Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/ynimg

Neuroimaging data sharing on the neuroinformatics database platform

Gregory A. Book [a,*], Michael Stevens [a], Michal Assaf [a,b], David Glahn [a], Godfrey D. Pearlson [a,b]

[a] Olin Neuropsychiatry Research Center, Hartford Hospital, Hartford CT, USA
[b] Yale University, Department of Psychiatry, New Haven, CT, USA

ARTICLE INFO

Available online xxxx

ABSTRACT

We describe the Neuroinformatics Database (NiDB), an open-source database platform for archiving, analysis, and sharing of neuroimaging data. Data from the multi-site projects Autism Brain Imaging Data Exchange (ABIDE), Bipolar–Schizophrenia Network on Intermediate Phenotypes parts one and two (B–SNIP1, B–SNIP2), and Monetary Incentive Delay task (MID) are available for download from the public instance of NiDB, with more projects sharing data as it becomes available. As demonstrated by making several large datasets available, NiDB is an extensible platform appropriately suited to archive and distributes shared neuroimaging data.
© 2015 Elsevier Inc. All rights reserved.

Background

Neuroinformatics Database (NiDB) was created to solve the problem of organizing and analyzing very large neuroimaging datasets and has since grown into a neuroimaging database platform (Book et al., 2013). When development of the platform began in 2005, a publication with a sample size of one hundred subjects was considered very large, while now sample sizes in the thousands are common (Kiehl et al., 2005; Meda et al., 2014). There are diminishing returns when using sample sizes larger than 1000 subjects; however, the ability to store and analyze data from multiple patient cohorts and longitudinal datasets is extremely valuable, especially when testing reproducibility (Kennedy, 2014).

Development of NiDB began as a system for searching and downloading of MRI scans collected in the previous 30 days, using flat file storage of meta-data. The system could only search by subject ID, protocol name, and scan date. However, as the amount of stored data grew, the system was re-written to use a SQL database and catalog more meta-data. As data sizes grew further, the system architecture was redesigned to be subject-centric, following a Subject → Enrollment → Imaging Study → Series hierarchy (Fig. 1). A subject-centric design allows association of multiple modalities of data with an imaging session, multiple imaging sessions with a subject's enrollment in a project, and enrollment of subjects in multiple projects. This architecture provides a standardized hierarchy into which new imaging modalities are stored in the database and makes the addition of project permissions and security straightforward. NiDB currently stores magnetic resonance (MR), computed tomography (CT), ultrasound (US), positron-emission tomography (PET), electroencephalography (EEG), pre-pulse inhibition (PPI), eye-tracking (ET), and genome data, but can expand with minimal effort to include any modality. NiDB is web-based, using PHP and JavaScript as the front-end, MySQL as the middle layer, and Perl as the backend. A separate uploader for large datasets is written in C++ and QT. Regular users access the system through the web-based GUI or QT based uploader, and administrators perform many maintenance operations through the web-based GUI. A small amount of maintenance is required in the back-end by a developer to backup data, add new modalities, fix bugs, or add enhancements. (See Fig. 2.)

Data importing, searching, and exporting features are available, as well as storage of subject demographics, system statistics, and project permissions. NiDB contains several features beyond data storage and searching, including: pipeline analysis, inter-instance sharing, and modular automated quality control (QC). Automated QC is 'modular', meaning a user can create a QC module/script which takes a data path as input, performs specified QC analysis, and inserts the results into the database. NiDB's pipeline system is connected to a compute cluster where analyses are automatically performed and results are imported back into the database to be associated with the original data. Data are analyzed using a normal bash script, with special NiDB variables that are replaced with full paths when the pipeline is run. Each pipeline has a set of data criteria, and all imaging studies that match the criteria are sent through the pipeline, which creates a custom cluster job with the correct paths and IDs for each imaging study. During the cluster job processing, output from the original bash script is logged and the status of an analysis can be viewed, along with summary statistics such as number completed, running, or in error state. Upon completion of each analysis, important results and figures (defined by the user) are automatically imported back into NiDB and are available for searching alongside the raw data. Data processed through a NiDB freesurfer pipeline was included in very large scale study of genetic association with subcortical brain structures (Hibar et al., 2015). The NiDB pipeline

* Corresponding author.
E-mail addresses: gregory.a.book@gmail.com, gregory.book@hhchealth.org (G.A. Book).

http://dx.doi.org/10.1016/j.neuroimage.2015.04.022
1053-8119/© 2015 Elsevier Inc. All rights reserved.

