Globus MEDICUS - Federation of DICOM Medical Imaging Devices into Healthcare Grids

Stephan G. ERBERICH[a,b], Jonathan C. SILVERSTEIN[c], Ann CHERVENAK[d], Robert SCHULER[d], Marvin D. NELSON[a], Carl KESSELMAN[d]
[a]*Department of Radiology, Childrens Hospital Los Angeles, Keck School of Medicine, University of Southern California (USC)*
[b]*Department of Biomedical Engineering, Viterbi School of Engineering, USC*
[c]*Departments of Surgery, Radiology and Computation Institute of the University of Chicago and Argonne National Laboratory*
[d]*Center for Grid Technologies, Division of Advanced Systems, Information Sciences Institute*

Abstract: The Digital Imaging and Communications in Medicine (DICOM) standard defines Radiology medical device interoperability and image data exchange between modalities, image databases - Picture Archiving and Communication Systems (PACS) - and image review end-points. However the scope of DICOM and PACS technology is currently limited to the trusted and static environment of the hospital. In order to meet the demand for ad-hoc tele-radiology and image guided medical procedures within the global healthcare enterprise, a new technology must provide mobility, security, flexible scale of operations, and rapid responsiveness for DICOM medical devices and subsequently medical image data. Grid technology, an informatics approach to securely federate independently operated computing, storage, and data management resources at the global scale over public networks, meets these core requirements. Here we present an approach to federate DICOM and PACS devices for large-scale medical image workflows within a global healthcare enterprise. The Globus MEDICUS (Medical Imaging and Computing for Unified Information Sharing) project uses the standards-based Globus Toolkit Grid infrastructure to vertically integrate a new service for DICOM devices – the DICOM Grid Interface Service (DGIS). This new service translates between DICOM and Grid operations and thus transparently extends DICOM to Globus based Grid infrastructure. This Grid image workflow paradigm has been designed to provide not only solutions for global image communication, but fault-tolerance and disaster recovery using Grid data replication technology. Actual use-case of 40 MEDICUS Grid connected international hospitals of the Childerns Oncology Group and the Neuroblastoma Cancer Foundation and further clinical applications are discussed. The open-source Globus MEDICUS project is available at http://dev.globus.org/wiki/Incubator/MEDICUS.

Keywords: DICOM Grid, Medical Image Grid, Protected Health Information, Grid PACS

Introduction

The Digital Imaging and Communications in Medicine (DICOM) standard (www.nema.org) defines the image and meta-data format and the network protocol between medical imaging devices. DICOM has become the de-facto standard for medical imaging adopted by all medical equipment vendors.

The latest DICOM standard, release v3, defines the image format and a peer-to-peer network transport protocol. DICOM objects contain meta information about patient, study, series, and image attributes. Together, image format, meta data, and network model, define a medical image specific application domain, referred here as DICOM domain.

Healthcare image management comprises primarily (i) storage, (ii) availability (fault-tolerance and disaster recovery), and more recently (iii) interoperability and information exchange. However DICOM does not provide image management, but leaves it to the vendors to implement. As a consequence, we see a variety of silo implementations today making it almost impossible to achieve interoperability. The Integrating the Healthcare Enterprise (IHE, www.ihe.net) initiative has be created to address the issue of interoperability with the approach to interface existing silos with significant success in recent years. The latest interface, the cross-enterprise document exchange (XDS) and its medical imaging cousin XDS-I are taking shape.

However XDS and XDS-I only address interoperability. Still missing is an open standards based architecture that provides an overarching and scalable infrastructure to implement all aspects of healthcare image management (i-iii). Such an infrastructure must keep the integrity of DICOM intact to preserve compatibility to existing and future investments into imaging devices, Picture Archiving and Communication Systems (PACS), and display workstations – a $1.6 Billion annual business.

We believe Globus Alliance Grid technology (1,2) is one such open architecture. It provides reliable industry standards for the most challenging problems associated with network collaborative environments: (i) high-speed reliable data transport utilizing high-bandwidth networks, (ii) enterprise level security (data, authentication, authorization), (iii) large scale data management and replication, and (iv) publication, discovery, sharing, and use of distributed, independently owned and operated computational, storage, and data resources federated in the Grid as web-services (WS, www.oasis-open.org). The Grid paradigm spawns a virtual organization (VO) over public and/or private networks between resource providers (e.g. imaging devices) and consumers (e.g. radiologists and scientists).

