Winter 6-1-2021

DSpace information retrieval system: a study using DICOM metadata standard

Hirak Jyoti Hazarika
North Eastern Hill University, India, hirakhazarika2014@gmail.com

Dr. S. Ravikumar
North-Eastern Hill University

Dr. Akash Handique
North Eastern Indira Gandhi Regional Institute of Health & Medical Sciences

DSpace information retrieval system: a study using DICOM metadata standard

Hirak Jyoti Hazarika
Research Scholar
Department of Library & Information Science
North-Eastern Hill University, Shillong: Meghalaya
E-mail: hirakhazarika2014@gmail.com

Dr. S. Ravikumar
Assistant Professor
Department of Library & Information Science
North-Eastern Hill University, Shillong: Meghalaya
E-mail: ravikumarnehu@gmail.com

Dr. Akash Handique
Associate Professor
Department of Radiology
North Eastern Indira Gandhi Regional Institute of Health and Medical Sciences
Shillong: Meghalaya
E-mail: akashhandique2019@gmail.com

Abstract:

Purpose:

The present paper talks about the use of institutional repository software (DSpace) in archiving DICOM images. In the current study, the authors have tried to integrate the DICOM metadata standard with DSpace, which was compatible with DC & OAI PMH. After integrating the DICOM standard with DSpace, and the repository tested with a sample of ten thousand images, the retrieval results using various DICOM tags was very satisfactory. This study paves for the use of OSS in storing and retrieving medical images**.** Digital Image Communications in Medicine (DICOM) standard has become a centre component for making an application nonpartisan programming layer between optional long haul stockpiling and essential application stockpiling. Thus, in exploring costs around long-term storage alternatives, we need to understand the roles and types of storage used in clinical applications. Ultimately, it is the volume of data that needs to be accessible for years, due to clinical and regulatory reasons that extend the cost of data management beyond the life of the technology used to keep it. Expenses related with clinical imaging keep on rising. Imaging offices regularly need to update gear each three to five years, which puts a gigantic weight on their monetary assets. Hence to overcome the ever increasing storage cost we tried OSS

(DSpace) which widely used in academic environment and tried to integrated DICOM standards by this process we have demonstrated that storage cost can be reduced substantially and the proposed imaged repository can be used as retrieval system in the health care industry.

Keywords: DSpace, DICOM standard, Metadata, Open Source Software, Image Retrieval.

Introduction:

Metadata plays a crucial role in a digital library/image repository/museum environment(Chen *et al.,* 2012). However, the development of metadata is not an easy task. Its formulation starts with analysing the attributes of collections as well as understanding the user information needs and information-seeking behaviour. Among the newly developing metadata standards for the World Wide Web, Dublin Core Metadata Element Set (Dublin Core) is the pioneer in the simple content description. There are various methods to deploy Dublin Core metadata, such as embedding it within a resource, linking it to a resource, or storing it in a database. General requirements of a modern metadata infrastructure for libraries are identified, including such qualities as versatility, extensibility, granularity, and openness(Isa, 2018). A new kind of metadata infrastructure is proposed that exhibits at least some of those qualities. Various institutions used to create metadata describing the content and properties of information resources as separate entities from the resource itself. Nowadays, however, given the web environment, the increasing number of published digital objects and the need to share the metadata describing those objects, it has become necessary to embed the metadata as an integral part of the objects. Embedding standard metadata in a file structure makes a file self-describing, helps identify the resource outside its home system and improves the searchability, discoverability and interoperability management of information objects.

The role of metadata in the library environment is now expanding to include managing electronic scholarly products and participating in the evolving scholarly communication process (Horwood *et al.,* 2004). Libraries replaced card catalogue with computer-based systems. This data is stored in the library management system using the MARC metadata standard. Metadata format in the digital libraries includes Dublin Core, METS, MODS, DDI, DOI, URN, PREMIS schema, EML, and OAI-PMH. There are many metadata formats to choose from that can be used for organizing, describing and providing access to image libraries or repositories i.e. IPTC (International Press Telecommunications Council), Dublin core metadata, Encoded Archival Description, DICOM (Digital imaging

communication in Medicine), PLUS (Picture Licensing Universal System), XMP (Extensible Metadata Platform), EXIF (Exchangeable image file format), VRA (Visual Resource Association) etc. In this study, the DICOM metadata standard will be introduced with Digital Library Software i.e. DSpace. In this paper, we present a case study for mapping DICOM Type vocabulary to DSpace. The mapping preserves the semantics of the DC records that correspond to different material types. In particular, this paper presents the problems that arise when creating semantic mappings from DC metadata to DICOM models. In detail, presents a semantic integration architecture, which considers DICOM as amediator schema and the mapping of DC (Dublin Core) to DICOM as part of this architecture.

