The Grid and the Biomedical Community: achievements and open issues

EGEE User Forum, March 1st, 2006

Pr. Isabelle E. Magnin
Head of CREATIS laboratory

CNRS-INSERM
Univ. Lyon1- INSA

Creatis

Inserm
Institut national
de la santé et de la recherche médicale

Information Society

www.eu-egee.org

With the collaboration of Johan Montagnat

I3S laboratory

Inserm

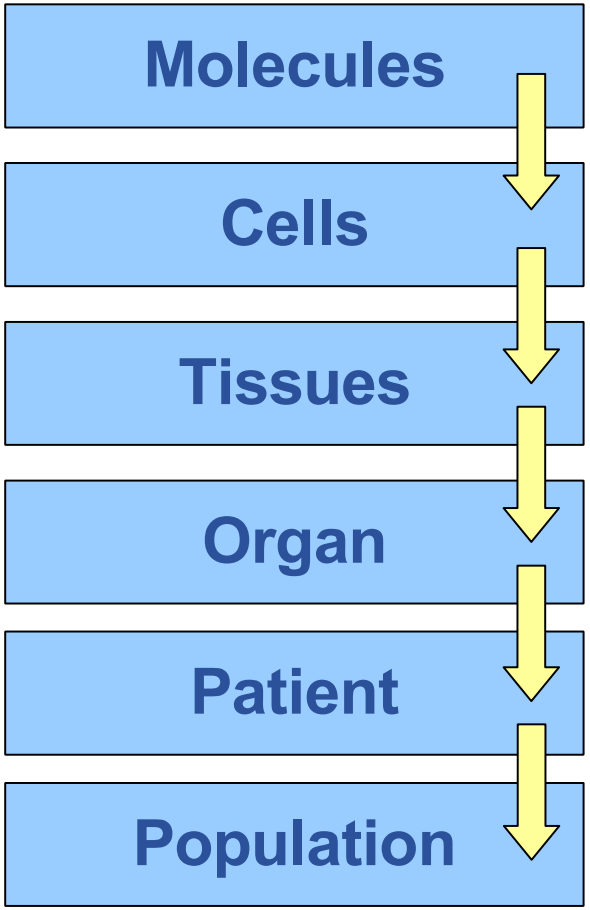
Institut national
de la santé et de la recherche médicale

Information Society

Biomedical Community

"Biomedical community"

- **Bioinformatics**
  - Genomics
  - Proteomics
  - Phylogeny...
- **Medical imaging**
  - Medical imaging
  - Computer Aided Diagnosis
  - Therapy planning
  - Simulation...
- **Life sciences**
  - Drug discovery
  - Epidemiology
  - ...

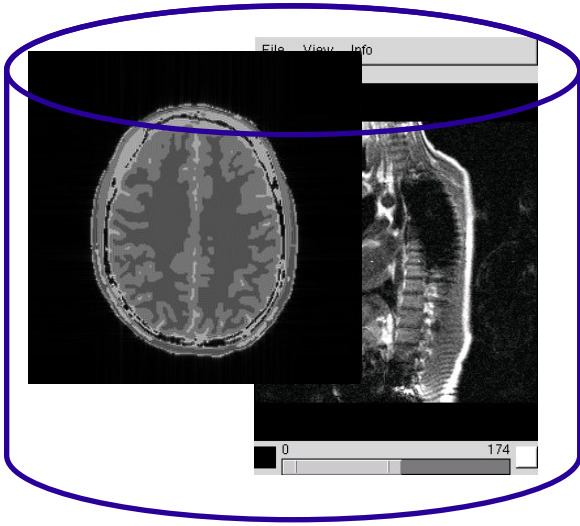


Medical images: the facts

- *Generalization of digital images processing for medical diagnosis and prognosis*
- *Huge amount of data produced by high resolution imagers*
  - Standard 3D volume: 20 Mb of data and more
  - High resolution image: 1024^3 voxels * 16 bits = 2 Gb of data
  - Sequences of medical images (heart imaging…): necessary tradeoff between spatial and temporal resolution (⟶ 5 Gb per second)
  - Ex: Lyon University Hospital about 12 Tbytes of data per year
- *Large databases of images or sequences*
  - 100's to 1000's of images for pharmacological / epidemiological studie
- *Need for remote processing (image transportation to computing centers)*

Medical databases

- **Data storage and archival**
  - ~PB / year / country
  - Need for long term archival (20 to 70 years)
  - Geographically distributed



- **Medical folders**
  - Patient / Image / Hospital-related metadata
  - Distributed patient records

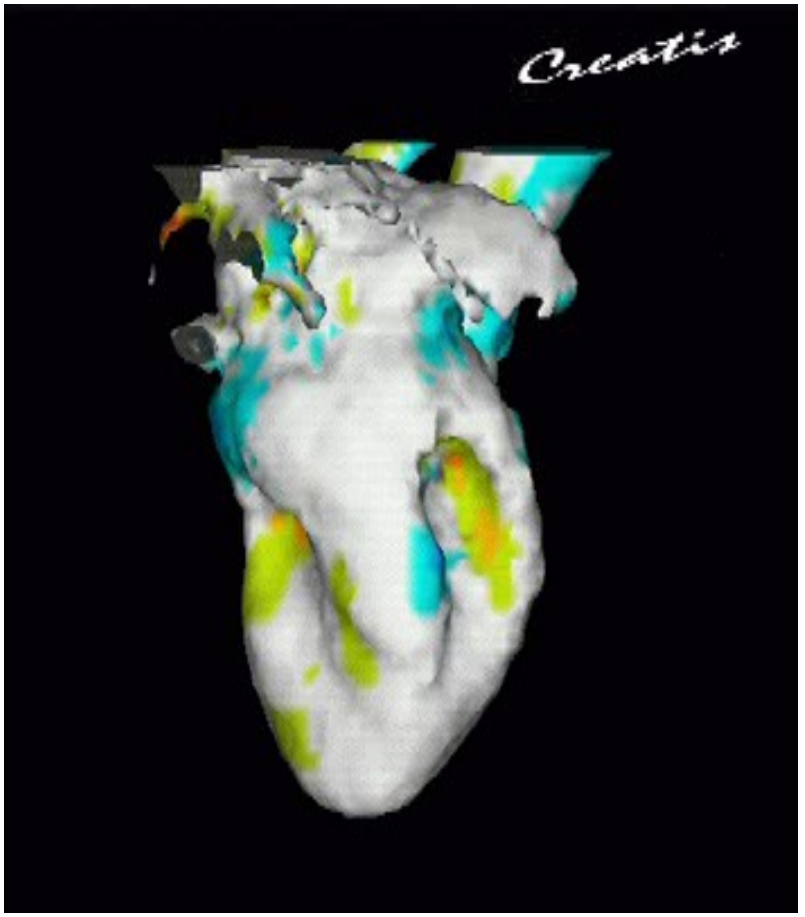
- **Large scale databases are needed**
  - Statistics
  - Epidemiology
  - Rare diseases
  - Personalized atlases construction
  - ...

Security needs

- Authentication **and** Authorization
  - Individual authentication
  - Multiple actors / roles
  - Fine grain authorization
- **Data** access control **at individual and group level**
  - Physicians
  - Patients
  - Researchers...
- Delegation: **Granting partial access rights**
- Encryption **for data storage and transfer**
  - Enforce patients privacy
  - Grant access right to accredited users only

Biomedical computing

- Embarrassingly parallel applications
  - databases processing
  - Bioinformatics
- Parallel computations
  - costly processings
- Interactive computations
  - resources reservation
  - user supervision and validation
- Emergency situations
  - resources preemption
- Algorithms warehouse
  - Algorithms sharing



3D left ventricle of dog's heart
(data: Mayo Clinic, Rochester)

Remote data storage and retrieval

- *Medical files organization inside/outside hospitals*

  database management, retrieval of medical information for visualization

- *Data exchange:* between several institutes involved in the care of the same patient → privacy

- *Data visualization*

  at interactive rate → compression, high bandwidth networks

- *Medical intervention planning:* video conferencing with image visualization, overlaid additional information → interactive tools

