

Document downloaded from:

<http://hdl.handle.net/10251/38991>

This paper must be cited as:

Salavert Torres, J.; Segrelles Quilis, JD.; Blanquer Espert, I.; Hernández García, V. (2012). Improving knowledge management through the support of image examination and data annotation using DICOM structured reporting. *Journal of Biomedical Informatics*. 45(6):1066-1074. doi:10.1016/j.jbi.2012.07.004.



The final publication is available at

<http://dx.doi.org/10.1016/j.jbi.2012.07.004>

Copyright Elsevier

Title

Improving Knowledge Management Through the Support of Image Examination and Data Annotation Using DICOM Structured Reporting

Authors

José Salavert Torres, J. Damià Segrelles Quilis, Ignacio Blanquer Espert, Vicente Hernandez García

Affiliation

Members of the Institute for Molecular Imaging Technologies (I3M), Universitat Politècnica de València (UPV)
Camino de Vera S/N, 46022
Valencia (Spain)

Abstract

An important effort has been invested on improving the image diagnosis process in different medical areas using information technologies. The field of medical imaging involves two main data types: medical imaging and reports. Developments based on the DICOM standard have demonstrated to be a convenient and widespread solution among the medical community. The main objective of this work is to design a Web application prototype that will be able to improve diagnosis and follow-on of breast cancer patients. It is based on TRENCADIS middleware, which provides a knowledge-oriented storage model composed by federated repositories of DICOM image studies and DICOM-SR medical reports. The full structure and contents of the diagnosis reports are used as metadata for indexing images. The TRENCADIS infrastructure takes full advantage of Grid technologies by deploying multi-resource grid services that enable multiple views (reports schemes) of the knowledge database. The paper presents a real deployment of such Web application prototype in the Dr. Peset Hospital providing radiologists with a tool to create, store and search diagnostic reports based on breast cancer explorations (mammography, magnetic resonance, ultrasound, pre-surgery biopsy and post-surgery biopsy), improving support for diagnostics decisions. A technical details for use cases (outlining enhanced multi-resource grid services communication and processing steps) and interactions between actors and the deployed prototype are described. As a result, information is more structured, the logic is clearer, network messages have been reduced and, in general, the system is more resistant to failures.

Keywords

Data Grid, DICOM, DICOM-SR, Medical Imaging, Knowledge Database.

Main Text

1. Introduction

Much effort has been invested on improving the medical image examination and data annotation process with the aid of computer technologies. The field of medical imaging involves two main data types: medical imaging and reports. Despite the commonalities between them, their differences lead to two different research lines in the literature: the storage and processing of images; and the integration of diagnostic reports through semantic interoperability.

With respect to the processing and storage of digital images, research lines concentrate on the use of Peer to Peer (P2P) and Grid technologies to allow the federation of distributed storages and computing resources through collaborative environments in the Virtual Organization [1] (VO) framework. A VO constitute a multi-institutional community of users sharing common organization policies and resources.

Picture Archiving and Communication Systems (PACS) and Radiology Information Systems (RIS) organise data according to patient-centric information, as they are oriented to healthcare delivery. This is insufficient for research and training activities, which require content-based searching. In addition, healthcare information systems tend to use security models with flat or user-based access permissions, which cannot be extended to multi-domain structures.

Regarding Data Grid technologies, Storage Resource Broker [2] (SRB) and Lightweight Middleware for Grid Computing [3] (gLite) provide general-purpose tools for storing large amounts of data in distributed environments. On one hand, several projects like Biomedical Informatics Research Network [4] (BIRN), a North American initiative that aims at building a virtual community of shared resources in the field of brain degenerative diseases, are based on SRB. On the other hand, NeuroLOG [5], which is a French project aimed at sharing and processing brain disease images, uses Medical Data Manager [6] (MDM) to manage information. MDM uses components from gLite such as ARDA Metadata Grid Application [44] (AMGA), LCG File Catalog (LFC) or Hydra. Additionally, the NeuGRID project [7] is a European initiative whose goal is to develop public electronic infrastructures needed by the European neuroscience community. NeuGRID is composed by several services that follow the Service Oriented Architecture (SOA) and it also uses gLite components such as glite User Interface (gliteUI), AMGA, Computing Element (CE), Storage Element (SE), Berkeley Database information Index (BDII), Disk Pool Manager (DPM) and Working Node (WN). Other projects have developed their own components based on Grid technologies. Cancer Biomedical Informatics Grid [8] (CaBiG) is creating a network that will connect the entire cancer community. This project has developed a Grid layer named CaGRID [9].

With respect to the second research branch (semantic interoperability for diagnostic reports integration), imaging repositories have strongly evolved, providing tools for sharing and processing studies of images and its associated metadata, and allowing data-mining techniques that improve diagnosis and therapy. In this sense, evidence-based medicine is a case-based methodology that relies on high quality and organized medical knowledge.

Data annotation is often performed by centrally storing metadata (MDM uses centralized AMGA databases). Some projects, like *Biomedical Informatics Research Network* project (BIRN), define their own lexicon in order to manage complex data representations, but it is not oriented to represent clinical reports. Moreover, there are also collaborative data collection projects that define diagnostic report templates like Health-e-Child [10] which focuses on creating a common database of pediatric disorders with the support of many medical centers.

Other approaches consider structured reports to enable context-based retrieval are Mammogrid [11] or NeuroBase [12][13]. In those approaches, metadata required for context-based searching is also located on central repositories that store the annotations extracted from the medical reports.

In both research branches, developments based on the *Digital Imaging and Communications in Medicine standard* [14] (DICOM) have demonstrated to be a convenient and widespread solution among the medical community. In fact, the majority of hospitals and vendors choose the DICOM standard to store and exchange digital images (CT, MRI, X-Ray ...). In the projects presented above, the DICOM connectivity is commonly provided through the development of DICOM plug-ins [15][16] to Grid standard protocols such as the Storage Resource Manager [17] (SRM) or the Globus Toolkit [18].

In addition, there is an extension of DICOM, namely DICOM Structured Reporting [19] (DICOM-SR), which defines the syntax and semantics of structured and standardised diagnostic reports. Notwithstanding DICOM-SR is not as widespread as DICOM for digital images, it is a quite convenient solution. DICOM-SR reports can be treated as general DICOM objects, so they can be managed by existing DICOM services and shared over the network along with DICOM studies of images.

Structured reports have many advantages: they follow a tree structure, minimize input typing, require less semantic effort and improve diagnosis efficiency. Structured reports offer a homogeneous way to structure report data, enhancing the capability of computer-enabled tools to extract knowledge, search and compare reports.