Literature Review:

Metadata is important for essential processes in the digital environment beyond mere description. Digital image metadata plays a crucial role in managing digital image repositories. TheInternational Press Telecommunications Council (IPTC) has developed a standard for storing descriptive metadata information for digital images. These metadata schemas, as well as other emerging standards, provide a standard format for creating, processing, and exchanging digital image metadata and enable image management, analysis, indexing, and search application (Allen and Schalow, 1999).(Zhang *et al.*, 2007)discussed in details the various technologies for image indexing and retrieval based on shape, colour, texture and spatial location.The DICOM meta data repository has been helpful in creating an awareness that the exposure levels in our county is far from normalised and possibility to collect and compare data from a centralised point of view has enhanced the possibilities to detect exposure differences (Källman *et al.*, 2009). DICOM metadata allow a remote and safe information analysis of millions of people, toward studies comparing clinical efficiency, dissemination of good practices, dissemination of medical technologies as well as the provision of healthcare quality (Santos *et al.*, 2011). This literature review discussed about the DICOM metadata important, metadata for images and metadata for retrieval. From the literature, it is found that DICOM metadata standard is not use with DSpace for Medical Image archive and retrievinginformation so far. So, the authors took the challenge for develop DICOM standard with DSpace and which will be beneficial for LIS community and Medical professionals.

About DSpace:

DSpace Software: DSpace IR software is an open source advanced library cum repository programming bundle ordinarily utilized for making open access vaults for academic and

distributed computerized content which is created by Hewlett-Packard and MIT. DSpace is a custom of teaming up Java web applications and utility projects that keep up an advantage store and a related metadata store. The web applications give interfaces to organization, store, ingest, pursuit, and access. DSpace is intended to make cooperation by investors simple. DSpace presently bolsters diverse kinds of record group for example pdf, doc,xls, psd, and, jpeg, png, mpg, and tiff. DSpace computerized library programming is additionally good with the common interoperability standards utilized in the Institutional repository space, for example,

Open Archives Initiative Protocol for Metadata Harvesting, OpenSearch, and RSS (https://duraspace.org/dspace/). The product has been utilized by exhibition halls, state chronicles, State and National Libraries, University Libraries, College Libraries, Medical Libraries Journal storehouses, consortiums, and business organizations to deal with their computerized resources everywhere throughout the world.

Be that as it may, DSpace isn't good with medical picture group, for example, DICOM record and Zip File. There is no medical picture watcher and peruser programming coordinate with a database to store and examination medical pictures just as DICOM pictures. Yet, it is an open source programming with the goal that source code is open for all to create and modify according to possess requirement. With the goal we have taken this product for improvement and alter this new component for the medical imaging world.

About DICOM Metadata:

Digital Imaging and Communication in Medicine (DICOM) is an established and non-proprietary standard for the storage and the exchange of information in medical imaging. DICOM defines not only a data format, but it specifies also a network protocol based on the ISO/OSI-model. DICOM (Digital Imaging and Communications in Medicine) is an application layer network protocol for the transmission of medical images, waveforms and accompanying

information. DICOM was originally developed by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology for computerized axial tomography (CAT) and magnetic resonance imaging (MRI) scan images. It is now controlled by the DICOM Standards Committee and supports a wide range of medical images across the fields of radiology, cardiology, pathology and dentistry. DICOM uses TCP/IP as the lowerlayer

transport protocol.

DICOM interfaces are available for connection of any combination of the following categories of digital imaging devices:

(a) Image acquisition equipment (e.g., computed tomography, magnetic resonance imaging,and computed radiography, ultrasonography, and nuclear medicine scanners).

(b) Image archives.

(c) Image processing devices and image display workstations.

(d) Hard-copy output devices (e.g., photographic transparency film and paper printers).

The DICOM Information Object Definitions encode the data produced by a widevariety of imaging device types, including: CT (Computed Tomography), MRI (Magnetic Resonance Imaging), Ultrasound, X-ray, Fluoroscopy, Angiography, Mammography, Breast tomosynthesis, PET (positron emission tomography), SPECT (single photon emission computed tomography), Endoscopy, Microscopy, Whole slide imaging, OCT (optical coherence tomography).

DICOM makes extensive use of Unique Identifiers. Almost every entity in the DICOM Data Model has a unique identifier. It is basically known as DICOM tags or elements. There are 5000 DICOM tags support by medical images. Mostly 11 DICOM tags are used in DICOM images processing and information retrieval. i.e.

