Published in final edited form as:

Neuroimage. 2016 January 1; 124(0 0): 1097–1101. doi:10.1016/j.neuroimage.2015.05.021.

Vanderbilt University Institute of Imaging Science Center for Computational Imaging XNAT: A multimodal data archive and processing environment

Robert L. Harrigan^{a,*}, Benjamin C. Yvernault^a, Brian D. Boyd^e, Stephen M. Damon^a, Kyla David Gibney^b, Benjamin N. Conrad^{b,c}, Nicholas S. Phillips^{b,c}, Baxter P. Rogers^{b,c,d,e}, Yurui Gao^{b,d}, and Bennett A. Landman^{a,b,c,d}

^a Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^b Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA, 37235

^c Radiology and Radiological Sciences, Vanderbilt University, Nashville, TN, USA, 37235

^d Biomedical Engineering, Vanderbilt University, Nashville, TN, USA, 37235

^e Psychiatry, Vanderbilt University, Nashville, TN, USA, 37235

Abstract

The Vanderbilt University Institute for Imaging Science (VUIIS) Center for Computational Imaging (CCI) has developed a database built on XNAT housing over a quarter of a million scans. The database provides framework for (1) rapid prototyping, (2) large scale batch processing of images and (3) scalable project management. The system uses the web-based interfaces of XNAT and RedCAP to allow for graphical interaction. A python middleware layer, the Distributed Automation for XNAT (DAX) package, distributes computation across the Vanderbilt Advanced Computing Center for Research and Education high performance computing center. All software are made available in open source for use in combining portable batch scripting (PBS) grids and XNAT servers.

1. INTRODUCTION

Philosophy

The Vanderbilt University Institute for Imaging Science (VUIIS) Center for Computational Imaging (CCI) maintains an active database with 206 projects, 28,322 subjects and 41,161 imaging sessions as of December 12, 2014. The goal of this database is to provide a new framework for algorithm development which allows for (1) rapid prototyping, (2) large scale batch processing of images and (3) scalable project management. This new framework

Corresponding author Robert L. Harrigan, Vanderbilt University EECS, 2301 Vanderbilt Pl., PO Box 351679 Station B, Nashville, TN 37235-1679, **Work:** 410-917-6166, rob.l.harrigan@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

requires minimal programming to run processing on new data sets by creating processing modules, which can be run on any number of data sets with minimal modifications. The structure forces pipeline versioning, minimizes human errors, and ensures that all data are processed in the same manner for large studies.

XNAT Architecture

The VUIIS CCI database is built on XNAT (Gao et al., 2013; Marcus et al., 2007), an open source imaging informatics platform developed by the Neuroinformatics Research Group at Washington University in St. Louis. Our XNAT is run in a group of virtual machines and provides a web interface for interaction with the database. The web interface allows viewing of datasets using snapshots. These snapshots can be loaded in the browser and provide an overview of image volumes.

We have multiple paths and tools for importing data into the VUIIS CCI XNAT database. The VUIIS human imaging center includes two 3T Philips Intera Achieva and one 7T Philips Achieva MRI scanner for research purposes. Any data generated on these scanners is automatically sent to a DICOM server. This server routes all data to the XNAT instance. Datasets are assigned to projects based on the principle investigator (PI) of the institutional review board (IRB) project under which they are acquired. Once data are owned by the PI's primary project, they can be shared or moved into other (sub)projects to facilitate study management. We also have command line tools, which can be used to upload batches of any supported data. These are especially useful for the upload of data from collaborators outside of Vanderbilt University and can support a large number of file types provided by collaborators. All uploaded data can then be converted into any number of formats for processing while maintaining the integrity of the originally uploaded data including, for example, DICOM properties on images from the DICOM server.

