

MRIdb: Medical Image Management for Biobank Research

Mark Woodbridge · Gianlorenzo Fagiolo ·
Declan P. O'Regan

Published online: 26 April 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Clinical picture archiving and communications systems provide convenient, efficient access to digital medical images from multiple modalities but can prove challenging to deploy, configure and use. MRIdb is a self-contained image database, particularly suited to the storage and management of magnetic resonance imaging data sets for population phenotyping. It integrates a mature image archival system with an intuitive web-based user interface that provides visualisation and export functionality. In addition, utilities for auditing, data migration and system monitoring are included in a virtual machine image that is easily deployed with minimal configuration. The result is a freely available turnkey solution, designed to support epidemiological and imaging genetics research. It allows the management of patient data sets in a secure, scalable manner without requiring the installation of any bespoke software on end users' workstations. MRIdb is an open-source software, available for download at <http://www3.imperial.ac.uk/bioinfsupport/resources/software/mridb>.

Keywords PACS · Digital imaging and communications in medicine (DICOM) · MR imaging

Background

Population-based epidemiological studies using whole-body magnetic resonance (MR) imaging, such as the UK Biobank

[1], present an opportunity to develop new approaches for quantitative phenotyping. MR data sets may be acquired and archived at multiple sites and then shared in various formats between research groups for analysis. This has created a need for an image database which supports simple but powerful search and retrieval facilities and which can be readily deployed in research facilities by non-specialists. Image management systems implementing the digital imaging and communications in medicine (DICOM) standard are commonly used for their flexibility and robustness; however, they are typically complex applications that present a steep learning curve to new users and can prove difficult to install [2, 3]. These issues can be addressed by presenting a consistent and intuitive interface to users of the database, providing bespoke tools to systems administrators, and including comprehensive yet concise documentation.

Methods

System infrastructure

MRIdb is a software application composed of a bespoke web application and a suite of utility scripts and tools. It depends on a number of other components, including an underlying clinical picture archiving and communications system (PACS), a scalable storage system, an authentication service and a relational database system, as shown in Fig. 1.

The foundation of MRIdb is DCM4CHEE [4, 5], a mature, highly configurable open-source PACS. It handles the vital, low-level functions of image archival from scanners using the DICOM protocol, metadata extraction from images into a relational database schema and raw image and thumbnail retrieval facilities. It natively provides web and DICOM interfaces, but the former is complex and provides extensive administrative and data manipulation facilities, whilst the

M. Woodbridge (✉)
Centre for Integrative Systems Biology and Bioinformatics,
Imperial College London, London SW7 2AZ, UK
e-mail: m.woodbridge@imperial.ac.uk

G. Fagiolo · D. P. O'Regan
Robert Steiner MRI Unit, MRC Clinical Sciences Centre, Imperial
College London, London W12 0NN, UK

