MICCAI tutorial on

grid services for medical image processing and registration

Introduction

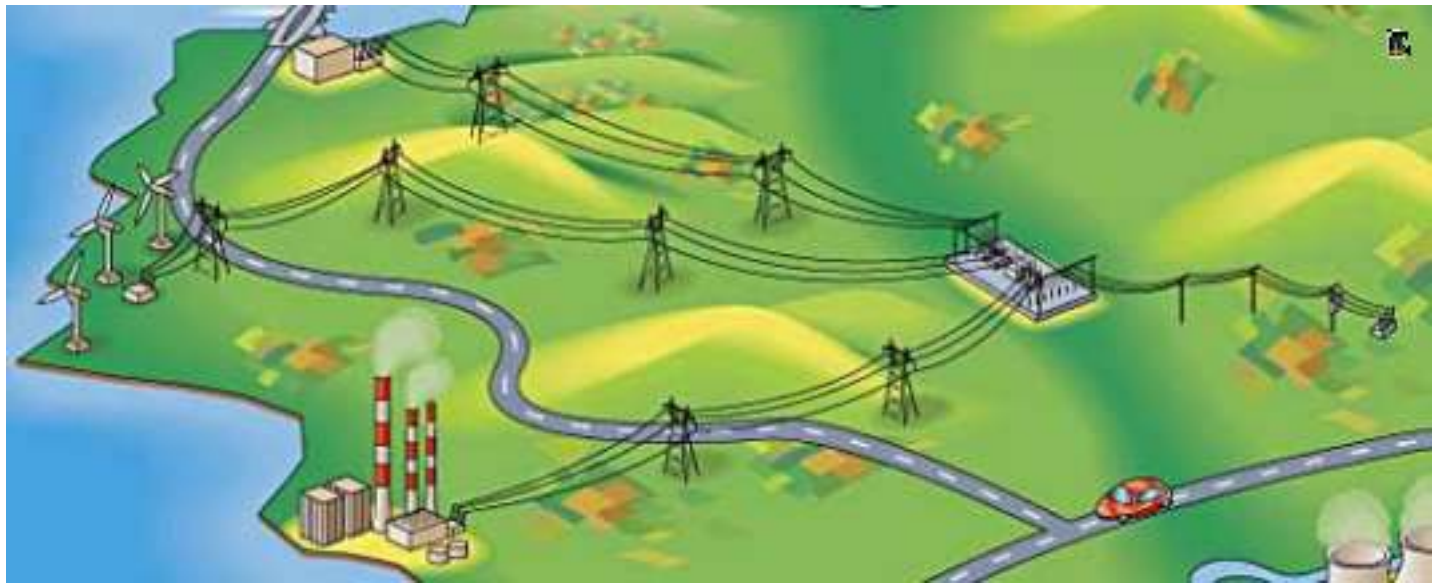
Johan Montagnat
I3S, CNRS
http://www.i3s.unice.fr/~johan/

Agenda

- 9h Grids, a tool for compute and data-intensive medical imaging applications
  Johan Montagnat
- 9h30 Mammography analysis on grids
  Mike Brady
- 9h50 Biomedical Informatics Research Network
  Ron Kikinis
- 10h10 IXI e-Science project
  Derek Hill
- 10h30 Coffee break
- 10h50 The case of medical image registration
  Daniel Rueckert
- 11h20 Round table: our vision of grid technology for medical image registration, trends and problems, opened discussion, call for participation
  X. Pennec, D. Hill, J. Montagnat
- 12h20 Lunch

What are grids?

- The myth: providing **unlimited** computing power by letting the user **transparently** access to **infinite** Internet resources.