- *Pathology studies*

  similar data retrieval for large scale pathology studies

  → images indexing

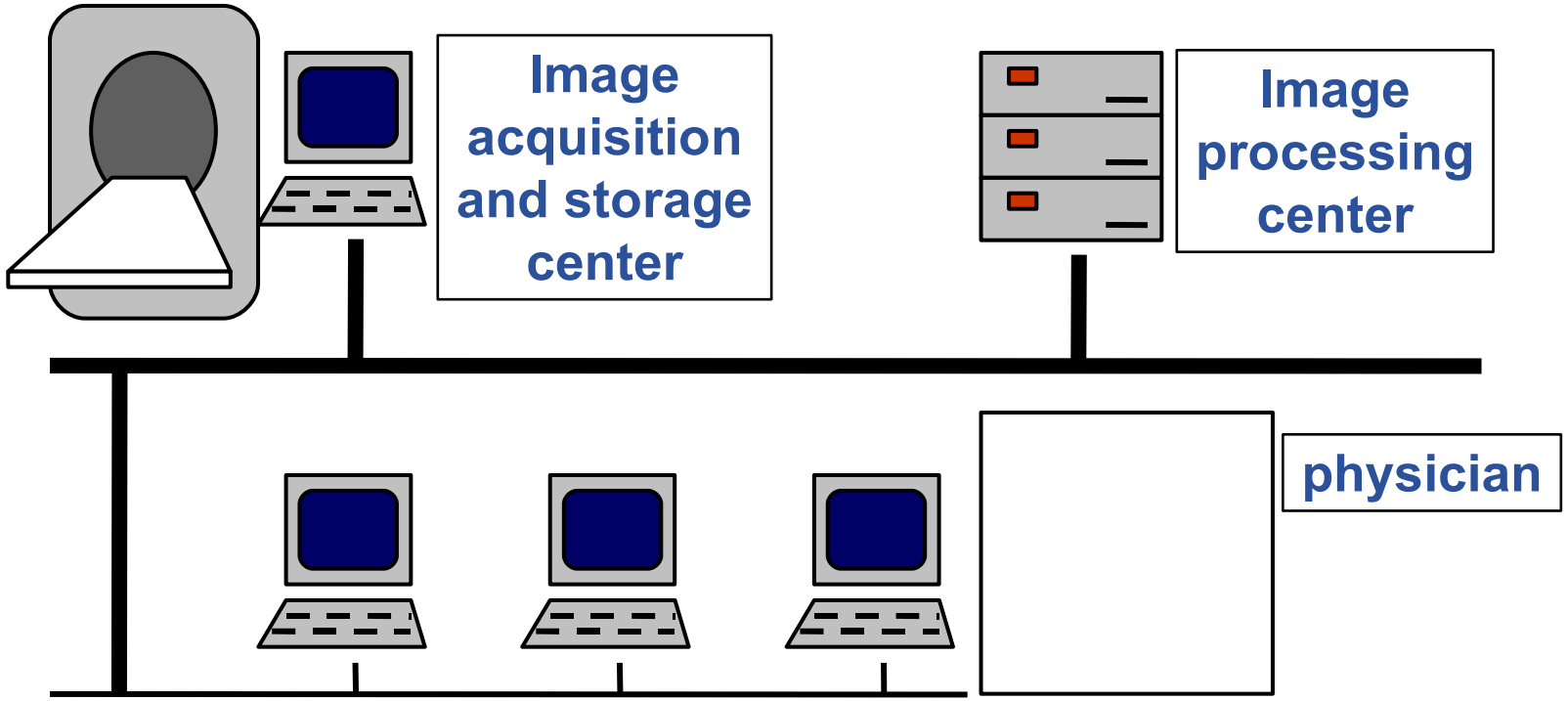
- *Property / privacy issues* → image traceability

Remote data processing

- ***Service producers over the grid managing requests from health institutes:*** costly algorithms (energy minimization procedures, stochastic algorithms…)
  → computation power and large memory requirements
- ***Data exchange format***
  → adaptability to algorithms input
- ***Available algorithms:*** knowledge of possible processing services
  → control parameters format
- ***Processing scheduling:*** distributing tasks over the grid
  → estimation of networking and processing costs
- ***Priority levels:*** for surgery rooms, emergency situations…
  → priority queuing, interruption of low priority jobs

Queries and computations on distributed databases

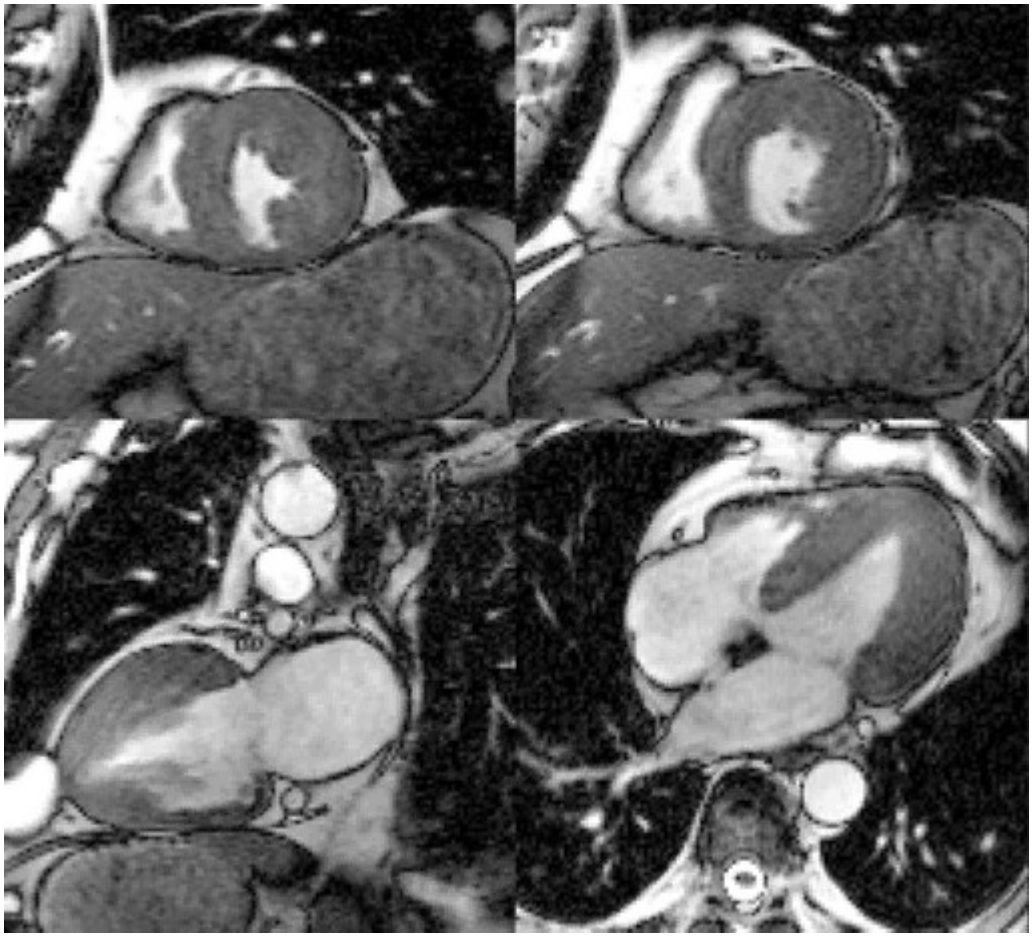
Objectives: remote data access and processing



Distributed database access and visualization

Remote processing

Magnetic resonance cardiac image sequences



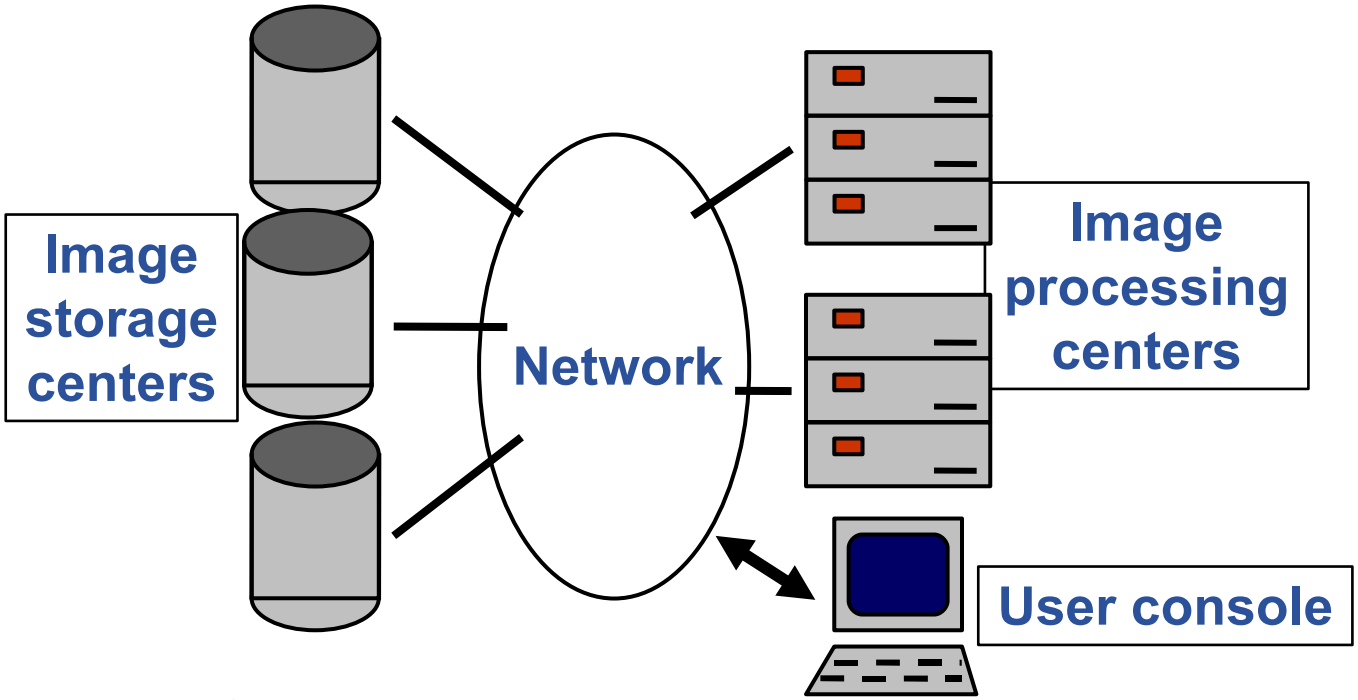
Courtesy of
P. Croisille)