Sl. No.	Elements/Tags	Tags Name
1	(0008 0020)	Study Date
2	(0008 0030)	Study Time
3	(0008 0050)	Accession Number
4	(0008 0090)	Referring Physician Name
5	(0010 0010)	Patient’s Name
6	(0010 0020)	Patient’s ID
7	(0010 0030)	Patient’s Date of Birth
8	(0010 0040)	Patient’s Sex
9	(0020 0010)	Study ID
10	(0020 0011)	Series Number
11	(0020 0020)	Patient Orientation

Comparison between Dublin Core and DICOM Metadata:

The Dublin Core Metadata Initiative (DCMI) Metadata terms is the current set of the Dublin Core vocabulary. This set includes the fifteen terms of the Dublin Core Metadata Element Set as well as the qualified terms.Dublin Core element set is a flexible and usable metadata schema enabling information exchange and integration between digital sources(Jackson *et al.*, 2008). Digital Imaging and Communication in Medicine (DICOM) metadata consolidation feasibility of images stored on a Picture Archiving and Communication System (PACS) archive for the production of indicators with different granularity levels, independently of medical imaging equipment manufacturers and information systems(Santos *et al.*, 2015). The DICOM metadata tags is known as NEMA (National Electrical Manufacturers Association) standard PS3 (https://www.nema.org/). There are 5000 tags available in Digital Imaging and Communications in Medicine (DICOM)metadata. Mainly 11 tags are used for image processing and information retrieval file indexing. The DC metadata and DICOM metadata elements name are listed below: -

Sl. No.	**Dublin Core Metadata**		**DICOM Metadata**	
	Element	***Tag Name***	***Element***	***Tag Name***
1	01	Title	0020	Study Date
2	02	Author	0030	Study Time
3	03	Subject	0050	Accession Number
4	04	Description	0090	Referring Physician Name
5	05	Publisher	0010	Patient’s Name
6	06	Contributor	0020	Patient’s ID
7	07	Date	0030	Patient’s Date of Birth
8	08	Types	0040	Patient’s Sex
9	09	Format	0010	Study ID
10	010	Identifier	0011	Series Number
11	011	Source	0020	Patient Orientation
12	012	Language		
13	013	Relation		
14	014	Coverage		
15	015	Copy Rights		

DICOM Metadata with DSpace

DSpace Repository software essentially uses to create to library file for uncommon books and research result. Till now, which isn't acquainted in the medicinal field with protect the DICOM therapeutic picture. Basically, Doctors and radiologist are utilizing working framework based software which are top expensive. Consequently, we have completed a pilot contemplate on DICOM pictures and DICOM Metadata to coordinate with DSpace Digital Repository (Figure 1). DSpace is free open source software and source code additionally unreservedly available. With the goal that we can create and modify according to our own needs.

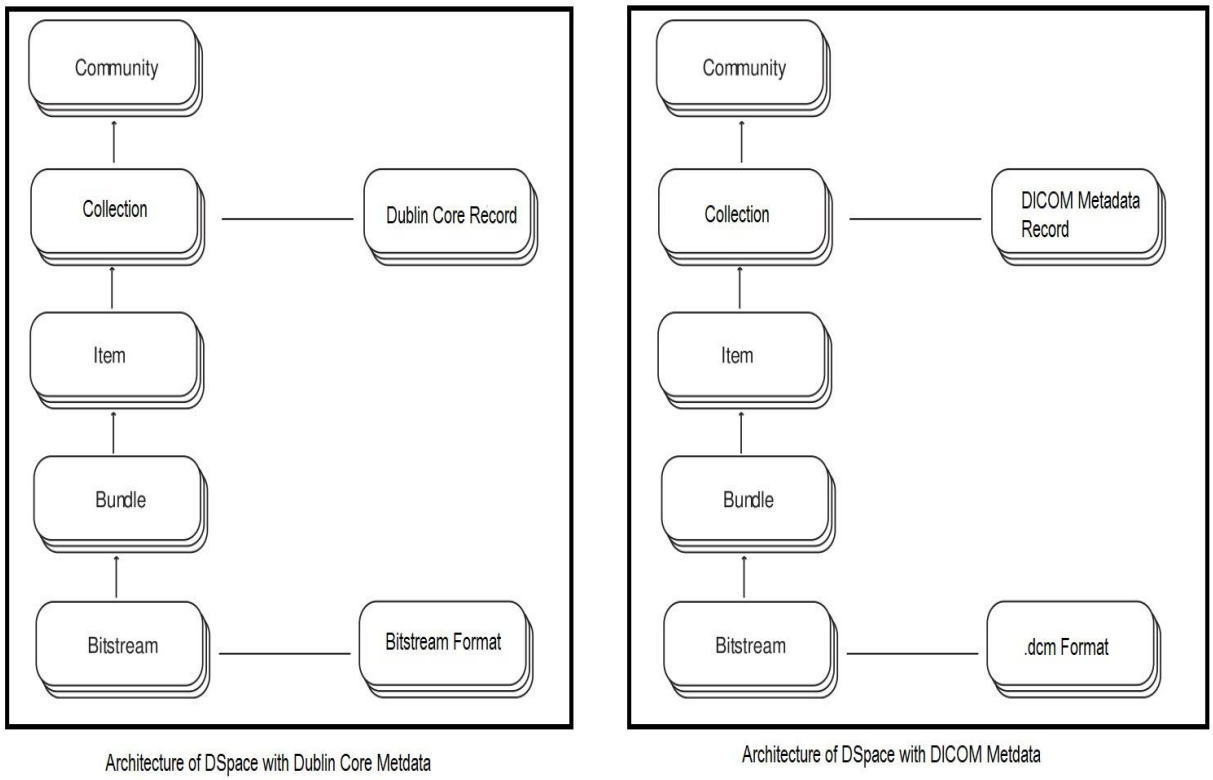

Fig 1: Architecture of DSpace

DICOM Metadata Integration with DSpace:

