# A Computational Platform for Radiology Image Retrieval and Analysis

Jarrel Seah<sup>1</sup>, Christopher Bain<sup>2</sup>, Christopher Mac Manus<sup>3</sup>, and Helen Kavnoudias<sup>4</sup>

Abstract— Many stakeholders in healthcare, and particularly in hospitals, have needs around the secondary use of radiology images which have been generated in routine care processes. We have previously constructed a large informatics platform at a major teaching hospital through the aggregation of massive and disparate data sets, and have reported on this work. This has been a strategic initiative in order to serve multiple needs across the organization including across the domains of operations, quality improvement and research—all of which are critical to the functioning of a large academic health centre. In this paper we describe the design, construction and operational utilisation of an image retrieval and analysis system (IAS). This leverages off the underlying informatics platform, and allows users to retrieve large numbers of DICOM images for quality assurance, audit or research purposes. We believe this to be a unique initiative in the Australian context.

*Keywords*— radiology, image, analysis, big data, hospital, DICOM, REASON.

### I. INTRODUCTION

In this paper, we outline the development and utilization of a mechanism designed to transfer high volumes of radiology scans and metadata in the Digital Imaging and Communications in Medicine (DICOM) [1] format, from existing production Picture Archiving and Communication Systems (PACS) to server storage for secondary use. This system mainly supports research involving external organizations where data has to be securely transferred, or research where computational analysis of images is undertaken.

Other possible uses for such a system include for teaching purposes, or the storage of images that would not otherwise be stored, for example the raw data provided from radiological modalities such as Computerized Tomography (CT) scanners. When linked with other systems such as the Alfred Health (AH) REASON Cohort Discovery Tool (CDT) [2] we have developed, interesting images of rare conditions can also be easily retrieved and stored for these purposes.

For example, a pilot project involving the computational analysis of cervical spine (neck) CT scans in trauma patients was proposed. The requisite cohort was trauma patients with and without a diagnosis of cervical spine pathology, with the associated cervical CT scans in that admission. About 2500 eligible patients fitting these criteria were found in the main data platform at our health service - the REASON platform. No satisfactory mechanism existed to securely and efficiently transfer DICOM data for such a large volume of scans, necessitating the development of novel infrastructure, namely the AH REASON Image Analysis System (IAS).

#### II. CONTEXT

## A. The Business Need

The pilot project on trauma patients is indicative of a burgeoning area of research that needs to draw on masses of image data for it to be successful. As the volume and quality of radiology scans increase, automation will be increasingly relied upon to improve the efficiency of scan interpretation by expert human radiologists. Computer aided diagnosis (CAD) [3] is a field which is ripe for development in radiology. Some sample papers which examine this field include work by Zhang et al [4] and Viswanath et al [5]. These papers focus upon CAD of lesions on radiological modalities which require large datasets for training and testing. Machine learning approaches, which have become popular in image analysis in recent years with the re-discovery of recurrent neural networks and stacked restricted Boltzmann machines, also require an extraordinarily large amount of data to perform well.

A system like the IAS is even more critical to research performed in a health service such as our one as it provides statewide referral services in the areas of adult trauma, burns and organ transplantation, creating a unique set of patients with interesting imaging that can often be used for research or teaching purposes. In fact, the range of services provided by this hospital was the impetus for developing the proposed project at this site, as the volume of cervical spine pathology would exceed that of most other health services. Along with the CDT that we have created, the IAS enriches the data sets provided to researchers and clinicians by serving up pictorial data in addition to structured and unstructured text data.

#### B. The Computational Context

As previously described in the literature [6], we have constructed a platform in a "big data" paradigm called The REASON Discovery Platform ®. As part of this platform, the IAS was created to handle the transfer of DICOM data using accession numbers or patient identifiers to select the required scans.

Jarrel Seah<sup>1</sup> is a final year medical student studying at Monash University, Melbourne, Australia. (e-mail: jarrelscy@gmail.com).

