

## **Infraestructura basada en Globus Toolkit para dar soporte a repositorios distribuidos de imágenes médicas**

*Juan-Carlos Guillermo<sup>1</sup>, Ronald Gualán<sup>1</sup>, Lizandro D. Solano-Quinde<sup>1</sup>, Diana Collaguazo-Montalvan<sup>2</sup>, Whasintong Ramírez-Montalvan<sup>2</sup>, Alexandra La Cruz<sup>1,3</sup>*

<sup>1</sup> Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010150.

<sup>2</sup> Facultad de Ingeniería, Universidad Politécnica Salesiana Isabel La Católica, Quito DC, Ecuador, 170143.

<sup>3</sup> Investigador Prometeo, Universidad de Cuenca Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010150.

Autores para correspondencia: [juan.guillermo@ucuenca.edu.ec](mailto:juan.guillermo@ucuenca.edu.ec), [ronald.gualan@ucuenca.edu.ec](mailto:ronald.gualan@ucuenca.edu.ec), [lizandro.solano@ucuenca.edu.ec](mailto:lizandro.solano@ucuenca.edu.ec), [alexandra.lacruz@ucuenca.edu.ec](mailto:alexandra.lacruz@ucuenca.edu.ec), [wramirez@ups.edu.ec](mailto:wramirez@ups.edu.ec), [dcollaguazo@ups.edu.ec](mailto:dcollaguazo@ups.edu.ec).

Fecha de recepción: 28 de septiembre 2015 - Fecha de aceptación: 12 de octubre 2015

### **RESUMEN**

La computación grid es una tecnología innovadora que permite utilizar de forma coordinada todo tipo de recursos (cómputo, almacenamiento y aplicaciones), actualmente se ha venido utilizando para el diseño de sistemas de teleradiología. La computación grid permite la integración y uso colaborativo de computadores de alto rendimiento, redes y base de datos, adecuada para el desarrollo de un sistema de teleradiología. En este trabajo se describe el diseño de una infraestructura basada en el uso de la computación grid como base para un sistema de teleradiología. Dicha infraestructura permite el acceso e intercambio de imágenes médicas almacenadas en los distintos repositorios ubicados físicamente en hospitales y centros de diagnósticos por imágenes. Globus Toolkit (v6.0) es usado como programa intermedio de comunicación en dicha infraestructura grid. Las imágenes son almacenadas en repositorios locales administrados y manipulados por servidores PACS (Picture Archiving and Communication System), sistemas utilizados generalmente en las estaciones de trabajo de los centros de adquisición de imágenes médicas. La mayor contribución de este trabajo es el diseño de una infraestructura grid que servirá para propósitos de búsqueda en cada uno de los nodos de la red ya sea por contenido semántico o visual, poniendo mayor énfasis en la integración de la seguridad de los datos, autenticación de usuarios y sobre todo manteniendo siempre la confidencialidad del paciente.

Palabras clave: Imagenología médica, sistemas distribuidos, computación grid, teleradiología.

### **ABSTRACT**

Grid computing is a recent and innovative technology; recently it has been used for designing teleradiology systems. Grid computing enables the integration and collaborative use of high performance computers, networks and databases, suitable for developing a teleradiology system. In this work we present a solution based on grid computing for a teleradiology system. This infrastructure allows accessing and exchanging of medical images stored physically located in different repositories like; hospitals and diagnostic imaging centers. Globus Toolkit (v6.0) is used as an intermediate communication program in this grid infrastructure. Images are stored in local repositories managed and handled by servers PACS (Picture Archiving and Communication System), usually used in workstations of the medical imaging centers. The main contribution in this paper is the design of a grid infrastructure for searching purposes either semantic or visual content, with greater

emphasis on the integration of data security, user authentication while maintaining patient confidentiality.

Keywords: Medical imaging, distributed systems, grid computing, teleradiology.