Structured reports are a mean to create and exploit knowledge databases of structured data, with the objective of providing a clinical decision support tool. Content-Based Image Retrieval (CBIR) applications have been revealed as powerful tools that improve diagnosis [20][21][22]. However, searching for similarities on images is not an effective task since very different images could lead to similar diagnosis. By using the CBIR over the information from the structured reports, the process becomes much more effective and clinically-relevant [23] and easy to exploit.

Another key feature of DICOM-SR is the use of a standard syntax, that enables referencing other reports, images, regions of

interest (inside images) and temporal coordinates (waveforms). This capability enables radiologists to accurately associate report findings with concrete image areas. In addition, DICOM-SR has support for unified lexicons like Radiology Lexicon [24] (RadLex), International Classification of Diseases [25] (ICD-10), Systematized Nomenclature of Medicine [26] (SNOMED), etc.... In this sense, storing information using standard semantic codes ends up with internationalization of reports, easy application of data-mining techniques and clearer information exchange between medical departments. Finally, DICOM-SR has defined report templates for various clinical domains. These include breast, chest, mammography, Ob-Gyn ultrasound, catheterization, vascular ultrasound and echocardiography.

Conversely, the major disadvantage of supporting structured reports is the productivity penalty. Filling out an electronic form requires more user interaction than plain report dictation supported by speech recognition systems. Currently, just as there are speech recognition systems that transcribe reports from voice to plain text, there are also tools that transcribe information from plain text to structured report format that would minimize this drawback [27][28][29].

The main purpose of all these applications is to increase the radiologist productivity and capability to store his observations in a structured, homogeneous and semantically standard way, while not distracting his eyes from patient images with complex menus and user interface options [30][31].

In summary, the use of structured reports has advantages and disadvantages over plain text, but the conclusion after examining the pros and cons is that the use of structured reports is desirable for both radiologists and clinicians [32], and fundamental to construct integrated knowledge bases.

As stated earlier, there exist many projects that use Grid technologies to successfully cover the challenge of sharing DICOM images between different user communities, but a very few of them deal with the DICOM-SR standard to enhance metadata annotations. The work presented in this paper combines both research lines. It aims at sharing medical images and its associated reports metadata among geographically distributed research institutions, while maintaining the confidentiality of the patient identity and the privacy of data.

We base this work on a middleware developed by our group namely Towards a Grid Environment to Process and Share DICOM objects [33][34] (TRENCADIS). It provides a knowledge-oriented storage model federating repositories of DICOM image studies and DICOM-SR medical reports. This middleware can use gLite [3] based compliant storage resources as back-ends such as European Grid Infrastructure [35] (EGI) or Spanish National Grid Initiative [36] (ES-NGI) SEs. Along with a real deployment of a Web application prototype in the Dr. Peset Hospital. This application runs on top of the TRENCADIS middleware and enables radiologists to input diagnostic reports based on breast cancer mammography explorations.

Section 2 describes the concrete objectives of this article. Section 3 contains a summary of the TRENCADIS technology, outlining the main features that support the distributed storage and the management of DICOM-SR metadata. Section 4 depicts

the infrastructure deployed in the University Hospital Dr. Peset. Section V outlines the use cases of the prototype application, including the workflows between actors and the system, and the internal interaction among the middleware grid services. Finally, the management of structured report, this article presents conclusions are exposed in Section VI.

2. Objectives

The main objective of this work is the design and implementation of a Web application prototype and the necessary TRENCADIS grid services to deploy it. The system will be able to federate structured report schemes and to instantiate a use case for the diagnosis and follow-on of breast cancer patients. The images will be based on DICOM format, and the diagnostic reports will follow the DICOM-SR standard, employing standard coding terminologies, procedures and schemes in breast cancer.

Furthermore, the prototype will act as a framework for sharing semantic annotations of breast cancer images. As a result, it will provide a comprehensive view of images and diagnostic reports along the whole infrastructure.

For the attainment of this general objective, the following targets have been defined:

- To develop the necessary infrastructure services to store and index DICOM information in the Grid Services that TRENCADIS middleware offers as resources.
- To deploy these resources among the institutions implied (Polytechnic University of Valencia (UPV) and University Hospital Dr. Peset, while leaving the door opened for the incorporation of new medical institutions.
- To identify the most important use cases that could happen in the proposed prototype and to define the flow of information among the Grid services involved in each one.
- To code and deploy the DICOM-SR templates associated to breast cancer diagrams and follow-on.
- To improve the expressivity of the Web Services Resources Framework (WSRF) grid services deployed.

3. TRENCADIS Technology

The TRENCADIS technology is the base of the work presented in this paper. TRENCADIS follows a Service-Oriented Architecture (SOA) model in which the management of resources is performed by Grid Services. These services follow the Open Grid Service Architecture [37][38][39] (OGSA) specification. TRENCADIS is structured into several layers, each implementing a different abstraction level (figure 1). Lower layers perform basic functionalities and interact directly with the resources, which can be gLite components [40]. This allows TRENCADIS applications to take advantage of the current infrastructures deployed

using gLite. Upper layers provide a friendly interface that hides internal middleware aspects to application developers. Each sub-layer of TRENCADIS is briefly described in the following sections.

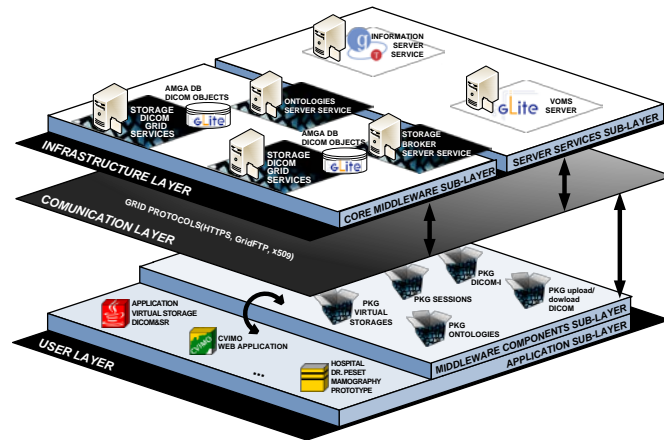


Figure 1. General Diagram of TRENCADIS Architecture.

A. TRENCADIS security architecture

Security is a key issue in any medical data processing environment. Confidential pieces of information must only be available to the patient and the medical team authorized to. Anonymity must be guaranteed while sharing data.

TRENCADIS security mechanisms are based on the Grid Security Infrastructure [41] (GSI), which has been enhanced with several tools and procedures that increase privacy. The whole security model and the infrastructure developed were presented in [42]. Three aspects are considered: Authentication and Service Access, Virtual Organization Membership System [43] (VOMS) and Privacy of data. TRENCADIS has been designed to implement encryption and decryption techniques that ensure that even administrators at local sites cannot access the information although they can read the data files in the storage elements.