The significance of the MEDICUS project (Medical Imaging and Computing for Unified Information Sharing) described here, extending previous work (3-5), is the seamless integration of DICOM devices into Grids. As such it builds on the overarching Grid infrastructure to implement fault-tolerant storage with disaster rccovery and interoperable exchange. In the Grid, medical images become transparently available anywhere in the VO (e.g. a Regional Health Information Organizations (RHIO) of hospitals or practices) and among VOs.

Because the images are available within the Grid infrastructure, standards based WS (e.g. storage, image processing, data mining) become easier to develop and to deploy. We believe that such an overarching open standards infrastructure, like

MEDICUS, will enable image availability at the point-of-care and thus helps to eliminate redundant imaging to the benefit of the patient and subsequently reduces cost.

1. Methods

The design of the Grid image management is based on the gateway concept, bridging between the DICOM domain and the VO (Figure 1). This concept provides the benefit of a lightweight extension, without changes to the existing DICOM equipment.

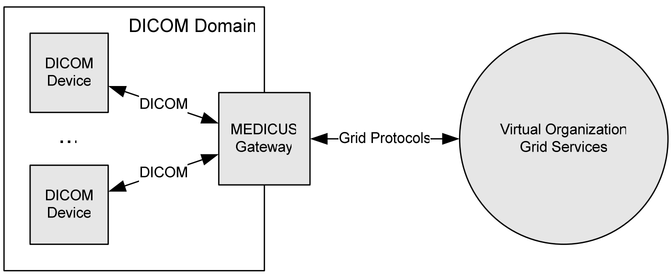

Figure 1. Gateway concept: Transparent connection between DICOM devices, e.g. PACS, and Globus Toolkit Grid services. The Gateway resides in the LAN or DMZ of the DICOM domain (e.g. hospital).

The MEDICUS gateway has four major functions: (i) DICOM to Grid Protocol translator, (ii) discovery (metadata and logical file), (iii) image data transport, (iv) image data caching. The components are vertically integrated from existing Globus technologies: (i) MyProxy for delegated credential creation and Grid authentication management (grid.ncsa.uiuc.edu/myproxy), (ii) Gridshib (6) and Internet2 Shibboleth (shibboleth.internet2.edu) to state user assertions for role-based service authorization, (iii) OGSA-DAI (7) as meta catalog to reference image attributes and storage keys, (iv) Replica Location Service (RLS, dev.globus.org/wiki/Replica_Location) to map image keys and physical storage locations, and (v) GridFTP for data transport and as storage provider (dev.globus.org/wiki/GridFTP).

1.1. DICOM Grid Interface Service

The DICOM Grid Interface Service (DGIS) (8) implements these functions and acts as the conduit between DICOM and Grid domains. Seamless communication between both domains requires carefully orchestrated image workflows and Grid methodologies (Figure 2), integrating security (data and access), data storage, image publication and discovery, and fault-tolerance. DGIS enables image workflows for (i) send (publication) and (ii) query/retrieve (discovery) of images which are described as follows.

