



Enabling Grids for E-scienceE

# HealthGrid: Select and Architect Services

*Yannick Legré*

*CNRS – HealthGrid association*

*([yannick.legre@healthgrid.org](mailto:yannick.legre@healthgrid.org))*

**Credit for the slides: F. Gagliardi – V. Breton**

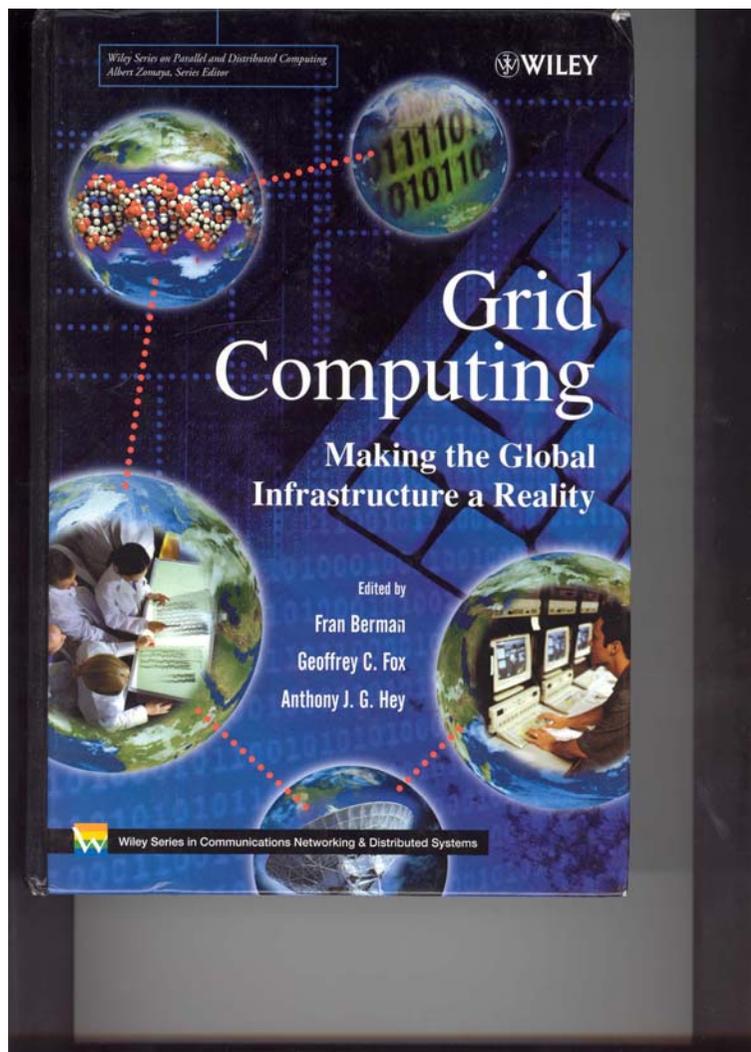
[www.eu-egee.org](http://www.eu-egee.org)