B. Core Middleware sub-Layer (Infrastructure Layer)

The infrastructure layer provides the basic resources of the environment, which are offered as standard interfaces defined in Web Services Definition Language (WSDL). Resources use protocols such as Simple Object Access Protocol (SOAP), LFC, AMGA and data formats like eXtensible Markup Language (XML) Schemes. It provides upper layers with a unique interface to all resources of the same type.

The objective of this sub-layer is to implement a wrap-up to communicate with the basic gLite resources of the medical infrastructure, enabling the storage of large amounts of data in distributed environments. The Core Middleware sub-Layer supports also other back-ends, such as Oracle databases or DICOM file systems, providing a uniform interface to the above layer above.

This paper presents the integration of gLite-based components for data storage, providing access to existing gLite infrastructures (EGI, ES-NGI). Three components are used to increase the availability of resources:

- SE [43] as backend for storing DICOM objects. Each medical center can have one or several SEs to host the data to be shared.
- LFC [43] to provide replica management and improve the availability of data and throughput when accessing from multiple simultaneous clients. A single LFC service is needed.
- AMGA [44] to share the metadata extracted from DICOM-SR. Its scalability and performance is comparable to direct database access [44].

These resources are made available through *Grid Services*, which have been implemented following the WSRF [46] specification. Data formats are specified using XML Schemes.

Grid Services can manage multiple states at runtime. Each state is stored into a *service resource*, which publishes its WSRF *resource properties* in the Index Information Service. Resource creation and destruction processes have been implemented in the *factory service* associated with each Grid Service. By deploying multi-resource grid services, middleware expressivity is increased, improving information retrieval through structures the information published in the IIS, and increasing the semantic scalability of the system (supporting easily the addition of new kinds of reports).

The Grid Services developed in this layer are the following:

1) *Ontologies Server Grid Service*

The Ontologies Server hosts and publishes the ontologies available in the system. In TRENCADIS, ontologies are templates that organize metadata for diagnostics. Each ontology contains a DICOM-SR content tree, as DICOM-SR semantic relations are mainly organized by means of tree-like structures. Through this schema, the whole report structure and contents can be used for indexing and searching DICOM images. TRENCADIS uses XML to specify ontologies, since it is expressive enough for the definition of terms and rules at the data level of DICOM-SR [45]. Templates contain attributes, restrictions and relations among fields, following the DICOM-SR standard rules. These fields are filled in by authorized users and uploaded (using the Storage DICOM service) as new reports in the system or used to construct metadata searches (through the Storage Broker service).

2) *Storage DICOM Grid Service*

The Storage DICOM Grid Service offers all the interfaces needed for sharing DICOM-SR and DICOM images. The Storage DICOM interacts through the gLiteUI [43] and the local AMGA service. A central LFC catalogue and a local SE to keep files

and references to the studies of DICOM image files are used as back-ends. There should be at least one instance of this service at each medical centre. New instances can be dynamically added as new centres join the infrastructure.

A multi-resource service publishes in the IIS an entry for each ontology type. In this way, information in the IIS is structured properly and queries are simplified. Moreover, adding new ontologies can be performed on the fly, as each Storage DICOM factory receives periodically the ontologies list creating new resources automatically.

Each Storage DICOM service resource creates directory trees in its associated AMGA repository, matching the tree structures of the ontology definitions kept in the Ontologies Server.

3) *Storage Broker Grid Service*

This Grid Service allows metadata searches over the DICOM-SR contents distributed among the system. It does not keep a full indexation of the content of all the DICOM-SR objects stored in each Storage DICOM, but just the minimum information to ensure scalability and a reasonable speed. Moreover, the architecture is able to handle multiple servers and distributed queries.

The search procedure is based on a mediator approach which consists on a component in the client side that transforms every query submitted by the user into several queries, one for each data repository. The mediator logic is implemented in the Storage Broker.

The Storage Broker is also a multi-resource grid service. Each resource manages the searching of an ontology type, asking the IIS about the Storage DICOM resources for this ontology type.

C. Server Services sub-Layer (Infrastructure Layer)

This sub-layer is composed of services needed by components of other sub-layers:

1) *Index Information Service (IIS)*

Core Middleware Services publish its *resource properties* in a central IIS. This information is commonly needed for the interaction with other services and client applications (e.g URIs). The IIS used in this case is part of the Monitoring and Discovery System 4 (MDS4) of GT4.

2) *VOMS Server:*

Authorisation in TRENCADIS is based on VOs. Users belong to one or more thematic VOs, and each thematic VO is authorized to search on one or more ontologies. Studies matching a specific ontology become accessible to the members of the associated VO.

VOs are managed through VOMS. The VOMS Service is deployed in the Server Service sub-layer and manages user memberships, groups and roles in one or more VOs. It is accessed by all the components that require authorization.

D. Communication Layer

The Communication layer defines the standards and protocols used in service communication. They depend on the backends (LFC or gridFTP in this case). GridFTP is used for transferring large amounts of data, SOAP on top of HTTPS is used for Grid service interaction and Secure Socket Layer (SSL) protocol is used to secure the transmissions with AMGA backends.

E. Middleware Components sub-Layer

This sub-layer contains high-level components distributed in different packages that interact with Core Middleware and Server Services sub-layers Grid Services. It provides applications with an object-oriented interface to access TRENCADIS functionalities, hiding middleware internals to the developer. This sub-layer implements five packages:

- *SESSIONS Package*, which implements the classes to open and close Grid sessions, needed to access the infrastructure.
- *DICOM-I / DICOM-SR Packages*, implementing the classes that manage, either individually or in groups, the DICOM medical images and structured reports.
- *DICOM Virtual Storage Package* deals with the storage and accesses to the DICOM objects in the Grid environment. All these DICOM images and structured reports are managed as a single virtual storage through DICOM repositories, regardless of its real location, offering a single access point to all the DICOM objects matching a given ontology. The Virtual Storage only offers searching criteria for the fields that are included in the ontology.
- *Upload /Download DICOM Packages* provide the classes for uploading DICOM Objects created using DICOM-I and DICOM-SR packages.
- *Ontologies Package* provides the classes for managing ontologies.

4. Infrastructure and Deployment

These components have been deployed in a medical scenario, implementing an infrastructure for the annotation and sharing of breast cancer medical images and reports.

The deployment involves two centres: UPV and University Hospital Dr. Peset. The TRENCADIS Grid services have been distributed as follows (figure 2):

- In UPV the VOMS Server and the Index Information Service from the Server Services Layer have been deployed. The Ontologies Server and Storage Broker services (Core Middleware sub-layer) have been deployed too.