Creatis

Image content-based search in medical image DB

Objective

Selection of a specific image by on line analyzing its content



Distribution of DB images over available computing elements

Remote processing of images

Selected images returned to user console

Workflows

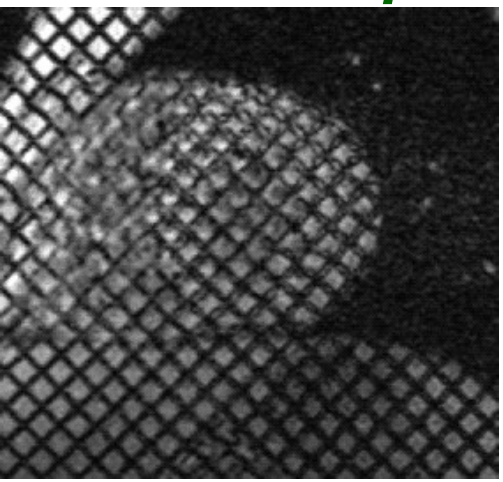
Objective: Chaining processes to build complex image analysis procedures

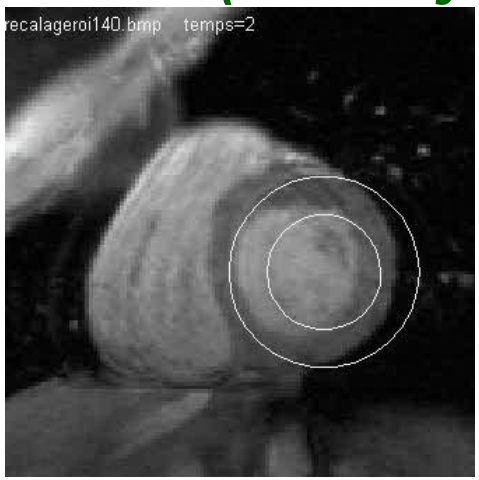
Example: images processing for cardiac activity quantification

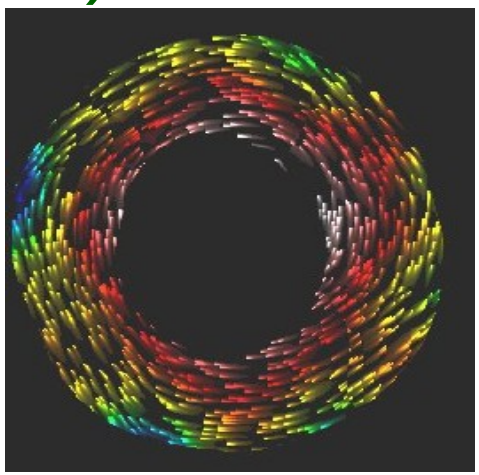
Input: tagged MRI sequences

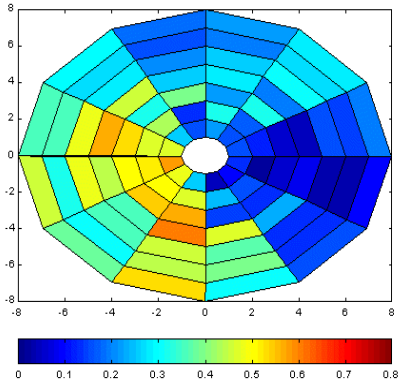
Chain of processes: (P. Clarysse)

Creatis









1. Tags and myocardium automatic extraction

2. Motion estimation

3. Quantification

Workflows

- *Medical trial of new drugs / large scale pathology studies*
  - very large scale studies (1000's of images)
  - statistical interpretation of results
- *Spreading over available computing nodes*
  - dataflow control
  - processing control
- *No real time constraint but large dataflow over the network*

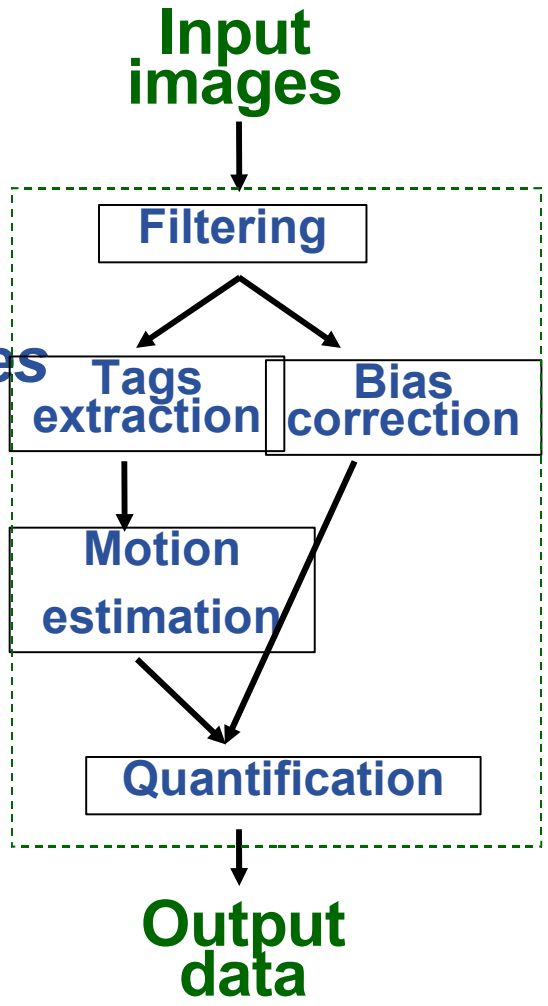
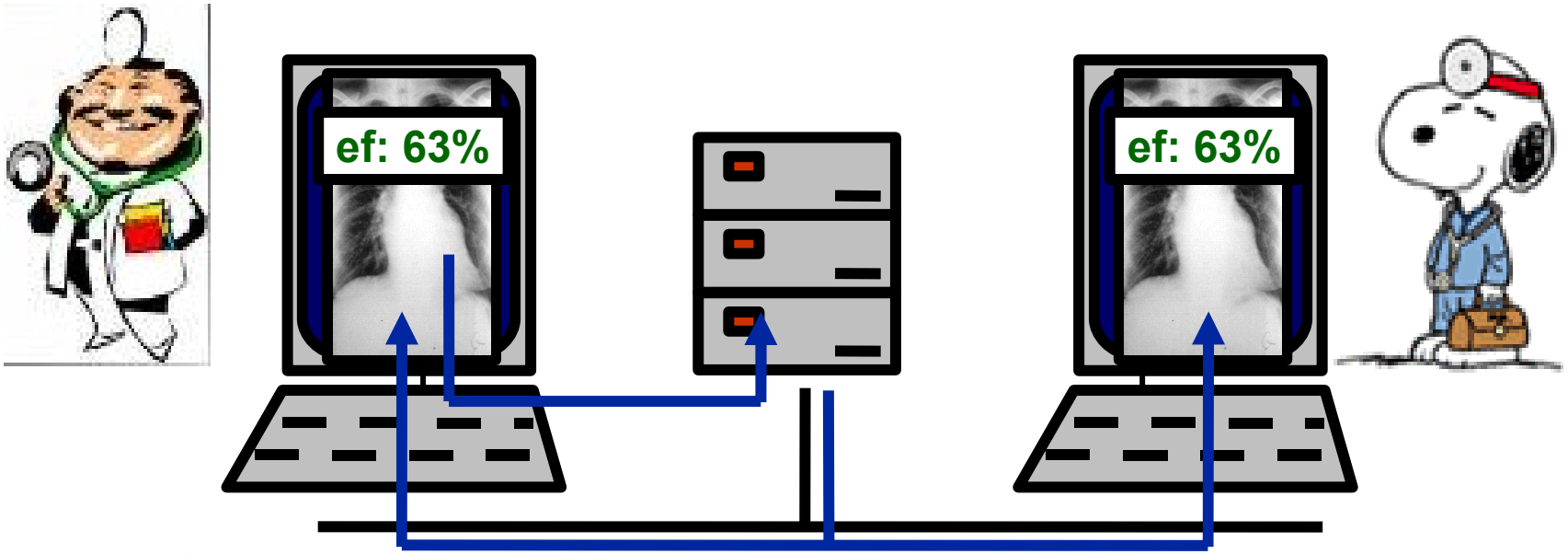