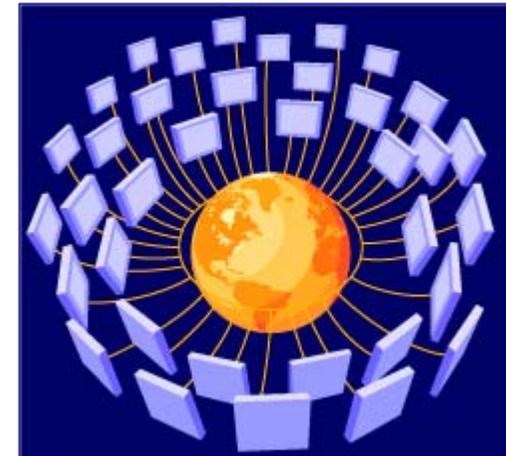
**I'm a frenchy!...**

**...so, feel free to interrupt me if you don't understand what I mean...**

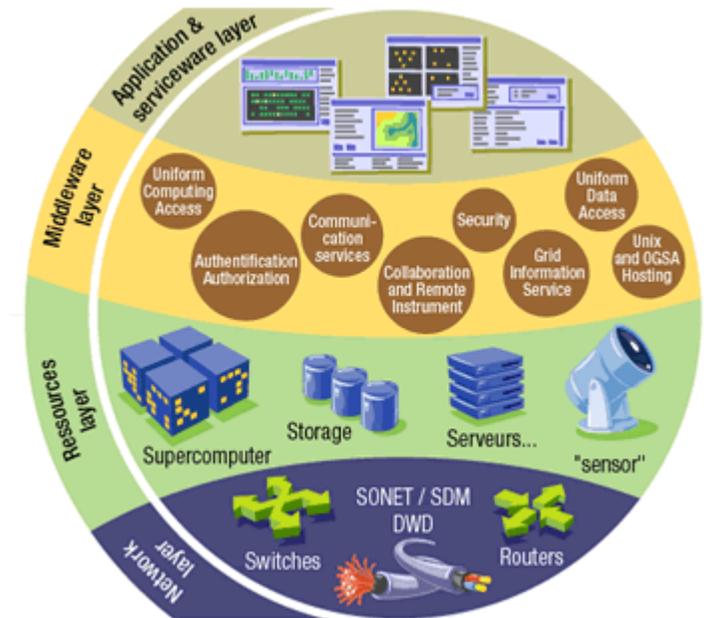
- **1. Intro GRID**
- **2. Why HealthGrid?**
- **3. Detail services**
- **4. Invite participation in HealthGrid**

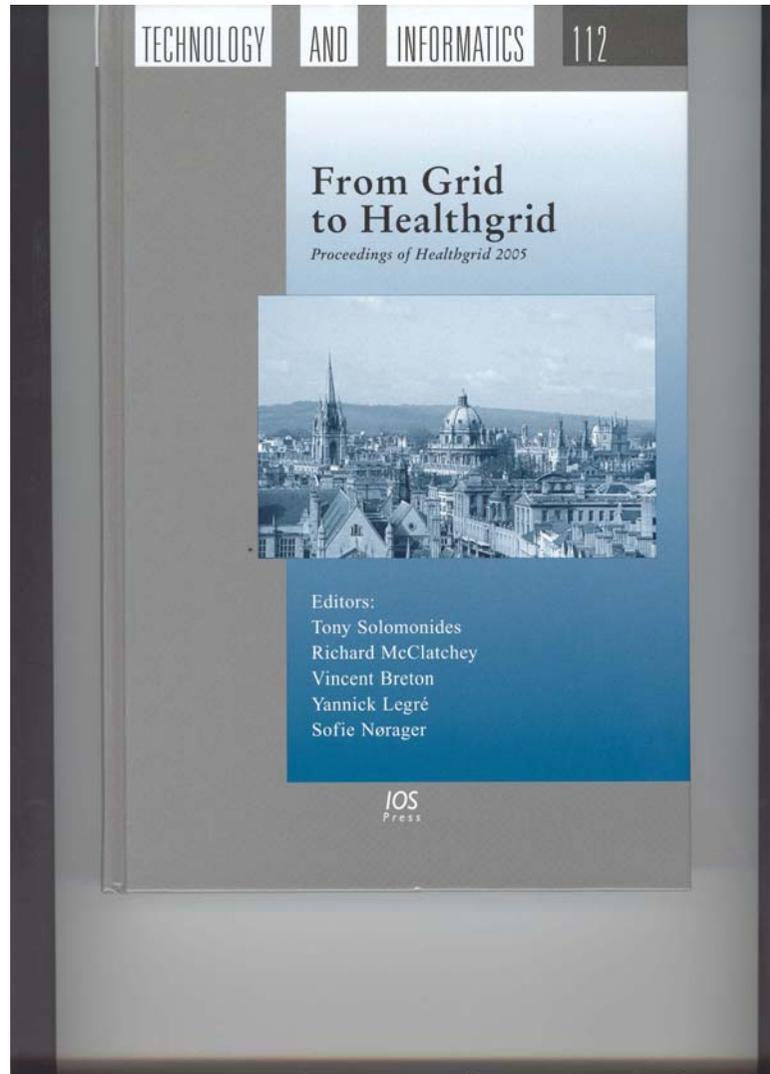


- **The World Wide Web provides seamless access to information that is stored in many millions of different geographical locations**
- **In contrast, the Grid is a new computing infrastructure which provides seamless access to computing power, data and other resources distributed over the globe**
- **The name Grid is chosen by analogy with the electric power grid: plug-in to computing power without worrying where it comes from, like a toaster**



- **The Grid relies on advanced software, called middleware, which ensures seamless communication between different computers and different parts of the world**
- **The Grid search engine not only finds the data the scientist needs, but also the data processing techniques and the computing power to carry them out**
- **It distributes the computing task to wherever in the world there is available capacity, and sends the result back to the scientist**





- **Connecting and understanding patient records across organization structures and even national borders**
- **Ensuring that information is secure and those accessing it are authenticated**
- **Discovering trustworthy sources of information for comparison**
- **Handling a huge volume of data in multiple formats**
- **Applying traditional information networks and technology into healthcare**