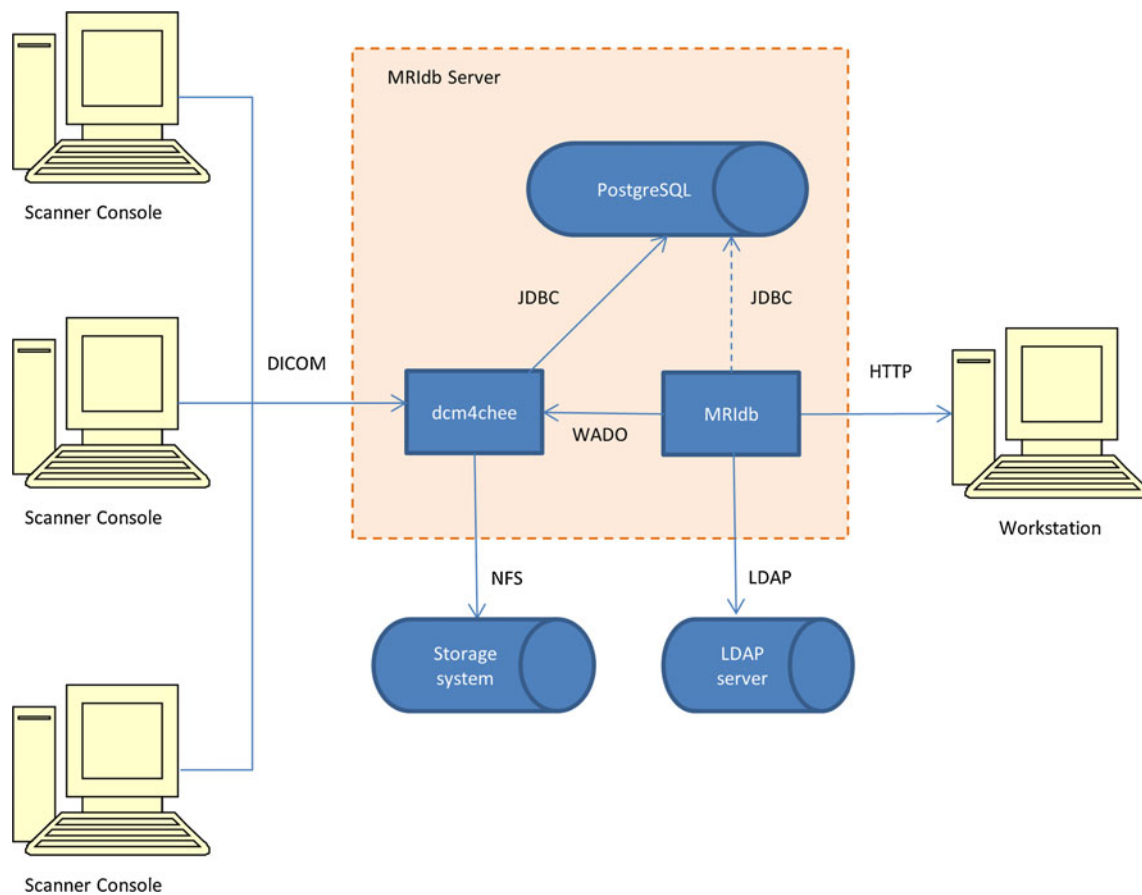


Fig. 1 MRIdb system architecture

latter enables access to unanonymised data. MRIdb’s primary function is, therefore, to provide an alternative, intuitive, read-only interface to image metadata, whilst offering study management and multi-format image export with enforced preservation of data integrity and anonymity.

Security

The MRIdb web application is intended as a simple, secure, centralised portal for users to retrieve images. It requires users to be authenticated using either an institutional lightweight directory access protocol (LDAP) server or a local password database and additionally enforces role-based access control for “visitors,” “researchers,” and “administrators.” Visitors are able to browse images, but it does not show (and does not allow search by) patient names, ages or dates of birth. Researchers are able to view patient information but cannot perform functions such as viewing usage logs or information about other users of the system. Administrators have full access and can create other administrators, as well as configure help material, perform import audits and check authentication and download logs. Importantly, no user, regardless of role, is able to export unanonymised data in any format, and

bulk downloads are password protected by default. The DCMTK [6] toolkit is used for DICOM anonymisation.

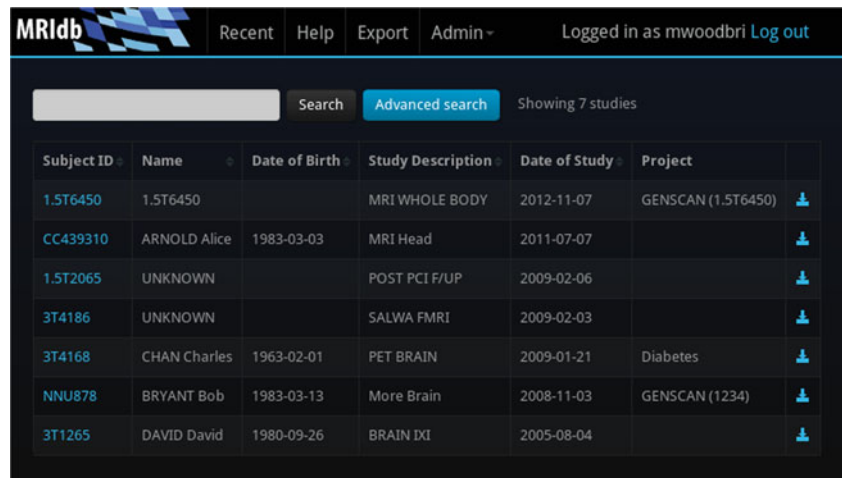
Search

Image search is performed in “simple” or “advanced” mode. The simple interface (Fig. 2) presents a single search box where the input terms are matched against a full-text index of (any combination of) subject identifier and name, study name, project name and participation identifier. The advanced mode allows individual criteria to be specified and adds subject gender and age, series protocol and acquisition date. Results are typically displayed in reverse chronological order by acquisition date but can also be sorted by subject identifier, date of birth or name, study description or import date (for import of legacy data). Studies and series are accompanied by thumbnail images and full metadata.