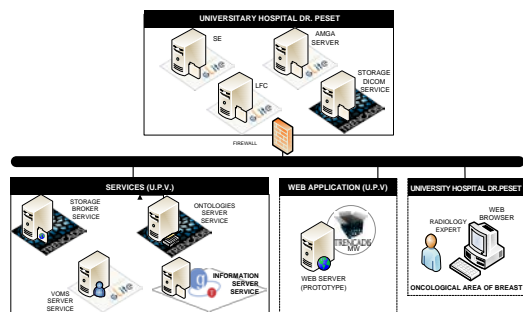


Figure 2. Scheme of the Infrastructure Deployment.

- In the University Hospital Dr. Peset, the Storage DICOM and its associated AMGA Server (Core Middleware sub-layer) have been deployed. A backend composed of a central LFC catalogue and a local SE has also been deployed in order to store studies of DICOM image files.
- The Web application prototype that manages the complete flow of breast cancer examinations is hosted in the UPV, and it is accessed by the Mammography Department of University Hospital Dr. Peset. This application uses TRENCADIS middleware components to access the Grid services deployed in the infrastructure.

5. Use Cases

In this section, the use cases of our Web application prototype are outlined. The workflow subsection describes the interaction of different actors with the system when performing the most common management tasks. The interaction subsection describes the communication messages between the different grid services that are transparent to the users.

A. Defining report templates in the VO

1) Workflow

The purpose of our prototype application is to manage diagnostic episodes of breast cancer patients. Medical episodes are composed of many medical studies including the associated DICOM-SR documents. Each diagnostic techniques (mammography

exploration, MRIs, scans, biopsies...), have a different XML report template. Each template defines an ontology that we have coded and deployed in the Ontologies Server.

In an ideal scenario, lead radiologists of a certain VO decide by consensus all the mandatory elements of a report structure. Then the structure is added to the system (Figure 3), and all the report elements are coded into a XML format designed by the TRENCADIS developers. This process is a programmatic and iterative task among radiologists and developers. This process is a programmatic and iterative task among radiologists and developers. Usually, hospitals may need to specify their own additional fields. In this sense, ontologies represent the common parts of the report shared by all hospitals. Moreover, ontologies provide distributed data management among all centres involved, as the common fields defined in them form a homogeneous knowledge base.

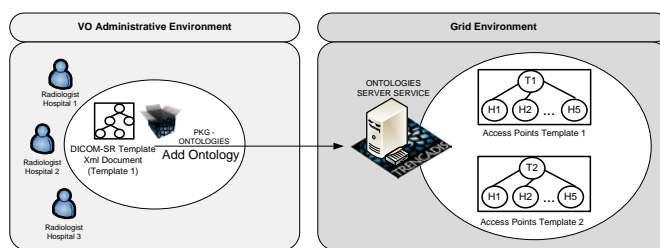


Figure 3. Workflow to define and send DICOM-SR templates to the Ontologies Server.

2) Interaction

The list of ontologies is stored in a centralised service called Ontologies Server. The storage and indexing services (Storage DICOM and Storage Broker) need a subscription to this service to query the distributed repositories. This subscription is performed by publishing the resource information in the IIS. Figure 4 describes this process. After the initial subscription, the Ontologies Server periodically queries the IIS (Figure 4, step 2) and sends the updated ontologies list to the service factories (Figure 4, step 4) by calling its UpdateResources() method (Figure 4, step 5 and 10).

It is important to outline that the functionality of the Globus Toolkit factories [47] has been extended. In Globus Toolkit 4.2 (GT4.2), a very simple factory component is included as a fundamental part of multi-resource grid services. The factory is defined using WSDL [48]. However, it is stateless by nature and its only purpose is to initialize and destroy resources using two standard methods. Moreover, the resources initialisation method cannot receive parameters, because GT4.2 does not expect resource properties to be initialised at creation time (user must set the data in further calls to the service resources).

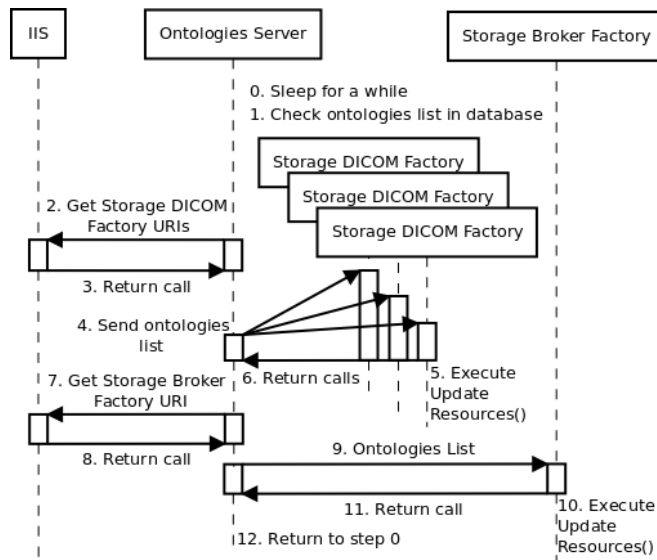


Figure 4. Interaction after the initial subscription of the Ontologies Server.

This simplistic factory has many disadvantages. First of all, the design becomes more complex as the client (the *Ontologies Server* in our case) has to establish multiple connections with the factory and the IIS in order to create new resources. A client of this type must check if the resources are still published, call the factory creation method and send the initial values of the resource properties (all of these once for each ontology in the system).

Secondly, the process becomes slow as these connections are performed between geographically distributed nodes. Moreover, calls must go through the whole GT4.2 stack.

Finally, it is hard to protect the system against failures as the information in the IIS is not automatically updated (Usually the information is kept during a few minutes to avoid network overheads). In addition, if a grid service shuts down and restarts, the GT4.2 IIS will contain temporal replicas of resource entries. These issues increase the complexity of the logic to avoid errors in the client side and make the system more fault-tolerant.

Conversely, we extended the factories by adding the *UpdateResources()* method and allowing resource properties to be initialised at creation time. Figure 6 illustrates the relationship between the Ontologies Server and the factories of the Storage DICOM and the Storage Broker services.

The method *UpdateResources()* is called by the Ontologies Server providing the ontologies list as a parameter (Figure 5, step 1). This list is stored internally in the factory and checked for any update on each call (Figure 5, step 4 and 7). Then, this new method creates or destroys the grid service resources whose ontologies have been added or deleted from the list (Figure 5, step 5 and 6).