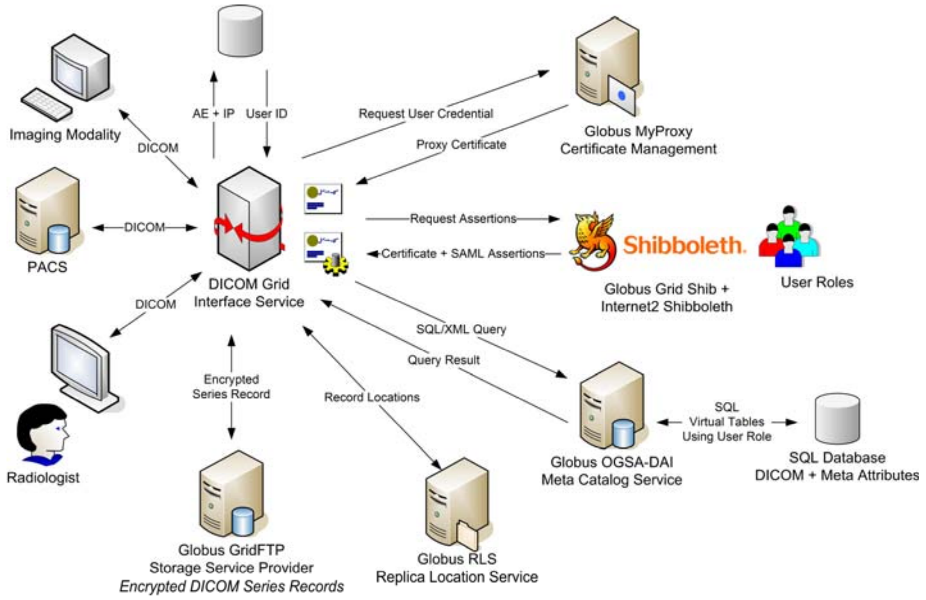

Figure 2. Healthcare image management with standards based Grid open architecture comprising user authentication (MyProxy), user-role authorization (SAML assertions), DICOM meta data query (OSGA-DAI), Grid image series discovery (RLS), and storage/retrieve (GridFTP).

1.1.1. Image Send Workflow

A typical image send workflow of a DICOM modality, e.g. a PACS, pushes images to DGIS using DICOM TCP/IP communication. DGIS accepts the connection and validates the application entity title and internet protocol address. Next the calling modality initiates a DICOM C-Store operation and DGIS acts as DICOM Storage Class Provider (SCP) to receive the images.

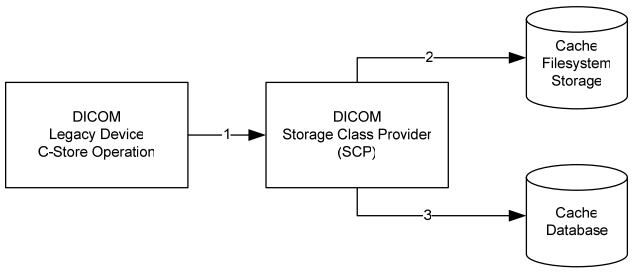

Figure 3. DICOM image C-Store operation from DICOM legacy devices stores images in study, series, and image hierarchy on the DGIS host file system.

DGIS is multithreaded and capable to receive multiple requests at a time (Figure 3.1). Images are stored as part-10 DICOM object on the host file system (Cache) (Figure 3.2) with meta-data reference in the Cache DB (Figure 3.3).

Next the Grid Scheduler finds and schedules new series for Grid publication. Each scheduler entry lists the series, Grid storage and the Meta Catalog URLs. Multiple storage and/or catalog servers at a time can be configured.

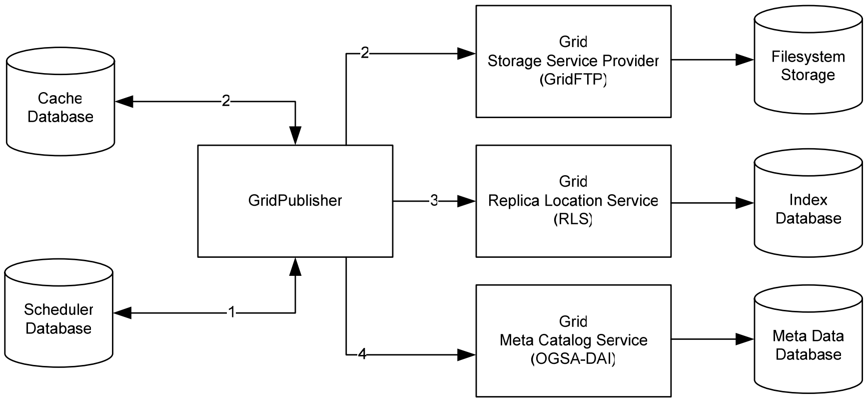

Figure 4. DGIS Grid Publisher thread executes scheduled DICOM image series updates to the Grid. Data (compressed series records) and meta data (DICOM attributes) are stored separately.

