

# A DICOM-based Software Infrastructure for Data Archiving

Dhaval Dalal, Julien Jomier, and Stephen R. Aylward

Computer-Aided Diagnosis and Display Lab  
The University of North Carolina at Chapel Hill, Department of Radiology  
27510 Chapel Hill, USA  
dalal@bnl.gov  
{jomier, aylward}@unc.edu

**Abstract.** In this paper, we present an open-source image access and archival system that merges the image transfer capabilities of DICOM with web-based digital library technology and the National Library of Medicine’s Insight Segmentation and Registration Toolkit (ITK). Data (DICOM objects, pdf files, tracker data, etc.) uploaded to a digital library is perpetually accessible and indexed on the web via a digital library “handle”. Via the presented integration of digital library and DICOM server technologies, DICOM objects uploaded to a digital library can also be downloaded directly into programs using standard DICOM query and retrieve utilities. Researchers can thereby discover data via web searches and download it via DICOM. Our current implementation uses DSpace as the open-source digital library system and DCMTK as the open-source DICOM library. Furthermore, we have developed a DICOM query and retrieve module and application for ITK, using DCMTK. These software are available in the InsightApplications repository.

## 1 Introduction

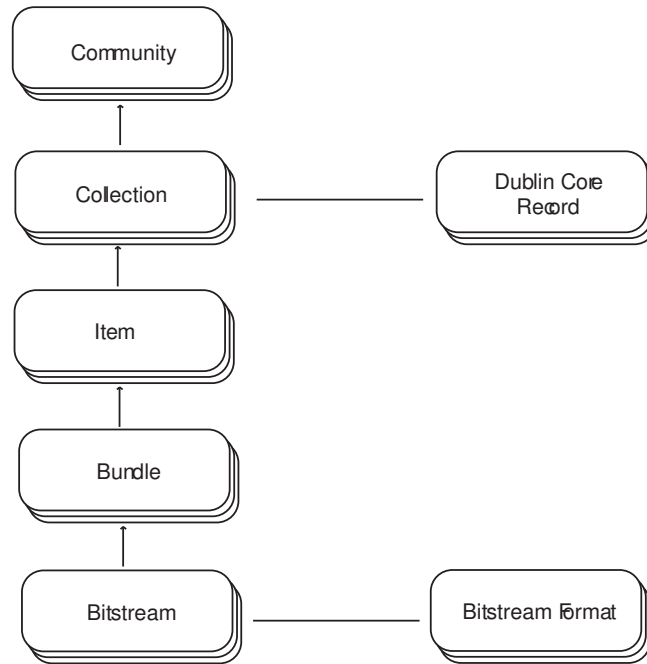
The goal of our project is provide the technology and infrastructure needed to make effective data sharing a reality. The NIH and individual researchers have recognized the multiple benefits to data sharing. Most NIH grant proposals are required to detail how the data generated by the grant would be shared. However, the diversity of computer science skills and networking infrastructure available to most researchers makes effective data sharing problematic. Unlike papers on publisher’s websites, data that is simply posted to a lab’s website is not likely to be well indexed or highly ranked by internet search engines. Furthermore, even if the data is discovered by another researcher, the important data formatting and acquisition information that is needed for that researcher’s project may not have been considered valuable by the researcher who generated the data and therefore may not have been stored with the data.

Two technologies exist that address the difficulties of data sharing: digital libraries and the DICOM image transfer protocol. Next we describe each of these technologies. The remainder of the paper then describes how we have merged them to provide a simple and effective tool for data sharing.

DSpace is an open-source digital library package developed by Hewlett-Packard and MIT for enterprise-level, web-based data dissemination and archiving. Researchers around the world may submit private and public collections of data, publications, presentations, and other electronic media to DSpace archives. A DSpace archive has a top-level community with any number of sub-communities for grouping its data. A community may contain one or more collections (as shown in figure 1) to further refine data groupings. Items are uploaded to a collection, and each uploaded item may contain multiple files (bundles) of arbitrary type, e.g., PDFs of publications, JPEG renderings, and slides from presentations. Over 50 file-types are “known” and their automatic conversion to related file-types is supported. Items may simultaneously exist in multiple collections. Metadata is collected for each item; DSpace uses the Dublin Core metadata standard [4] - it allows for the specification of multiple authors, tracking of items via group/project-specific codes, and referencing an uploaded paper’s appearance in print. By allowing multiple files to be uploaded per item, the data, software, intermediate results, and final analysis used in generating a paper can be grouped and published along with the paper. Furthermore, digital library systems such as DSpace employ the Library of Congress’ web-based handle system for Internet-based searching and cataloging of archived electronic media. These tags are the Internet equivalent of an ISSN. Via handles, digital library data is indexed by Yahoo! and Google and will be accessible even if it is moved. Digital library data, however, is generally only accessible via interactive web sessions and web-based downloading, and using digital library data often involves data format conversion.