Image enhanced video-conferencing

Objective: Interactive discussion on medical images



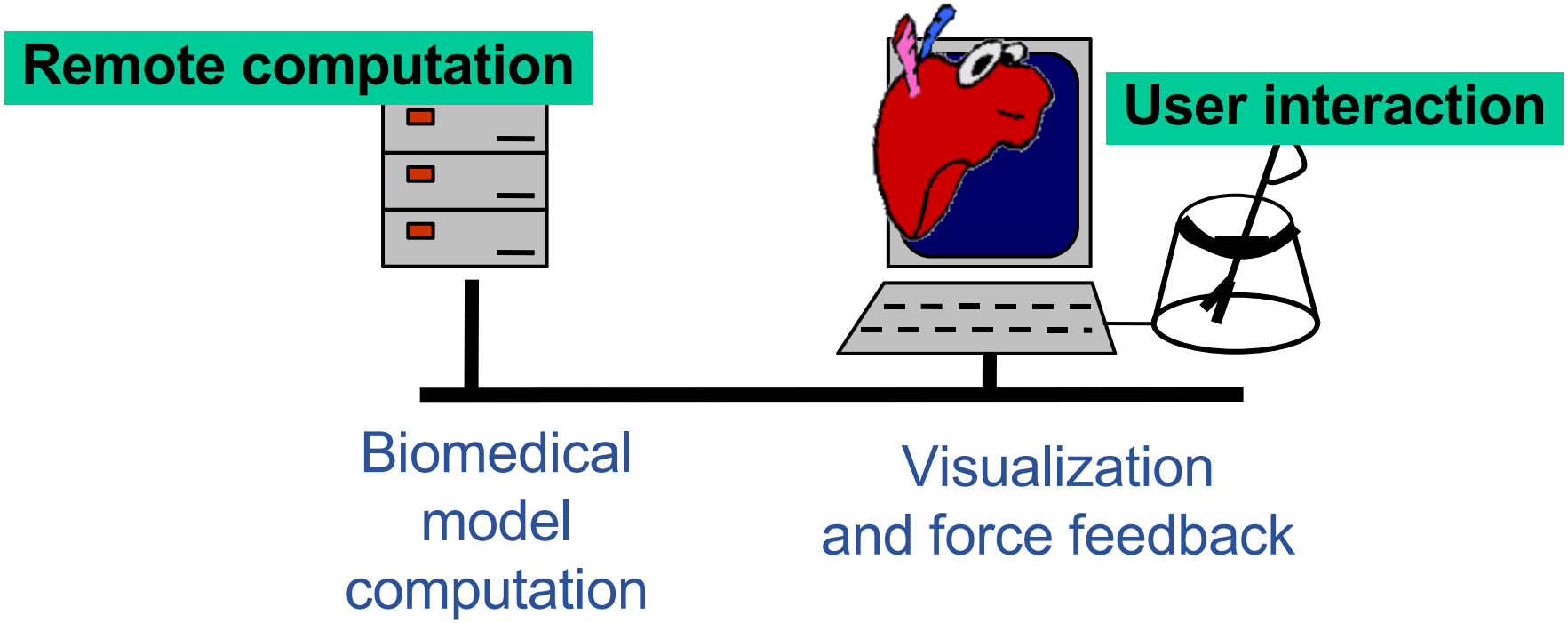
Interactive rate

High quality medical images transmission

Access to computation resources

Interactive application (e.g. surgery simulation)

Objective: real-time model interaction



Position tracking
Biomedical model-deformation
Real time visual (25 Hz) and force (300 Hz) feedback

Parallel applications (e.g. FEM)

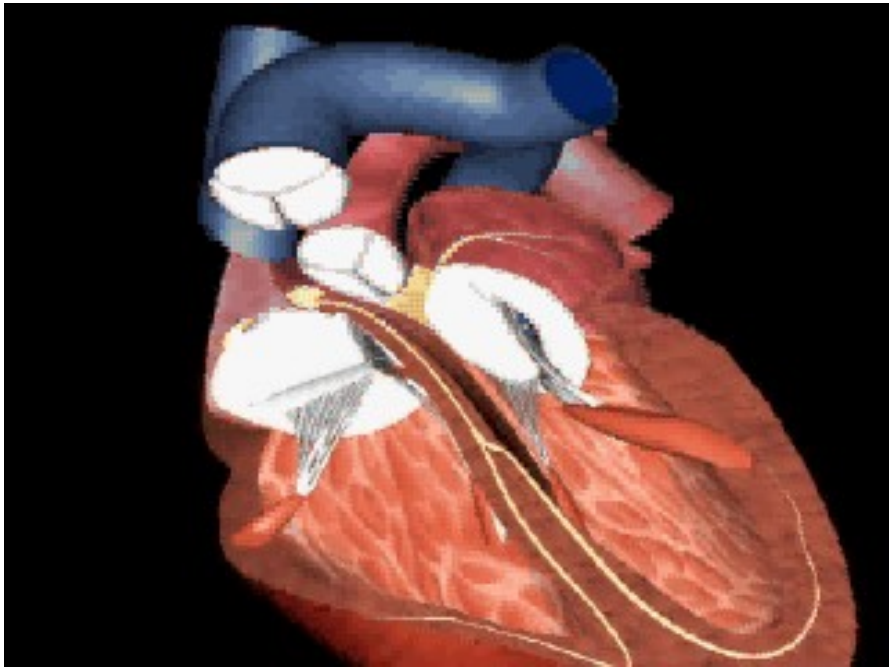
- ***Objectives: modeling heart anatomy, dynamics and physiology for heart image processing***

  bio-mecanical model

  electrical model

  very complex structure

  biological scale out of range



- ***Finite Element modeling***

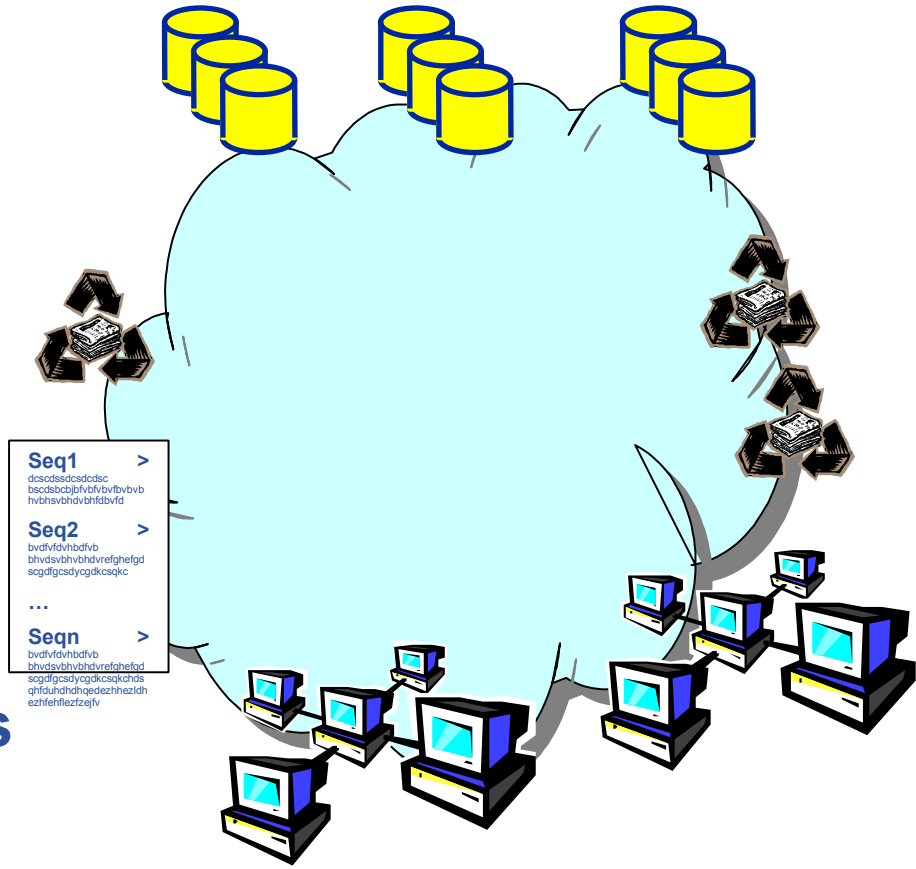
  elements oriented in heart fibers direction: fine resolution

  electrical propagation model

  MPI implementation for linear analysis optimization

Medical algorithms assessment

- **Sharing** data
  - Common datasets
- **Sharing** algorithms
  - Testing others algorithms
- **Sharing** procedures
  - Common test suites
- **Sharing** computing **resources**
  - Larger assessment studies
- **Empowering testing and comparisons**



Achievements

On different layers

Applications level

Applications
12 biomedical applications ported on the EGEE grid

High-level interfaces
Generic portals
Application specific interface

Specific biomedical services
Medical Data Management
Data-intensive workflow management
Interactive access

