International Journal of Population Data Science

Journal Website: www.ijpds.org

Secure provision of unconsented radiological imaging for research

Mcallister, Gordon¹

¹University of Dundee

Objectives

- Design and implement an architecture for managing unconsented DICOM imaging
- Maintain sufficient data to define research cohorts when data quality is unknown
- Perform project-level linkage and extraction into a Safe Haven (SH) environment
- Extract large image volumes for multiple projects with limited storage constraints
- Provide applications for an imaging research workflow within the SH environment
- Serve as a prototype for the Farr/NHS Scotland project to create a research dataset from Scotland's national PACS

Approach

The software architecture builds on the Research Data Management Platform (RDMP) developed at Dundee's Health Informatics Centre (HIC) within Farr@Dundee. The RDMP provides core services common to loading any dataset, with configuration and extensibility points for dataset-specific implementations. This architecture augments the RDMP with scalable micro-services performing peripheral functions.

Images are sourced from the local PACS server in Ninewells Hospital and cached securely within HIC using an implementation for the RDMP with a custom server to query/retrieve data.

Data stored in the catalogue should be anonymous, according to the Scottish SH model. The imaging dataset is poorly understood, with several potentially identifiable free-text fields which may contain information required for defining suitable research cohorts. The load process only permits verified metadata fields into the anonymised catalogue; a Mongo database stores other data for later analysis, should a field subsequently be required for cohort definition.

A DICOM extraction implementation is provided, using DICOM Confidential for anonymisation and a project-specific remapping of DICOM GUIDs.

Two provisioning methods have been designed. A basic copy when sufficient storage is available, and a more sophisticated method using a custom filesystem to provide separate project-specific views onto shared image files.

Results

A full end-to-end solution has been developed, from initial caching through to provisioning anonymised images. Two imaging cohorts have been loaded, one with over 5000 studies. NHS Tayside CT and MR data since 2008 is currently being loaded.

Two projects have had anonymised extracts released using the 'copy' method. The custom filesystem method has been developed and tested with limited amounts of data.

This work has highlighted anonymisation, cohort creation and SH issues which require further exploration.

Conclusion

A production system for securely providing linked DICOM imaging to researchers has been implemented, serving as a testbed for a national system which will provide a unique population-level resource for researchers.

*Corresponding Author:

Email Address: g.i.mcallister@dundee.ac.uk (G. Mcallister)