Digital Imaging and Communications in Medicine (DICOM) is an image communication standard for medical devices. We have made use of DCMTK [2], an open-source DICOM package for medical image interchange, storage, and printing that was developed at OFFIS, Germany and that implements a large portion of the DICOM standard. DICOM objects (images) contain an extensive set of information regarding the imaging protocol used to acquire the image, sufficient details for nearly any future use of that image. However, DICOM searching is limited to queries that must be directed to a specific DICOM server and that are limited to finding matches based on image acquisition descriptors and patient identifiers. A detailed or research-oriented description of the patient/pathology in the images is generally not integrated into a DICOM object and cannot be used in DICOM searches.

We have merged the strengths of DSpace (grouping data with papers, supporting internet-based searching, and ensuring data persistence) with the strengths of DICOM (downloading images and their acquisition descriptors directly into applications) to facilitate the distribution and use of image data. That merger is discussed next.



**Fig. 1.** DSpace architecture.

## 2 Implementation

In this section we present our current implementation of a “dicom digital library.” The two extensions our implementation provides are (1) the creation of DSpace Media Filters to support common medical imaging file formats and (2) the implementation of a generic, open-source, DICOM file browser.

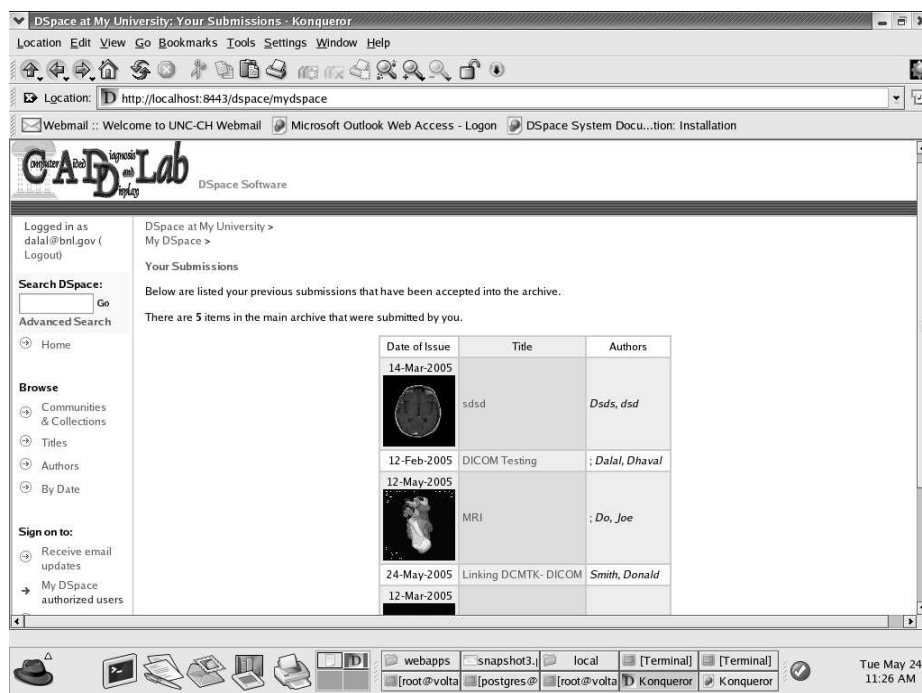
### 2.1 Media Filters

We have extended DSpace to support the anonymization and uploading of DICOM, Analyze, MetaImage and other common image file formats supported by ITK.

DSpace currently supports over 50 different file types for uploading. A “supported” file type has a corresponding media filter that implements the standard for accessing data in that format. With the existence of a media filter, conversion to other known file types is possible, so data in that format can be preserved beyond the useful lifetime of its particular format. For items containing images in supported types, the corresponding media filters produce thumbnail representations for display on the items’ web pages. When an uploaded file type is not recognized/supported, the author is required to enter a description of the format so that others may make use of the data. Such an ad hoc description of

a format, however, may not be sufficiently enabling and will not support the export the information within the file for discovery by search engines or subsequent preservation.

As a 3D image may consist of hundreds of slices, each stored in a separate DICOM object file, our media filters and anonymization methods support the batch processing of multiple DICOM files stored in a single compressed file. Such grouping and compression greatly simplifies uploading data to and the management of a DSpace site. As does ITK, DSpace uses a file's suffix to determine its format. We designate a compressed collection of DICOM files using the suffix ".dgz." If another suffix is used, the uploading process allows a user to manually specify that the file is in the compressed-DICOM format.



**Fig. 2.** Example of output of the new Media Filters. An item's webpage from CADD-Lab's DSpace system. It contains images in the supported DICOM format. The mediafilters automatically generated thumbnail images from the uploaded data.

## 2.2 Linking DSpace/DCMTK

We have implemented methods for linking DSpace with DCMTK's DICOM Server so that 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. Uploads to the DICOM server are disabled since DICOM files do not contain sufficient detail to complete a DSpace database entry. For example, DSpace descriptions should include author, companion papers, diagnosis, treatment, and outcome information. 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's DICOM images available from a DICOM server, those images can be read directly into an image analysis program or downloaded as a batch and automatically converted to another (3D) image file format, as described in the next paragraph.

