HOPE, an open platform for medical data management on the grid

M. Diarena, S. Nowa, Jean-Yves Boire, V. Bloch, D. Donnarieix, A. Fessy, B. Grenier, B. Irrthum, Yannick Legre, L. Maigne, et al.

HOPE, an open platform for medical data management on the grid

M. Diarena[1], S. Nowak[1], J.Y. Boire[5], V. Bloch[1], D. Donnarieix[3], A. Fessy[1], B. Grenier[2], B. Irrthum[4], Y. Legré[1], L. Maigne[1], J. Salzemann[1], C. Thiam[1], N. Spalinger[2], N. Verhaeghe[2], P. de Vlieger[1,5] , and V. Breton[1]

[1] *LPC Clermont-Ferrand, Université Blaise Pascal, CNRS-IN2P3, campus des Cézeaux, 63177 Aubière cedex, France*

[2] *HealthGrid association, 36 rue Charles de Montesquieu, 63430 Pont-du-Château, France*

[3] *Unité de physique médicale, département de radiothérapie-curiethérapie du Centre de Lutte Contre le Cancer Jean Perrin, 63000 Clermont-Ferrand, France*

[4] *CHU Gabriel Montpied, 58 rue Montalembert 63000 Clermont-Ferrand, France*

[5] *ERIM, 28 Place Henri Dunant, BP38, 63001 Clermont-Ferrand Cedex, France*

Abstract. The paper describes a platform developed for the secure management and analysis of medical data and images in a grid environment. Designed for telemedicine and built upon the EGEE gLite middleware and particularly the metadata catalogue AMGA as well as the GridSphere web portal, the platform provides to healthcare professionals the capacity to upload and query medical information stored over distributed servers. A job submission environment is also available for data analysis. Security features include authentication and authorization by grid certificates, anonymization of medical data and image encryption. The platform is currently deployed on several sites in Europe and Asia and is being customized for applications in the field of telemedicine and medical physics.

Keywords. Grid, telemedicine, medical data, data management

1 – Introduction

The application of grid technology to healthcare has been explored within the framework of a white paper introducing and explaining the concept of healthgrid [3]. A healthgrid is an innovative use of information technology to support broad access to rapid, cost-effective and high quality healthcare. Recognized as an emerging aspect of eHealth, it is an environment where data of medical interest can be stored, processed and made easily available to the different actors of healthcare, physicians, healthcare centres and administrations, and of course citizens.

A number of projects have been exploring the use of grids for medical data and image management. Among others, MAMMOGRID [12], MEDIGRID [13] and Health-e-Child [14] have been developing middleware and tools for this purpose.

The platform described in this paper was designed to enable telemedicine in a grid environment. The concept of telemedicine covers all the potential means to exchange medical data or images whatever the distance between two physicians. The sharing of knowledge and the exchange of diagnosis between physicians contribute to the improvement of the standard of medical knowledge, particularly in developing countries [1]. As stressed in a recent report from the European Commission and the World Health Organization, "Information and communication technologies are changing health care delivery and are at the core of effective, responsive health systems. These technologies are key to connecting people, information and research to improve health in countries. They are also vital in enabling rapid response to global threats to health." [2].

One of the main added values of grids for telemedicine is to offer to healthcare professionals the capability to query distant databases in a transparent way. Today, whenever a physician wants to share medical data with a colleague for the purpose of a second diagnosis, he or she has to send him the information by email or to make it available in a common repository.

The grid technology allows changing this procedure. Indeed, let us take the example of two teams of clinicians in two different countries in need of exchanging medical information for a second diagnosis on patients treated in country B. As illustrated on Figure 1, the main advantages of the grid-enabled telemedicine are the following:

- Patient data are kept in databases inside hospitals where patients are treated. As a consequence, the physicians can keep full control on the data they make available for second diagnosis. At any time, they can close external access.
- Access for second diagnosis can be strictly controlled through the authorization and authentication mechanisms of the grid. Only a subset of the patient record can be exposed through the grid.
- In case additional information is needed, the team of clinicians requesting the second diagnosis can provide access to more patient information or perform new examinations and store the results on their local database.