Next the Grid Publisher thread (Figure 4) executes the scheduled Grid requests (Figure 4.1). First a compressed series record using the Java zip container is created using the loss-less Lempel-Ziv-Welch (LZW) algorithm. Compressing the images of one series into a single series record provides about 60% in data reduction in average and a new atomic data element used in the Grid. This is of importance because Grid data transport is optimized for large files and single small image file transfer used by DICOM proved to be inefficient. However DICOM query and retrieve at the image level from a modality is supported by DGIS in conformance to the DICOM standard.

The unique identifiers (UID) of study and series are used as keys in MCS and in RLS to reference a series. In future versions of Globus MEDICUS, these UIDs will be replaced with new DICOM compatible ones to ensure anonymity of the series information (even when unassociated with specific identifiers, these identifiers also exist in the individual hospitals' PACS, so trail re-identification is possible).

Now the Grid Publisher publishes series records via GridFTP to the remote Grid SSP (Figure 4.2). Next the RLS is updated with the new Grid location (Figure 4.3). The VO's MCS (Figure 4.4) is invoked to execute an XML/SQL update statement of relevant DICOM attributes (e.g. patient, study, series, images attribute levels). After successful Grid publication the series request is removed from the DGIS Scheduler Database. Now the series is discoverable in the Grid.

1.1.2. Image Receive Workflow

A DICOM legacy device (e.g. a PACS Display Workstation) queries DGIS using standard DICOM C-Find operation to obtain images from the Grid. DGIS translates C-Find into a Grid request, an XML embedded SQL query schema, and sends it to the MCS. The MCS executes the statement on the back-end database (e.g. MySQL, Oracle, or DB2). The query result is returned to DGIS and translated into a DICOM C-Find respond object and sent to the calling device. The device parses the result object and lists the query results to the application.

Once a C-Find is obtained, a selection of images at the patient, study, series or image levels is prepared and returned to DGIS. Both synchronous DICOM C-Get and asynchronous DICOM C-Move operations (Figure 5.1) are handled by DGIS and translated into XML embedded SQL queries. Study and series UIDs are obtained from the MCS (Figure 5.2) and compared to existing series in the Cache DB (Figure 5.3). If not in cache, the Grid storage location is requested from RLS (Figure 5.3) and a GridFTP transfer is invoked (Figure 5.4). Then a C-Get/Move response (C-Store SCU role) is created and images are send to the destination device of the DICOM retrieve (Figure 5.5). DGIS provides streamed operations - series images are sent immediately to the DICOM device while remaining requested series are still in Grid transfer. This allows immediate image review before the entire study is retrieved. DGIS preserves study and series order for C-Find and C-Get/Move operations.

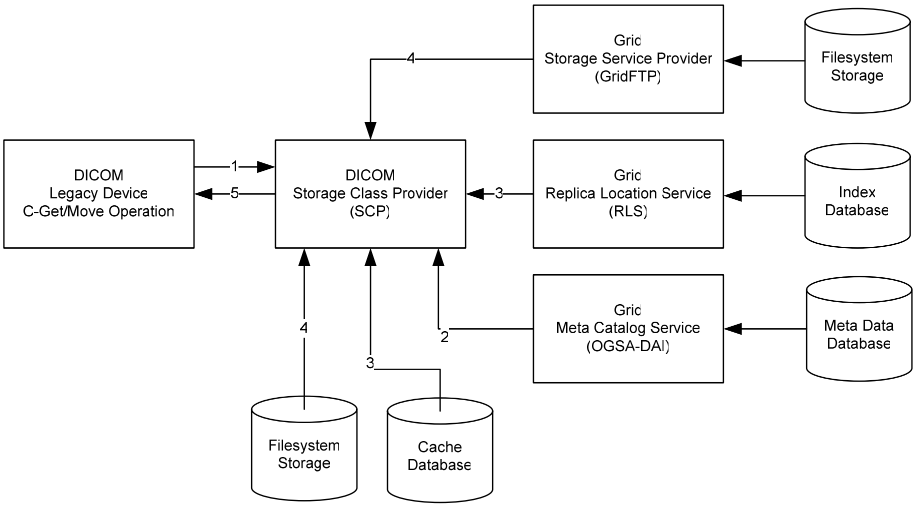

Figure 5. DGIS C-Get/Move DICOM operation. Image locations are discovered from the Grid MCS and delivered either from the local file system or from a Grid SSP.

