NeuroImage 144 (2017) 287–293

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

The Washington University Central Neuroimaging Data Archive☆

Jenny Gurney [a], Timothy Olsen [b], John Flavin [a], Mohana Ramaratnam [c], Kevin Archie [d], James Ransford [e], Rick Herrick [a], Lauren Wallace [a], Jeanette Cline [a], Will Horton [a], Daniel S. Marcus [a]

[a] Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[b] Deck5 Consulting, Normal, IL, USA
[c] NRG India, Pune, Maharashtra, India
[d] DICOM Grid, Phoenix, AZ, USA
[e] Department of Radiology and Radiological Sciences, Vanderbilt University, Nashville, TN, USA

ARTICLE INFO

Article history:
Accepted 28 September 2015
Available online 9 October 2015

Keywords:
Central Neuroimaging Data Archive
CNDA
Washington University
Imaging
Informatics
Neuroinformatics
Data
Sharing
Database
Archive
Multi-site
Multi-center
De-identify
Anonymization
Pipeline
JAAT
XNAT
Alzheimer's disease
Brain metabolism
Controls
Cancer
HIV
Sickle-cell anemia
Tourette syndrome

ABSTRACT

Since the early 2000's, much of the neuroimaging work at Washington University (WU) has been facilitated by the Central Neuroimaging Data Archive (CNDA), an XNAT-based imaging informatics system. The CNDA is uniquely related to XNAT, as it served as the original codebase for the XNAT open source platform. The CNDA hosts data acquired in over 1000 research studies, encompassing 36,000 subjects and more than 60,000 imaging sessions. Most imaging modalities used in modern human research are represented in the CNDA, including magnetic resonance (MR), positron emission tomography (PET), computed tomography (CT), nuclear medicine (NM), computed radiography (CR), digital radiography (DX), and ultrasound (US). However, the majority of the imaging data in the CNDA are MR and PET of the human brain. Currently, about 20% of the total imaging data in the CNDA is available by request to external researchers. CNDA's available data includes large sets of imaging sessions and in some cases clinical, psychometric, tissue, or genetic data acquired in the study of Alzheimer's disease, brain metabolism, cancer, HIV, sickle cell anemia, and Tourette syndrome.

Introduction

Washington University has a long and rich history of neuroimaging research, including seminal contributions in the development of positron emission tomography (PET), fMRI, resting state fMRI, and human connectomics (Raichle, 1998; Snyder and Raichle, 2012; Van Essen et al., 2013). Since the early 2000's, much of this work has been facilitated by the Central Neuroimaging Data Archive (CNDA), an XNAT-based imaging informatics system. The CNDA is uniquely related to XNAT, as it served as the original codebase for the XNAT open source platform.

☆ Author note funded in part by the Neuroimaging Informatics and Analysis Center (Principal investigator: Daniel S. Marcus; P30 NS048056), the Dominantly Inherited Alzheimer Network (Principal investigator: John C. Morris; U01 AG032438), ACS, Healthy Aging and Senile Dementia (Principal investigator: Tammie Benzinger; 5P01AG00399130), the Amyloid Imaging in the Adult Children Study (Principal investigator: John C. Morris; 5P01AG026276-02), Washington University's Intellectual and Developmental Disabilities Research Center (Principal investigator: Terrie E. Inder; 1P30HD062171), and the XNAT Imaging Informatics Platform (Principal investigator: Daniel S. Marcus; 5R01EB009352).

E-mail address: gurneyj@mir.wustl.edu (J. Gurney).

http://dx.doi.org/10.1016/j.neuroimage.2015.09.060
1053-8119/

History

The CNDA was initiated in 2002 to manage the Washington University (WU) Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight ADRC) imaging data. It was designed to be both an active study management tool as well as a long term imaging archive. The system was implemented as an XML database that provided a streamlined mechanism for developers to implement new clinical and behavioral data types. Its rudimentary imaging workflow depended largely on FTP-based data transfer and a semi-structured directory and file naming convention. A web application provided basic data navigation and reporting. By 2004, the CNDA had evolved into a Java web application with a unique hybrid XML/relational database. With its increasing sophistication, additional WU research studies began relying on the CNDA, and its core software was extracted into the generalized open source Extensible Neuroimaging Archive Toolkit (XNAT). XNAT is now widely used within neuroimaging (see reports in this issue by Alpert et al. and Yvernault et al. for example systems) and the broader imaging community (Doran et al., 2012; Gao et al., 2013; Winslow et al., 2011). For several years, the CNDA and XNAT continued to grow in lock step, with the XNAT development team also functioning as the development, operations, and user support team for the CNDA. Support for DICOM was incorporated by 2006 and project-based access control was developed soon after. Since 2012, the CNDA has served as the official repository for all research imaging at Washington University with automated archiving of scans obtained on all the University's standard dedicated research scanners (see Hodge 2016, for management of data collected by the Human Connectome Project). In addition to hosting WU-generated data, the CNDA also supports a number of multi-center studies, including the Dominantly Inherited Alzheimer Network (DIAN) (Morris et al., 2012) and the INTRuST Post-traumatic Stress Disorder/Traumatic Brain Injury Clinical Consortium (http://intrust.sdsc.edu). The CNDA continues to serve as a primary test bed for ongoing XNAT development. Most recently, scripting tools have been incorporated into the CNDA infrastructure to enable de-identification and import of large batches of retrospective patient data from the WU Radiology Department's diagnostic division.