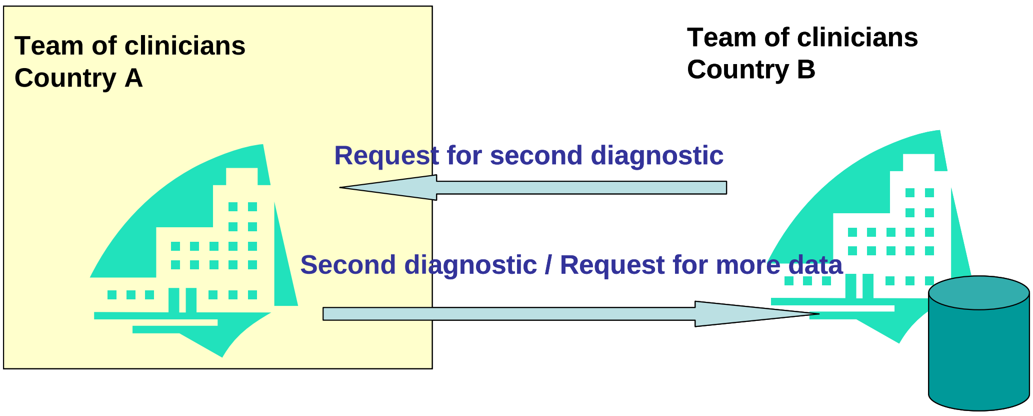


Figure 1. Schematic representation of grid-enabled telemedicine between two teams of clinicians.

Telemedicine is often initiated through collaboration between two clinicians but telemedicine networks are now being set up involving teams of clinicians in several countries.

As illustrated on Figure 2 in the simple case of 3 groups of clinicians, the grid approach adds a very significant added value to enable telemedicine networks:

– Several teams of physicians can easily be involved as the data are always kept where they are produced. Adding the diagnosis of another clinician only involves granting him/her access to a view of the data and can be interrupted at any time.
– The data being distributed over the participating hospitals, there is no single point of failure. A network breakdown should only have local impact.

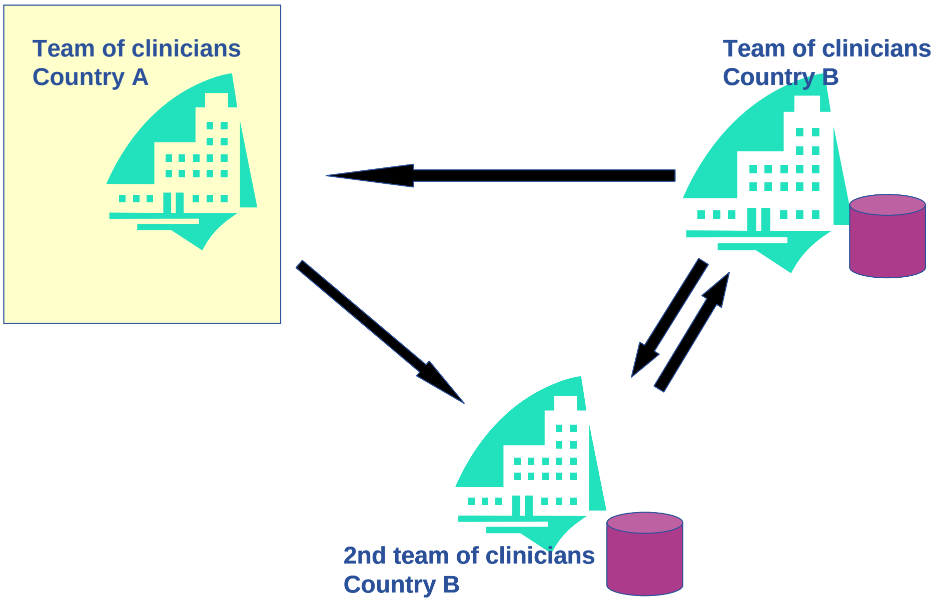


Figure 2. Schematic representation of grid enabled telemedicine between 3 clinician teams.

Grid-enabled telemedicine is particularly appealing for training. The example shown on Figure 2 illustrates how two teams from Country B are able to exchange diagnosis on medical cases with the supporting expertise of clinicians from country A. In the case of a network of hospitals, a subset of the medical records can be selected and tagged in each hospital for training purposes. These medical records can be used for training sessions through e-learning or videoconference.

The paper is organized as follows:

– Following this introduction, section 2 describes the platform structure and architecture. Details are provided on the data, images and job management features as well as the security framework

- Section 3 describes how the platform has been customized for applications in medical physics and neurosurgery within the framework of collaborations with local clinicians.
- Section 4 draws the main conclusions and proposes some perspectives

2 - Material and methods

To allow physicians to manage and exchange medical data and images, the platform uses web services technology and grid services provided by gLite middleware [4]. Physicians access the platform using a web portal developed with the GridSphere [5] portlet container that presents to them a user-friendly interface to access several distributed medical services that manage images and medical information. Medical information is stored locally in the hospital where it is produced using the AMGA metadata catalogue [6]. Information between services deployed in different locations is exchanged using the SOAP messaging protocol [7]. Medical images are stored anonymized and encrypted on the grid while their corresponding metadata are stored in the local AMGA server. The proposed medical platform allows submitting, monitoring, and managing medically-related jobs such as for instance treatment planning for radiotherapy treatments.