## 1. INTRODUCCIÓN

Actualmente la Universidad de Cuenca ha desarrollado una aplicación web de repositorios de imágenes médicas, que permite buscar y recuperar información contenida en la imagen médica, ya sea de forma visual o semántica. Las consultas en este sistema se realizan utilizando Sistemas Gestores de Bases de Datos Federadas (SGBDF). Si bien podemos utilizar consultas federadas para manejar diferentes bases de datos de manera transparente al usuario. Se propone el uso de sistemas distribuidos (Grid), a fin de aprovechar las ventajas que nos ofrece la computación grid. Watson (2003) describe la manera de combinar consultas federadas en bases de datos y la computación grid. En este artículo se tratan aspectos fundamentales que hacen posible la implementación de una infraestructura de computación grid, la misma que permite la integración de grandes repositorios de imágenes médicas y que son utilizados para realizar búsquedas en base a su contenido visual y semántico (Milano *et al.*, 2009). Por otro lado, las imágenes médicas son las que generan mayor impacto en el almacenamiento y procesamiento, y es por esto que se ha optado por trabajar con la computación grid (Noreña, 2013).

La principal herramienta utilizada como capa intermedia de comunicación en este trabajo es Globus Toolkit (v6.0), el cual hace posible la implementación de la infraestructura grid (Foster & Kesselman, 1999), que presentamos en este trabajo, poniendo mayor énfasis en la capa de seguridad GSI (Grid Security Infrastructure) (Filipovic & Straub, 2006), la capa de administración de datos GridFTP (Allcock *et al.*, 2003) y administración de credenciales que se encuentran almacenadas en el servidor MyProxy (Novotny *et al.*, 2001). Teniendo en cuenta una visión global acerca del tratamiento que se les da a las imágenes médicas como son almacenamiento, consulta y recuperación a través del servicio PACS (Picture Archiving and Communication System) basado en la web (Warnock *et al.*, 2007).

El artículo se distribuye de la siguiente manera, en la Sección 2 se presenta un resumen de los trabajos relacionados a este tema. En las Secciones 3 y 4 se describen los componentes y su implementación para la creación de la infraestructura grid basada en Globus Toolkit que brinda soporte a repositorios de imágenes médicas distribuidos. Luego en la Sección 5 se describe la funcionalidad del Servidor MyProxy, quien almacena la información de los usuarios de la infraestructura grid. Finalmente en las Secciones 6 y 7 se presentan los resultados obtenidos de las pruebas preliminares que se hicieron sobre la arquitectura y las conclusiones y trabajos futuros en base a la experiencia obtenida durante el desarrollo del proyecto.

## 2. TRABAJOS RELACIONADOS

En la actualidad existen proyectos y desarrollos en el área de la computación grid, enfocados e integrados en diversas áreas biomédicas como: MammoGrid (Amendolia *et al.*, 2004), MediGrid (Krefting *et al.*, 2009), VirtualPACS (Sharma *et al.*, 2009), caGrid (Saltz *et al.*, 2006), MEDICUS (Erberich *et al.*, 2007), mantisGRID (Ruiz *et al.*, 2011), etc. Gualán *et al.* (2015) describen las características, comparativas, ventajas y desventajas más importantes de las plataformas grid de MEDICUS, mantisGRID y VirtualPACS, a su vez plantea una propuesta de la arquitectura para una plataforma grid DICOM, que sirve para realizar consultas de imágenes basadas en estudios DICOM, anotaciones semánticas y búsquedas basadas en contenido (CBIR-Context Based Image Retrieval). Generalmente las arquitecturas grid como las que hemos encontrado en la literatura presentan la funcionalidad de realizar únicamente consultas sobre estudios DICOM, sin embargo la contribución de esta nueva arquitectura que se propone en este trabajo, es extender esa única funcionalidad a otros

tipos de consultas como consultas sobre estudios DICOM por contenido visual (CBIR) y semántico, a través de consultas federadas.