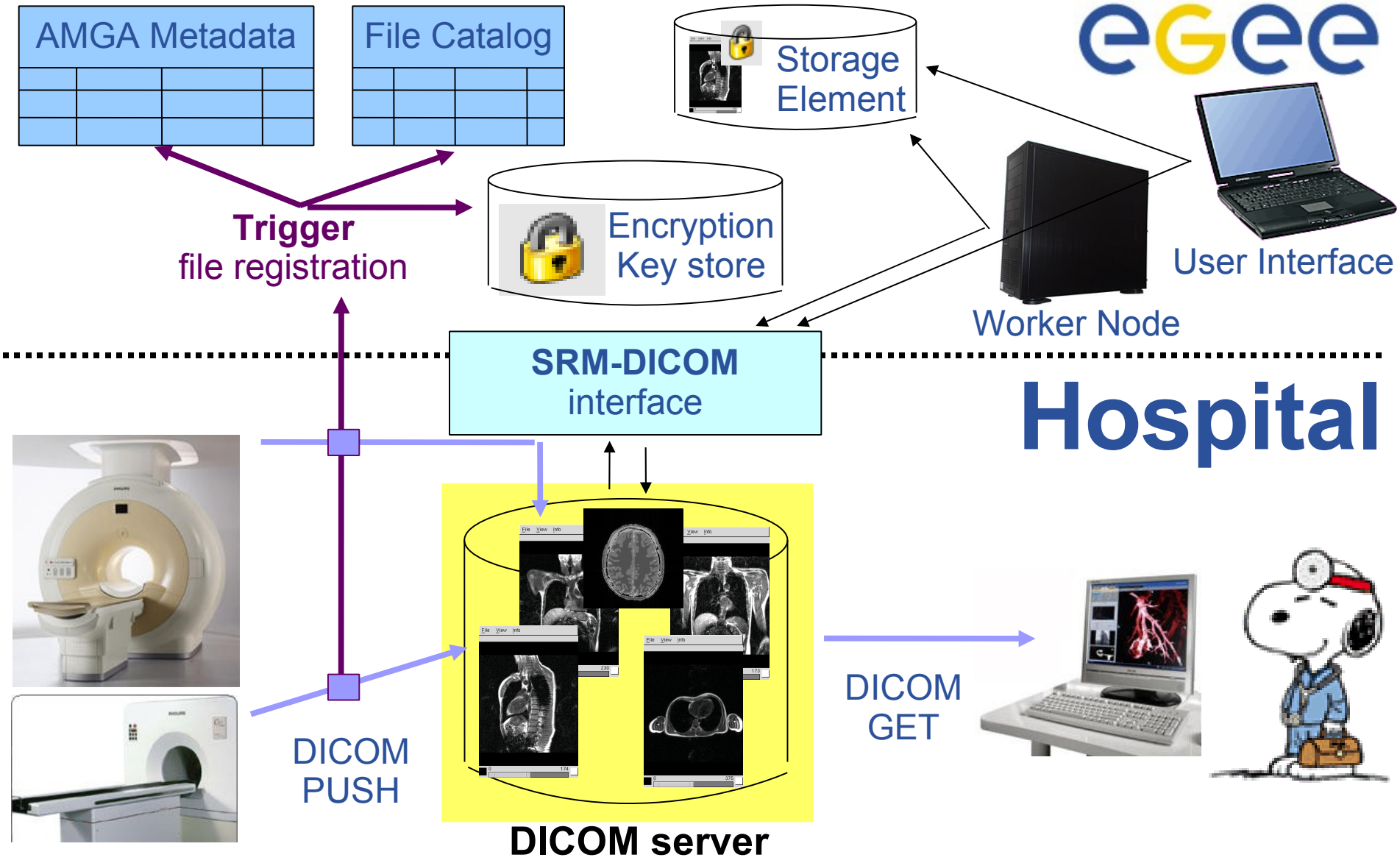
Infrastructure level

Middleware

Resources

Communication layer

Medical Data Management on Grids



Medical Data Management

Medical Imager

SRM DICOM
- MDM Trigger
- GridFTP
- gLite I/O

Grid Catalogs
- File Catalog (Fireman) — JRA1
- Encryption Keystore (Hydra) — JRA3
- Metadata Catalog (AMGA) — ARDA

Trigger:
- Retrieve DICOM files from imager.
- Register file in Fireman
- gLite EDS client: *Generate encryption keys and store them in Hydra*
- Register Metadata in AMGA

NA4

MDM Client Library

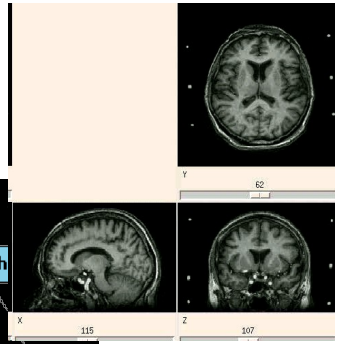
Application

Client

Client Library:
- Lookup file through Metadata (AMGA)
- Use gLite EDS client:
  - *Retrieve file through gLite I/O*
  - *Retrieve encryption Key from Hydra*
  - *Decrypt data*
- Serve it up to the application

Data-intensive workflows

100's to 1000's of medical images



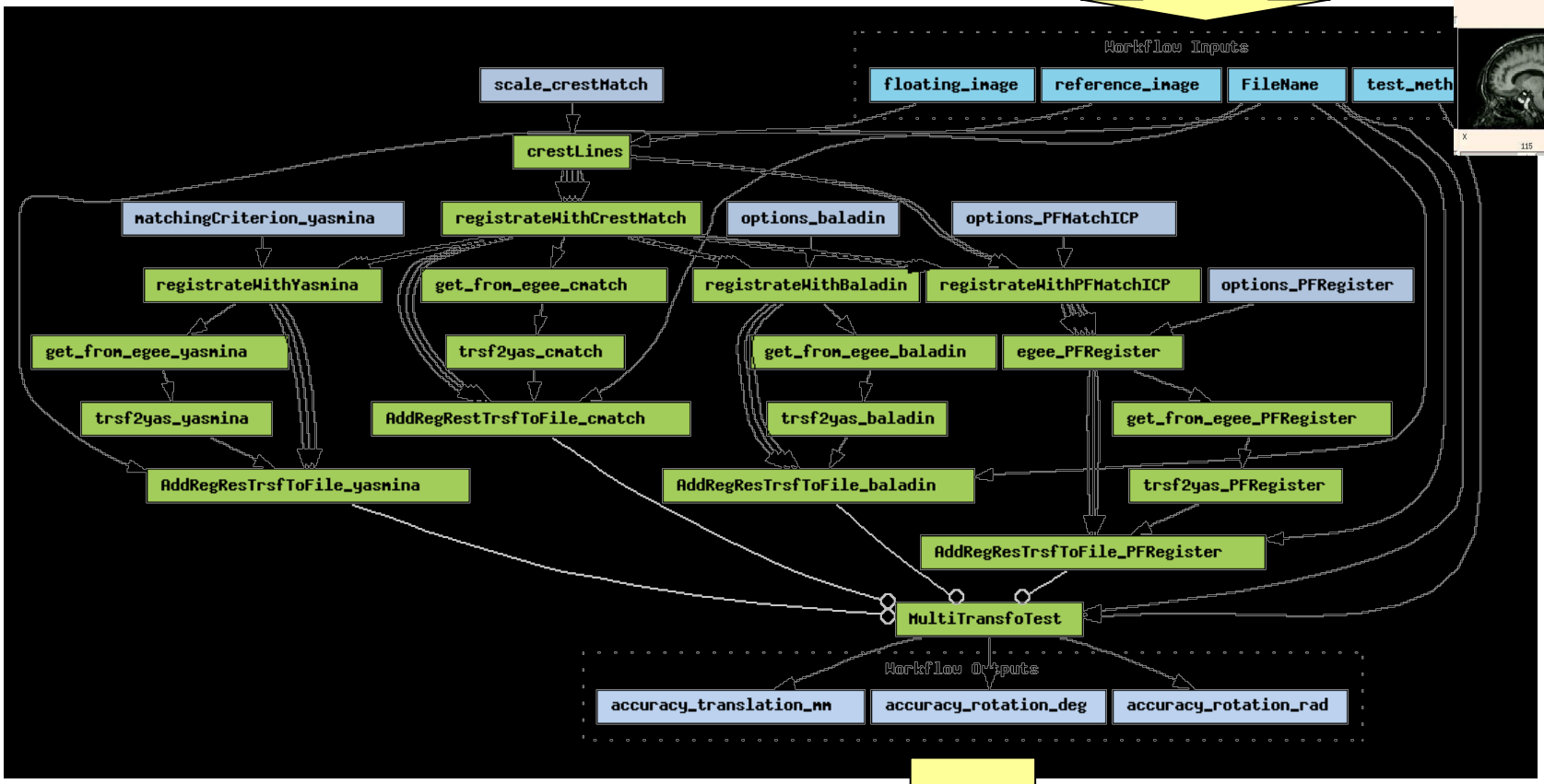


Image registration algorithms assessment

MOTEUR workflow engine

- **Data-intensive workflow manager**
  - compatible with Taverna workflow description language (Scufl)
  - Exploit grid parallelism
  - Implementing result traceability http://www.i3s.unice.fr/~glatard

- **Interfaces**
  - Web Services
  - GridRPC (DIET middleware)