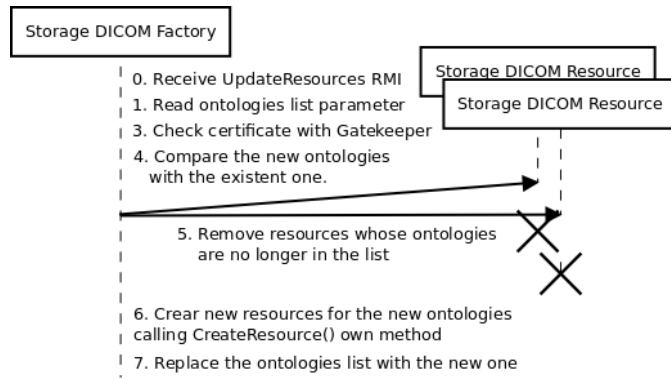


Figure 5. Interaction of UpdateResources method.

This approach is clearly better as it only needs a single network connection between the Ontologies Server and the factories. Moreover, the Ontologies Server is the unique entity involved in the resource creation process, avoiding concurrence. In addition, the calls to the standard methods that create new resources are executed internally into the factory services, which is faster and less error prone. The error management is also easier as the factory keeps the list of ontologies and does not need to connect to the IIS to perform checks. Moreover, the Ontologies Server sends the ontologies list every certain time so, after a connection failure, the factories will still receive the ontologies list in a posterior call and the system will reach a correct and stable state.

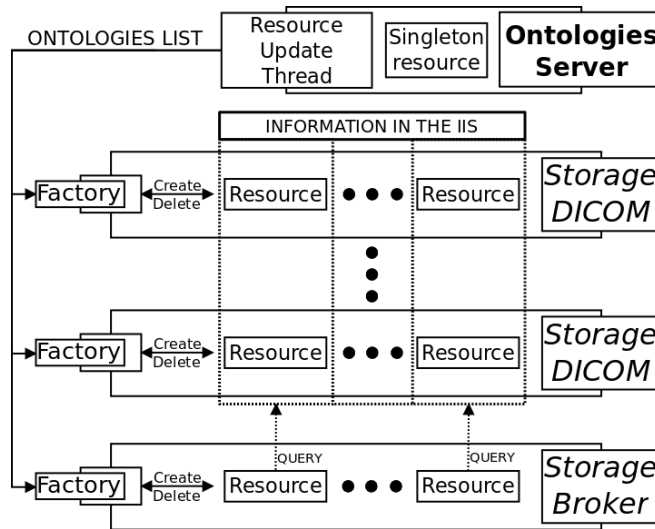


Figure 6. TRENCADIS Services Interaction.

The process described here involves actions in several *Core Middleware* and *gLite Core* components, but it is totally transparent to the user, who only has to concern about inserting or deleting XML templates in the *Ontologies Server*.

B. Inserting DICOM-SR data from Medical Centres

1) Workflow

Radiologists access the web application portal in order to write the diagnostic reports of a patient episode (Left-Figure 7, step 1). Depending on the report, the application selects and reads the associated template (written by consensus) and dynamically creates a form to input report data (Left-Figure 7, step 2).

To send a report to the DICOM Storage Service, the physician must validate and mark the report as complete. While a report is not marked as complete, it can be updated or even extended with new data. After marking a report as complete, it is sent to a DICOM Storage Service, becoming in a read-only report without possibility of change. If it is necessary to amend a previously sent report, the user must generate a new report, in order to keep a history of complete reports and subsequent updates.

The Web application prototype produces the report information in XML format. XML offers more advantages than DICOM-SR, as there are more XML tools and parsers available and files are human readable. However, DICOM-SR provides compatibility with DICOM-compliant storage and communication protocols.

We found that in order to validate a DICOM-SR structure tree we needed to implement our own mechanisms. This is because common XML validation tools (such as XML schemas) have limitations [49].

The task of adding and deleting reports in the local repository of a medical centre is managed by the Storage DICOM service (Left-Figure 7, step 3). Each repository has an associated instance of the Storage DICOM service, which contacts a local AMGA repository holding the reports metadata (Left-Figure 7, step 4).

It is worth to outline that in order to use the diagnosis information contained in the structured reports stored in the StorageDICOM when delivering care to patients, this information needs to be converted into a "human readable" report and then distributed downstream to Electronic Medical Report (EMR) systems. In this sense, Health Level Seven (HL7) organization has developed the Clinical Document Architecture (CDA) standard, which has been mainly designed to provide an exchange model for clinical documents. A key use case for CDA is making these electronic documents available within an EMR [50]. CDA include a fully DICOM-SR compatible model to enable DICOM-SR documents to exchange information with non-DICOM devices, such as information systems. The DICOM standard committee has an agreement with HL7 renewed at 2008 [51].

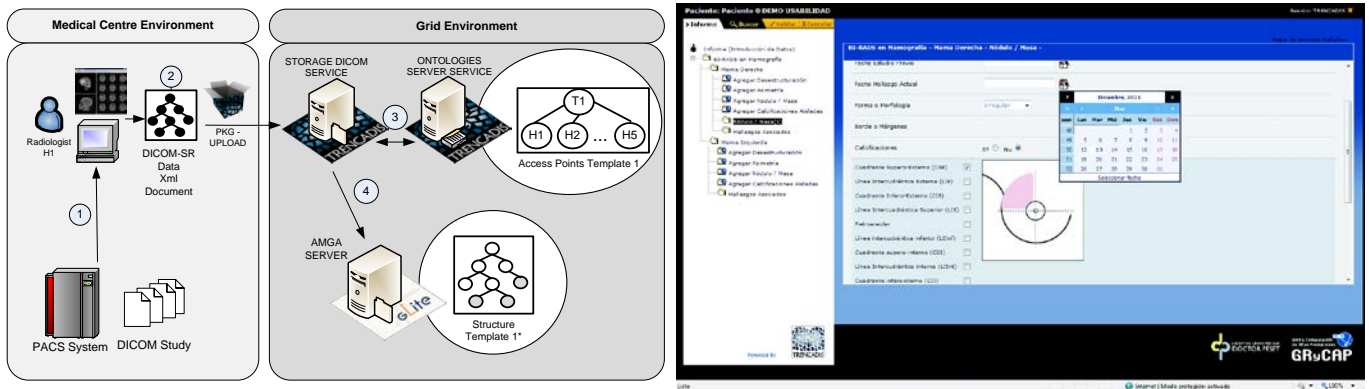


Figure 7. Left describe steps for Adding DICOM-SR in the DICOM Storage service. Right shows a screen capture of the user interface in this use case.

2) Interaction

The DICOM Storage is in charge of inserting, deleting and retrieving medical reports. It is a stateful grid service in which each resource manages the report data of its corresponding ontology type by communicating directly with the distributed data repositories (currently AMGA backends).

When inserting a new report, clients must first launch a query to the IIS, indicating the storage location and the ontology type of the report. The IIS will answer with the URI of the corresponding DICOM Storage resource.