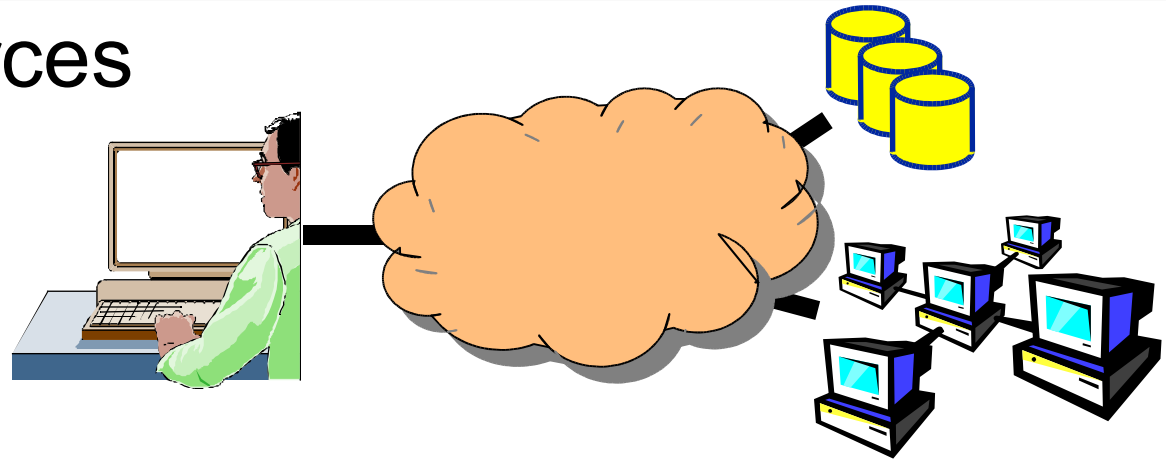
The infamous electrical network analogy

- Early grid adopters
  - Data storage: napster
  - Computation: SETI@home
  - Information: web

} Client-server technology

What are grids?

- **Assembling** resources
  - Storage
  - Computing
  - Network
- **Federating** users
  - Large scale user communities (Virtual Organizations)
  - Ease exchanges
- Pushing **standards**
  - Communication protocols
  - Data representation and formats
  - Computation control languages
- The real grid potential is in **sharing** resources, data, and knowledge.



New bottles around old wine?

- Scalability/Extensibility problems
  - Load balancing
  - Parallelization
  - Decentralization
  - Fault tolerance
  - Security...
- **Distributed computing** has addressed these kind of problems for decades!

- New wine
  - Interoperability — high level interfaces, standards
  - Authentication and autorization — certificates and policies
- Peer-to-Peer (P2P) or agent technologies mature
  - Alternative to the client-server approach

The new factors empowering grids

- Faster **networks**
  - Growth of network bandwidth vs growth of computing/storage: cheapest data exchanges
- Cheap **PCs**
  - Successful development of cluster computing
  - Some degree of standardization of hardware and software
- **Standardization** bodies
  - W3C http://www.w3c.org/ (HTTP(S), *ML, SOAP, WSDL...)
  - OASIS http://www.oasis-open.org/ (WS, security...)
  - Global Grid Forum http://www.ggf.org/ (OGSA, WS-RF...)
  - ...
- Grid **marketing**
  - Industry adoption

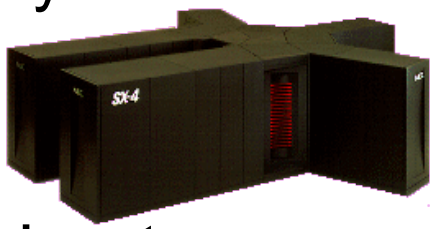
All is becoming *gridified*

- **Overuse** of “grids”
  - grid = parallelism
  - grid = cluster computing
  - grid = just name it
  - grid or grid**S**?
- Over-expectations
  - in grids deployment schedule
  - in grids capabilities
  - in transparency of grids from the user point of view
- PCs clusters are:
  - cheap... but administrators are not
  - able to deal with embarrassingly parallel applications
- Supercomputers and dedicated systems have their own playground

From PCs to supercomputers

- Several kind of grids:
  - network of computers
  - network of clusters
  - network of supercomputers

Medical imaging community



- Single PC capabilities
  - powerful, yet limited capabilities of a each host
  - general purpose network connection
- Parallelism
  - trade-off between splitting and network overhead
- Supercomputers
  - lowest communications overhead, large splitting
- Global Grid
  - Network of clusters and supercomputers
  - Resource brokering among available resources

Grids are emerging technologies

- Some components available
  - User authentication
  - High performance data transfer
  - Data managers with replication and metadata
  - Resource brokering...
- Still a lot to be done
  - Performance issues in scalability
  - High level representation of data
  - Transparency of the underlying infrastructure...
- Medical data processing is very complex
  - This community has one of the most challenging requirements list
  - Confidentiality of data and security requirements
  - Parallel processing, interactive jobs, emergency situations...