2.1 Platform architecture

The platform goal is to enable the exchange and treatment of medical information over the grid in a secure way, which means exposing the information to the grid users which are allowed to access it and treating the medical information with analysis and job management facilities.

The platform was designed to fulfil additional requirements:

- The applications running on the platform must be easy to use as healthcare professionals are too busy to learn complex tools. In particular, the grid complexity must be completely hidden.
- The platform should be adaptable to any medical field. It must be flexible so it can be customized to adapt to the “modus operandi” of its users.
- The platform must be able to interact with existing hospital data and image archiving systems (PACS).
- The platform should be fault tolerant in the perspective of clinical routine
- The platform should not require global connectivity. Any network failure outside the hospital should not affect data treatment in the hospital, nor the exchange and treatment of the medical information outside the area affected by the failure.
- The platform should fulfil all legal requirements in relation to security

The platform architecture is built on four pillars:

- All the low-level data management, communications and grid interaction features are developed using web service technology (business logic layer);
- The healthcare professional accesses all the platform services through a web portal developed with the GridSphere [5] portlet container (presentation layer);

– The medical data are stored in an AMGA server inside the hospital. AMGA is the metadata catalogue developed within the framework of the EGEE project;
– The grid infrastructure provides the storage and computing resources needed to store the medical images and to treat the data according to the needs of the users.

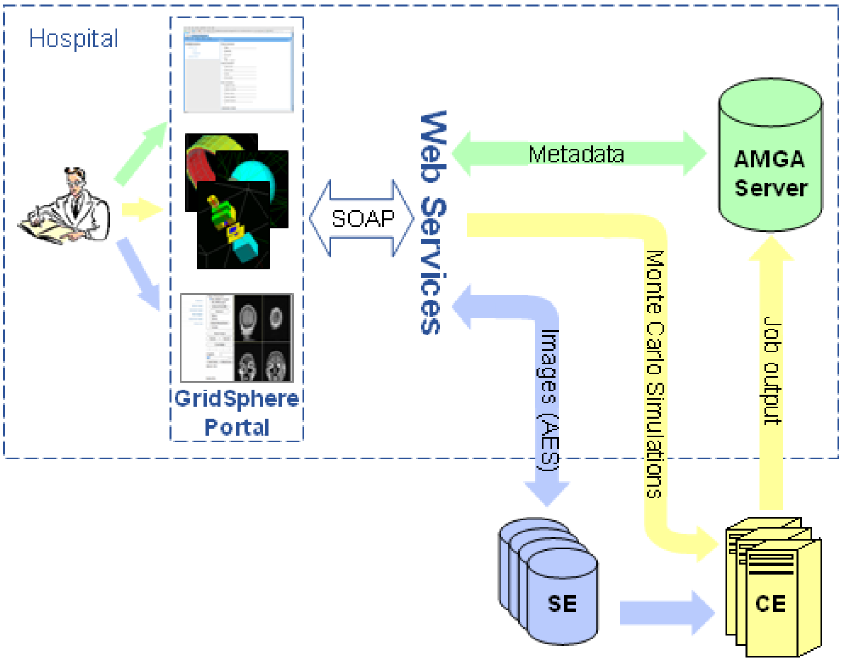


Figure 3. Information and image management inside a hospital

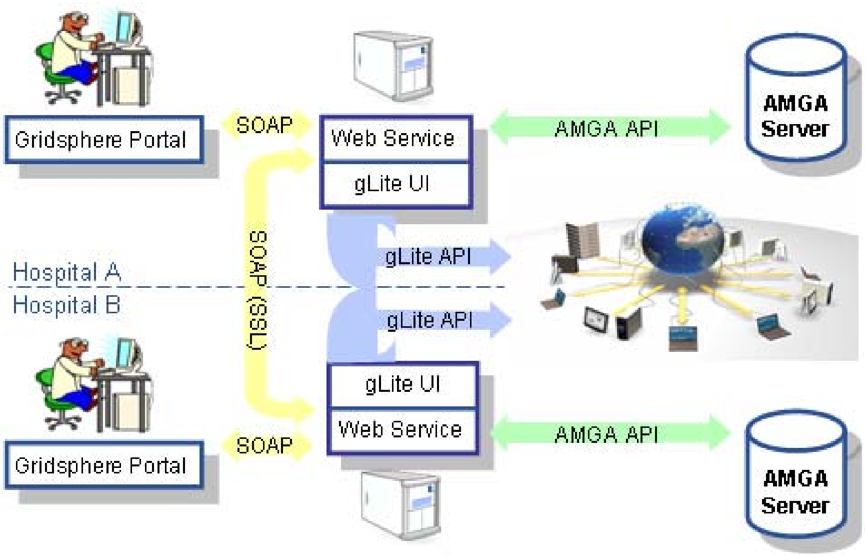


Figure 4. Information and image sharing between hospitals in different locations

2.2 The web service layer

As already said, speaking about the telemedicine platform technical organization, the first big distinction that can be made regarding its architecture is the separation between the presentation layer, developed using Gridsphere portal container, and the business logic part which is developed using web services and manages all the low level interaction with different grid services.