### 2.3 Query/Retrieve Application

We have developed a cross-platform DICOM query-and-retrieve tool that operates as a stand-alone 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 file-dialog or are saved to disk when used as a stand-alone application. Our application named "FIDicom-QueryChooser" is based on FLTK [6] and DCMTK and is freely available from the InsightApplications package [3].

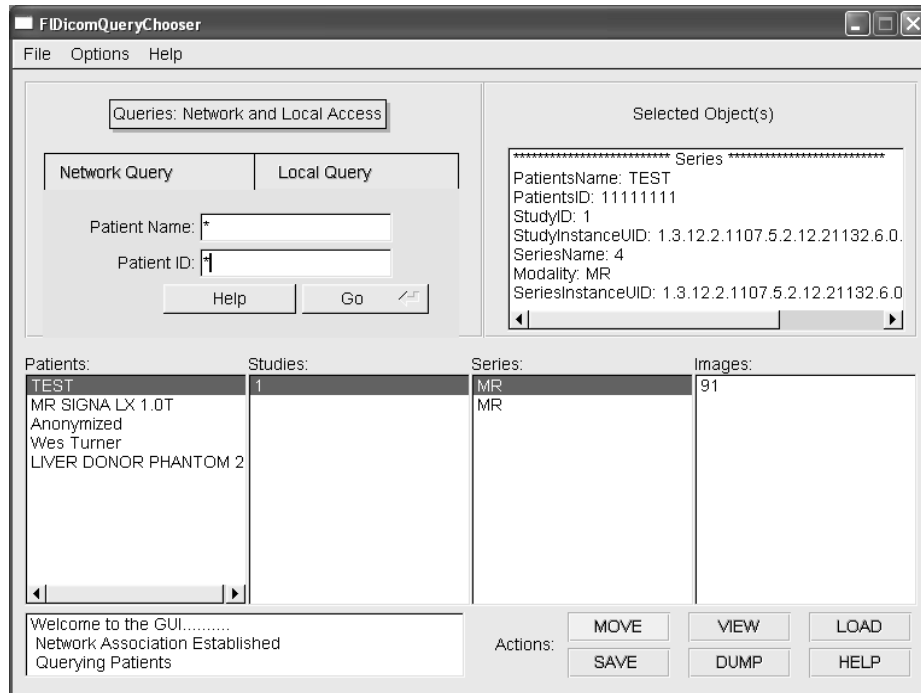
## 3 Discussion and Conclusions

The Computer-Aided Diagnosis and Display Lab (CADDLab) at the University of North Carolina operates a DSpace digital library call the Medical Image Data Archive System (MIDAS). It has been assigned the base handle NA0.1926. Using the presented DSpace/DCMTK extension, DICOM objects uploaded to MIDAS can be:

1. grouped with other files such as journal articles or web pages.
2. viewed as thumbnails.
3. made publicly or privately available for download via a web browser.
4. searched using Google, Yahoo!, and other search engines.
5. downloaded directly into software applications via DICOM query/retrieve application.

This system has proven to be an effective tool for remote collaboration and data sharing within and beyond UNC. It exceeds NIH data sharing requirements.

One demonstration of the effectiveness of MIDAS is the Insight Journal which is hosted by MIDAS. The data sharing and open-science archival capabilities of MIDAS are ideally suited for the open-source focus of the Insight Journal. The IJ's initial application, managing the papers and reviews for the 2005 MICCAI Open-Source Workshop, went very smoothly. Thumbnails were automatically generated for uploads that included image data. Only one of 37 papers experienced uploading difficulties, and those difficulties stemmed from the authors



**Fig. 3.** Main window of the FIDicomQueryChooser. Tabs show results from a network query; local directory browsing is also supported. DICOM objects can be processed individually or as a series batch.

not providing details required by DSpace’s Dublin descriptors; the upload “failure” was actually a feature of DSpace to ensure that all uploads are sufficiently described for their effective indexing and use.

Most importantly, the methods in this paper promote and are available as open-source software. They build upon existing open-source libraries (DCMTK and DSpace) and standards (DICOM and the Open Archives Initiative [5]). They work 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., CTN and Eprints) can also be integrated using the same framework. They also provide much-needed DICOM query-and-retrieve capabilities to ITK. Furthermore, they directly support the open-science sharing of data, software, and results in a sufficiently descriptive manner so that published work can be replicated and science truly advanced.

This work is funded by is NLM Contract, DICOM and Digital Libraries for Digital Archiving, NLM N01 467-MZ-402070.

## References

1. DSpace: A digital repository system <http://www.dspace.org>
2. DCMTK: A DICOM toolkit <http://dicom.offis.de/index.php.en>
3. The insight toolkit: Segmentation and Registration toolkit <http://www.itk.org>
4. Dublin Core Metadata Initiative: <http://dublincore.org/>
5. Open Archives Initiative: <http://www.openarchives.org>
6. The Fast Light Toolkit: <http://www.fltk.org>