Current services

The CNDA provides investigators with a permanent archive for imaging data. All data is securely stored on a modern ZFS file system with a full offsite backup. New research imaging sessions acquired at Washington University can either be sent directly to the CNDA from the scanner or uploaded through the website using a Java-based upload wizard. To ensure secure transmission of all imaging sessions, outside institutions are restricted to uploads through the CNDA website unless specialized DICOM relay hardware has been installed at their scan facility.

All data in the CNDA is organized into projects. A project can be marked as belonging to one or multiple investigators. Generally one project is created for each new study, except in the case of multi-center studies, for which one project is often created for each institution contributing data.

Researchers can configure their projects to automatically apply session anonymization to scrub PHI or recode their DICOM data before it is permanently stored. As the data is archived, CNDA extracts and stores to the database a limited set of information about the session, including time and date of scan, patient name, patient id, and the scanner make and model, which can be used later for searching or generating reports.

Automated imaging data analysis is implemented in CNDA using XNAT pipelines (Table 1). Investigators can add a pipeline to a project by submitting a short form on the pipeline configuration tab to provide default processing variable values for their images. Currently, the processing pipelines available to researchers in the CNDA are all neuroimaging-based. Data analysis pipelines are available for structural and functional MR as well as for PET. Also available is a set of utilitarian pipelines for tasks like atlas registration, DICOM conversion, structural MR defacing and quality control.

The CNDA offers a variety of services to support multi-center projects. Access to data can be restricted by site for study coordinators and site investigators, while the overall investigator and administrative core are provided with a pooled view of all the sites' data. To enable image data transfer directly from an external site's scanner, the CNDA can provide a relay device, a Mini PC running RSNA Clinical Trial Processor (CTP) (http://mircwiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor) which receives images directly from a scanner via the unsecure DICOM protocol and forwards them to the CNDA over secure HTTPS protocol, while also de-identifying and recoding the image headers. CNDA's Protocol Validation pipeline and Manual QC form can help to ensure that study imaging quality is maintained.

For researchers conducting retrospective studies, the CNDA provides an automated way to retrieve clinical scans from the Washington University Medical Center's clinical image archive (often referred to as a PACS). The Joint Anonymization and Archive Tool (JAAT) (http://nrg.wustl.edu/software/jaat) simplifies the transfer of DICOM for the researcher, shifting the download and de-identification responsibilities to CNDA staff (Fig. 1). Once IRB approval for the study has been established, the investigator only needs to provide a spreadsheet specifying the patient exams to be pulled from the clinical PACS, the CNDA subject and session labels to be used in recoding and the researcher's project. To date, the CNDA has retrieved close to 5000 clinical scans using JAAT. A project currently underway will allow qualified CNDA users to deploy the JAAT at remote institutions and securely transfer anonymized patient exams to the CNDA.

The CNDA provides numerous routes for the upload of non-imaging data. The DIAN and ADRC projects regularly retrieve and store data into the CNDA using automated custom tools built upon the XNAT RESTful

Table 1
CNDA pipelines.

Pipeline	Description	Input
FreeSurfer v5.1	Runs FreeSurfer v5.1 recon-all with qcache	MR with T1
FreeSurfer v5.3	Runs FreeSurfer v5.3 recon-all with configurable parameters	MR with T1 Optional: Flair or other T2
Generic bold preprocessing	Prepares (BOLD) images for seed-based data analysis	MR with BOLD, T1 and one of the following: T1W, TSE, or PDT2 Optional: resting state scans
Benice	Computes the Resting State Networks from BOLD studies for pre-surgical use	MR with BOLD and T1
HOF	Prepares sessions from clinical tumor protocol studies for ROI analysis	MR with T1, T2 and DWI
PET unified pipeline	Generates binding potentials for regions of interest (ROI)	PET Recommended: FreeSurfer or hand-drawn ROIs
Register MR to Atlas	Registers MR scans to specified atlas	MR with T1
Dicom to Nifti	Uses dcm2nii to convert some or all DICOM series to NIfTI	Any neuroimaging DICOM of modalities MR, PET, CT or SPECT
Face Masking	De-identifies a subject's face by blurring recognizable facial characteristics	MR with T1 or T2
Protocol validation	Checks the experiment for pre-defined project acquisition parameters	Any DICOM session
FBIRN phantom	Creates Automated Phantom QA data using 64bit version 1.9.5 of BIRN tools	MR with BOLD acquired on fBIRN phantom