national health related projects and the NoE (French CNRS, Edinburgh Parallel Computing Centre, Moscow Telecom. Centre, Tech. Univ. Valencia, …)
- European organizations connecting HEAVEN to other major grid initiatives (CERN, European Molecular Biology network, …)
- Middleware developers and healthcare IT providers (CEA, IBM, NEC, Siemens, …)

... all partners will connect biomedical communities willing to use grid technology through the other projects they are involved in.
HEAVEN goals

- Integration of grid technology into health practice
- Integration of health specific needs, requirements and legacy into the new generation of grid services
- Integration of the different health-care levels with the help of grid technology
- Integration of grids for health into the European legal and ethical framework
- Integration of private companies into the health grid framework
- Integration of health-related grid projects at regional, national and European levels
Deployment of pilot medical applications

• Portfolio of 30 different applications
• Clustering according to
  – Their usage of the grid (data or computing centered) and the requested services
  – The medical level addressed (from molecule to population)
  – The end users (healthcare professionals, health care administrations, researchers, …)
• HEAVEN open to collaborations outside Europe
  – Joint pilot medical applications
  – Common standards and requirements (GGF)