This architecture keeps separate the user-oriented part of the platform from the grid-oriented one allowing also the design of a personalized platform interface for each different customer without changing anything on the data management or grid interaction side.

The choice of web service technology comes from the need to use standard and well-supported communication protocols (SOAP in our case) to guarantee interoperability and security between different instances of the platform installed in remote hospitals.

The web service layer manages all routing mechanisms used to exchange SOAP messages between remote platform instances allowing physicians belonging to different and remote hospitals to share patient files, clinical exams and medical images.

The adoption of standard communication protocols provides high-level interoperability, worldwide support and all the flexibility needed for a distributed telemedicine platform. Moreover, it also guarantees a high level security not easily achievable using ad-hoc developed communication protocols.

The grid offers a huge amount of computational resources that can be exploited to speed up complex image analysis algorithm. The proposed platform is able to submit, monitor and manage jobs on the grid thanks to the experience gained developing job management web services for the new WISDOM production environment [8].

2.3 The Gridsphere web portal

The user interface built on top of the web service layer has been developed using the GridSphere web portal [5]. Web technologies are well-adapted to the project, especially to the environment in which it runs: grid or communication can be achieved only if a rather developed network exists. Moreover, using a web-based user interface has several benefits: first, almost every computer can use a web browser that can display the platform, which means little or no installation process for client machines. Second, it allows an easy access to the platform, as any computer connected to the network can access it. As a side-note, Gridsphere is a servlet and can also be installed on almost every kind of system.

More specifically, Gridsphere is a portlet container, which means it is able to display portlets along with regular markup code. Portlets are servlets that generate mark-up code fragments. A portlet container (e.g. the Gridsphere portal) is able to use these fragments to build a complete web page. The use of portlets brings a more modular user interface: every fragment can be integrated the way you want, without any concern for their content, allowing modification of the user interface to adapt it to specific needs.

The platform offers numerous possibilities on its pages which contain one or more portlets. Portlets we developed communicate through the HTTP session for two reasons:

- to perform complex tasks requiring several portlets
- to ensure that the user choices (such as the patient currently managed) are stored and kept whatever portlet is displayed.

2.4 AMGA server and medical information management

Medical information and DICOM [9] metadata extracted from medical images are managed using AMGA metadata catalogue. This grid service provides access to different database back-ends in a uniform way and independently of the particular database system used. This makes AMGA a really interesting service to develop software solutions in a distributed and heterogeneous environment.

In our deployment plan an AMGA metadata catalogue is installed in each hospital to allow our platform to access and manage in a uniform and coordinated way both medical information and distributed physician's credentials.

As already explained, a strong requirement for the telemedicine platform architecture is to be as generic and flexible as possible so that it can be easily adapted to manage database structures which are created from real medical sheets used in hospitals and which are different for each medical disciplines. To adapt to the existing hospital information systems, the architecture should allow different hospitals to specify a different (and usually personalized) database structure for the same medical field.

To make this possible we decided to use the XML [10] mark-up language to represent all the information managed and exchanged in the platform. XML is a standard and generic mark-up language that allows defining and representing in a clean and unambiguous way any kind of structured data, including medical information (from database table structure up to any kind of patient or clinical exam sheet).

Following is an example of an XML document describing a possible patient table inside the platform:

```
<?xml version="1.0"?>

<table name="Patient" type="data (patient)">
   <group name="PATIENT_INFO"/>
   <group name="FAMILY_INFO"/>
   <group name="OTHER_INFO"/>
   <attribute name="NAME" type="varchar(40)" group="PATIENT_INFO" priority="1"/>
   <attribute name="BIRTH_DATE" type="date" group="PATIENT_INFO" priority="0"/>
   <attribute name="BIRTH_WEIGHT" type="float" group="PATIENT_INFO"/>
   <attribute name="SEX" type="sel:M/F" group="PATIENT_INFO" priority="0"/>
   <attribute name="FATHER_NAME" type="varchar(40)" group="FAMILY_INFO"/>
   <attribute name="MOTHER_NAME" type="varchar(40)" group="FAMILY_INFO"/>
   <attribute name="ADDRESS" type="varchar(256)" group="FAMILY_INFO"/>
   <attribute name="PHONE_NUMBER" type="varchar(18)" group="FAMILY_INFO"/>
   <attribute name="NEIGHBOUR_NAME" type="varchar(40)" group="OTHER_INFO"/>
   <attribute name="NEIGHBOUR_ADDRESS" type="varchar(256)" group="OTHER_INFO"/>
   <attribute name="AUTHORITY_NAME" type="varchar(40)" group="OTHER_INFO"/>
   <attribute name="AUTHORITY_ADDRESS" type="varchar(256)" group="OTHER_INFO"/>
   <attribute name="GENERAL_COMMENTS" type="comment" group="OTHER_INFO"/>
</table>
```

Starting from such a file the platform generates automatically both the database structure inside AMGA server and the user interface in the Gridsphere web portal.