The hospital, service URI, ontology type and AMGA replica URI are the properties that each Storage DICOM resource publishes in the IIS. In this case, performing a correct use of stateful services and resource properties provides an IIS database of the online resources in the system (figure 6). In addition, service calls will have less parameters, as many of them are implicit in the resource properties of the resources managed by the grid services (Figure 8, step 1).

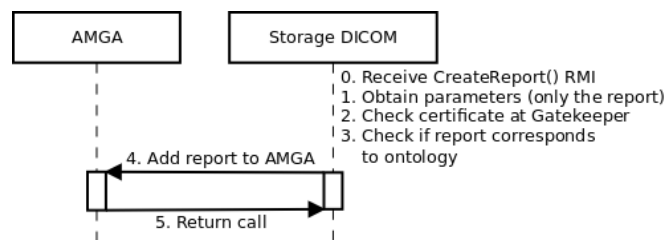


Figure 8. Interaction of Inserting DICOM-SR data from Medical Centres.

Once the URI of the resource is retrieved, the report is sent by calling the *CreateReport()* method of the StorageDICOM service.

The system checks if the report type matches the ontology type managed by the Storage DICOM resource. If the client selects the wrong resource (Figure 8, step 3) the process will be aborted at this point. If the condition is satisfied, the DICOM-SR will be

added to the AMGA backend (Figure 8, step 4). The process of adding the report to the AMGA backend is transactional, so the state of the system will remain unchanged if an error occurs.

C. Searching DICOM-SR Data

1) Workflow

Users can query DICOM-SR content in order to retrieve diagnostics reports through the Web portal (Right-Figure 9). Query forms are constructed from the report templates available in the Ontologies Server (Left-Figure 9, step 1). The Web application generates an AMGA query that is sent to the Storage Broker service (Left-Figure 9, step 2), which communicates with the AMGA backends of the Storage DICOM services.

The Storage Broker implementation uses a mediator approach, which is composed of a component in the client side that transforms every query submitted by the user into several queries, one for each data repository (Left-Figure 9, step 3). These queries are performed concurrently. The mediator waits for the results of all the queries, integrates them and presents the result to the user (Left-Figure 9, step 4).

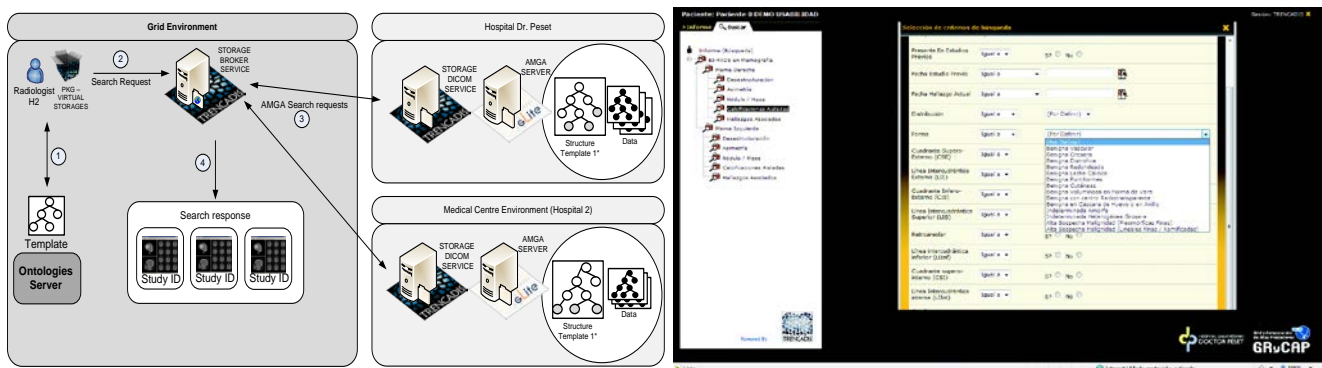


Figure 9. Left describe steps for Distributed searches with the Storage Broker service. Right shows a screen capture of the user interface in this use case.

2) Interaction

TRENCADIS middleware allows distributed metadata queries that can be submitted to the Storage Broker service. This grid service is responsible of converting a single query into many, one for each DICOM Storage resource of a particular ontology structure.

Before calling the service, the client must ask the IIS for the URI of the Storage Broker resource in charge of the ontology type related to the search. After that, the *QueryAMGA()* method of the proper Storage Broker resource is called.

The *QueryAMGA()* method checks how much time has passed since the last query and contacts the IIS if necessary, updating the information about the available Storage DICOM resources (Figure 10, step 3). Then it examines the resource properties of the resources and reads the URIs of the AMGA backends of each StorageDICOM service, creating a different search for each

backend (Figure 10, step 5). Finally, the service reads the search results and presents them to the user.

If a concurrent search fails, the result will contain a notification of the failure along with the successful searches. This means that a user has the illusion of performing a single query to the system, but still receives information about the individual fails of the backends.

In [52] the authors performed a technical validation of the AMGA service, showing that it is able to process 25 structured report queries per second. Additionally, a usability test has been performed with 11 expert radiologists, the response times are similar to other systems employed in clinical practise.

Recently, AMGA added support for federation, allowing parallel metadata searching into a single view and handling partial failures through the user-defined policies [45]. We plan to take advantage of these new features.

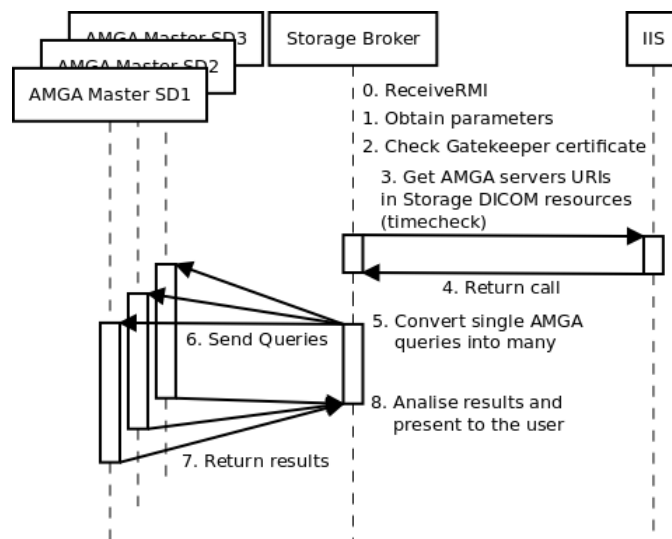


Figure 10. Interaction of Searching DICOM-SR Data

6. Conclusion

The application designed in this work enables sharing DICOM studies (including DICOM images and DICOM-SR reports) in a geographically distributed environment. The infrastructure has been deployed in different centres and uses gLite components for interoperability with existing Grid infrastructures (EGI, ES-NGI). Currently, it is fully operational as a prototype.