Using DB back-ended handlers for publication and discovery workflows ensures fault tolerance of the DGIS. Thus a system failure and restart of the system does not affect the operational state of DGIS which will continue where interrupted. DGIS logs all activities in its DB providing an audit trail.

1.2. Security

Sharing medical images with embedded meta-data across medical facilities raises concerns about protected healthcare information (PHI) defined by HIPAA (Health Insurance Portability and Accountability Act of 1996, www.hhs.gov/ocr/hipaa).

Therefore Grid image workflows must be strictly controlled by user authentication and access authorization with audit logging. Globus Grid technology provides two industry standard methodologies (i) X.509 certificates for authentication and (ii) SAML certificate extensions (www.oasis-open.org) for access authorization.

MEDICUS integrates these methodologies to ensure privacy-preserving through a three-layered process (Figure 2). The first layer is that DGIS obtains authentication from an identity provider (IdP). Here we use the Internet2 Shibboleth as trust anchor with GridShib translation and obtain SAML (www.oasis-open.org) assertions (Figure 2). The second layer is the MyProxy X.509 certificate authority (CA) of the VO which creates a limited lifetime credential with or without SAML embedded assertions. Here the CA is operated by the VO administrator (e.g. the hospital network), the IdP at the federal level.

The third layer is that all images stored in the Grid (series records) are de-identified in compliance to DICOM using a message digest one-way encryption. Now the original patient identifiers (e.g. patient name, patient ID) are stored in the MCS database along with the encrypted identifiers. To access the original identifiers one needs to bear matching SAML assertions in the X.509 certificate. GridShib receives the proxy credential of the first-layer and adds SAML assertions about the user, e.g. "Dr. No is member of the Keck School of Medicine USC, Los Angeles".

The layered security mechanism is comparable to a passport (X.509) and a visa (SAML) which allows entrance and privileges. A user needs to apply for both, passport and visa, separately increasing the overall security.

Table 1. Example of SAML assertions stored in the MCS database.

ID	User	Department	Institution
0815	Dr. No	Radiology	Keck School of Medicine, USC, Los Angeles, USA
0816	Dr. Yes	Neurology	Sick Childerns Hospital, University of Toronto, Toronto, Canada

Now DGIS implements this workflow by either obtaining a user credential (e.g. PACS administrator credential) or by matching the calling DICOM device. The latter uses the calling application entity title (AE) and internet address (IP) as user identification (e.g. Dr. No uses Display Workstation "Dr. No's WS" with a static IP "192.168.0.42"). Now SAML assertions and proxy credentials are obtained. The SAML embedded proxy credential is now presented to all MCS activities, publication and discovery. MCS exposes WS operations – update/insert and query. In both cases SAML assertions are incorporated into the SQL statement at the MCS. Therefore the DGIS SQL statement is never directly executed, but only used to define the update/search space.

DGIS removes PHI when performing publication and restores them when needed for query/retrieve. When images are published to the MCS, user, department, and institutional assertions of the publisher are associated with the patient. If data of the same patient from multiple institutions is added, additional entries exist for all image providers allows them to access the PHIs (Table 1). Now accessing RLS or SSP one only requires first-layer security - a valid proxy credential of the VO, because the information stored here does not contain PHI.

The layered security model has two significant benefits: (i) it is PHI preserving under the control of a VO independent trust anchor allowing multiple security domains on-top of VOs, (ii) it allows access to the image data (with encrypted PHI) without PHI clearance. The latter is particularly useful if images must be shared, but PHI should not (e.g. commercial SSP, commercial image processing services, teleradiology, and VO external image review). Because SAML assertions are added to the certificate, one can collect assertions from other security domains and aggregate access to several VOs.