This architecture allows managing complex data without any a priori restriction on the data structure to use to represent patients and associated clinical exams. Together with SOAP protocol and the Web Service technology, the use of XML documents to store and exchange information makes the platform completely independent and decoupled from any particular database structure deployed inside each hospital and makes data sharing between physicians belonging to different hospital even easier to implement.

2.5 Security aspects

Medical data management demands high level standards in terms of security and data privacy both concerning data access policy and data security on the grid.

To guarantee this high-level of security, we developed our platform keeping in mind some main requirements:

- User authentication using grid certificates;
- Local and inter-hospital communications encrypted using SSL;
- Fine-grained user authorization policy based on AMGA ACL mechanism;
- Medical images stored anonymized and encrypted on the grid.

Using grid certificates issued by authorized certification authorities is sufficient to grant access to the platform only to users owning a valid certificate. This step establishes a first level check on user identity and authentication. Data access policies inside the platform are managed using security capabilities made available by AMGA metadata catalogue. In particular we use username and password authentication to access the medical data and to create data access control lists (ACLs).

Presently, we are also working to integrate in our platform a new authentication and authorization system based on professional smart cards granted by a public consortium to all professional physicians. When this integration will be ready, the authentication and authorization process will be even more easy and secure.

All information concerning patients belonging to a given hospital are kept inside the hospital itself and are never replicated outside. This allows to increase security because the only access point to the sensitive patient information is through the web portal that is the only component in the platform able to communicate, through the web service, with the private database inside the hospital.

Communications between physicians and the platform as well as all the inter-hospital communications are secured using SSL protocol to avoid information theft during their transport.

All the medical images, stored on the grid, are anonymized and encrypted using AES encryption algorithm before they are sent outside the hospital; their decryption key is stored inside the local AMGA server and is accessible only to authorized users.

3 - Application to telemedicine and medical physics

The platform was designed from the beginning to address the needs of telemedicine which requires sharing image and information between hospitals in different locations. As a first medical use case, the platform was tested for a neurosurgery application between French and Chinese clinicians.

The platform functionalities to access the hospital's Picture Archiving Communication Systems and to submit jobs over the grid were also soon identified to be particularly relevant for radiotherapy treatment planning.

In this section, we are going to present how the platform is currently being customized and deployed for these two applications.

3.1 Application to telemedicine

3.1.1 Introduction

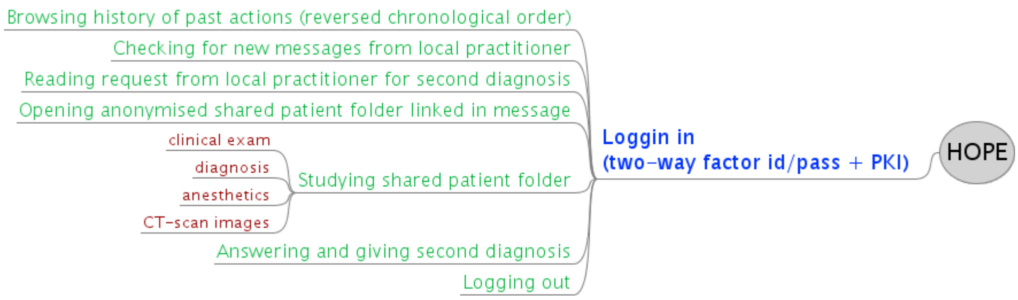
Chain of Hope [11] is a humanitarian organization which sponsors surgical missions in developing countries to train local surgeons as well as treat children. As these missions are limited in time, key issues for their success are the identification of patients needing surgery before the mission and their post-surgery follow-up after the mission. Another issue is related to the transfer of skills and techniques which should not be limited to the mission durations. Communication means are needed to keep exchange of information between European and local clinicians for remote consulting.

Since March 2000, humanitarian missions were conducted in the neurosurgical unit of the Regional Hospital of District of Chuxiong (Yunnan, PRC) by French clinicians in order to treat hydrocephalus and some other neurosurgical pathologies in children (tumours, malformations, …) which were not treated previously in this hospital.