Retrieval

MRIdb supports several means of retrieving images. Individual studies or series can be downloaded in DICOM, NIfTI or Analyze, format with anonymisation and format conversion

Fig. 2 MRIdb simple search interface



Subject ID	Name	Date of Birth	Study Description	Date of Study	Project	
1.5T6450	1.5T6450		MRI WHOLE BODY	2012-11-07	GENSCAN (1.5T6450)	↓
CC439310	ARNOLD Alice	1983-03-03	MRI Head	2011-07-07		↓
1.5T2065	UNKNOWN		POST PCI F/UP	2009-02-06		↓
3T4186	UNKNOWN		SALWA FMRI	2009-02-03		↓
3T4168	CHAN Charles	1963-02-01	PET BRAIN	2009-01-21	Diabetes	↓
NNU878	BRYANT Bob	1983-03-13	More Brain	2008-11-03	GENSCAN (1234)	↓
3T1265	DAVID David	1980-09-26	BRAIN IXI	2005-08-04		↓

performed automatically. A custom-built tool is used to convert multi-frame series to NIfTI format, with the XMedCon [7] and MRICron [8] toolkits used for other conversions. Regardless of format, the downloads are packaged, compressed and named in a consistent manner, reflecting details of the date of acquisition, protocol, scanner and project identifier (if available). Series can be interactively added to a clipboard, enabling bulk download of scans of interest from various subjects. These downloads are password protected. Administrators are able to perform a non-interactive batch download by uploading a spreadsheet specifying the relevant series. These are automatically downloaded to the researcher's computer by a bespoke tool with progress reported in the web interface.

Visualisation

Image thumbnails are shown on the study preview and series view pages (Fig. 3). Users are able to explore series using the Weasis DICOM viewer [9] or ImageJ image processing tool [10], both of which can be launched from inside MRIdb. They are cross-platform Java applications distributed from the server and therefore do not require software installation before use, and updates are automatically downloaded as required. DICOM images are anonymised and stacked into a format compatible with the standard ImageJ configuration before being transferred.

Study management

MRIdb allows subjects to be assigned to research projects with an optional participation identifier. In order to preserve anonymity, these subject identifiers are not permitted to contain any part of a patient's name or hospital identifier. The identifiers are visible to all users and are used throughout the user interface. They can be searched for in order to retrieve and export subsets of subjects, and images downloaded for offline analysis are named using both project and subject identifiers (when present). Project assignments are shared

between users of the system and the action of assigning or modifying identifiers is logged in the MRIdb audit log. Projects can only be deleted or renamed by administrators.

Administration

MRIdb provides a series of administrative features and tools. For systems administration, these include interruptible data migration and audit scripts to import DICOM files from legacy systems to DCM4CHEE, customized initialisation (init) scripts to automatically start and stop DCM4CHEE and MRIdb as required, and scheduled (cron) jobs to simplify systems monitoring and remove temporary files. The web application logs all user actions, allows bulk upload of system initialisation data (such as user and project records) and optionally reports system errors to a specified user via e-mail. MRIdb is accompanied by a reference guide that provides a full description of how to download, install and configure the system, as well as describing the security, backup and monitor procedures that should be adhered to. It is also possible for administrators to rebrand the system, providing a customised name and logo for a local installation.

Results

MRIdb is an integrated solution for MR clinical research image management. It combines mature open-source software tools for image archival, format conversion and visualisation with a consistent, intuitive, cross-platform user interface and a suite of administrative tools. The user interface is implemented using HyperText Markup Language and is, therefore, platform-independent and accessible using any modern web browser. The server component is written in Java and Python and designed to run on Linux. It is an open source under the GNU General Public License v3.0 [11] and is freely available from the MRIdb website [12] in source and binary form.

Fig. 3 MRIdb series view

A turnkey distribution of MRIdb that can be deployed by researchers without deep technical knowledge is available in the form of a virtual appliance. Usage of this machine image eliminates the lengthy installation process that is common to most PACS as it only requires minimal configuration. This includes specification of the location of the storage space allocated for image archival, the address of the LDAP server used for user authentication and the e-mail address of the system manager (to whom errors are automatically reported). The image is based on a minimal installation of CentOS Linux [13] and can be imported into any virtualisation container supporting the standard open virtualization format (OVF).