AGIR

- **Execution infrastructures**

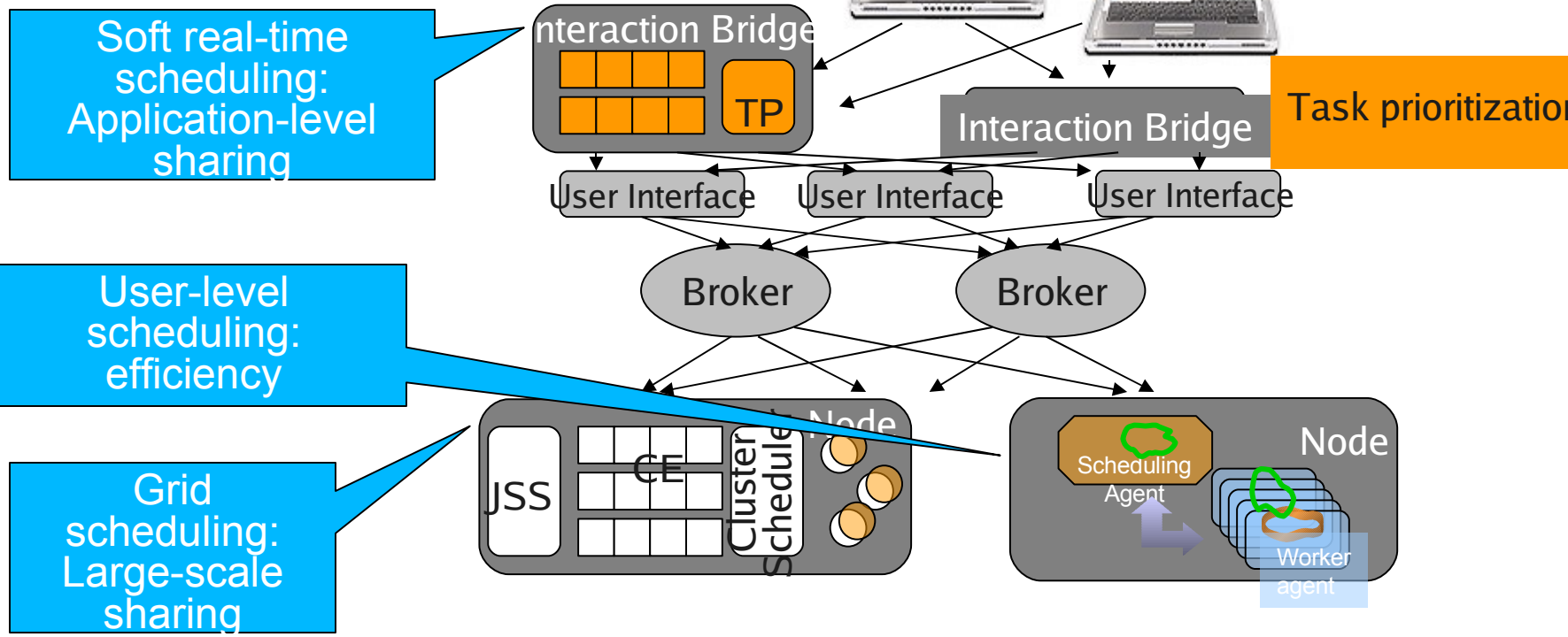
> 1000 procs
OAR batch submitter
research infrastructure

Enabling Grids for E-sciencE

> 18000 procs, 5 PB
LCG2 middleware (migration to gLite)
production infrastructure

Scheduling policies & services

- **EGEE Working group Short Deadline Jobs (chair C. Germain): Biomed, HEP, Digital Libraries,…**
- **Growing awareness:**
  - Support in gLite 3.2
  - Session at eIRG April meeting



On different layers

Applications level

Applications
12 biomedical applications ported on the EGEE grid

High-level interfaces
Generic web portals
Application specific interface

Specific biomedical services
Medical Data Management
Data-intensive workflow management
Interactive access

Infrastructure level

Middleware

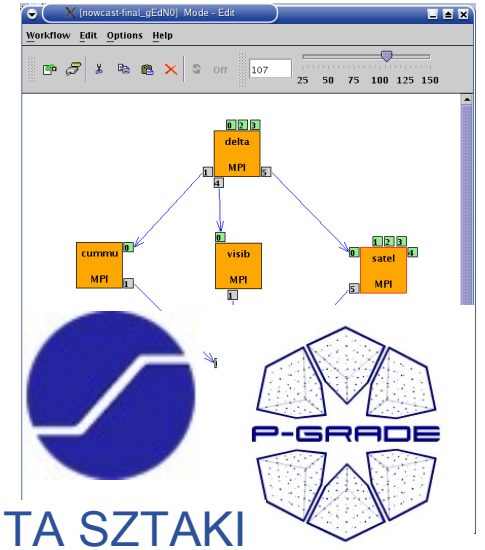
Resources

Communication layer

Generic web portals

- **GENIUS web portal,** https://genius.ct.infn.it/
  - LCG2 and gLite1 intefaces
  - Job, Data management and Information services
  - Built on the engineframe framework

- **PGRADE portal,** http://www.lpds.sztaki.hu/pgportal/
  - Multi-grids (including EGEE interface)
  - Legacy code wrapper
  - Parallel and sequential codes
  - Workflows support (DagMan, MOTEUR)