1.3. Data-Storage, Fault-Tolerance, and Disaster Recovery

Because image storage is external and supplemental to existing PACS in the MEDICUS model, FT and DR is inherited by the core replication capabilities of Globus Toolkit (9). As such smaller healthcare providers without PACS (> 400 beds) can use Globus MEDICUS as open-source Grid PACS with image caching at the gateway. Larger providers with PACS can deploy their SSP(s) on campus or remote with high-speed network access. In both scenarios FT and DR is achieved by the flexible and scalable nature of the Grid model. A GridFTP SSP is a light-weight storage unit requiring only the GridFTP server on a standard PC hardware (currently ~$800/TB). This allows proportional scaling with application requirements (e.g. Linux based PC up to multi-processor data center). Obviously one can achieve significant increase in data safety and FT, when using data replication at multiple deployed Grid connected SSPs in combination with RLS indexing the SSPs content, shifting PACS operations toward a pure network based approach - already the trend in the industry.

1.4. Web-Service Image Management

Besides image publication and discovery through DICOM legacy devices, one can take advantage of the fact that images are now discoverable directly with in the Grid - direct access to the data (meta and image data) through standards based WS interfaces. Therefore no separate SDK or API is needed for application development.

In order to provide a global image management paradigm, one wants to be able to control DGIS from the Grid. The next release of DGIS will provide a WS interface accessible from Grid allowing reverse – Grid initiated – workflows. This is of particular interest in combination with GridPortal technology (e.g. GridSphere), because it enables a pure WS image management.

2. Results

We made two observations using MEDICUS in the research image workflows within the Children's Oncology Group (COG) and Neuroblastoma Cancer Foundation (NANT) Grids, connecting 40 international medical centers.

(i) MEDICUS was well accepted by PACS technologists and physicians because it hides Grid specifics and does not require changes to the existing DICOM environment and workflow. As a result MEDICUS has been incorporated seamlessly into the research image workflow at the deployed sites.

(ii) One anticipates better network performance using the more efficient compressed series records and GridFTP vs. DICOM for WAN Enterprise operations. We setup a benchmark test to measure these network performance differences using six representative nodes using Linux 2Ghz systems within the COG Grid (CHLA/USC Los Angeles; Stanford Univ. Medical Center, Palo Alto; Univ. of Washington, Seattle; Mayo Clinic, Rochester; Univ. of Toronto, Toronto; St. Justine Hospital, Montreal).

A test suit comprising real image data representing typical clinical and cancer research data types was used: CT: 1 chest series, 42MB; MRI, 13 brain series, 255MB; PET/CT, 2 abdomen series, 45MB. The test suit was transferred 10 times (about 3.4GB) between nodes (one at a time) during office hours at 9am, 12pm, and 3pm pacific standard time. The average performance gain measured for MEDICUS Grid transfer using compressed series record compared to direct DICOM transfer to the six nodes is about factor 2.3 (avg.: 20.6 Mbps Grid; 8.6 Mbps DICOM). No data encryption was used for both protocols.

We also investigated the performance difference per data type (MRI, CT, PET/CT) to compare DICOM vs. Grid protocol overhead. Data types with small data elements result in many small 2D files. These have a proportional large meta-data overhead (e.g. low resolution PET data) (10). Initial results measured between Los Angeles-Seattle showed Grid performance remains more constant (std.dev: 2.75 Mbps Grid; 3.19 Mbps DICOM using public networks).

3. Discussion

Sharing of clinical radiological image data across institutions has proven challenging for clinicians, clinical scientists and even patients. Although the DICOM standard has proven very useful for vendors to build compatible systems and components within radiology departments at hospitals and other large clinical care and research settings, the standard is bulky or limited in regard to scalability, fault-tolerance, security, and communication performance. Difficulties faced by users due to these limitations include the inability to routinely and predictably manage clinical image data obtained from one institution at another institution. Images are often transported by hand on films or on compact disk, released only via individual patient consent for that particular image series and thereby are viewable only in the limited settings in which the patient physically participates in obtaining them and providing them to the user or workstation. We believe this is below current expectations of patients. Similarly, in the research setting, separate archives (usually de-identified, which limits future usefulness) are typically created under IRB protocol for each particular use. This creates redundancy and waste. Ideally, clinical image users would participate in virtual organizations in which integration and sharing of images across institutions would be easier, more universal, and still meet the strictest security standards (11-13).