### 3. COMPONENTES DE HARDWARE Y SOFTWARE

Los principales componentes de la infraestructura grid que se presenta en este trabajo está conformada por tres capas principales: Presentación, Seguridad y Mediación. Para propósitos de prueba se implementaron tres nodos dentro de la infraestructura grid, mismos que representan a los hospitales. Se toma un hospital como nodo central (en donde están instalados la mayoría de los servicios que luego son utilizados a través de un Web Service (Erl, 2004), y los otros dos nodos actúan como nodos secundarios. Se utilizaron recursos y herramientas de software tales como: Globus Toolkit, Aplicación Web, Servidores PACS dcm4chee y Servidor Fuseki (Jena & Fuseki, 2004); tres computadoras (Linux/Centos), sobre las cuales se instalaron los componentes de Globus. En la Figura 1 se describe el esquema general de los principales componentes de la infraestructura grid desarrollada en este trabajo. Donde se define un nodo central y los nodos secundarios. El nodo central contiene todos los servicios que ofrece la arquitectura grid y los nodos secundarios actúan como consumidores de esos servicios. En la Figura 2 se presenta la descripción de los componentes y servicios del nodo central de la infraestructura médica grid, organizada como una arquitectura de tres capas: Presentación, Seguridad y Mediación.

#### 3.1. *Capa de presentación*

Proporciona los mecanismos necesarios para soportar un conjunto de servicios de consultas. El primer mecanismo es una aplicación web que permitirá procesar consultas DICOM, a través de la librería dcm4chee. El segundo mecanismo cuenta con dos portales web que permitirá al usuario realizar consultas CBIR y consultas basadas en semántica respectivamente a través de una interfaz gráfica amigable. Cada portal web tiene su propio motor que realiza la lógica de negocio y que es necesario para realizar consultas en los repositorios PACS y RDF a través del Web Service (ver Figura 2).

#### 3.2. *Capa de mediación*

Contiene los recursos distribuidos (servidores PACS, RDF Stores y servidor Fuseki) y los protocolos correspondientes que permiten la comunicación entre los distintos nodos de la plataforma grid. Utilizamos un Web Services como intermediario para las comunicaciones distribuidas entre los distintos hospitales de la infraestructura grid (ver Figura 2).

#### 3.3. *Capa de seguridad*

La capa de seguridad es la base fundamental de la arquitectura que se presenta en este trabajo, permitiendo la autenticación y la autorización de los usuarios (médicos, especialistas, investigadores, pacientes) en la plataforma grid. Como herramienta de software se utiliza el GSI de Globus como lo hace MEDICUS o mantisGRID, debido a que es una herramienta ampliamente utilizada, catalogada por muchos como el estándar de facto para la construcción de plataformas grid (Foster & Kesselman, 1999). Esta capa de seguridad funciona de forma transparente para los usuarios dentro de la plataforma grid. A continuación se describen los servicios utilizados para la implementación de la seguridad GSI de Globus. Globus Toolkit es un *Middleware* de código abierto, desarrollado por Global Alliance, que permite el desarrollo de arquitecturas grid, los servicios se ejecutan en plataformas Unix incluyendo Linux. Esta herramienta incluye librerías que permiten a los usuarios ejecutar trabajos, transferir y replicar archivos, la motorización o administración de los recursos, publicar información de la infraestructura grid, etc. Todos estos servicios comparten un elemento común de seguridad llamada GSI, que habilita a los usuarios el uso de los recursos dentro de la infraestructura grid a través de certificados de usuario (Foster & Kesselman, 1999).