UNCORRECTED PROOF

ARTICLE IN PRESS

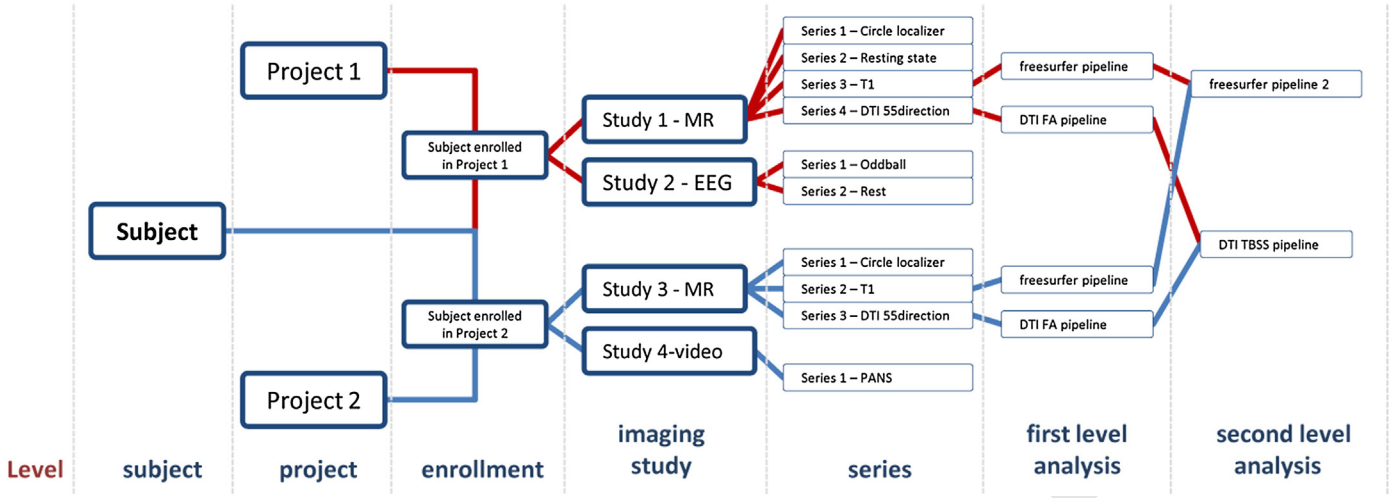

Fig. 1. Neuroinformatics database hierarchy. This is a near universal format into which any modality of imaging can be stored.

system includes inter-pipeline dependencies to allow efficient processing of data. Examples of tested NiDB pipelines include freesurfer, SPM fMRI processing, FSL DTI, FSL fMRI, and Human Connectome Pipeline (HCP) analyses (Fig. 1).

A feature important to future data sharing is NiDB's export to National Database on Autism Research (NDAR) format. NDAR is a large scale data repository hosted by the NIMH, and the underlying database system is now used for the Research Domain Criteria (RDoC) project which seeks to archive data collected under NIMH sponsored projects. NiDB compatibility with NDAR allows for direct, seamless data export.

Design

The current iteration of NiDB was designed as an active study management system for neuroimaging and clinical research data. Data is imported from one of several sources (DICOM receiver, web-based importer, GUI based importer, or inter-instance sharing), archived and QC'd. The data is then searchable to users. Archived data is associated with existing or new subjects, at which time demographic data may be imported from meta-data (eg DICOM header) or manually entered. Data stored on the public server is static for the main projects listed in the available data section, but may be dynamic for other projects.