Usage of computerised medical files including clinical and radiological data shared by each medical team was identified as a relevant tool for telediagnosis of new patients medical problems, to solve some difficulties of diagnosis and to specify the indications for operation. Such usage provides also French neurosurgeons with information concerning the follow-up of patients seen by local neurosurgeons between two missions. The sharing of these computerised medical files between several medical teams (Clermont-Ferrand, Shanghai, Chuxiong) working together to achieve these missions could also improve knowledge and education of medical teams who less frequently treat some pathologies.

To achieve this goal, a standard medical file was designed and translated in Chinese language. This file, subdivided in six sheets, gathers the relevant information at the different steps of the medical procedure (parental agreement for surgery, pre-operative status, operative report, follow-up, complications and re-operation, anaesthetic examination).

In the case of complication where some children had to be operated again between two missions of the French clinicians, local neurosurgeons had to fill in the same medical files.



3.1.2 Neurosurgery telemedicine interface on HOPE

Special care has been given to the design of a user-friendly interface dedicated to neurosurgery on the basis of feedback from healthcare professionals. The goal is to create a natural workflow that will enable easy searching, adding, updating, sharing and discussing clinical exams and their attached images. The Figure 5 describe the various functionalities available on the platform so that both local and remote practitioners can provide a first and second diagnosis to dramatically improve patient care in remote areas and solve together difficult neurosurgery conditions.

Figure 5. Use case scenarios from the perspective of the remote expert practitioner

Figure 6. Screen shot of the HOPE platform

3.2 Application to medical physics

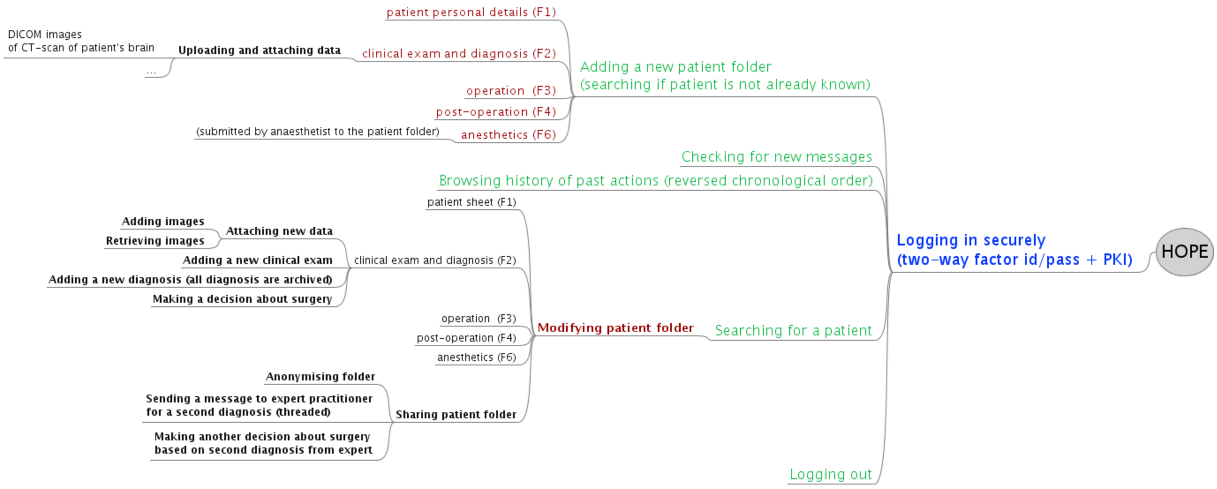


Figure 7. Use cases

3.2.1 Enabling accurate and fast Monte Carlo treatment plannings

With the objective of making Monte Carlo dose computations become standard for radiotherapy quality assurance, planning and plan optimisation, we participate to the development of a Monte Carlo platform dedicated to SPECT, TEP, radiotherapy and brachytherapy simulations Together with 21 other research laboratories which are involved in the international collaboration OpenGATE (http://www.opengatecollaboration..org) : the goal is to develop a Monte Carlo modelling software which is more efficient and long lasting. This software, GATE, with its accuracy and flexibility was published publicly in 2004 and now has a community of over 1000 users worldwide. The limiting issue of GATE right now is its time consuming simulations for modelling realistic scans or treatment planning.

In order to reduce the computing time of the GATE simulations, we first studied the advantage to use the grid to partition the simulations on distributed processors in order to perform accurate treatment planning.

Figure 8 illustrates the computing time in minutes of a Monte Carlo simulation running on a single 3 GHz processor locally and the same simulation splitting in 10, 20, 50 and 100 jobs on multiple processors on the grid. In this case, the lowest computing time is obtained for 50 jobs running in parallel.