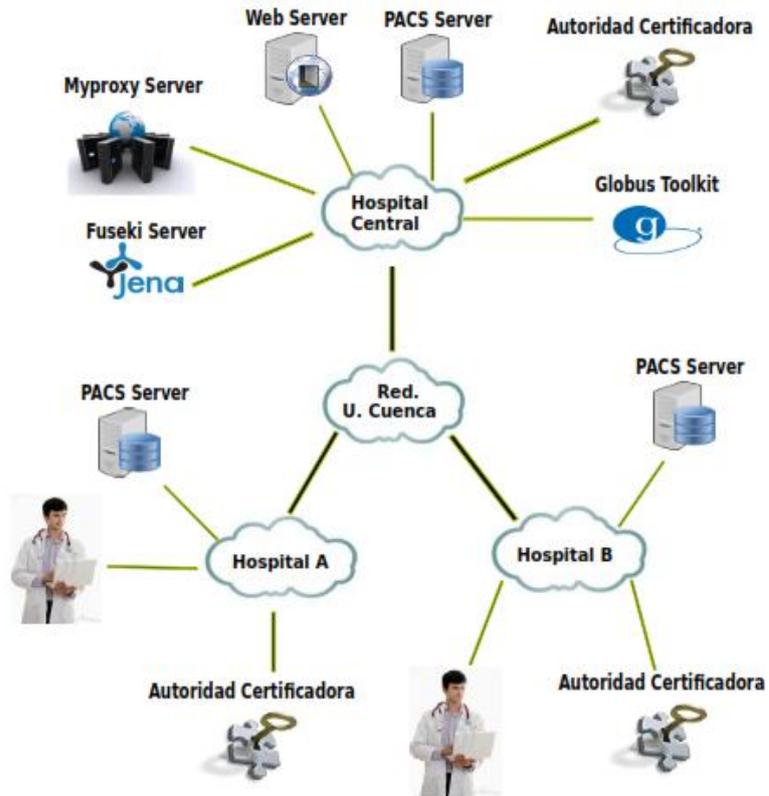


Figura 1. Esquema general de la infraestructura grid.