Semantics

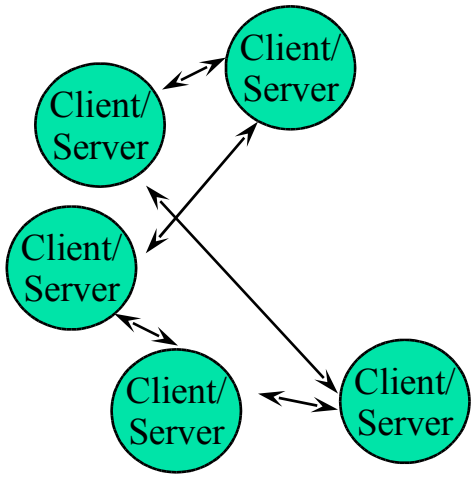
- Too much information kills information
  - To find some information on the web: google
  - Try “Mike Brady” keyword...
- Semantic web
  - Common framework that allows data to be shared and reused across applications
  - Give explicit meaning to information
  - Resource Description Framework
- Semantic grids
  - Propagate to grids what may become the web
- **Ontologies**, OWL
  - Formally describe the meaning of terminology used in a domain
  - Enable processing of information rather than just delivery to humans

Real grid enabled applications

- Very few decentralized, scalable algorithms
  - such as Domain Name Service
- How does Google backup the web?
  - 6 centers
  - 20000 PCs
  - Load distribution

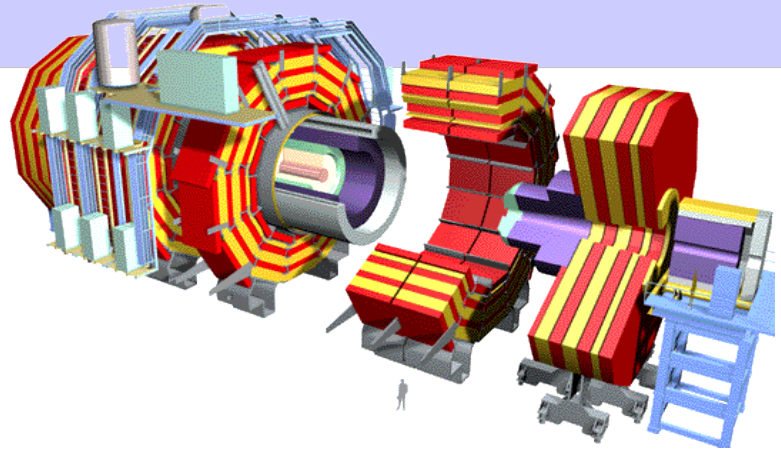
Google™

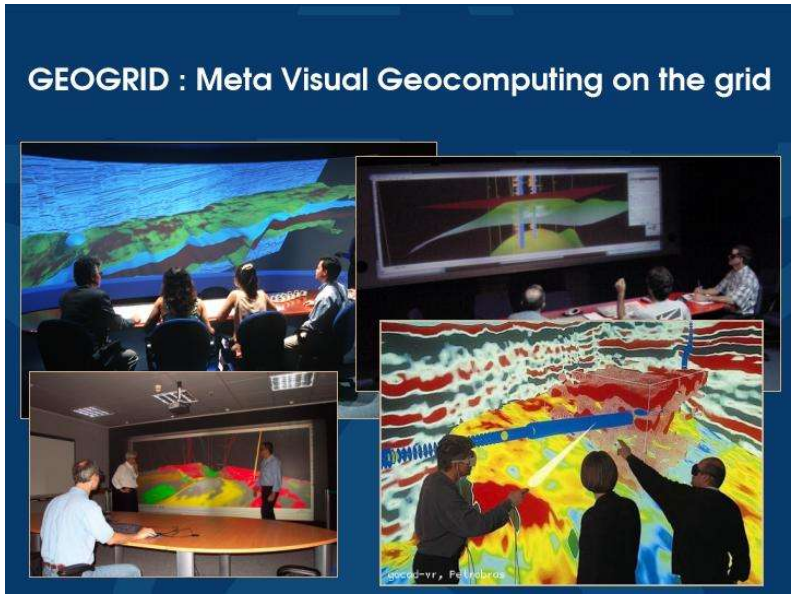
- Embarrassingly parallel applications
  - Easily distributed computations
  - Just bring in more CPU and bandwidth
- More to learn from P2P and agent technologies
  - Lot done in the field of data management
  - What about computations, knowledge extraction?