Christopher Bain<sup>2</sup> is with Alfred Health and Monash University,

Melbourne. (+61 3 9076 3079; e-mail: Christopher.bain-info@alfred.org.au). Christopher Mac Manus<sup>3</sup>, is with Alfred Health, Melbourne. (+61 3 9076 5433; email: C.Macmanus@alfred.org.au).

Helen Kavnoudias<sup>4</sup>, is with Alfred Health, Melbourne. (+61 3 9076 3606 email: h.kavnoudias@alfred.org.au)

With these increasing demands for images and labelled data, an efficient method of transferring this data must be made. Images are have traditionally been transferred between institutions using physical media, such as CDs and DVDs, with an inbuilt viewer into each one. This consumes a large amount of resources and is highly impractical - for example, 2,500 DVDs, one for each patient, would be required in this pilot project. These methods are also highly insecure and are also slow, as the media must be physically transported to the receiver, and then "the data" re-entered into their local system. In conclusion, there is a pressing need for a system that can, with little or no human input, provide a dataset of labelled, anonymized scans for research and other purposes — a need that will only increase with the growth of CAD in the field of radiology.

#### III. SYSTEM OVERVIEW

As outlined above, the aim of this work was to establish a prototype electronic access system via the REASON Platform [6], whereby researchers can acquire and store diagnostic imaging files from the PACS for subsequent analysis. Having established that such a service can be created, the further aim of this work was to explore options for how it can be embedded for ongoing use in the health service.



Fig. 1 DICOM files viewed in the MicroDICOM viewer

In the DICOM hierarchy, each radiology scan or test consists of multiple series, each containing multiple images. Each scan is associated with a unique accession number which can be retrieved by a range of means – for example by requesting the relevant accession number or numbers from the Radiology Department (RD), or by searching their systems. Put simply, the aim of developing the IAS system was to allow a user to request DICOM files attached to certain studies on certain patients, from the source system – the PACS – without the need for human intervention.

The subsequent requirement of the system was to allow these selected files to be transferred to another form of storage from where they could be viewed – eg – with a DICOM viewer like MicroDICOM (see Figure 1), or accessed in some other way – eg – by specific computer code for the purposes of computational analysis.

#### IV. METHODOLOGY

### A. System Architecture

The hospital PACS administrator arranged for a specific DICOM server to be a point of contact for the IAS, to be utilized between 00:00 and 07:00 hours daily, to limit the load upon the network.

The overall architecture of the system can be seen in Figure 2. But is as follows: a Windows 2008 Server R2 machine with a dual core 3.4Ghz processor, 4GB of RAM and Broadcom Gigabit NetXtreme connection was utilized for this purpose to provide an Application Programming Interface (API) as well as a Hyper-text Transfer Protocol (HTTP) [7] front-end to enable the simple retrieval of radiological examinations via the DICOM communication standard from PACS.

Authentication with Active Directory (AD) is provided to restrict access as well as to provide network sharing capabilities where users can access the required files via a network share on the IAS. Email notifications are also enabled as image retrieval often takes a long time due to the large data transfers involved. This means that users can go about their daily work and only come back to the IAS when they are notified by email that their request has been filled, and the desired images are ready to be viewed.

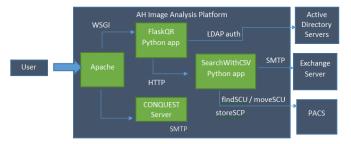


Fig. 2 System Architecture

Technically, the platform comprises of a Python server created in Flask, which communicates with the PACS servers via an open source library known as DICOM Toolkit (DCMTK). Push requests are sent to the PACS servers, and the ConQuest DICOM server is used to receive the images. An APACHE server is used to communicate via a Web Server Gateway Interface (WSGI) with both the ConQuest server and the Python Flask app.