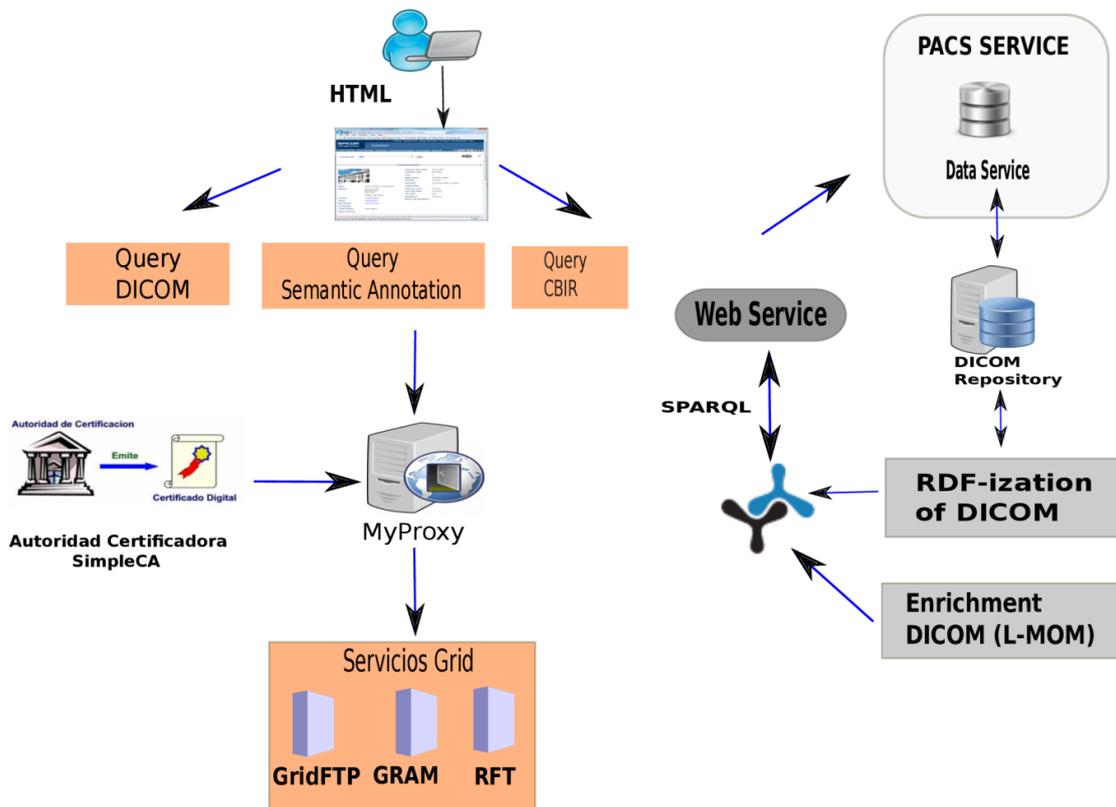


Figura 2. Descripción de los componentes y servicios del nodo central.

### 3.4. Seguridad GSI

Los estudios DICOM, aparte de contener información referente al estudio médico, contienen información relacionada directa o indirectamente a la identidad del paciente. Esto hace que dichos estudios sean documentos altamente confidenciales. Para garantizar esto se implementó un nivel de seguridad basado en la autenticación de usuarios y acceso autorizado. Globus proporciona dos metodologías estándar para realizarlo mediante GSI (Foster & Kesselman, 1999): (a) Mediante certificados X.509 para la autenticación (Sotomayor & Childers, 2006). (b) Mediante certificados SALM para la autorización a través de shibboleth (Morgan *et al.*, 2004). Adicionalmente se implementó el Servidor MyProxy que sirve para almacenar credenciales de usuarios pertenecientes al grid (Basney *et al.*, 2005).

## 4. IMPLEMENTACIÓN DE LA ARQUITECTURA

En ésta sección se describen los pasos a seguir para la instalación y configuración de la seguridad GSI de Globus en 3 máquinas con Linux/Centos. Al finalizar la construcción de la plataforma grid, se ejecuta el servicio de GridFTP para mover un estudio DICOM entre los distintos nodos del grid.

- **Instalación de la Autoridad Certificadora SimpleCA.** Permite garantizar la seguridad GSI en la plataforma médica grid, la cual emite un certificado (CA-Certification Authority). Este certificado se debe crear en el nodo central, siendo el nodo central la autoridad certificadora de todos los nodos conectados en el grid. Este proceso genera un archivo (.targz o .rpm) el cual contiene la información de la entidad certificadora, incluyendo el certificado público, privado, políticas de firmado y algunas extensiones. Este archivo se instala en los otros nodos del grid, con el fin de que estos otros nodos puedan realizar peticiones de certificación.
- **Certificación de Nodo/Usuario.** Dentro de la plataforma grid se instala el paquete generado por la CA (el archivo .targz o .rpm) y a su vez se crean los usuarios. Para fines de pruebas en nuestro caso se generaron tres usuarios; nodogrid1, nodogrid2 y nodogrid3. Cada uno de estos usuarios pertenece a un nodo distinto dentro del grid. Estos Nodos/Usuarios pueden enviar trabajos y transferir estudios DICOM dentro de la infraestructura grid una vez que hayan obtenido su certificado de autenticación.

Una vez que se generan los certificados (CA) tanto para los nodos, como para los usuarios, la información generada se guarda en un archivo llamado *Gridmap-file*, con fines de validar la autorización de usuarios y nodos dentro de la arquitectura grid. Por cada usuario se genera un certificado *proxy*, el cual contiene una cadena de seguridad formada por: *CA + certificado de usuario + certificado proxy*. Una vez que el usuario recibe este certificado *proxy* entonces podrá acceder a los servicios que brinda el servicio grid.

## 5. SERVIDOR MYPROXY

MyProxy es un repositorio para portales grid que permite almacenar credenciales *proxy X.509*, mejora la usabilidad, dando a los usuarios acceso a sus credenciales a través de la red usando autenticación a través de contraseñas, permitiendo a los usuarios delegar sus credenciales a través de un navegador web (Basney *et al.*, 2005). En la infraestructura grid se implementó una configuración típica con un servidor MyProxy dedicado para el hospital central y clientes MyProxy para el resto de hospitales.