Five templates for mammography, magnetic resonance, ultrasound, pre-surgery biopsy and post-surgery biopsy have been defined to code diagnosis and treatment of breast cancer patients. The definition process of these templates has been explained in section V. The process has involved specialists in breast cancer from University Hospital Doctor Peset.

The most important use cases in the diagnosis process have been outlined (defining report templates, inserting report data, and searching report data) and the flow of information among the Grid services has been described.

The need to work with several ontologies forced us to improve some of the Grid services in the infrastructure to support multiple resources efficiently. Thanks to this, queries to the IIS are simpler and separate views can be created easily. Moreover, the logic inside the grid services is clearer, with less network messages and even increased stability.

In addition, we concluded that enabling the definition of custom service resource constructors would be a great feature to include into APIs of upcoming releases of Grid Middlewares. We simulated this feature by calling an internal resource function instead of sending the parameters into a network service call and the performance of the whole process is significantly faster.

An infrastructure has been deployed in two centres (UPV and Hospital Doctor Peset), creating a knowledge database that could be extended by adding new Storage DICOM services. New centres would join the VO through VOMS memberships.

Acknowledgment

The authors wish to thank the financial support received from The Spanish Ministry of Education and Science to develop the project “CodeCloud”, with reference TIN2010-17804.

References

- [1] Ł. Skitał, R. Słota, D. Nikolow, J. Kitowski. “Methodology for Virtual Organisation Design and Management”. EGEE User Forum, CERN, Geneva, Switzerland, March 1-3, 2006.
- [2] Moore, R.W, Wan, M., Rajasekar, A. “Storage resource broker; generic software infrastructure for managing globally distributed data”. Local to Global Data Interoperability - Challenges and Technologies, 2005. ISBN: 0-7803-9228-0
- [3] gLite. Lightweight Middleware for Grid Computing. <http://glite.cern.ch/introduction>, visited on January 2011.
- [4] Jeffrey S. Grethe, Chaitain Baru, Amarnath Gupta et. al. Biomedical Informatics Research Network: Building a national collaboratory to hasten the derivation of new understanding and treatment of disease. Studies in health technology and informatics, Vol. 112, pp. 100-109, 2005.
- [5] Johan Montagnat, Alban Gaignard, Diane Lingrand, Javier Rojas Balderrama, Philippe Collet, Philippe Lahire. NeuroLOG: A community-driven middleware design. HealthGrid 2008.
- [6] Johan Montagnat, Ákos Frohner, Daniel Jouvenot et al. A Secure Grid Medical Data Manager Interfaced to the gLite Middleware. Journal of Grid Computing 6(1):45-59, March 2008.
- [7] University of the West of England, Bristol UK. Neugrid deliverable: D6.1b Distributed Medical Services Provision (Design Strategy). Technical Report, 2009.
- [8] “CaBIG. Cancer Biomedical Informatics Grid”. <https://cabig.nci.nih.gov> . visited on January 2011.
- [9] Infrastructure Overview Page. <https://cabig.nci.nih.gov/workspaces/Architecture/caGrid> . visited on January 2011.
- [10] Jörg Freund, Dorin Comaniciu, Yannis Ioannidis, Peiya Liu, Richard McClatchey, Edwin Morley- Fletcher, Xavier Pennec, Giacomo Pongiglione and Xiang Zhou. Health-e-Child: An Integrated Biomedical Platform for Grid-Based Paediatrics. HealthGrid 2006.
- [11] S R Amendolia, F Estrella, W Hassan, T Hauer, D Manset, R McClatchey, D Rogulin, T Solomonides. “MammoGrid: A Service Oriented Architecture based Medical Grid Application.”. Lecture Notes in Computer Science. ISBN 978-3-540-39090-9. Volume 3251, pp. 939-942 (2004).
- [12] C. Barillot, R. Valabregue, J-P. Matsumot, F. Aubry, H. Benali, Y. Cointepas et al.”NeuroBase : Management of Distributed and Heterogeneous Information Sources in Neuroimaging.”. DiDaMIC-2004, a satellite workshop of Miccai-2004.
- [13] “myGRID project.” <http://www.mygrid.org.uk>, visited on March 2007.
- [14] Digital Imaging and Communications in Medicine (DICOM) Part 10: Media Storage and File Format for Media Interchange. National Electrical Manufacturers Association, 1300 N. 17th Street, Rosslyn, Virginia 22209 USA.
- [15] J. Montagnat, Á. Frohner, D. Jouvenot, C. Pera, P. Kunszt, B. Koblitz, N. Santos, C. Loomis, R. Texier, D. Lingrand, P. Guio, R. Brito Da Rocha, A. Sobreira de Almeida and Z. Farkas, “A Secure Grid Medical Data Manager Interfaced to the gLite Middleware”. Journal of Grid Computing (JGC). ISSN: 1570-7873 (Print) 1572-9814 (Online). Volume 6, Number 1 / March de 2008. Pages 45-59. Springer Netherlands, 2007
- [16] Dagmar Krefting, Julian Bart, Kamen Beronov, Olga Dzhimova, Jürgen Falkner, Michael Hartung, Andreas Hoheisel, Tobias A. Knoch, Thomas Lingner, Yassene Mohammed, Kathrin Peter, Erhard Rahm, Ulrich Sax, Dietmar Sommerfeld, Thomas Steinke, Thomas Tolxdorff, Michal Vossberg, Fred Viezens,