To maximize the amount of data that can be processed, observed, and managed by our engineers, we have implemented an XNAT control system housed in REDCap (Harris et al., 2009). REDCap is a secure web application, developed at Vanderbilt, for building and managing online surveys and databases. Figure 1 shows the REDCap control panel which simplifies project management. This database houses information on an XNAT-project specific basis about what processing algorithms are to be executed for each sequence. This information is pulled automatically to generate a project "settings" file which is launched on a regular basis by our processing infrastructure, Distributed Automation for XNAT (DAX, <https://github.com/VUIIS/dax>). The control system is responsible for snapshot generation, end-to-end processing and interfacing with the supercomputing environment.

Data processing occurs on the Vanderbilt Advanced Computing Center for Research and Education, which houses more than 6000 Linux CPU cores. DAX has been developed with a series of settings making the package portable on any portable batch scripting (PBS) system. It has been tested on the Sun Grid Engine and SLURM platforms as well as a MOAB scheduler with a TORQUE queue manager. Each customized module is a spider that performs an image processing task using a variety of open source tools such as the Java Image Science Toolkit (Lucas et al., 2010), FSL (Jenkinson et al., 2012), FreeSurfer (Fischl, 2012), ANTs (Avants et al., 2011), NiftyReg (Ourselin et al., 2001; Ourselin et al., 2002) as

well as others. In alignment with our philosophy, all collaborators are encouraged to work with our engineers to implement their own processing pipelines as XNAT spiders. To add spiders to the system, one should request write access to the NITRC (project id: masimatlab). Write access is needed to prevent vandalism and spam attacks, but is freely given to verified users. Each spider is written to a different directory within the repository to help manage code conflicts. The first step of each spider is to download data from XNAT to a processing node using the package Pyxnat (Schwartz et al., 2012). Statistics from each process are added to a REDCap database. All DAX interfacing with REDCAP is done through the PyCap python package, which allows programmatic access to the REDCap databases through the REDCap application programming interface (API) (<https://github.com/sburns/PyCap>). Once data has been downloaded, the Python spider can then call other languages such as MATLAB or command line tools to process the data.

Once processing is complete, all pipelines produce a PDF report, which details results of the processing pipeline and can be quickly reviewed by a human. These reports are tailored for each pipeline to serve as a tool to quickly inspect the result of processing to determine the quality. Any other files generated by the processing pipeline can be uploaded to XNAT after processing. If results need to be downloaded for further analysis, this can be facilitated through the web interface, which allows the download of items from the database. DAX also provides command line tools for bulk download and a Python API for interfacing with the XNAT database.

2. Database Design

Our database was designed for active study management and control. We place an emphasis on efficient data processing and make sure that our infrastructure not only allows for large scale imaging studies, but also is optimized for them. Our data archive is constantly growing and changing. New scans are added daily as they are acquired on our local scanners as well as new data from collaborating institutions being added and processed.

The software underlying the database was designed to be portable and readily reproduced. We have deployed private instances of this system to handle protected health information data and sensitive institutional projects. Some users have brought large external projects into the common processing environment and we support this.

3. Data Availability

The VUIIS CCI XNAT contains 206 projects including over 258,635 scans and holds both magnetic resonance imaging and computed tomography on human and non-human primates. Support for near-infrared spectroscopy data is in the process of being implemented. The database also contains three data sets for use by the wider community: Multi-Modal MRI Reproducibility Resource, OASIS and IXI as well as completed processing results. The Multi-Modal MRI Reproducibility Resource (MMRR) contains scan-rescan imaging sessions on 21 healthy volunteers with a selection of multi-modal MRI acquisitions. The Open Access Series of Imaging Studies (OASIS) data set contains a cross-sectional collection of 416 right-handed subjects aged 18 to 96. Each subject contains 3 or 4 T1 MRI scans. One hundred of the included subjects over the age of 60 have been clinically