Available data

NiDB is currently hosted in two different instances, each with different data and accessibility. The internal instance of NiDB, only accessible within the Olin Neuropsychiatry Research Center network, contains 12.9 TB of raw data from 199,951 imaging series from 22,464 imaging sessions from 11,147 subjects in 158 projects. In total, 366 days of CPU time have been used to compute QC metrics and 16.6 TB of data have been requested. While these data are not all available publicly, it attests to the scalability of NiDB. The external (public) instance of NiDB contains the data described in this paper, available at http://olinnidb.org (Table 1). Five major projects comprise the data currently shared on the public server: ABIDE, B–SNIP1, PARDIP, B–SNIP2, MID.

Autism Brain Imaging Data Exchange (ABIDE) data was aggregated by the International Data-sharing Initiative (INDI) and imported into NiDB (Di Martino et al., 2014; Mennes et al., 2013). The ABIDE dataset contains resting fMRI, structural MR, and phenotypic data from 16 projects (sites) examining autism spectrum disorder. The original downloads from ABIDE were single blocks of data from each site, but after importing into NiDB, subsets of the data can be searched for and downloaded. Part one of the Bipolar–Schizophrenia Network on Intermediate Phenotypes (B–SNIP1) study, examines multiple phenotypes in individuals with schizophrenia, psychotic bipolar disorder, and schizoaffective disorder, and their first-degree relatives. Data were col-

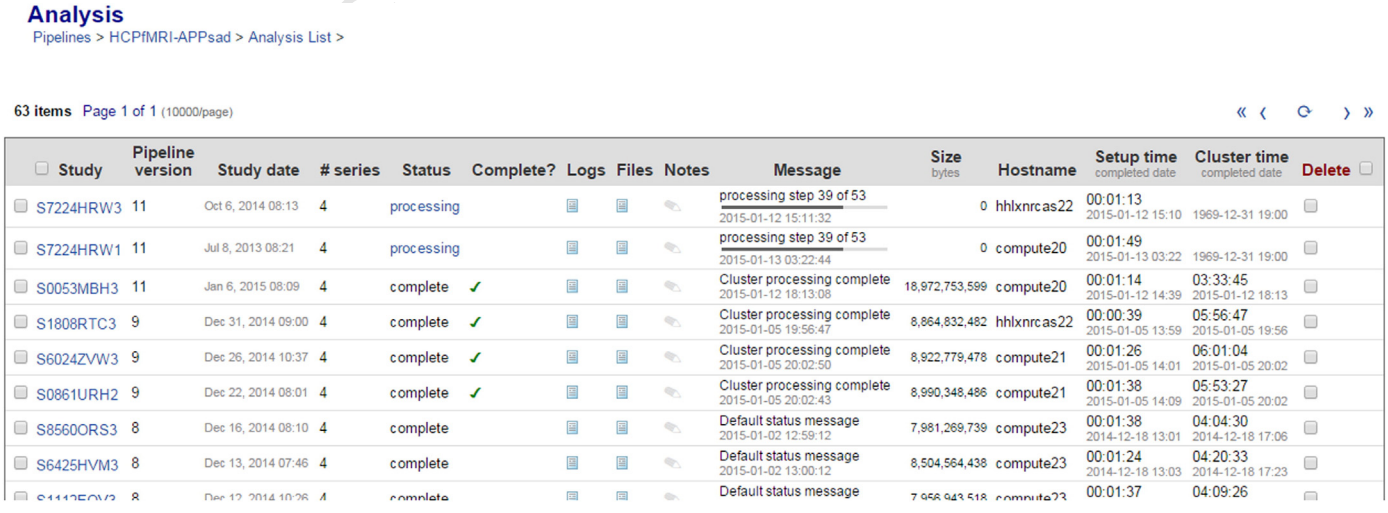

Fig. 2. NiDB pipeline system, analysis list. Imaging studies that meet the data criteria are processed through the pipeline and their status is displayed.

Please cite this article as: Book, G.A., et al., Neuroimaging data sharing on the neuroinformatics database platform, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.04.022

Download English Version:

https://daneshyari.com/en/article/6023825

Download Persian Version:

https://daneshyari.com/article/6023825

Daneshyari.com