The usage of the platform is straightforward – users first authenticate using their Lightweight Directory Access Protocol (LDAP) [8] username/password pair. They then have the option of uploading a Comma Separated Value (CSV) file containing a list of accession numbers and patient identifier numbers that correspond to the desired images, or manually typing in those numbers on the webpage itself. The webpage (see Figure 3) is JavaScript enabled to automatically add more input boxes if required, improving the usability of the system. Users are also given the option to anonymise their scans, an option which will be mandatory for users with a lower authentication level. This anonymisation function is critical for research applications of the system, but not so for quality assurance, audit or teaching applications.

By anonymising the scans, all identifiable metadata is stripped from DICOM files and the files themselves are renamed. A master sheet is created such that the files can be re-identified if need be. When the request is completed, an email is sent to users notifying them and providing them with a Server Message Block (SMB) link to a file location (see Figure 4) which is protected by Access Control List (ACL) permissions, limiting access to only the user who requested the data. Users are also limited to read-only permissions, preventing them from altering data on the server.

# Browse for a CSV file to upload:



# Alternatively enter numbers here

Accession number:	18247231	Patient ID: 27465632
Accession number:	93846125	Patient ID: 29573635
Accession number:	19284766	Patient ID: 29375651s
Accession number:		Patient ID:
Find patients Anonymize		

Fig. 3 System web page interface

# B. System Development Approach

The approach used in this work has been one of evolutionary prototyping underpinned by a heavily user-centred design philosophy, as has been the case for much of our work. The department leading the work (the Information Services Department (ISD)) is a heavy user and provider of data itself, and so has been acting as test users of the evolving system along with the developer, and whilst guiding next steps through each iteration of development. This fits with broader mission of the department in providing better access to data across the organization. The other key stakeholder group with input to the design and construction of the system was the RD.

# C. System Development Approach

In the context of the proposed pilot project, this system was used in conjunction with the CDT. The CDT was used to retrieve all patients admitted via the Emergency Department (ED) for trauma, who also had a CT scan of the cervical spine. Their respective diagnoses were also retrieved using the system, classifying them by cervical level (from cervical vertebrae 1 through 7 as well as occipital condyle injuries) as well as the type of injury (dislocation/subluxation vs fracture). Using this information as the gold standard, the images provided were able to be used by the project for training as well as testing.

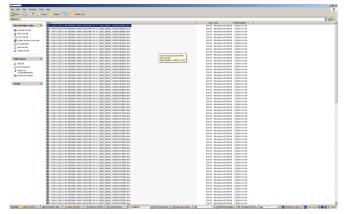


Fig. 4 Individual DICOM files after transport into a restricted access folder

#### V. OUTCOMES

Accession numbers were retrieved from the CDT and were assembled into a CSV file where they were uploaded over the course of multiple nights to the system. Retrieval speed was adequate, retrieving up to 1000 scans per 7 hours.

The large number of series in each scan proved to be a bottleneck and in subsequent retrievals over the course of this project, the type of series was hardcoded in to the retrieval process (only sagittal (midline view) reformats of soft tissue were extracted. See Figure 5). A future improvement would be to expose this setting to end users to reduce the waiting time for scan sets to be downloaded.

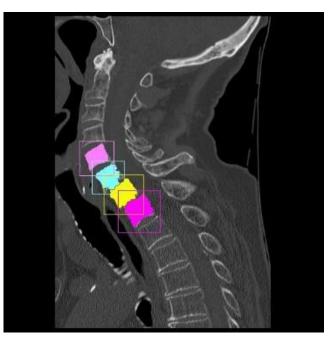


Fig. 4 Computer enhanced images of anatomical structures in the cervical spine (neck) – sagittal (midline) view

Sagittal reformats were used by the project to perform 3D reconstruction and vertebrae identification and segmentation. The segmented vertebrae were used to calculate widely used metrics in radiology such as vertebral width, height and prevertebral soft tissue thickness. These metrics closely

correlated with the limited human measured sample, as well as literature reported averages [9].

The system performed well in this pilot project, and no errors or disruption of normal services to the hospital occurred. The project's intermediate outcomes showed that as per recent literature in the medical space, pre-vertebral soft tissue thickness, a traditional marker of injury to the cervical spine, is not sensitive enough to exclude all cervical injuries.