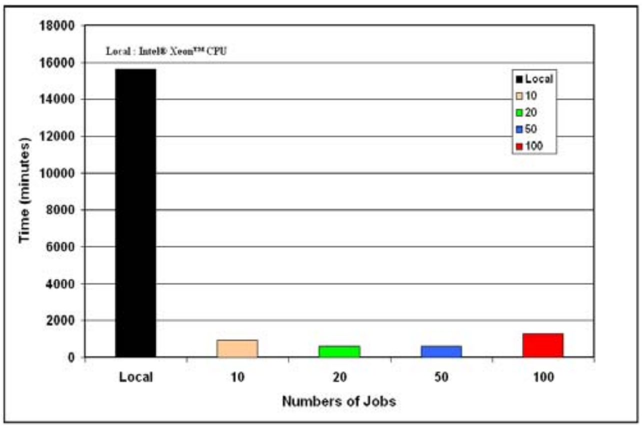


Figure 8. Reduction of the computing time of a dosimetry calculation on the grid

The speed-up coefficient for this submission is 26: in other words, a simulation running in about 10 days on a single 3 GHz processor, now runs in 10 hours using the grid computing facility.

3.2.2 A secured web portal for Monte Carlo dosimetry calculations

As the grid showed a significant added value in reducing the computing time, the next step was to develop a secured web portal to enable medical physicists and physicians to use the grid to compute treatment planning using GATE Monte Carlo simulations. The huge number of commands with many options and rigid sequences, as well as the specific language to describe the jobs (Job Description Language, JDL), makes the Grid approach and usage difficult for inexperienced users. Providing quick, secure and easy to use tools to physicians or medical physicists is critical to convince them to perform treatment planning on the Grid infrastructure.

Figure 9 shows how the HOPE portal is used for dosimetric studies. Its functionalities are the following:

- An authentication module allowing the user to obtain the rights to use the platform and access the grid using grid certificates. In a near future, the medical staff will have the possibility to use their own professional card to be identified on the HOPE portal and to use the grid.
- When the user is logged in, he/she has the possibility to upload or access medical data located on PACS server at hospital. In the case of medical imaging for radiotherapy, AMGA services located at hospital collect metadata as attributes

like the name of the patient, the characteristics of the disease, etc. SSL connections are used for the transfer of data in addition with encryption systems. Authentication using ACLs (Access Control Lists) are used for the access to metadata in the database. The AMGA server provides a replication layer which makes databases locally available to user jobs and replicate the changes between the different participating databases.

- The anonymized medical images are registered on the grid. In order to enable advanced security transfers, GridFTP is used.
- The user can visualize the medical images directly from the patient sheet.
- By visualizing the pathology, the physician can choose what kind of material he wants to use to treat the patient using ionizing particles (linear accelerator, brachytherapy sources, ...)
- Then, the GATE simulations are submitted on the grid. Medical images are copied from SE to CE to be used in the GATE simulations to model the patient body. When the simulations are finished, the user receives a notification.
- The treatment plans can be directly visualized from the HOPE portal and downloaded on the personal computer of the user.