## 6. RESULTADOS

Una vez que realizada la instalación y la configuración de seguridad GSI en cada uno de los nodos de la arquitectura grid de prueba utilizada, con un nodo central y dos nodos interconectados, simulando dos hospitales o centros de imágenes, se procedió a verificar que los certificados de los usuarios (claves públicas y privada estuvieran correctos), para luego generar un certificado proxy X.509 en cada uno de los nodos con una clave privada adicional y un tiempo de validez de 7 días. Es necesario indicar que las claves generadas durante el proceso de certificación de nodos dentro de la arquitectura grid fueron debidamente encriptados usando los algoritmos de encriptación RSA y sha1-RSA. Una vez obtenido el certificado proxy X.509, se realizaron pruebas de transferencias de imágenes DICOM entre los tres nodos de la infraestructura grid, a través del protocolo GridFTP de Globus Toolkit de alto rendimiento, seguro y confiable, optimizado para trabajar con grandes cantidades de datos. Al utilizar los componentes que ofrece el Middleware Globus Toolkit se pudo obtener altas tasas de transferencia para el envío de imágenes DICOM (reducción del tiempo de transferencia con respecto a una arquitectura tradicional), es decir que si queremos trabajar con imágenes de gran tamaño el protocolo GridFTP es ideal.

## 7. CONCLUSIONES Y TRABAJOS FUTUROS

Las tecnologías grid en general y Globus Toolkit en particular, se han diseñado para permitir que las organizaciones virtuales puedan integrar y compartir datos a través de varias instituciones. El objetivo principal de implementar esta infraestructura grid es que este enfocado principalmente en la integración de una herramienta para la seguridad que sea eficiente y transparente para el usuario, además de que sea capaz de dar soporte a repositorios distribuidos a través de la autenticación y autorización de usuarios dentro de la infraestructura grid. En general esta infraestructura brinda un fuerte énfasis en las consultas hacia repositorios DICOM, consultas sobre estudios DICOM por contenido visual (CBIR) y anotaciones semánticas sobre imágenes médicas almacenadas en servidores PACS. La implementación de esta infraestructura grid sirve como modelo para futuras implementaciones orientadas a plataformas médicas basadas en el uso de la tecnología grid. Para su implementación se utilizó Globus Toolkit, especialmente por los componentes de alto nivel que ofrece este *Middleware* (seguridad GSI y transferencia de archivos GridFTP). Además se plantea la utilización de un Web Service para compartir los recursos heterogéneos en la arquitectura grid. Brindando así altas tasas de transferencia de datos entre un PACS remoto y una estación local de trabajo.

## AGRADECIMIENTOS

Parte de este proyecto fue soportado por el programa PROMETEO de la Secretaria de Educación Superior, Ciencia, Tecnología e Innovación de la Republica del Ecuador, SENESCYT, la DIUC (Dirección de Investigación de la Universidad de Cuenca), Universidad Técnica particular de Loja, la Universidad Politécnica Salesiana de Quito y CEDIA (Consortio Ecuatoriano de Internet Avanzado) con el proyecto CEPRAVIII-Teleradiología.

## REFERENCIAS

Allcock, W., J. Bester, J. Bresnahan, A. Chervenak, L. Liming, S. Tuecke, 2003. *Gridftp: Protocol extensions to ftp for the grid*. Global Grid Forum GFD-RP, 20, 1-21. Disponible en <http://grid.desy.de/docs/GridFTPprotocol.pdf>.