DSpace Repository software essentially uses to create to library resources for books and research result. Fundamentally, Doctors and radiologist utilized working framework based software which was over the top expensive. Consequently, we have completed a pilot contemplate on DICOM metadata structure to coordinate with DSpace Digital Repository. DSpace is free open source software and source code additionally unreservedly available. With the goal that we can create and modify according to our own needs.

```
1505 metadata.dicom.contributor.*           = Consulting Physician Name
1506 metadata.dicom.contributor.physician    = Consulting Physician Name
1507 metadata.dicom.contributor.physician    = Referring Physician Name
1508 metadata.dicom.date.issued              = Issue Date
1509 metadata.dicom.description             = Description
1510 metadata.dicom.description.report       = Reports
1511 metadata.dicom.identifier              = Other Identifiers
1512 metadata.dicom.identifier.study        = Study Date
1513 metadata.dicom.identifier.accession    = Accession No
1514 metadata.dicom.identifier.id           = Study ID
1515 metadata.dicom.identifier.patientid    = Patient ID
1516 metadata.dicom.identifier.studyuid     = Study UID
1517 metadata.dicom.identifier.uri          = URI
1518 metadata.dicom.institute               = Institute/ Hospital
1519 metadata.dicom.relation.ispartofseries = Series/Report no.
1520 metadata.dicom.subject                 = Keywords
1521 metadata.dicom.imagetitle              = Image Title
1522 metadata.dicom.patient                 = Patient Name
1523 metadata.dicom.imagetype               = Image Source
```

Fig: 2 DICOM Metadata compilation with DSpace

DSpace provides two options for presenting data on the web: JSPUI, a user interface based on Java Server Pages (JSP) and XMLUI, another user interface developed by Texas A&M University and based on the Apache Cocoon framework. Because of the need to create a custom look and feel for the collection, we decided to use JSPUI. (Elias, 2012;Hazarika and Ravikumar, 2019)has provided an overview of DSpace and its development. But it is not discussed any metadata in his experiment. So, the major challenge in integrating the DICOM metadata into DSpace was accessing a set of *dcm* files (bitstreams) associated with a single record; this was mainly because of how DSpace stores beat streams in its asset store directory. A workaround was to create a separate copy of the dcm files and saved them under the DICOM directory; then a Java script takes the date variable and uses it to match the corresponding folder, reads folder content, and generates a drop-down menu which allows users to navigate/ view the pages of that particular issue. When DSpace loads a page for a given record, an JSPUI template sends a URL with the variable to call the JavaScript and embeds the metadata into DSpace. Figure 2 shows the basic JavaScript DICOM Metadata structure.

Therefore, the authors have developed the new metadata structure for DSpace Software inside the backend of JSPUI interface (**Location**: /dspace-api/src/main/resources/Messages. Properties) to integrate DICOM Metadata standard (Figure 2). DICOM picture document comprises of several cuts, each put away in a different DICOM object record and record size will be 200 MB to 3 GB. However, its includes parsing and

displaying of DICOM object's metadata (See Figure 3). Through this, medical professionals can view DICOM metadata with the file. After this development, it will be helpful for store and retrieve medical images as like CT Image, X-ray Image and MRI Image (See Figure 3). DICOM files uploaded to a DSpace are available via DICOM query and retrieves. One of the strengths of this process is that data can only reside in the DICOM image archive if it has been described at the level of detail required for it to be uploaded into DSpace. DICOM files do not contain sufficient detail to complete a DSpace database entry. For example, DSpace descriptions includes title, author, and publisher etc. By requiring the use of DSpace for uploading, DICOM images will be discoverable via Internet/DSpace searches and will persist via Internet/DSpace handles. By simultaneously making the DSpace DICOM images available from a DICOM server, images can be read directly into an DICOM metadata standard program (Figure 3).

