NeuroInformatics Database (NIDB)

A Portable Database for Storage, Pipeline Processing, and Sharing of Neuroimaging Data

Gregory A Book[1], Pawel Skudlarski[1,2], Michael Stevens[1,2], Godfrey Pearlson[1,2]

[1] Hartford Hospital, Olin Neuropsychiatry Research Center, Hartford, CT, USA; [2] Psychiatry, Yale University School of Medicine, New Haven, CT, USA

Background

Neuroinformatics history: The field of neuroimaging has benefited from advances in EEG, PET and MRI imaging technology since the early 1990s. Developing alongside these imaging methods were inexpensive commercial hardware which allowed storage, analysis, and sharing of neuroimaging data. The amount of data collected data has increased exponentially since the early days of neuroimaging because of many factors: MRIs moved from 1.5T to 3T and TRs shrank which producing more volumes, there was an increase in installations of research dedicated scanners and a subsequent increase in study enrollment and more funding for neuroscience research, all leading to larger and larger stores of neuroimaging data. MRI datasets can vary in size from a few MB to several GB. A 3T MRI scanning 5 participants per day could produce 35 imaging sessions per week, each 500MB in size, for a total of nearly 1TB of raw MR data in one year. Add to that the EEG, PET, CT, videos, SNP, genome, and exome data collected for the same participants, and a single imaging center can produce several TB of raw data per year.

What to do with all this data? Many labs and institutions create their own home grown solutions to storing and processing their data. These solutions are often elegant and effective, but are very specific to the development site, ie: they are not portable, and do not make it easy to share the data they store. It's also rather difficult to import data from someone else's site. Labs are often focused more on their own projects and techniques, especially in a developing field like neuroimaging, and not on developing common software to store the raw data. Not until the mid-2000s, were databases, such as xNAT, LONI, and COINS available to the neuroimaging community.

So, why create a new database? Time, money, and extent. Current open-source systems are time consuming to install, and customize. Commercial systems are not free. Most systems are built one modality at a time, with extensive work needed to add new modalities. NIDB addresses these problems and adds new functionality. **The NIDB project establishes a common platform for collaboration and data sharing using an easy to install, portable database.**

The Software

NiDB Neuroinformatics Database

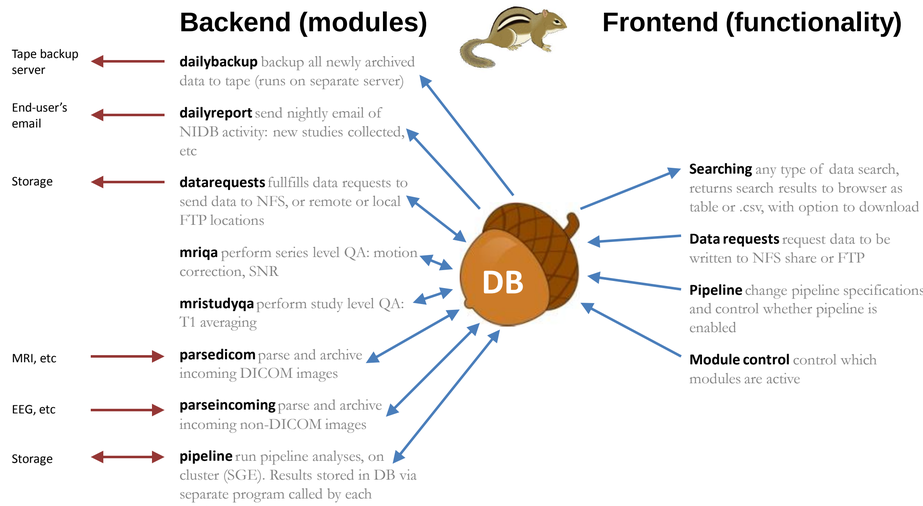
NIDB is a web based database system with support to automatically store, perform QA, and analyze any modality data. Data can be searched at the series or analysis level.

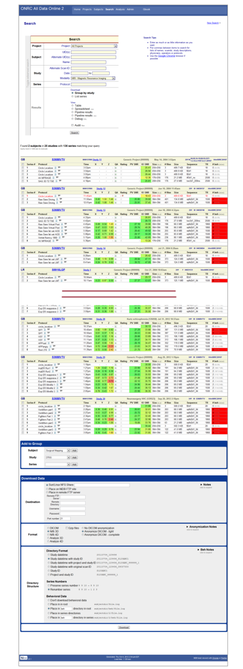
Features:

- Installation in 1 hour, on CentOS 6.3
- Data stored at your site
- Series level data access
- Download data directly to compute servers
- Automatic QA: motion, SNR
- Store ANY modality of data
- Automated pipeline analysis (raw MR to analysis results instantly)
- Export search results to spreadsheet
- Can create groups of subjects, studies, series
- DICOM receiver
- Support for family relationships
- Support for paper-based assessments
- User supplied data quality rating system

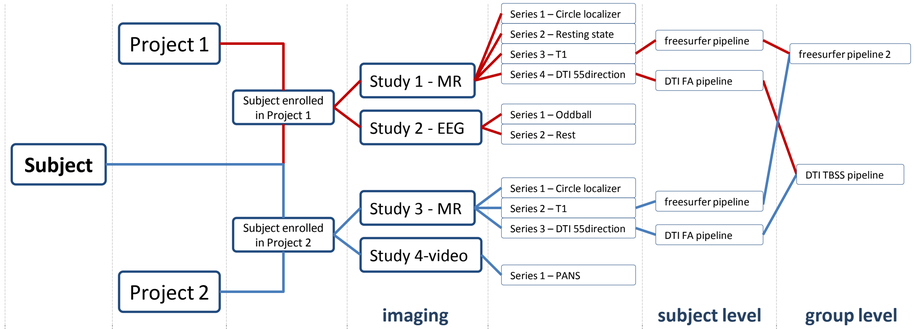
Module Flow & Frontend/Backend Interaction

NIDB is modular. Backend processing functions are separated and utilize multiple cores. Website functions are separate from the backend and essentially form a queue for backend processing so the frontend does not lag.





Storage Hierarchy



Level	subject	project	enrollment	imaging study	series	subject level analysis	group level analysis
Searchable?	✓	✓	✓	✓	✓	✓	✓
Groupable?	✓			✓	✓	✓	
Limitations	175,760,000 (7 char UID) 4,569,760,000 (8 char UID)	1,000,000 projects (000000 – 999999)	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
Granularity	Coarse	Coarse	Coarse	Medium	Fine	Fine	Coarse
Current ONRC size & avg search time for 1000 items	9,300 0.1s	77 0.1s	11,000 0.1s	18,500 1 sec	135,000 1 sec	8,000 analyses, 4.6 million result values, 10-15 sec	2 0.1s

Development Site

NIDB was developed at the Olin Neuropsychiatry Research Center, at the Institute of Living, Hartford Hospital, in Hartford, CT USA. Development began in 2005 and is currently in its 4th generation. NIDB currently hosts 6TB of raw data from 9,300 subjects on mirrored Silicon Mechanics servers each with 40TB capacity, 16-cores, and 24GB RAM.

Software Distribution

The key to expanding NIDB's user base is to provide a simple installer so that anyone with minimal Linux experience (and root access to a CentOS 6.3 system) can install the system. A fully configured Oracle VirtualBox image is available for download.

Future Work

The primary future goal of NIDB is to create a common data exchange format, so one NIDB installation can send data directly to another installation. A second goal is to create import scripts for many types of data. Both features will allow sites to create massive neuroinformatics databases.

Download NIDB (.tar.gz installer and VM image) from http://nidb.sourceforge.net