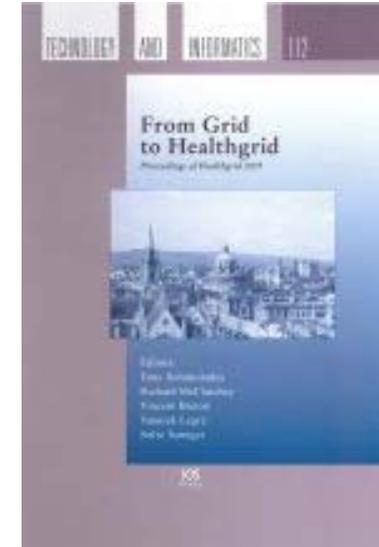
- **A HealthGrid is an innovative use of emerging information technology to support broad access to rapid, cost-effective and high quality healthcare**
- **A HealthGrid is a global ICT infrastructure that will enable a coordinated, flexible, and secure sharing of diverse resources including**
  - Networks
  - Computers
  - Storage
  - Data
  - Applications
- **... across dynamic and geographically dispersed organizations and communities**

- **HealthGrid is not only about number crunching...**
  - Although it provides computing resources for CPU demanding medical applications
- **HealthGrid is about offering services to healthcare professionals to enable evidence-based medicine**
  - Bringing all relevant data and tools needed for evidence-based decision making
- **Technical challenges**
  - Developing standards and interfaces to enable sharing of information
  - Developing technology to secure information

- **Summarize and formalize the discussions during the second HealthGrid conference in Clermont-Ferrand (January 2004)**
- **Produce a reference document to**
  - Clarify the concept of HealthGrid
  - Analyze the technical challenges
  - Identify healthcare issues that could be better addressed TODAY using grid technology
  - Propose a business model for HealthGrids
  - Address legal, ethical and security issues
- **Document produced end of 2004**
  - Contribution from > 40 experts from Asia, Europe, USA

- **From Grid to HealthGrid: prospects and requirements**
- **Creating a compelling business case for HealthGrid**
- **Application areas**
  - Medical imaging and medical image processing
  - Computational models of the human body for therapy planning and computer assisted intervention
  - Grid enabled pharmaceutical R&D: pharma grids
  - Grids for epidemiological studies
  - Genomic medicine grid
- **From grid to HealthGrid: confidentiality and ethical issues**
- **Legal approaches of the HealthGrid technology**

- **“long” version (70 pages)**
  - Published with HealthGrid 2005 conference proceedings
    - From Grid to HealthGrid – Studies in Health Technology and Informatics (vol. 112) – IOS Press – T. Solomonides et al. (eds)
    - <http://www.iospress.nl/html/158603510x.php>
  - Contact: Y. Legré ([yannick.legre@healthgrid.org](mailto:yannick.legre@healthgrid.org))
  
- **“short” version (35 pages) for governing bodies and health authorities**
  - Editor: CISCO
  - Available now as booklets
  - Contact: K. Dean ([kevin.dean@cisco.com](mailto:kevin.dean@cisco.com))



- **Areas of healthcare provision and research that could benefit TODAY of grid:**
  - Medical imaging and image processing
  - Modeling the human body for therapy planning
  - Pharmaceutical research and development
  - Epidemiological studies
  - Genomic research and treatment development



Enabling Grids for E-science

# State of the art of HealthGrid applications



HEALTHGRID

- **Provide a forum for GRID projects in the medical, biological and biomedical domains as well as for grid projects that seek to integrate these.**
- **Demonstrate the functionalities developed within such projects.**
- **Reinforce and promote the awareness of the deployment of grid technology in health.**
- **Encourage involvement across the spectrum of bio/medical informatics from technology developers and researchers through research network representatives and regional health authorities to healthcare clinicians and administrators**

- **Applications using grid computing:**
  - SEED (TERAGRID customer)
  - CDSS (EGEE customer)
  - ROKKY (Japanese Biogrid customer)
  - Fibre tracking (Dutch grid customer)
  - Neurophysiology (UCL inner grid)
  - Cardiac MRI analysis (inner grid)
- **Grid infrastructures hosting biomedical applications**
  - BIRN
  - EGEE
- **Applications deploying their own grid infrastructure**
  - eDiamond
  - Mammogrid
  - Medical image processing (MAGIC-5 infrastructure in Italy)
  - Grid-enabled radiotherapy surgery planning (GEMSS)
- **Virtual laboratories**
  - ProGenGrid

- **Emergence of multidisciplinary grid infrastructures**
  - EGEE, Grid3, Open Science grid, ....: grids of Linux clusters
  - DEISA, TeraGrid, ....: grids of supercomputers
- **Difficult transition from old generation (Globus Toolkit 2.n) to new generation middleware**
- **All infrastructures identified life sciences as key application areas**
- **Healthcare is the next step provided legal issues are addressed**