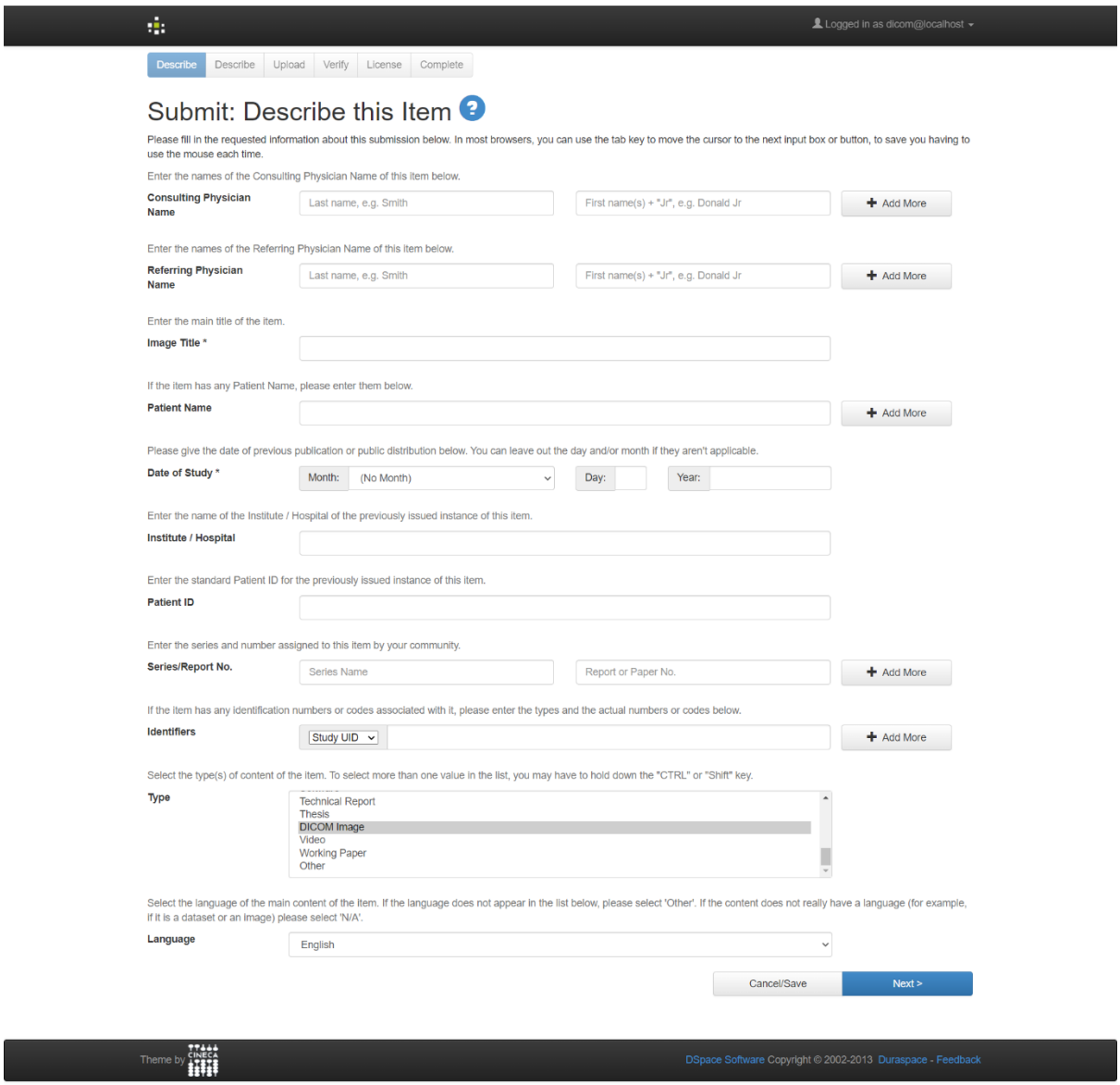

Fig: 3 DICOM Metadata with DSpace

In the recent development of DSpace (Version 6.3) as of now supported more than 75 no's distinct records types or file format types is acceptable. With the existence of a media filter, conversion to other known file types is possible, so data in that format can be preserved and retrieve beyond the useful lifetime of its particular format. But DICOM Metadata standard is a new standard for DSpace. So, it is not recognised by the search box of DSpace. Basically DSpace can search and retrieve information about Document **Title, Author, Publisher and ISBN**. So, the author configures media filter with DSpace by running cornjob (**Example: *@daily /home/dspace/bin/dspace filter-media***). Figure 4 shows the DICOM Metadata standard retrieval function after compilation DICOM metadata with DSpace. Now medical professionals can search data with different variable as like **Image Title, Consulting Physician name, Study ID and Report ID** etc.