- Anette Weisbecker. "MediGRID: Towards a user friendly secured grid infrastructure". *Future Generation Computer Systems (FGCS)*, Volume 25, Issue 3, March 2009, Pages 326-336
- [17] Arie Shoshani, Alex Sim, Junmin Gu. "Storage Resource Managers: Middleware Components for Grid Storage". MSS 2002.
- [18] S. G. Erberich, J.C. Silverstein, A. Chervenak, R. Schuler, M. D. Nelson, C. Kesselman, "Globus MEDICUS – Federation of DICOM Medical Imaging Devices into Healthcare Grids" – *From Genes to Personalized Healthcare: Grid Solutions for the Life Sciences*. pp. 269-278. Ios Press, Studies in Health Technology and Informatics, 126. ISBN 978-I-58603-738-3.
- [19] D. Clunie. "DICOM Structured Reporting". ISBN 0-9701369-0-0.
- [20] Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, et al. Automated storage and retrieval of thin-section CT images to assist diagnosis: system description and preliminary assessment. *Radiology* 2003 Jul;228(1):265-70.
- [21] Giger ML, Huo Z, Vyborny CJ, Lan L, Bonta IR, Horsch K, et al. Intelligent CAD workstation for breast imaging using similarity to known lesions and multiple visual prompt aids. In: Sonka M, Fitzpatrick JM, editors. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*; 2002. p. 768-73.
- [22] Zheng B, Mello-Thoms C, Wang XH, Abrams GS, Sumkin JH, Chough DM, et al. Interactive computer-aided diagnosis of breast masses: computerized selection of visually similar image sets from a reference library. *Acad Radiol* 2007 Aug;14(8):917-27
- [23] Napel SA, Beaulieu CF, Redriguez C, Cui J, Xu J, Grupta A, et al. Automated Retrieval of CT images of Liver Lesions Base don Image Similarity: Method Preliminary results. *Radiology* 2010 jul; 256(1)
- [24] Sanjoy Kundu, Maxim Itkin, Debra A. Gervais, Venkataramu N. Krishnamurthy, Michael J. Wallace, John F. Cardella, Daniel L. Rubin, and Curtis P. Langlotz. The IR Radlex Project: An Interventional Radiology Lexicon—A Collaborative Project of the Radiological Society of North America and the Society of Interventional Radiology. Volume 20, Issue 7, Supplement Pages S275-S277, July 2009
- [25] International Statistical Classification of Diseases and Related Health Problems 10th Revision Version for 2007. <http://apps.who.int/classifications/apps/icd/icd10online> . Visited on 2011.
- [26] International Health Terminology Standars Development Organisation. SNOMED CT Technical Reference Guide. January 2010 International.
- [27] Sevenster M, van Ommering R, Qian Y. Automatically Correlating Clinical Findings and Body Locations in Radiology Reports Using MedLEE. *J Digit Imaging*. 2012 Apr;25(2):240-9.
- [28] Zimmerman SL, Kim W, Boonn WW. Informatics in radiology: automated structured reporting of imaging findings using the AIM standard and XML. *Radiographics*. 2011 May-Jun;31(3):881-7. Epub 2011 Feb 25.
- [29] Soysal E, Cicekli I, Baykal N. Design and evaluation of an ontology based information extraction system for radiological reports. *Comput Biol Med*. 2010 Nov-Dec;40(11-12):900-11. Epub 2010 Oct 20.
- [30] Gale B, Safriel Y, Lukban A, Kalowitz J, Fleischer J, Gordon D. Radiology report productions times: voice recognition vs. transcription. *Radiol Manage* 2001;23(2):18-22.
- [31] David L. Weiss and Curtis P. Langlotz, "Structured Reporting: Patient care enhancement or productivity nightmare?" in *Radiology*, vol. 249, num. 3, RSNA 2008, pp. 739–747.
- [32] Hayt DB, Alexander S. The pros and cons of implementing PACS and speech recognition systems. *J Digit Imaging* 2001;14(3):149-157.
- [33] I. Blanquer, V. Hernandez, D. Segrelles. TRENCADIS. A Grid Architecture for Creating Virtual Repositories of DICOM Objects in an OGSA-based Ontological Framework. ISSN: 0302-9743, ISBN-10: 3-540-68063-2, ISBN-13: 978-3-540-68063-5. Issue: Volume 4345, 2006. Pages 183-194.
- [34] I. Blanquer, V. Hernández, J.E. Meseguer, D. Segrelles. "Content-Based Organisation of Virtual Repositories of DICOM Objects". *Future Generations Computer Systems*. FGCS Journal. ISSN-0167-739X. DOI:10.1016/j.future.2008.12.004. 2009.
- [35] "European Grid Infrastructure (EGI). Towards a sustainable grid infrastructure". <http://www.egi.eu>

- [36] "National Grid Initiative home page". www.es-ngi.eu
- [37] "Open Grid Services Architecture (OGSA)", <http://www.globus.org/ogsa>. Visited on May 2007.
- [38] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the Grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 15(3), 2001.
- [39] I. Foster, C. Kesselman, J. Nick, S. Tuecke. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.
- [40] I. Blanquer, V. Hernández, J. Salavert, D. Segrelles. "Integrating TRENCADIS Components in gLite to Share DICOM Medical Images and Structured Reports". *Healthgrid applications and core Technologies*. pp. 64-75. Ios Press, *Studies in Health Technology and Informatics*, 159. ISBN 978-I-60750-582-2.
- [41] R. Oppliger. "Security technologies for the World Wide Web". Second edition. Computer Security Series. Artech House. ISBN: 1580533485.
- [42] I. Blanquer, V. Hernández, D. Segrelles, E. Torres. "Enhancing Privacy and Authorization Control Scalability in the Grid through Ontologies". *TITB*. ISSN-1089-7771. Vol 12. Iss. 1. pp. 16-24. 2009.
- [43] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, Á. Frohner, K. L@rentey, F. Spataro, From gridmap-file to VOMS: Managing authorization in a Grid environment, *Future Generation Computer Systems (FGCS)* 21 (4) (2005) 549_558.
- [44] B. Koblitz, N. Santos, V. Pose. "The AMGA Metadata Service". *JGC*. ISSN: 1570-7873 (Print) 1572-9814 (Online). Volume 6, Number 1 / March de 2008. Pages 61-76. Springer Netherlands, 2007.
- [45] "Globus 4 toolkit". <http://www.globus.org/toolkit>. Visited on 2011
- [46] "The WS-Resource Framework". www.globus.org/wsrp. Visited on 2011
- [47] Borja Sotomayor. The Globus Toolkit 4 Programmers Tutorial. <http://gdp.globus.org/gt4-tutorial>. Visited on 2011.
- [48] "Web Service Description language (WSDL)". <http://www.w3.org/TR/wsdl>. Visited on 2011.
- [49] K. P. Lee, PhD and Jingkun Hu, MS. "XML Schema Representation of DICOM Structured Reporting". *Journal of American Medical Informatics Association*. Volume 10, number 2, Mar-Apr 2003. Pages 213-223. doi: 10.1197/jamia.M1042.
- [50] Bilykh, I. Jens H. Jahnke, Glen McCallum, Morgan Price. Using the Clinical Document Architecture as Open Data Exchange Format for Interfacing EMRs with Clinical Decision Support Systems. 19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006. 855-860. ISSN : 1063-7125. Print ISBN: 0-7695-2517-1 .DOI: 10.1109/CBMS.2006.166.
- [51] Health Level Seven Imaging Integration. <http://www.hl7.org/Special/committees/imagemgt/overview.cfm>. Retrieved 14 April, 2012.
- [52] I. Blanquer, V. Hernández, J. Salavert, D. Segrelles. "Using Grid-Enabled Distributed Metadata Database to Index DICOM-SR". *Healthgrid Research, Innovation, and Business Case*. pp. 117-126. Ios Press, *Studies in Health Technology and Informatics*, 147. ISBN 978-I-60750-027-8.