Application areas

- High Energy Physics
  - Pb of data generated per sec as the output of the Large Hadron Collider
- Astronomy
  - Astronomy images
- Earth Observation
  - Satellite images
- Geophysics
  - Sismic data revealing underground structure
- Aeronautics, industry
  - Finite Element Modeling, etc.
- ...



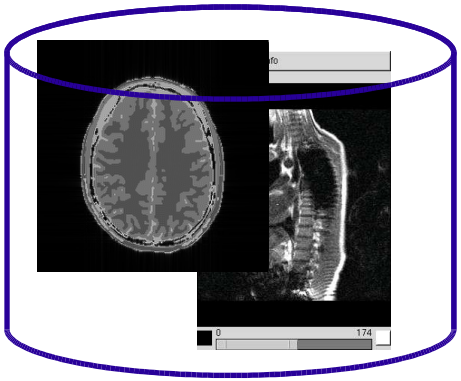


Sharing to get stronger

- Unlike electrical power, adding computers is not equivalent to add volts
  - Distributed computing conjecture : a parallel machine performance is lower than the sum of individual processors performance
- Humans are selfish
  - Every user want to use other's resources without seeing others using its resources
- The real grid potential is in **sharing**
  - datasets: to create virtual databases, esp. for rare data
  - algorithms: to foster reusability and comparisons
  - resources: to face computation picks
  - enforce standards

Medical data on grids

- Data storage and archival
  - 10 Tb of medical images/hospital/year
  - Need for long term archival (20 to 70 years)
- Datasets
  - Large scale data sets
  - Statistics, epidemiology
  - Rare diseases
  - Personalized atlases construction
- Data representation
  - Medical images
  - Metadata
  - Ontologies