diagnosed with very mild to moderate Alzheimer's disease. A reliability data set of 20 non-demented subjects is also included. The Information eXtraction from Images (IXI) data set contains nearly 600 MRI images from normal, healthy subjects acquired at 3 different institutes. Each subject has T1, T2, PD-weighted, MRA and Diffusion data. Moreover, the XNAT instance is serving as a source repository for Research Domain Criteria (RDoC) data that is being publicly shared through the National Institute of Mental Health (NIMH) National Database for Autism Research (NDAR) (Insel et al., 2010). These raw imaging data are available from other repositories (e.g., the previously released MMMRR dataset on NITRC). The value of hosting these resources on XNAT is the availability of extensive image processing with consistent approaches across public, semi-private (e.g., ADNI investigators at Vanderbilt share a collaborative ADNI mirror following the data license terms), and private data. All sets have been processed using multiple XNAT spiders and the processing results are available along with the data. We have developed spiders to handle segmentation using Freesurfer (Fischl, 2012), FSL First (Patenaude et al., 2011), Multi-atlas (Asman and Landman, 2012, 2013) and SPM 8 (VBMQA)(Mechelli et al., 2005). Spiders for DTI analysis have been developed including DTI quality analysis (Lauzon et al., 2013a), EVE white matter stamper, Bedpost (Behrens et al., 2007), TBSS (Smith et al., 2006) and Tracula (Yendiki et al., 2011). Two separate lesion segmentation pipelines have also been implemented, Lesion TOADS (Shiee et al., 2010) and Lesion Segmentation Tool (LST) (Schmidt et al., 2012). fMRI quality analysis (Friedman and Glover, 2006) is also implemented as a spider. Table 1 shows a summary of publicly available data and the available processing results.

Of the 258,635 total scans, 189,202 of them have a neuroimaging informatics technology initiative (NIFTI) format volume which has become our file format of choice for image volumes. This consistency helps support our mission for large scale batch processing across data sets. The database contains 192,241 snapshots which are preview images of the NIFTI volumes allowing for easy evaluation of large datasets. Ancillary data including but not limited to respiratory or cardiac information can also be sent through this pipeline. Demographic and other phenotypic information is available for some data sets, the inclusion of this data is at the discretion of the collaborator. All protected health information is stored in REDCap or a separate PI-controlled database. As of March 2015, XNAT is not certified to be a compliant datastore for protected health information data. To avoid potential privacy concerns, we store all health identifiers in REDCap and use XNAT study identifiers (involving random patient ids) to link the PHI to the images. For easy of data federation, we advocate storing image-derived scalar quantitative measures in REDCap instead of XNAT (however, the final decision is left to the PI). These data come from multiple scanners and centers across the United States. Both processed and analyzed data is also available for download once project access has been granted. All processing algorithms are freely available as open source through our masimatlab project on the Neuroimaging Informatics Tools and Resources Clearinghouse (<http://www.nitrc.org/projects/masimatlab>). A feature of the XNAT system is that all data sets have unique uri resources.

4. Quality Control

A feature of our system is that we automatically run first level processing for most data types (as illustrated in Figure 2). The summary measures are typically exported to REDCap and made available in PDF form. The PDF documents enable rapid manual review and data perusal. The quantitative measures from these documents are typically exported to a REDCap datasets for filtering and quality analysis. Additionally, a snapshot viewer has been incorporated into the XNAT web interface, allowing 3D image sequences to be viewed in multiple planes online, thus sparing local storage capacity. The project managers then inspect these reports to make a determination of the quality of the results. More complicated image processing pipelines can be split into intuitive modules and quality can be assessed at all levels of the pipeline to ensure final product quality. Some processing workflows require manual quality assessment for progression through the pipelines. The reports simplify this process by removing the need to download processed data.

5. Data Access

Data access is managed on a per-project basis and users must register with an email address and password to be granted access to public data projects. Once a user has registered with XNAT, they can request access to any private projects they need to access. Access to a project allows the user to view all imaging data, ancillary data and processed data. This includes the snapshots for quick data preview from a web browser as well as the PDFs generated by the processing pipelines showing results.