Grid technologies in general and Globus in particular were designed to enable virtual organizations to integrate and share data across institutions by decomposing the systems into specialized compatible components. Globus MEDICUS addresses the underlying technical issues that enable integration of today's PACS across institutions via a technical decomposition of them into resource consumers and providers. Specifically, Globus MEDICUS can be deployed as a master archive into which multiple sites can provide data with specific access privileges gaining FT and security features and improved performance only; or by deploying DGIS and SSPs at each institution allowing the user to virtually aggregate images across multiple SSPs (without pre-negotiated inter-institutional sharing agreements, rather with sharing agreements with individuals or VOs). In these architectures, no specific individual identifier is required for each patient because fuzzy logic can be applied to identify patients at multiple institutions or data can be requested only from the specific places in which it is expected to exist.

Naturally, building an enabling infrastructure alone is insufficient to meet user demands. Social issues such as legal requirements and work-flow preferences within each institution and across institutions must be addressed to build workable global healthgrids. However, what MEDICUS does for DICOM is enable virtually any management policy imaginable to be executed while providing FT and improved performance.

4. Conclusions

The Globus MEDICUS project addresses several major barriers to establishing global clinical image sharing: (i) Integration of multiple sites into distributed scalable, FT, global PACS via trivial cost and open technology; (ii) Role-based, user-based and/or anonymous data security according to Grid and other open HIPAA compliant standards; (iii) Integration of DICOM devices and PACS via DICOM protocols; (iv) Loss-less image compression to substantially increase communication performance. Initial deployment across 40 medical centers in North America is concluded and will serve as ample test-bed to prove the value of the Grid model. Others are encouraged to use Globus MEDICUS and contribute toward its open development at: dev.globus.org/wiki/Incubator/MEDICUS

5. Acknowledgment

The authors would like to acknowledge Gaurang Mehta for his generous help with GT4 coding, installing, and advice. Thanks to Dr. David Clunie for his support with DICOM and his open-source PixelMed Java Dicom toolkit (www.pixelmed.com). This work is supported by a grant of the National Institute of Health, UO1-CA97452, and by an award to the New Approaches to Neuroblastoma Therapy (NANT) consortium from the Children's Neuroblastoma Cancer Foundation (cncf-childcancer.org).

References

[1] 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

[2] Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, pp 2-13, 2005

[3] An OGSA Middleware for managing medical images using ontologies. J Clin Monit Comput. 2005 Oct;19(4-5):295-305

[4] MammoGrid: large-scale distributed mammogram analysis. Stud Health Technol Inform. 2003;95:194-9

[5] Grid Based Medical Image Workflow and Archiving for Research and Enterprise PACS Applications. Proceedings SPIE Medical Imaging 2006, Volume 6145, pp. 203-208 (2006)

[6] Identity Federation and Attribute-based Authorization through the Globus Toolkit, Shibboleth, Gridshib, and MyProxy. In *5th Annual PKI R&D Workshop*, April 2006.

[7] The design and implementation of Grid database services in OGSA-DAI. Concurrency and Computation: Practice & Experience, Vol17 , Issue 2-4 (February 2005), p357-76

[8] DICOM grid interface service for clinical and research PACS: a globus toolkit web service for medical data grids. Int Journal of Computer Assisted Radiology and Surgery, 2006, 1:87-105; p100-102, Springer, Heidelberg

[9] Wide Area Data Replication for Scientific Collaborations. Proceedings of 6th IEEE/ACM International Workshop on Grid Computing (Grid2005), November 2005

[10] PACS-based functional Magnetic Resonance Imaging. SG Erberich. Comput Med Imaging Graph. 2003 Mar, 27(2-3):229-40

[11] New trends in the virtualization of hospitals--tools for global e-Health. Stud Health Technol Inform. 2006;121:168-75

[12] An extensible framework for sharing clinical guidelines and services. Conf Proc IEEE Eng Med Biol Soc. 2004;5:3155-8.Stud Health Technol Inform. 2006;120:381-91

[13] TRENCADIS--a WSRF grid MiddleWare for managing DICOM structured reporting objects. Stud Health Technol Inform. 2006;120:358-67