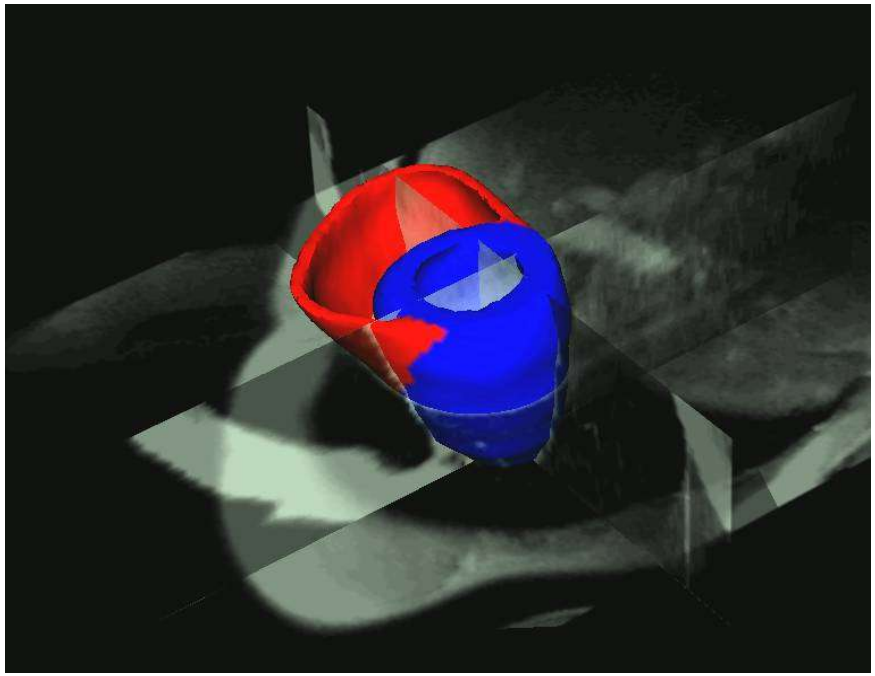
Security of medical applications

- **Authentication** and **Authorization**
  - Certificate authorities issuing certificate pairs
  - Asymmetric certificate based authentication
  - Authorization policies enforcement
- Data **access control** at individual level
  - Physicians
  - Patients
  - Researchers
- **Delegation**
  - Granting access rights
- **Encryption** for data storage and transfer
  - Best effort privacy protection
  - Key-based data encryption

owner

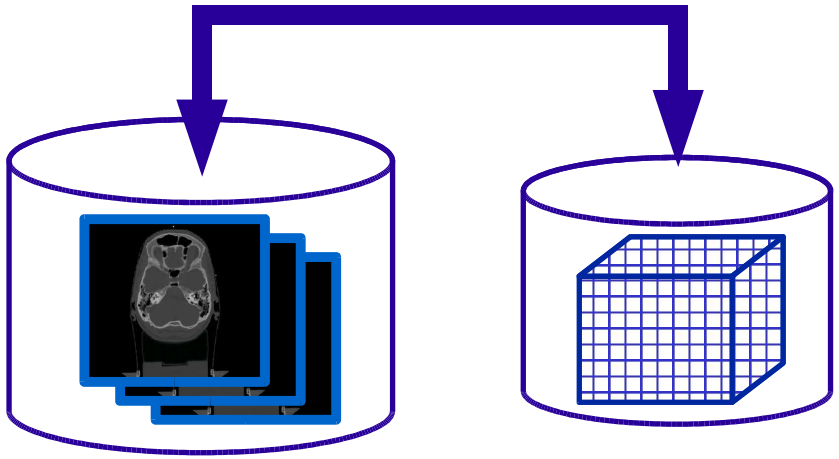
Biomedical computing

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



Standardization

- Medical data representation
  - File level: DICOM, Analyse...
  - Database level: Data storage and mediators
  - Inter-sites: ontologies