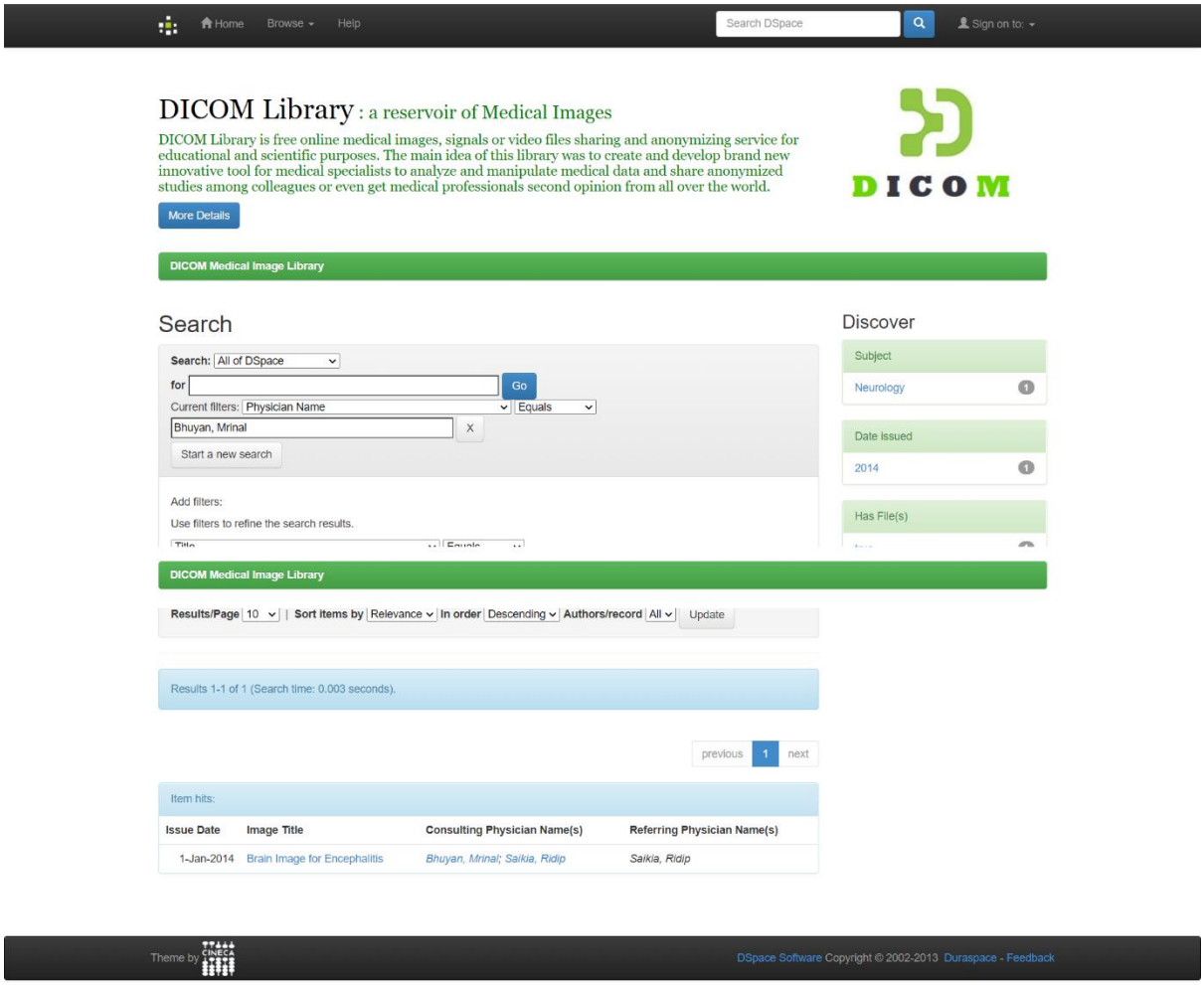

Figure 4: DICOM Metadata Retrieval Function

The authors developed a cross-platform DICOM query-and-retrieve tool that operates as a standalone application or as a file dialog that allows applications to read DICOM files from a remote DICOM server or from a local media storage. The results remain in memory when used as a filedialog or are saved to disk when used as a stand-alone application(Figure 5).