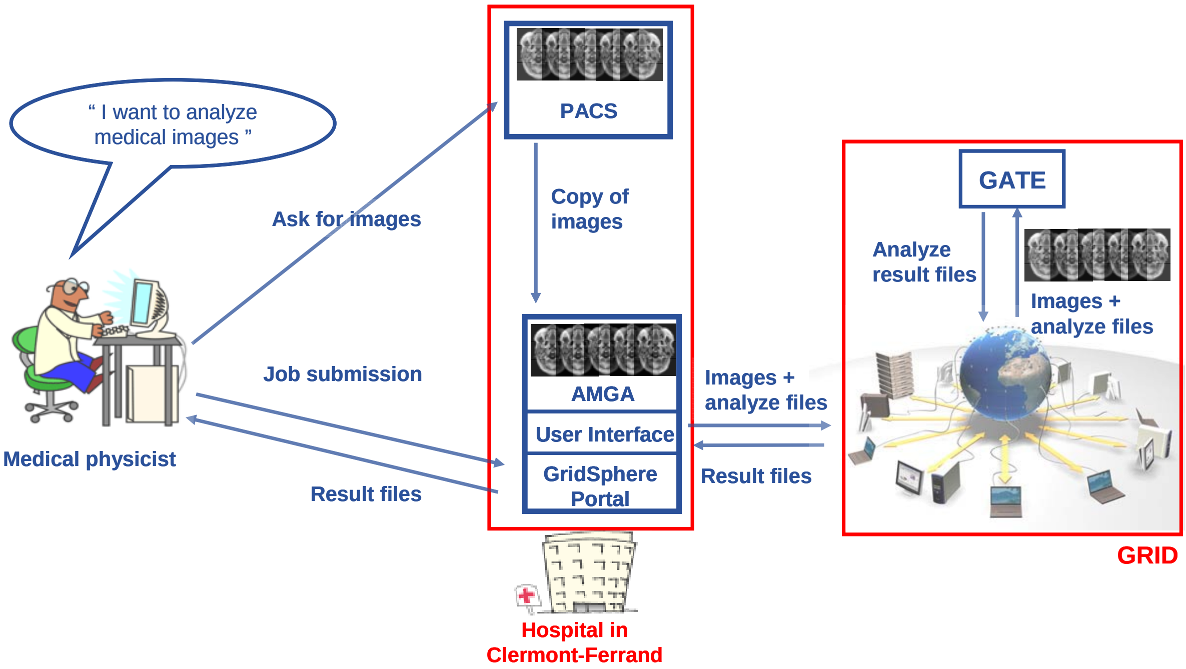


Figure 9. HOPE application in medical physics

The web portal offers to the user a transparent and secure way to create, submit and manage GATE simulations using realistic scans in a Grid environment. The gain in computing time obtained by splitting the simulations. The conviviality of the web portal and the Grid performances could enable, in a near future, the usage of Monte Carlo simulations from clinical centres or hospitals to treat patients in clinical routine for specific radiotherapy treatments.

4-Conclusion and perspectives

We have described in this article a secure platform to handle, query and analyze medical data and images in a grid environment. The platform has been customized for 2 applications in the field of telemedicine and medical physics and is currently deployed on four sites in Asia and Europe.

In the near future, we are going to implement a service allowing the healthcare professionals to authenticate themselves to the platform using their professional electronic card. With this service, the user owning already a grid certificate will just have to enter the card in a card reader device to get access to the platform.

Acknowledgements

The work described in this article was partly supported by grants from the European Commission (BioinfoGRID, EGEE, Embrace), the French Miinistry of Research (AGIR, GWENDIA) and the regional authorities (Conseil Régional d'Auvergne, Conseil Général du Puy-de-Dôme, Conseil Général de l'Allier).

The Enabling Grids for E-sciencE (EGEE) project is co-funded by the European Commission under contract INFSO-RI-031688. The BioinfoGRID project is co-funded by the European Commission under contract INFSO-RI-026808. The EMBRACE project is co-funded by the European Commission under the thematic area "Life sciences, genomics and biotechnology for health", contract number LHSG-CT-2004-512092. The SHARE project is co-funded by the European Commission under contract number FP6-2005-IST-027694.

Auvergrid is a project funded by the Conseil Regional d'Auvergne. The AGIR and GWENDIA projects are supported by the French ministry of Research.

We acknowledge the support of AMGA developers team and particularly Birger Koblitz who has been always available to answer our questions and address the issues we kept raising during the development of HOPE.

References

[1] F. Jacq, F. Bacin, N. Meda, D. Donnarieix, J. Salzemann, V. Vayssière, M. Renaud, F. Traoré, G. Meda, R. Nikiema, N. Jacq, V. Breton, Towards grid-enabled telemedicine in Africa, proceedings of IST-Africa 2006 conference, Pretoria, May 2006

[2] J. Dzenowagis, G.Kernen, WSIS report "Connecting for Health: global vision, local insight", ISBN 92 4 159390

[3] V. Breton, K. Dean and T. Solomonides, editors on behalf of the Healthgrid White Paper collaboration,"The Healthgrid White Paper", Proceedings of Healthgrid conference, IOS Press, Vol 112, 2005

[4] E. Laure, S.M. Fisher, A. Frohner, C. Grandi, P. Kunszt, et al. , Programming the Grid with gLite. Computational Methods in Science and Technology, 12(1):33–45, 2006.

[5] GridSphere: a portal framework for building collaborations, J Novotny, M Russell, O Wehrens, Concurrency and Computation: Practice & Experience, 2004 Volume 16, Issue 5 (2004) 503 - 513 , http://www.gridsphere.org

[6] AMGA metadata catalogue : N. Santos and B. Koblitz, Distributed Metadata with the AMGA Metadata Catalog, in Workshop on Next-Generation Distributed Data ManagementHPDC-15, Paris, France, June 2006

[7] SOAP messaging protocol : http://www.w3.org/TR/soap/

[8] WISDOM production environment

[9] DICOM : http://medical.nema.org/
[10] XML : http://www.w3.org/XML/
[11] Chain of Hope: http://www.chainedelespoir.org
[12] D. Manset et al, "MammoGrid: A Service Oriented Architecture based Medical Grid Application", Lecture notes in Computer Science, 2004 – Springer H. Jin, Y. Pan, N. Xiao, and J. Sun (Eds.): GCC 2004, LNCS 3251, pp. 939–942, 2004. © Springer-Verlag Berlin Heidelberg 2004
[13] J. Montagnat, Vincent Breton, Isabelle E. Magnin "Partitioning medical image databases for content-based queries on a grid". *Methods of Information in Medicine*, vol 44(2), pp 154-160, 2005, Schattauer
[14] Health-e-Child: http://www.health-e-child.org/