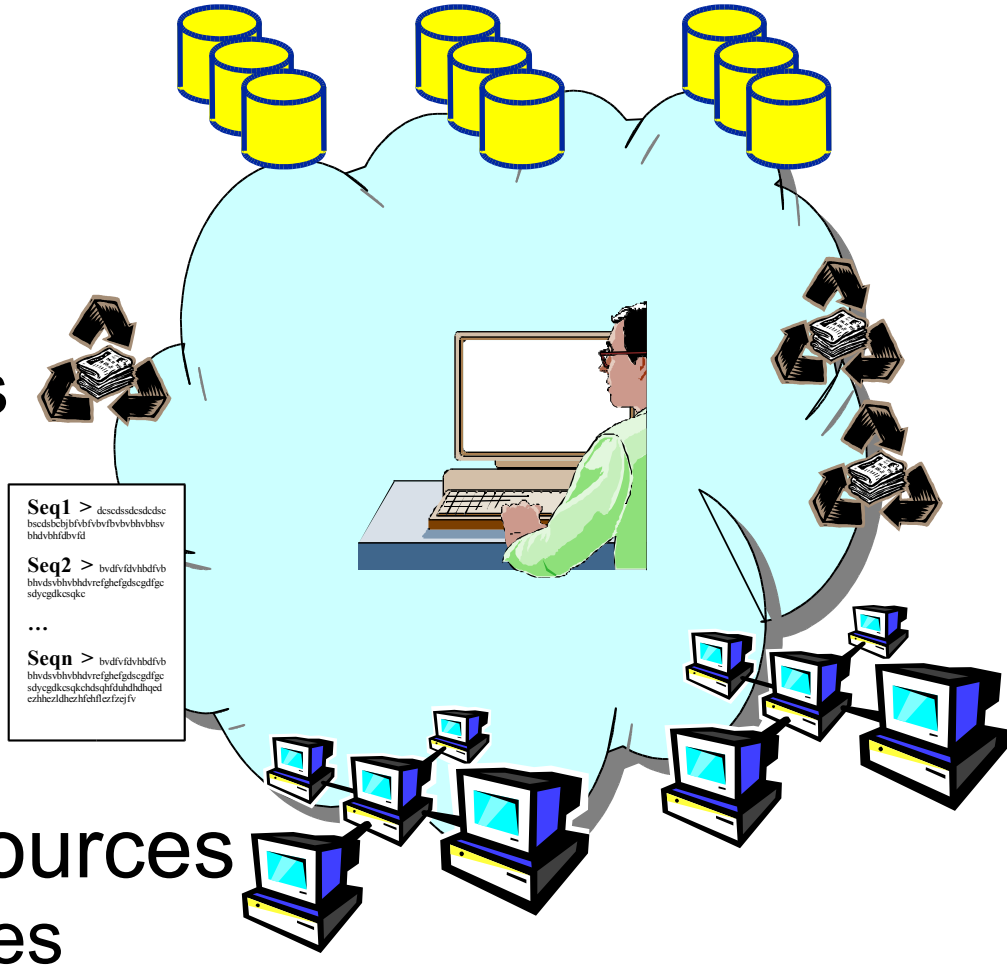


- Algorithms reusability
  - Image formats
  - Input/output
  - Parameters

- Unique opportunity to enforce standards

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



Grid infrastructures in Europe

- **EGEE**
  - http://www.eu-egee.org/
  - 27 countries, 70 partners
  - Production platform
  - 8000 PCs to be deployed in 2 years
  - LCG2 middleware (Globus+DataGrid based)

Enabling Grids for E-science in Europe

- **DEISA**
  - http://www.deisa.org/
  - consortium of supercomputing centers
  - Build a distributed terascale computing facility
  - Tight coupling of supercomputing centers by high bandwidth networks

Distributed European Infrastructure for Supercomputing Applications

References and related events

- **HealthGrid** association 
  - http://www.healthgrid.org/
  - HealthGrid conference (next event in Oxford, 2005)
- **BIRN**: Biomedical Informatics Research Network
  - http://www.nbirn.net/
- **TeraGrid**: http://www.teragrid.org/
- **UK e-Science**: http://www.rcuk.ac.uk/escience/
  - **myGrid**: http://www.mygrid.org.uk/, In silico biology
  - **IXI**: http://www.ixi.org.uk/, medical imaging
- French **ACI-GRID**, http://www-sop.inria.fr/aci/grid/public/
- **Mammogrid**, http://mammogrid.vitamib.com/
- **GEMSS**, http://www.ccrl-nece.de/gemss/
- **Crossgrid**, http://www.eu-crossgrid.org/
- …

Conclusions

- Grids are **emerging** technologies
  - Early adopters
  - Evolving field
- Yet, they are addressing medical imaging challenges
  - Standards and tools arising
  - Some successful applications deployed
  - The tool creates the need: new applications will emerge from grid infrastructures
- Need to **federate** the biomedical community
  - Strengthen the community
  - Deploy large scale infrastructures
- Need **standards**
  - To enable flexibility and interoperability

Meeting follow-up

- http://www.i3s.unice.fr/~johan/miccai/

Agenda

- 9h Grids, a tool for compute and data-intensive medical imaging applications
  Johan Montagnat
- 9h30 Mammography analysis on grids
  Mike Brady
- 9h50 Biomedical Informatics Research Network
  Ron Kikinis
- 10h10 IXI e-Science project
  Derek Hill
- 10h30 Coffee break
- 10h50 The case of medical image registration
  Daniel Rueckert
- 11h20 Round table: our vision of grid technology for medical image registration, trends and problems, opened discussion, call for participation
  X. Pennec, D. Hill, J. Montagnat
- 12h20 Lunch