Data can be mirrored between XNATs using our Xnatmirror tool which is in the DAX package. This tool has the capability of mirroring all image formats, metadata and data types, including custom fields, as long as they are present in the destination database. Large data request downloads are handled through the command line tools and python API included in DAX. Specifically, we provide Xnatdownload and Xnatupload. Xnatupload can automatically create the hierarchy for a project and upload individual files to appropriate locations. Xnatdownload allows the user to select data based on project, subject, session, scan, or processing type and filter based on file format or data status. Xnatdownload can also handle failed-connections and maintain a synchronized copy of a dataset. These tools interface with the XNAT REST API which handles large data downloads.

PIs manage all data usage agreements and institutional review board compliance. The database is setup to not contain any PHI. Image meta-data is de-identified upon addition to the database. This is done because the primary goal of this database is to allow for large scale image processing algorithm applications. There is no system in place for updating users if data is withdrawn, revised or added to. Collaborators can choose to be notified about processing of data including when a scan begins processing, finishes processing or fails processing.

All data acquired at VUIIS and by VUIIS investigators is automatically added to the database. We also support the inclusion of data acquired elsewhere when it is part of collaborations with VUIIS investigators.

6. Looking Forward

The VUIIS XNAT data and DAX have enabled close integration of imaging science with high performance computing. In the past 5 years, we have transition from using 1 CPU-year to 200+ CPU-years on the shared campus infrastructure. By realizing synergies between diverse projects, we have reduced the level of effort required to explore emerging analysis approaches. Since our initial deployment on a single server, we have grown into an enterprise cloud environment. We plan to continue growth and support more investigators with diverse neuro- and non-neuro research interests.

Acknowledgements

Research reported in this publication was supported by the National Institutes of Health Award Numbers R21EY024036, 5T32EY007135 and 5R01MH098098. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This project was supported by ViSE/VICTR VR3029. The project described was supported by the National Center for Research Resources, Grant UL1 RR024975-01, and is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06. Additional support was received from the Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

Grant Support

Bennett A. Landman: 5R21EY024036, David Zald: 5R01MH098098

References

- Asman, AJ.; Landman, BA. Information Processing in Medical Imaging. Springer; 2011.
Characterizing spatially varying performance to improve multi-atlas multi-label segmentation; p. 85-96.
- Asman AJ, Landman BA. Formulating Spatially Varying Performance in the Statistical Fusion Framework. Medical Imaging, IEEE Transactions on. 2012; 31:1326–1336.
- Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. Med Image Anal. 2013; 17:194–208. [PubMed: 23265798]
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage. 2011; 54:2033–2044. [PubMed: 20851191]
- Behrens T, Berg HJ, Jbabdi S, Rushworth M, Woolrich M. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage. 2007; 34:144–155. [PubMed: 17070705]
- Fischl B. FreeSurfer. NeuroImage. 2012; 62:774–781. [PubMed: 22248573]
- Fonov V, Evans A, McKinsty R, Almlí C, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009; 47:S102.
- Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. Journal of Magnetic Resonance Imaging. 2006; 23:827–839. [PubMed: 16649196]
- Gao, Y.; Burns, SS.; Lauzon, CB.; Fong, AE.; James, TA.; Lubar, JF.; Thatcher, RW.; Twillie, DA.; Wirt, MD.; Zola, MA.; Logan, BW.; Anderson, AW.; Landman, BA. Integration of XNAT/PACS, DICOM, and Research Software for Automated Multi-modal Image Analysis. Proc SPIE 8674. 2013.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009; 42:377–381. [PubMed: 18929686]

- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American journal of psychiatry*. 2010; 167:748–751. [PubMed: 20595427]
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. *Fsl. NeuroImage*. 2012; 62:782–790. [PubMed: 21979382]
- Lauzon CB, Asman AJ, Esparza ML, Burns SS, Fan Q, Gao Y, Anderson AW, Davis N, Cutting LE, Landman BA. Simultaneous analysis and quality assurance for diffusion tensor imaging. *PloS one*. 2013a; 8:e61737. [PubMed: 23637895]
- Lauzon CB, Crainiceanu C, Caffo BC, Landman BA. Assessment of bias in experimentally measured diffusion tensor imaging parameters using SIMEX. *Magnetic Resonance in Medicine*. 2013b; 69:891–902. [PubMed: 22611000]
- Lauzon CB, Landman BA. Correcting power and p-value calculations for bias in diffusion tensor imaging. *Magnetic Resonance Imaging*. 2013; 31:857–864. [PubMed: 23465764]
- Lucas BC, Bogovic JA, Carass A, Bazin PL, Prince JL, Pham DL, Landman BA. The Java Image Science Toolkit (JIST) for rapid prototyping and publishing of neuroimaging software. *Neuroinformatics*. 2010; 8:5–17. [PubMed: 20077162]
- Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*. 2007; 5:11–34. [PubMed: 17426351]
- Mechelli A, Price CJ, Friston KJ, Ashburner J. Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews*. 2005; 1:105–113.
- Ourselin S, Roche A, Subsol G, Pennec X, Ayache N. Reconstructing a 3D structure from serial histological sections. *Image and vision computing*. 2001; 19:25–31.
- Ourselin, S.; Stefanescu, R.; Pennec, X. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002*. Springer; 2002. Robust registration of multi-modal images: towards real-time clinical applications; p. 140–147.
- Papadakis NG, Martin KM, Wilkinson ID, Huang CL-H. A measure of curve fitting error for noise filtering diffusion tensor MRI data. *Journal of Magnetic Resonance*. 2003; 164:1–9. [PubMed: 12932449]
- Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage*. 2011; 56:907–922. [PubMed: 21352927]
- Plassard, AJ.; Hinton, KE.; Venkatraman, V.; Gonzalez, C.; Resnick, SM.; Landman, BA. *SPIE Medical Imaging*. Orlando, FL: 2015. Evaluation of atlas-based white matter segmentation with eve.
- Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C. An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage*. 2012; 59:3774–3783. [PubMed: 22119648]
- Schwartz Y, Barbot A, Thyreau B, Frouin V, Varoquaux G, Siram A, Marcus DS, Poline J-B. PyXNAT: XNAT in python. *Front Neuroinform*. 2012; 6
- Shiee N, Bazin P-L, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*. 2010; 49:1524–1535. [PubMed: 19766196]
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*. 2006; 31:1487–1505. [PubMed: 16624579]
- Whitcher B, Tuch DS, Wisco JJ, Sorensen AG, Wang L. Using the wild bootstrap to quantify uncertainty in diffusion tensor imaging. *Human brain mapping*. 2008; 29:346–362. [PubMed: 17455199]
- Yendiki A, Panneck P, Srinivasan P, Stevens A, Zöllei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front Neuroinform*. 2011; 5

Author Manuscript

Figure 2.

Quality analysis reports generally perform first level analyses, including diffusion tensor imaging (A)(Asman and Landman, 2013; Lauzon et al., 2013b; Lauzon and Landman, 2013; Papadakis et al., 2003; Whitcher et al., 2008), functional MRI (B)(Friedman and Glover, 2006), multi-atlas brain segmentation (C)(Asman and Landman, 2011, 2013; Ourselin et al., 2001; Ourselin et al., 2002), white matter labeling (D)(Plassard et al., 2015), registration to template spaces(E)(Fonov et al., 2009), FreeSurfer (F)(Fischl, 2012) and TRACULA (G) (Yendiki et al., 2011).

Table 1

Publicly available data and processing

Data Set	Disease	Modality	Scan Types	Number of Subjects	Number of images	Completed Processing
Kirby 21 Reproducibility Resource	Healthy	MRI	MPRAGE, FLAIR, DTI, fMRI, B0 and B1 maps, ASL, VASO, quantitative T1 and T2 mapping, MT	21	618	FSL First, Freesurfer, Lesion TOADS, multi atlas, TBSS, VBMQA, EVE, dtiQA, fMRIQA
Open Access Series of Imaging Studies (OASIS)	Alzheimer's Disease	MRI	T1	416	436	FSL first, freesurfer multi-atlas
Information extraction from Images (IXI)	Healthy	MRI	T1, T2, PD-weighted, MRA, Diffusion	584	2707	Multi-atlas