#### VI. THE FUTURE

## A. Technical Developments

One of the limitations of the prototype that was soon discovered was the large storage volume likely to be required for concurrent projects or usage. The cases acquired in the test project were enough to nearly fill the 300 GB of allocated server storage. An option currently being explored to allow transition to a routine, sustainable service, is the use of "rentable" capacity in a private cloud so that capacity can be expanded to suit, and paid for, by individual projects, once base level capacity is exceeded. An automatic compression routine could also help bypass this bottleneck.

Much future work remains to improve the usability of the platform. For example, interfacing the software with other tools such as the CDT would help to retrieve images without explicitly defining accession numbers or patient identification numbers – e.g. to help retrieve all images of patients with a diagnosis of pneumonia without having to explicitly identify each of the individuals in that cohort if that was not otherwise needed.

Automatic scheduling and load balancing will also need to be considered in scaling up the system, as high volume queries should only occur at quiet hours to prevent excessive network usage. Furthermore, queries can be prioritized and cached to improve the efficiency of the system. Automatic scheduling can be useful for prospective studies which wish to capture scans forward of a particular time. Further work also includes automatic removal of expired data and allowing users to set an expiration date on their images, as doing so will prevent the system from using excessive memory.

Currently, the system only considers one server as the point of contact, and does not interact with modalities and scanners directly. By adding these connections, the system could provide storage for non-clinical purposes at a much cheaper rate than would normally occur – meaning that high fidelity data could be captured and retained prospectively rather than relying on compressed data that is currently stored in PACS due to storage constraints. So for example, "raw" data could be pulled direct from CT or magnetic resonance imaging (MRI) machines.

Security hardening and penetration testing should also be considered, as the authentication process via cookie handling may be prone to tampering and may lead to unauthorized access of the system if not careful. Although access to the IAS is currently restricted to intranet-only access, this is still not an ideal scenario with wider future use in mind.

# B. Business Usage

The ability to extract image data directly from DICOM data or convert it to other forms of 3D data, would also be useful for researchers approaching the problem from a non-medical perspective, as they may be unfamiliar with the DICOM standard.

There are also other future use cases that the IAS may be able to support. One key example is the acquisition and storage of digital pathology images. There is evidence that historic physical storage processes for the various kinds of pathology images (eg – microscopic tissue samples from excised tumours) are progressively moving to the digital realm [10]. Equally, clinical photography images such as those of eye diseases [11] or the changes in skin wounds as they heal or worsen [12], could be shipped en-masse and stored for analysis, in an approach similar to that used in this case.

#### VII. CONCLUSION

We believe that this initiative is the first of its kind in an Australian hospital. Whilst intended for small scale use at this stage, the service is potentially quite scalable to large scale production use, so it can support many simultaneous projects and researchers, be they medical, science or technology researchers. As part of looking to the future, the issues around storing large volumes of image data in a private cloud are being explored.

This pilot project usage of the AH REASON IAS demonstrates that it is feasible to use this as a pipeline for transmitting DICOM data for more than viewing purposes. Novel computational approaches to radiology often necessitate the transfer of DICOM data between computers without human intervention, or between institutions. In order to facilitate this, alternatives such as the IAS offer increased speed and ease of use over traditional approaches such as burning CDs/DVDs.

Security and confidentiality considerations are also key in the healthcare space. The DICOM standard provides for authentication and encryption via the Advanced Encryption Standard (AES), which can also be added to the system to prevent man-in-the-middle attacks in the future, ensuring that end users are properly authorized to access sensitive information.

# ACKNOWLEDGMENT

We would like to acknowledge the contributions of the following ISD staff members to this work: Bismi Jomon and Annie Gilbert. In addition we would like to acknowledge Professor Ken Thomson and Ms Jo Chidley from Radiology, and Mr Joel Macatangay from the IT Infrastructure team, for their important contributions to this work.