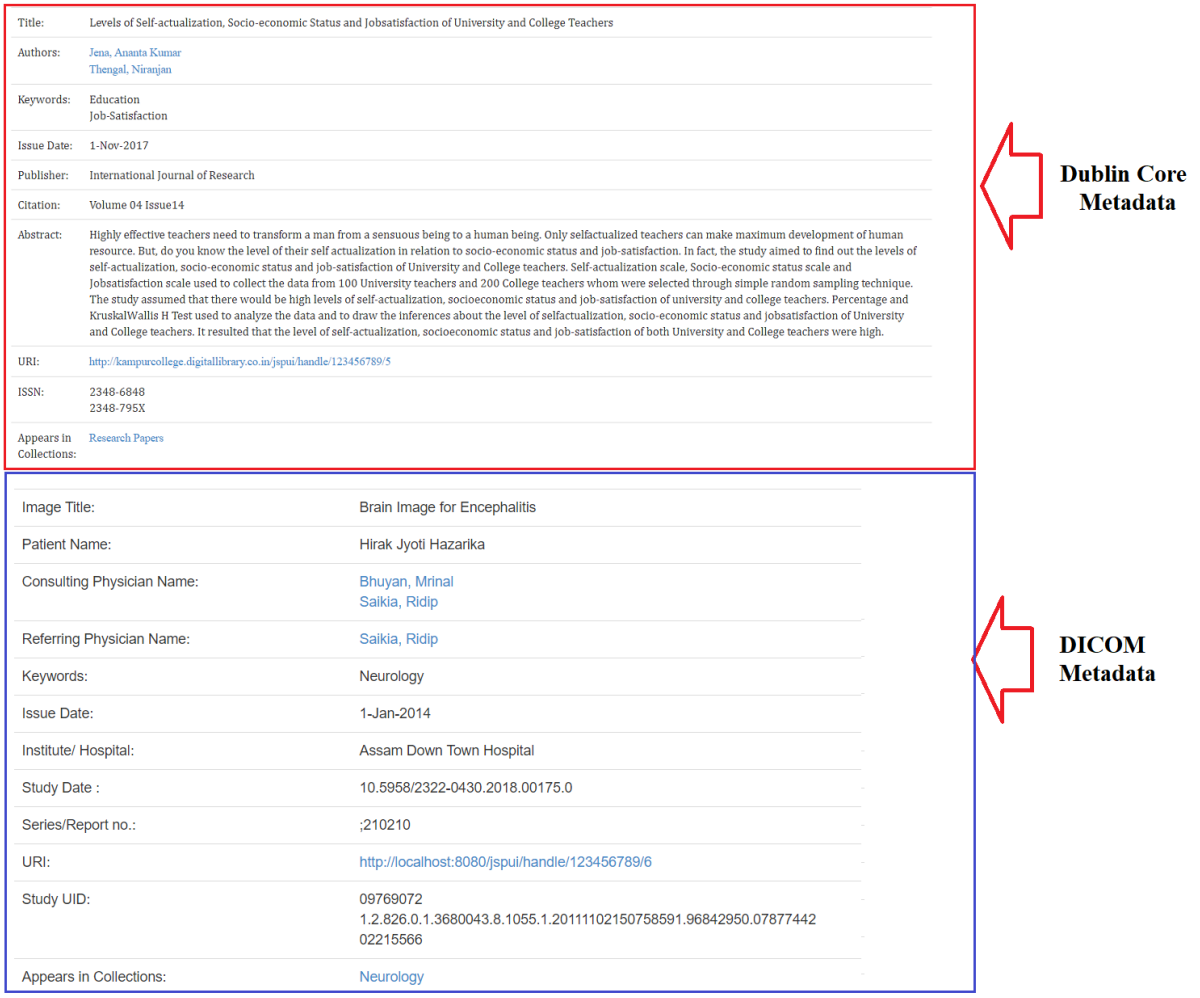

Fig 5: Dublin Core Metadata Vs DICOM Metadata Standard

Finding:

Most importantly, the methods in this paper promote and are available as open-source software. The present study was an attempt to find out use and develop an architecture for Medical Image Library using Open Source Software i.e. **DSpace**. From the study, it is revealed that DSpace has more capability to accommodate DICOM Image file. The results indicated that the current technologies combined with JavaScript, are sufficiently mature to permit the implementation of professional medical images metadata i.e. DICOM with DSpace web-based Institutional Repository Software (See Figure 5).

The DICOM file produced by a wide variety of diagnostic devices that are embedded with information. From the present study, we proposed a new platform to preserved and conserved **dcm file** for a long period of time without converting the image to other file formats like **PNG/ JPEG** extension which leads to loss of embedded information. The cost of digital preservation is still a non-developed matter (Bote, Fernandez-Feijoo and Ruiz, 2012). Software maintenance cost is typically more than fifty percent of the cost of the total software life cycle (Eski and Buzluca, 2011). Our present approach of using OSS platforms for storing **dcm** file are in this direction which will reduce the financial burden of healthcare industries in preserving DICOM images for future generations.

The system is based on a client-server intranet system in which the client needs only a Web browser which enables DICOM users to access these images seamlessly without any hindrance. In addition, if a security policy is in place, the system can easily be used from outside the hospital as this IR System runs on Linux platform which is robust in security compliance compared to the proprietary operating system.

This article is based on DSpace indexing system which is robust in nature in return it enables precise information retrieval. Retrieval in DSpace is very responsive because of the metadata standard followed and the metadata captured by the system for indexing a particular entity. Hence the proposed repository can be considered as an online information retrieval system for DICOM images.

Discussion:

The main objective of this study is to developed and designed the DICOM Based Medical Image Repository. Presently, there is no Open Source DICOM Image Archive or Medical Image Repository software for LIS Community as well as Medical Community. Therefore, Medical Professionals need to store all images in one PC or Hard Disk which is not globally access available or secure. In this study, the authors built up a fresh or out of the box new creative platform for medical experts to store, analyze and control medical information and offer anonymized considers among associates or even get medical experts from everywhere throughout the world in the web-enabled platform.