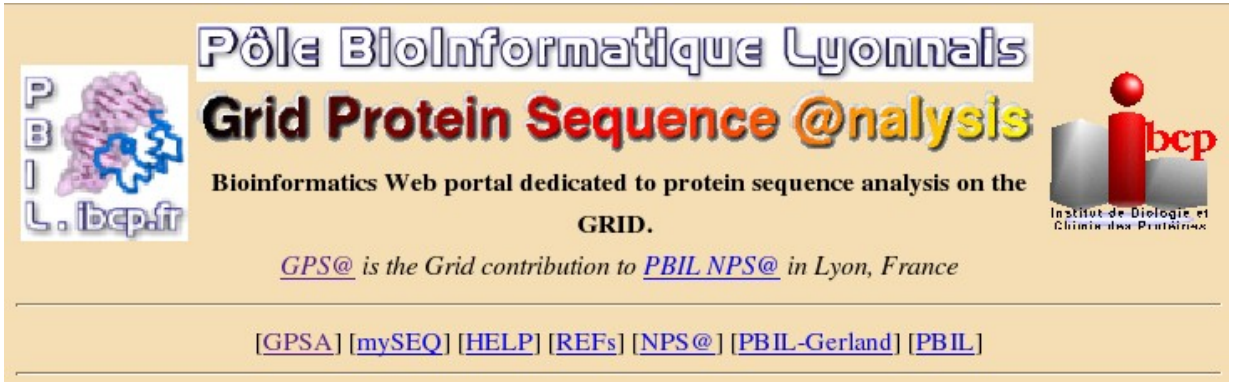


MTA SZTAKI

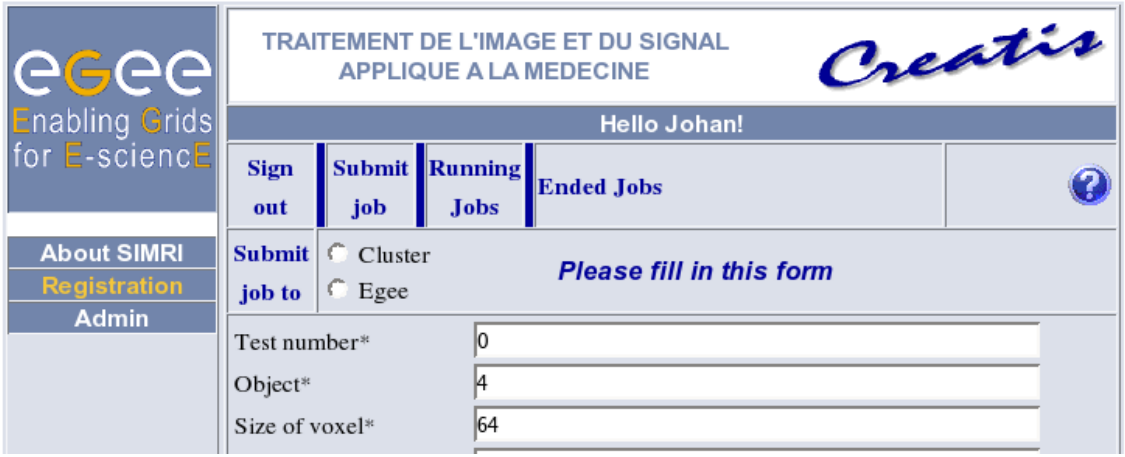
- **Still low-level for end-users**

Application specific web-based portals

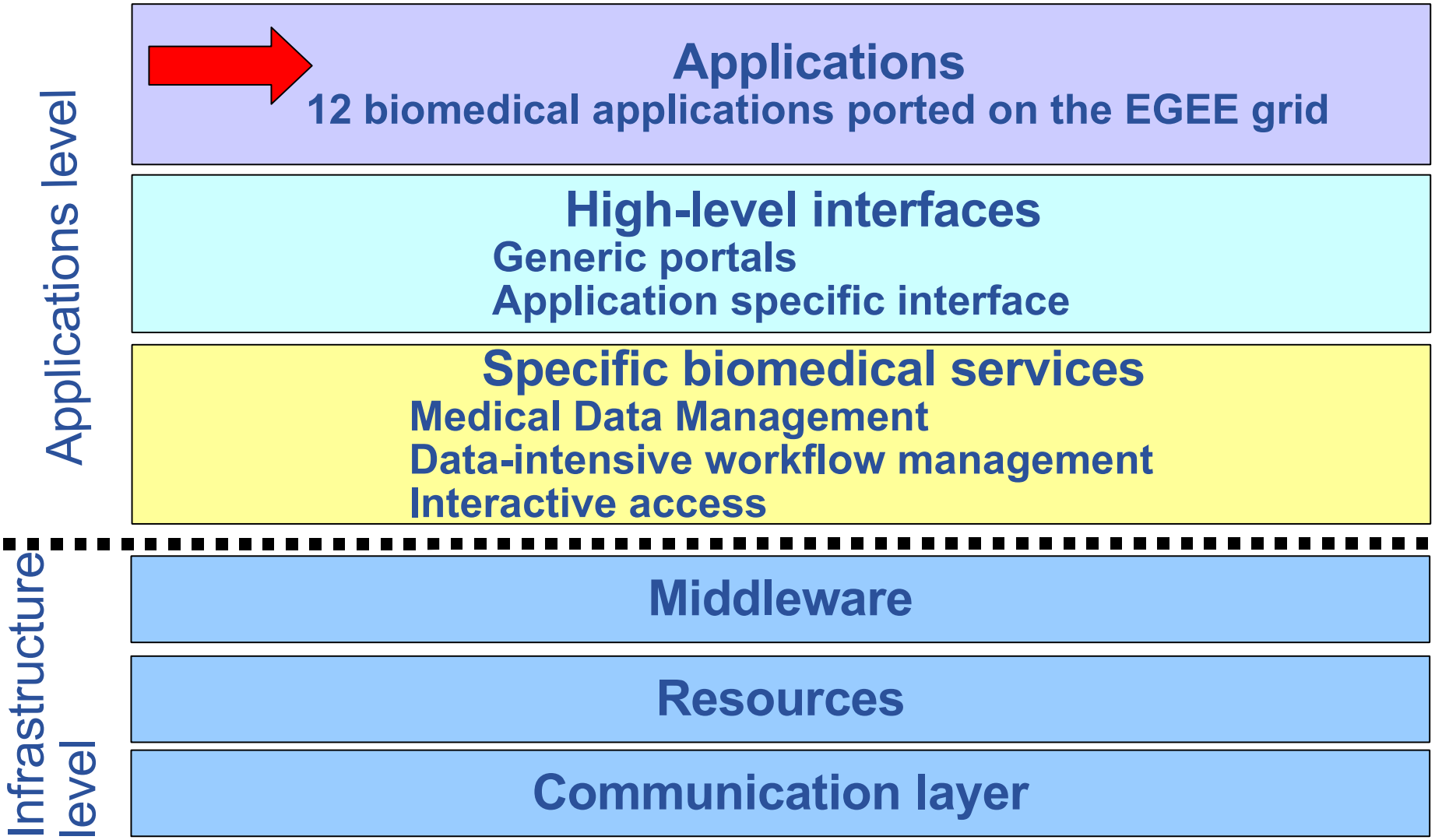
- **GPS@ portal,** http://gpsa.ibcp.fr/
  - 9 bioinformatics applications
  - EGEE interface
  - Application-specific visualisation



- **3D MRI simulator portal,** http://simri.creatis.insa-lyon.fr/
  - Transparent access to EGEE or local resources
  - Application-specific parameterization
  - Image production history