# Status of deployment of HealthGrid applications

## Computing Grid For data crunching applications

- Most of the applications showed at HG conference are using grids for computing
  - Users are happy of inner grid as well as large scale grid infrastructures (EGEE)
  - Next step: legal issues, SLAs,...
- Very few data grid applications
  - Deployment is complex in Europe (Mammogrid)
  - On the other hand, BIRN shows very impressive deployment
- No knowledge grid operational yet
  - Talks on ontology
  - Tools being developed (Virtual labs, workbenches)
  - Need for usage of these ontology and deployment of these workbenches on grid infrastructures
  - First step: life sciences and biomedical research (Chicago Systems Biology Architecture, BIRN, EGEE,...)

**Data Grid**  
Distributed and optimized storage of  
large amounts of accessible data

**Knowledge Grid**  
Intelligent use of Data Grid for  
knowledge creation and tools  
provisions to all users

- **Users Management**

- Authentication:

- Managed users by users

- Each of them should

- *Own a x509 certificate delivered by a one of the Certificate Authority (~30) member of the EUGridPMA (EUropean Grid Policy Management Authority) trust domain*
      - *Belong to a Virtual Organisation*
      - *Have an account on a User Interface*

- Authorization:

- Managed at a Virtual Organisation level or at least for a subset

- Up to now we are using Lightweight Directory Access Protocol servers but we're moving to Virtual Organisation Management Service which will allow use of more complex features such as groups, role, ...

- **Resources Management**

- Managed per site
- By site administrators
- According to their own site policy
- At a Virtual Organisation level (or a subset of it)
- For one or multiple Virtual Organisations

- **Data Management**

- One and only one Replica Location Service per Virtual Organisation
- Data stored on the Grid are accessible by the whole Virtual Organisation (at least until VOMS will be massively used) – no Access Control Lists for now
- Need to deal with common standards used in Life Science / HealthCare domains