The virtual appliance contains a full installation of DCM4CHEE and the PostgreSQL [14] database and does not disable or restrict access to any of its functionality. It therefore provides a simple means to deploy a proven PACS, in addition to the user-friendly visualisation, retrieval and study management options provided by the MRIdb web application. The instance of DCM4CHEE in the MRIdb VM is configured to pre-cache image thumbnails for retrieval via WADO and to index additional series attributes to optimise performance of the bespoke web interface.

Discussion

An evaluation of available free/open-source medical imaging data handling software was undertaken prior to the

development of MRIdb, which itself replaces an existing in-house solution. The available solutions can be divided into three different categories: (1) DICOM data and protocol handling software suites such as the DCMTK toolkit or DCM4CHEE, (2) self-contained or virtualised PACS such as the DCMTB DICOM Toolbox [15] or CDMEDIC [16] and (3) comprehensive image-based clinical research data handling solutions such as the XNAT [17] imaging informatics platform. Given the underlying support that all these systems have for the DICOM protocol, it will be possible to develop integration connectors as required in the future.

XNAT, the most similar solution to MRIdb, is a powerful, flexible system that offers a project-centric storage system. XNAT can sort incoming data according to project specific tags that have to be specified using the scanner terminal during image acquisition. This, in turn, means that a change of data entry procedure has to be enforced. If project tagging is not done at scan time, XNAT still allows the data to be sorted after acquisition by using a temporary data storage facility. Once data are assigned to a project in XNAT, those data are only accessible by the project members. MRIdb is more flexible and open in the sense that the acquisition procedure is not affected, i.e. the data do not need to be tagged at scan time. Once the data are on MRIdb, the data can be accessed by all the research centre members, and they can be also tagged to a specific project. This open and simple approach, adopted in both the data acquisition

process and the user interface, allows users to transfer and retrieve data without having to go through a steep learning curve or change in operating modalities.

MRIDb has been in use at a medium-sized clinical research department of a large university for 6 months. This installation currently contains 7 TB of data and 15,000 studies, primarily from Philips MR scanners, including a large corpus of legacy images as well as scans acquired since its development. It has been reliable, required very minimal system administration and has been well received by its users.

Acknowledgments This work was funded by the Medical Research Council, UK, the Imperial NIHR Biomedical Research Centre, UK, and a British Heart Foundation, UK, project grant (PG/12/27/29489).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. UK Biobank <http://www.ukbiobank.ac.uk/> Last viewed January 2013
2. Paré G, Trudel MC: Knowledge barriers to PACS adoption and implementation in hospitals. *Int J Med Info* 76(1):22–33, 2007
3. Honea R: How many people does it take to operate a picture archiving and communication system? *Journal of Digital Imaging* 14(1):40–43, 2001
4. DCM4CHEE <http://www.dcm4che.org/confluence/display/ee2/Home> Last viewed January 2013
5. Warnock MJ, et al: Benefits of using the DCM4CHE DICOM archive. *Journal of Digital Imaging* 20:125–129, 2007
6. Eichelberg M, et al.: Ten years of medical imaging standardization and prototypical implementation: the DICOM standard and the OFFIS DICOM Toolkit (DCMTK). *Medical Imaging 2004*. International Society for Optics and Photonics, 2004
7. Nolf E: XMC: An open-source medical image conversion toolkit. *Eur J Nucl Med Mol Imaging* 30(Issue 2 Supplement):246, 2003
8. MRICron <http://www.mccauslandcenter.sc.edu/mricro/mricron/> Last viewed January 2013
9. Weasis <http://www.dcm4che.org/confluence/display/WEA/Home> Last viewed January 2013
10. Schneider CA, Rasband WS, Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9:671–675, 2012
11. The GNU General Public License v3.0 <http://www.gnu.org/licenses/gpl.html> Last viewed January 2013
12. MRIDb <http://www3.imperial.ac.uk/bioinfsupport/resources/software/mridb> Last viewed January 2013
13. CentOS <http://centos.org/> Last viewed January 2013
14. PostgreSQL <http://www.postgresql.org/> Last viewed January 2013
15. Langer S, Charboneau N, French T: DCMTB: a virtual appliance DICOM toolbox. *Journal of Digital Imaging* 23(6):681–688, 2010
16. CDMEDIC http://cdmedicpacsweb.sourceforge.net/CDMEDIC_PACS_WEB.html Last viewed January 2013
17. Marcus DS et al.: XNAT: a software framework for managing neuroimaging laboratory data. *Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping*, Florence, 2006