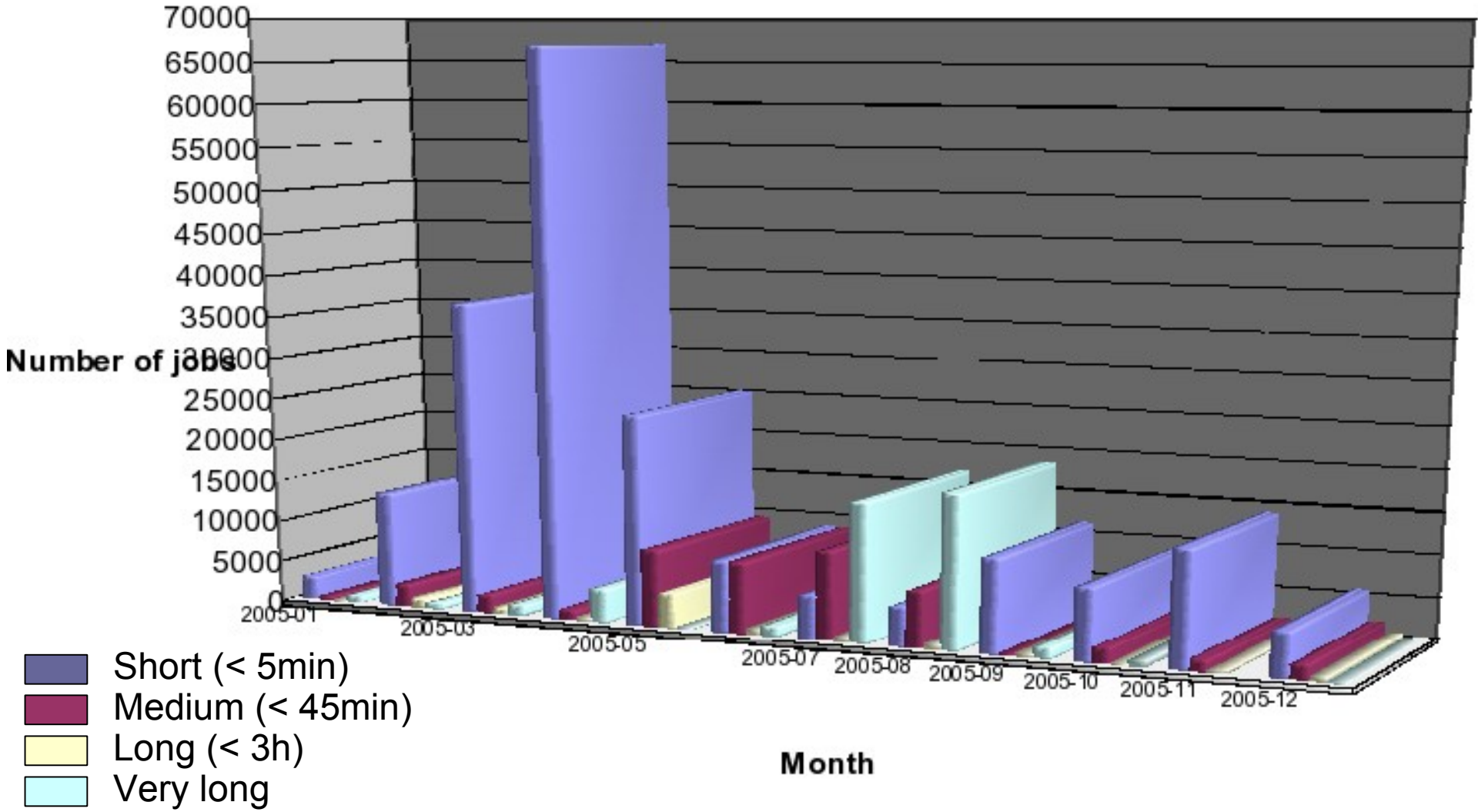


On different layers



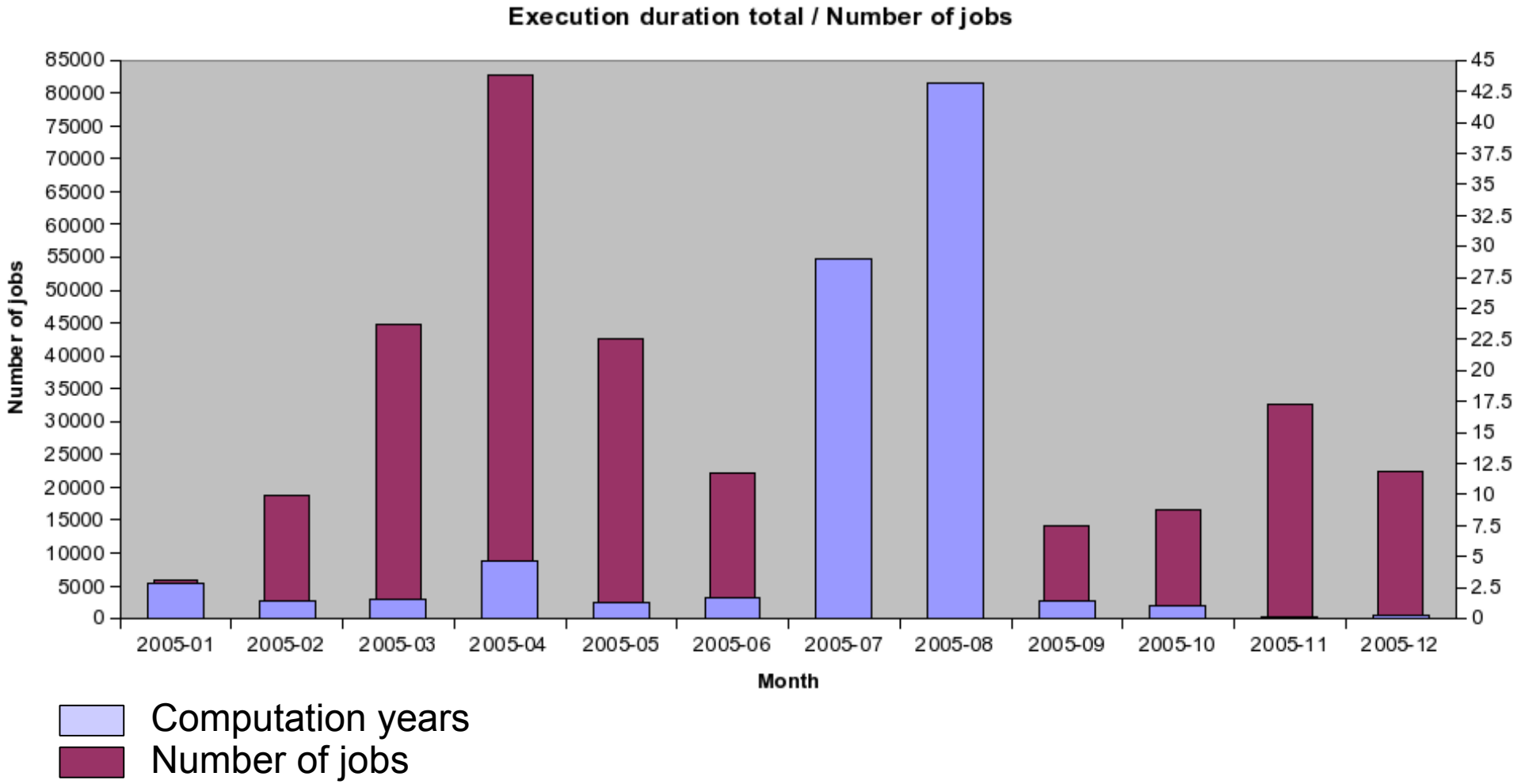
EGEE Biomedical VO activity

- **Number of jobs per month in 2005: ~25,000 jobs / month**



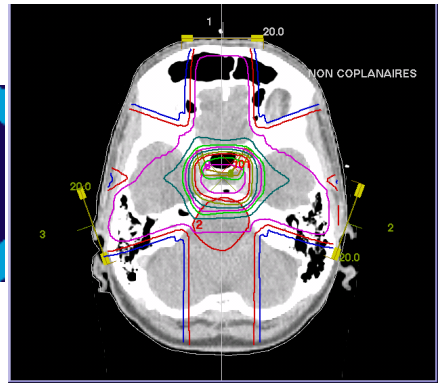
EGEE Biomedical VO activity

- CPU years / month

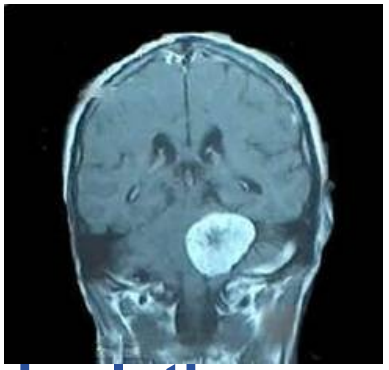


Medical imaging applications

- **GATE – radiotherapy planning**
  - Monte-Carlo simulation
  - Parallel runs
  - Ported on EGEE and DEISA infrastructures

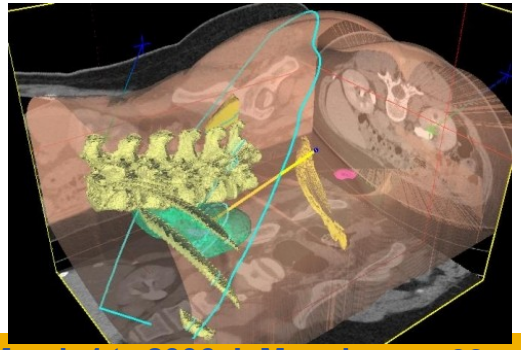


- **CDSS – Clinical Decision Support System**
  - 7 classifiers to take decisions
  - 1000'image in expert databases
  - R-GMA based



- **GPTM3D – Interactive radiological image manipulation**
  - Interactivity
  - Agent-based scheduling
  - Demonstrated at the 1st EGEE review

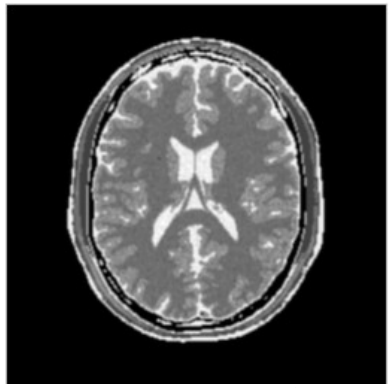
AGIR



Medical imaging applications

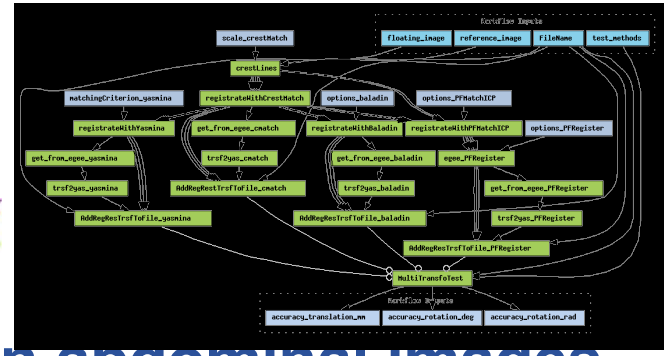
- **SiMRI3D – 3D MRI parallel simulator**
  - Parallel simulation kernel
  - Image and artifacts simulation
  - Web portal

Creatis

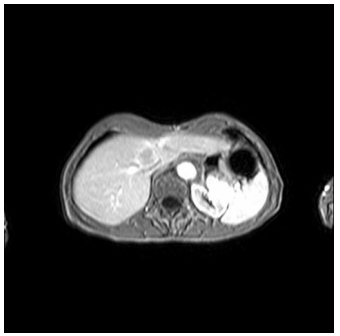


- **Bronze Standard – Registration algorithms assessment**
  - Complex workflow
  - Full image databases processing

AGIR
INRIA
i3S sophia antipolis



- **Pharmacokinetics – Perfusion study in abdominal images**
  - Image time series processing
  - Two level prallelism
  - Demonstrated at the 2nd EGEE review



Bioinformatics applications

- **GPS@ - Bioinformatics web portal**
  - Anonymous portal to bioinformatics algorithms
  - Provides gene databases
  - Serves a large user community

bcp
Institut de Biologie et Chimie des Protéines

Grid Protein Sequence @nalysis

- **Xmipp – Electron microscopy images analysis**
  - 3D molecular structure analysis
  - Compute-intensive reconstruction
  - Parallel computations

CNB

- **SPLATCHE – Gene evolution simulation**
  - Simulate past demography of human populations
  - Rejection sampling Bayesian framework
  - Parallel execution of 100,000's run

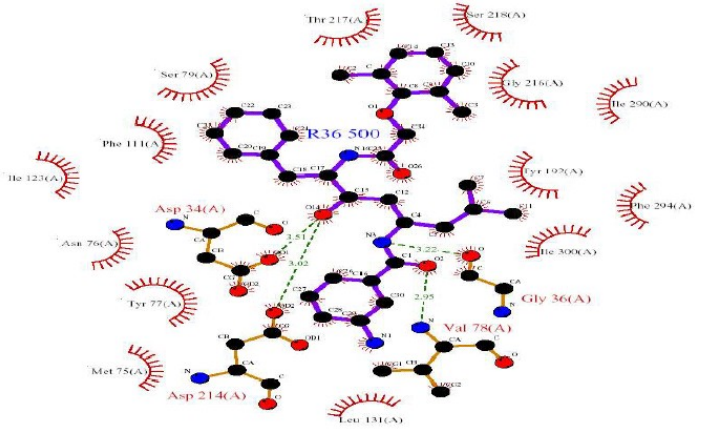
UNIVERSITÄT BERN

Drug discovery applications

- **WIDSOM – In-sillico drug discovery**
  - Autodock and Flexx docking algorithms
  - Biomed Data Challenge last summer
  - WISOM open day last December

Institut
Algorithmen und Wissen-
schaftliches Rechnen

- **GROCK – Molecular docking**
  - GRAMM and FTDOCK docking algorithms
  - Robust jobs submitter for intensive load conditions
  - Web interface



Open issues

Application requirements

- **A challenging application area**
  - Enormous amounts of data
  - Data privacy
  - Complex data sets
  - Heterogeneous data sources
  - Interactive and real-time tasks
  - Short response times requirements, medical emergency
  - Worklflows and dataflows
  - ...
- **Data security is a key point**
  - Legal: to respect (heterogeneous) national regulations
  - Technical: to ensure integrity and validity of computations
  - Sociological: to be accepted by the medical community

High level interfaces

- **On of the next challenges for biomedical application**
  - To enable access to grids for non experts
  - To provide relevant interfaces depending on the applications
- **Still a lot to be done**
  - Transparent access to medical resources
  - Compatible with clinical practice
- **Access medical resources**
  - Heterogeneity of formats
  - Different medical information systems
- **Compatible with clinical practice**
  - Non expert users
  - “push-button” algorithms

Conclusions

Grid related events

HEALTHGRID
Association

www.healthgrid.org

- **HealthGrid conferences, BioGrid conferences, Workshops...**