### REFERENCES

- [1] DICOM Website http://dicom.nema.org/ Accessed 26/2/2015.
- [2] Bain, C., MacManus, C. and Seah, J. (2014). Web based cohort identification across large healthcare data sets – opening the treasure chest. Proceedings of ICDMME 2015, Bangkok, Thailand. April 2015. (In press).
- [3] Salehi, L., & Azmi, R. (2014). A novel method based on learning automata for automatic lesion detection in breast magnetic resonance imaging. Journal of Medical Signals and Sensors, 4(3), 202–10. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4187355&to ol=pmcentrez&rendertype=abstract.

- [4] Zhang Z, Srivastava R, Liu H, Chen X, Duan L, Wong DWK, Kwoh CK, Wong TY and Liu J. (2014) A survey on computer aided diagnosis for ocular diseases BMC Medical Informatics and Decision Making, 14:80.
  - http://dx.doi.org/10.1186/1472-6947-14-80
- [5] Viswanath, S., Bloch, B. N., Rosen, M., Chappelow, J., Toth, R., Rofsky, N. and Madabhushi, A. (2009). Integrating Structural and Functional Imaging for Computer Assisted Detection of Prostate Cancer on Multi-Protocol In Vivo 3 Tesla MRI. Proceedings Society of Photo-Optical Instrumentation Engineers, 7260, 72603I. doi:10.1117/12.811899. http://dx.doi.org/10.1117/12.811899
- [6] Bain, C., & MacManus, C. (2014). Advancing data management and usage in a major Australian health service: The REASON Discovery Platform TM. In Proceedings of the International Conference on Data Science and Engineering. Cochin, India.
- [7] W3C Website HTTP. http://www.w3.org/Protocols/ Accessed 26/2/2015.
- [8] Microsoft Developer Website: LDAP https://msdn.microsoft.com/enus/library/aa367008%28v=vs.85%29.aspx Accessed 26/2/2015.
- [9] Seah J. (2014) Computer Assisted Diagnosis of Cervical Spine Fractures in Computed Tomography Scans of Trauma Patients. Monash University, BMedSci Thesis.
- [10] Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL, Kuzmishin JA, Montalto MC and Parwani AV (2014). Can Digital Pathology Result In Cost Savings? A Financial Projection For Digital Pathology Implementation at a Large Integrated Health Care Organization. Journal of Pathology Informatics, 5:33. http://dx.doi.org/10.4103/2153-3539.139714
- [11] Mitry, D; Peto, T; Hayat, S; Blows, P; Morgan, J; Khaw, KT, and Foster, PJ. (2015). Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. PLoS One. Feb 18;10(2):e0117401.
- [12] Landa, DL; van Dishoeck, AM; Steverberg, EW and Hovius, SE. (2014) Quality of measurements of acute surgical and traumatic wounds using a digital wound-analysing tool. Int Wound J. Aug 14. doi: 10.1111/iwj.12330 http://dx.doi.org/10.1111/iwj.12330.
- **Jarrel S.** is a final year medical student at Monash University in Melbourne, Australia. In addition he has completed a Bachelor of Medical Science on computational diagnostic support of cervical spine injuries, and is an accomplished software developer, having developed the internationally known and award winning Eyenaemia app. (www.eyenaemia.com)
- **Chris B.** is the Director of the Information Services Department at Alfred Health in Melbourne, and was the inaugural Director of Health Informatics there. In addition he is an Adjunct Associate Professor in the Faculty of IT at Monash University also in Melbourne. Chris has professional IT accreditation with the ACM in the US and the ACS in Australia. He is also an Associate Fellow of the Australasian College of Health Service Management and a Fellow of the Australasian College of Health Informatics.
- **Chris M.** is the Manager of Data Services in the Information Services Department at Alfred Health, and is currently completing a Master of Public Health. Chris is a Computer Scientist and Software Engineer by training and has been working in related roles for 12 years.
- **Helen K.** is the Manager of Radiology Research at Alfred Health. She is a widely published, PhD qualified researcher in her own right, and leads the major research projects run out of the Radiology Department at the health service.