This is helpful for the community to develop Open Source DICOM Medical Image Repository with the help of the Open Source Platform. This study helps to all Medical Librarian Community, Laboratories, Radiologist and LIS Professionals to build and store an

open-source Digital Library. Presently, there is no open-source Digital Library software to preserve this type of image and real-time DICOM metadata retrieval system. Most importantly, the methods in this paper encourage and are available as open-source software. They worked without modifying those libraries or standards so that each may grow independently and yet benefit one another. By not requiring direct modifications, they set a standard whereby related software libraries (e.g., E-Prints, Omeka & Greenstone) can also be integrated using the same framework. They also provide much-needed DICOM query and retrieve proficiencies.

Conclusion:

DSpace based image repository is not a new idea in the area of the institutional repository. Being an open-source platform it has given us a free hand to modify the source code to accommodate the new .dcm file format and DICOM metadata standard in Dspace. After incorporating these changes; the system response to keyword search and other tag search was very precise. In conclusion, all of the above presents a great opportunity for further discussions and research in this field DICOM based image repository. Incorporating medical image metadata standard in DSpace opened a new window for archiving healthcare dataset which is having embedded information with each image. We acknowledge that our work will moreover add to the initiation of cooperation among various experts to decide the Dspace betterment aspects. In the long run, we believe that all presented tries and proposition will help archivists, institutional repository custodian, customers, specialists, researchers, teachers similarly as understudies and various people from the general populace to find what they need in virtual spaces like progressed documents even more quickly and adequately. With everything taken into account, the whole of the above presents an inconceivable open entryway for extra discussions and investigation in this field.

Acknowledgment:

We acknowledge **Editor and Reviewer** for his valuable comments and suggestions which helped to improve our manuscript always.

Reference:

Allen, R. B. and Schalow, J. (1999) 'Metadata and data structures for the historical newspaper digital library', in *Proceedings of the eighth international conference on Information and knowledge management - CIKM '99*. New York, New York, USA: ACM Press, pp. 147–153. doi: 10.1145/319950.319971.

Bote, J., Fernandez-Feijoo, B. and Ruiz, S. (2012) 'The Cost of Digital Preservation: A Methodological Analysis', *Procedia Technology*, 5, pp. 103–111. doi: 10.1016/j.protcy.2012.09.012.

Chen, C.-M. *et al.* (2012) 'Developing a Taiwan library history digital library with reader knowledge archiving and sharing mechanisms based on the DSpace platform Article information', *The Electronic Library*. Emerald Publishing, 30(3), pp. 426–442. doi: 10.1108/02640471211241681.

Elias, T. (2012) 'Document Viewers for Non-Born-Digital Files in DSpace', *Journal of Digital Information*, 13(1).

Eski, S. and Buzluca, F. (2011) 'An Empirical Study on Object-Oriented Metrics and Software Evolution in Order to Reduce Testing Costs by Predicting Change-Prone Classes', in *2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops*, pp. 566–571. doi: 10.1109/ICSTW.2011.43.

Hazarika, H. J. and Ravikumar, S. (2019) 'Document viewer with DSpace: a practical approach', *Library Hi Tech News*. Edited by R. S. Emerald Publishing Limited, 36(5), pp. 22–26. doi: 10.1108/LHTN-05-2019-0031.

Horwood, L. *et al.* (2004) 'Library Management OAI compliant institutional repositories and the role of library staff"The librarian's role in institutional repositories: A content analysis of the literature OAI compliant institutional repositories and the role of library staff', *Library Management OCLC Systems &amp Iss Reference Services Review*. PT, 25(23), pp. 170–176.

Isa, S. E. (2018) 'Image embedded metadata in cultural heritage digital collections on the web: An analytical study', *Library Hi Tech*. Emerald Publishing Limited, 36(2), pp. 339–357. doi: 10.1108/LHT-03-2017-0053.

Jackson, A. S. *et al.* (2008) 'Dublin Core Metadata Harvested Through OAI-PMH', *Journal of Library Metadata*, 8(1), pp. 5–21. doi: 10.1300/J517v08n01_02.

Källman, H.-E. *et al.* (2009) 'DICOM Metadata repository for technical information in digital medical images', *Acta Oncologica* , 48(2), pp. 285–288. doi: 10.1080/02841860802258786.

Santos, M. *et al.* (2011) 'DICOM and Clinical Data Mining in a Small Hospital PACS: A Pilot Study', in Cruz-Cunha, M. M. et al. (eds) *ENTERprise Information Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 254–263.
Santos, M. *et al.* (2015) 'DICOM Metadata Access, Consolidation and Usage in Radiology Department Performance Analysis. A Non-proprietary Approach', *Procedia Computer Science*, 64, pp. 651–658. doi: 10.1016/j.procs.2015.08.579.
Zhang, J. *et al.* (2007) 'DICOM Image Secure Communications With Internet Protocols IPv6 and IPv4', *IEEE Transactions on Information Technology in Biomedicine*, 11(1), pp. 70–80. doi: 10.1109/TITB.2006.879606.