**→ this is the weaker point as confidentiality and privacy are so important in our domains**



Enabling Grids for E-science

# Example of a HealthGrid application running today



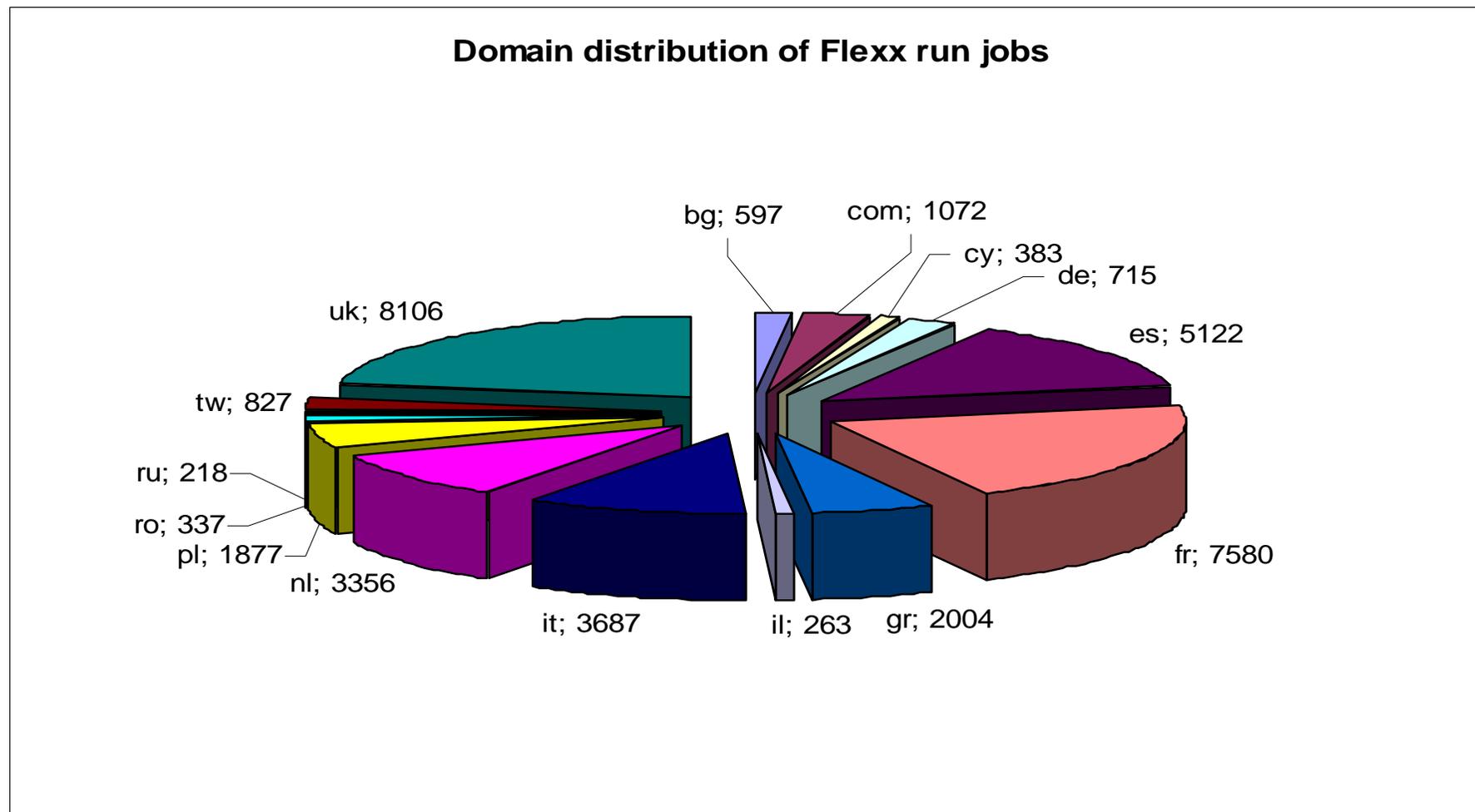
HEALTHGRID

- **Grid technologies promise to change the way organizations tackle pharmaceutical research and development**
- **Grids can significantly impact the drug discovery workflow**
- **WISDOM (Wide In Silico Docking On Malaria) is a drug discovery initiative on an open grid infrastructure**
  - Focus on malaria
  - Partners: Fraunhofer Institute, CNRS-IN2P3,
  - Grid infrastructures used: EGEE, AuverGrid
- **First deployment of in silico docking during the summer**

- **Biological information**
  - **Plasmepsin** is a promising aspartic protease target involved in the hemoglobin degradation of *P. falciparum*. 5 different structures are prepared (PDB source)
  - **ZINC** is an open source library of 3,3 millions selected compounds. They are made available by chemistry companies and are ready to be used
- **Biomedical informatics tools**
  - **Autodock** is free for academic, with grid based empirical potential and flexible docking via MC search and incremental construction
  - **FlexX** is a licensed software originally developed by Fraunhofer Institute, available for the data challenge during 2 weeks, with Boehm potential and fragment assembly energy function
- **Grid tools**
  - Wisdom\_env is an environment for an automatic, optimized and fault tolerance workflow using the grid resources and services
  - EGEE biomedical Virtual Organization will be dedicated/no-dedicated resources

- **Duration: 11 July – 31 August**
- **Significant parameters**
  - Two different molecular docking applications (Autodock and FlexX)
  - About one million virtual ligands selected
  - Five target proteins and five parameter variations for the docking process
- **First numbers (<http://wisdom.healthgrid.org>)**
  - Total of about 46 million ligands docked
  - Equivalent to about 80 CPU years
  - Up 1000 computers in 15 countries used simultaneously
  - Millions of files adding up to a few TB of data
  - Average crunching factor ~600
- **Consortium identified willing to synthesize the most promising hits for further studies (biochemical assays; cell-based assays)**

## International distribution of jobs





Enabling Grids for E-science

# Invite participation in HealthGrid



- To provide a place of dialog and exchange between European and international projects
  - Web site: [www.healthgrid.org](http://www.healthgrid.org)
- To produce collaborative documents
  - HealthGrid White Paper (<http://whitepaper.healthgrid.org>)
- To organize conferences and workshops on HealthGrids

**4<sup>th</sup> HealthGrid Conference**  
**Location : Valencia (Spain)**  
**Dates: during the week of June**  
**5-9, 2006 (tbc)**

- **Act as a catalyzer for collaboration and dissemination**
  - Foster collaboration between projects
  - Disseminate information through its web site [www.healthgrid.org](http://www.healthgrid.org)
  - Produce collaborative documents such as the HealthGrid White Paper
  - Organize conferences and workshops on HealthGrids
- **Potential roles beyond Europe**
  - Foster collaboration between projects
  - Common repository for HealthGrid related projects
    - Place to download middleware components for deployment of biomedical grid applications
    - Such services exist for national initiatives (OMII)
    - Requirements: interoperability and security
- **You are more than welcome to join!**
  - Send an email to [contact@healthgrid.org](mailto:contact@healthgrid.org)

**You see things; and you say “Why?”  
But I dream things that never were; and I say “Why not?”**

**George Bernard Shaw**

**Any questions?**