- Amendolia, S.R., W. Hassan, T. Hauer, D. Manset, R. McClatchey, D. Rogulin, T. Solomonides, 2004. MammoGrid: A service oriented architecture based medical grid application. In: Grid and Cooperative Computing-GCC 2004, pp. 939-942. Springer Link.
- Basney, J., M. Humphrey, V. Welch, 2005. The MyProxy online credential repository. *Software: Practice and Experience*, 35(9), 801-816.
- Erberich, S.G., J.C. Silverstein, A. Chervenak, R. Schuler, M.D. Nelson, C. Kesselman, 2007. Globus MEDICUS-federation of DICOM medical imaging devices into healthcare Grids. *Stud. Health Technol. Inform.*, 126, 269-78.
- Erl, T., 2004. *Service-oriented architecture: A field guide to integrating XML and web services*. Publisher: Prentice Hall, 560 pp.
- Filipovic, B., T. Straub, 2006. *Grid security infrastructure - Ein überblick*.
- Foster, I., C. Kesselman, 1999. *The globus toolkit. The grid: blueprint for a new computing infrastructure*. Morgan Kaufmann Publishers, pp. 259-278.
- Gualán, R., J. Guillermo, W. Pérez, L. Solano-Quinde, W. Ramírez, A. La Cruz, 2015. Grid platform for medical federated queries supporting semantic and visual annotations. In: SIPAIM 2015, Cuenca, Ecuador (in press).
- Jena, J., P. FUSEKI, 2004. *Semantic web frameworks*. In: Oaks, R., University of Sterling. Disponible en <http://www.cs.stir.ac.uk/courses/31Z7/posters/2014/rao.pdf>, 1 pp.
- Krefting, D., J. Bart, K. Beronov, O. Dzhimova, J. Falkner, M. Hartung, A. Hoheisel, T.A. Knoch, T. Lingner, Y. Mohammed, 2009. MediGRID: Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems*, 25(3), 326-336.
- Milano, F., J.F.G. Eijó, A. Gómez, F.G.B. de Quirós, M. Risk, 2009. *Procesamiento de señales médicas en una arquitectura grid*. XVII Congreso Argentino de Bioingeniería, VI Jornadas de Ingeniería Clínica.
- Morgan, R., S. Cantor, S. Carmody, W. Hoehn, K. Klingenstein, 2004. Federated security: The shibboleth approach. *Educause Quarterly*, 27(4), 12-17.
- Noreña Tatiana, R.E., 2013. Compresión de imágenes medicas. *Revista del Instituto Nacional de Salud Biomédica*, 33(1), 137-151.
- Novotny, J., S. Tuecke, V. Welch, 2001. *An online credential repository for the grid: Myproxy*. In: Proc. 10th IEEE Int. Symp. On High Performance Distributed Computing, pp. 104-111.
- Ruiz, M.G., A.G. Chaves, C.R. Ibañez, J.M.G. Mazo, J.C.R. Giraldo, A.P. Echavarría, E.V. Díaz, G.P. Restrepo, E.N.M. Munera, B.G. Loaiza, 2011. mantisGRID: A grid platform for DICOM medical images management in Colombia and Latin America. *Journal of Digital Imaging*, 24(2), 271-283.
- Saltz, J., S. Oster, S. Hastings, S. Langella, T. Kurc, W. Sanchez, M. Kher, A. Manisundaram, K. Shanbhag, P. Covitz, 2006. caGrid: Design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics*, 22(15), 1910-1916.
- Sharma, A., T. Pan, B.B. Cambazoglu, M. Gurcan, T. Kurc, J. Saltz, 2009. VirtualPACS - A federating gateway to access remote image data resources over the grid. *Journal of Digital Imaging*, 22(1), 1-10.
- Sotomayor, B., L. Childers, 2006. *Globus Toolkit 4: Programming Java Services*. Morgan Kaufmann Publishers, 506 pp.
- Warnock, M.J., C. Toland, D. Evans, B. Wallace, P. Nagy, 2007. Benefits of using the dcm4che Dicom archive. *Journal of Digital Imaging*, 20, 125-129.
- Watson, P., 2003. *Databases and the grid*. In: Berman, F., G. Fox, A.J.G. Hey (Eds.). *Grid computing: Making the global infrastructure a reality*. John Wiley & Son Ltd., 1060 pp.