

HCP Informatics: ConnectomeDB and Cloud- based Processing

Daniel Marcus, PhD

**Associate Professor & Director, Neuroinformatics Research
Group @ Washington University School of Medicine**

Outline

- ConnectomeDB
 - Activities from the browser
 - Activities from command line
- Cloud computing
 - HCP data on the cloud
 - HCP pipelines and analysis on the cloud
- Running your own HCP Informatics System

What is ConnectomeDB?

- Public resource for accessing HCP data
- Includes WU-Minn data along with MGH and other data sets
- Future home of Connectome Coordination Facility (CCF) data

The screenshot displays the ConnectomeDB website interface. At the top, there's a navigation bar with 'CONNECTOME db' logo, 'All Datasets', 'HCP Subject Keys', and a search bar. Below the navigation bar, the 'Current Project' is set to 'WU-Minn HCP Data - 500 Subjects + MEG2'. The main content area features a section titled 'Public Connectome Data' with a sub-header 'Updated Nov 25, 2014: Unprocessed, source and channel-level processed MEG data from WU-Minn HCP available on over 60 subjects.' To the right of this section is an Amazon Web Services logo and text indicating that HCP data is mirrored on Amazon Web Services. Below this, there are three data release sections: 'WU-Minn HCP Data - 500 Subjects + MEG2' (526 subjects with MRI data, 67 with MEG data, 542 with behavioral data), 'WU-Minn HCP Lifespan Pilot Data' (27 subjects with MRI data), and 'MGH HCP Adult Diffusion' (35 subjects with MRI data). Each section includes a brief description of the data, a 'Last Updated' date, and a 'Keywords' field. The interface also includes buttons for 'Open Dataset', 'Explore Subjects', and 'Download Image Data'.

CONNECTOME db All Datasets HCP Subject Keys Search by ID Search

Current Project: WU-Minn HCP Data - 500 Subjects + MEG2

Logged in as: dan2 Auto-logout in: 0:28:50 - renew Logout

Public Connectome Data

Updated Nov 25, 2014: Unprocessed, source and channel-level processed MEG data from WU-Minn HCP available on over 60 subjects.

Update Log View Other Projects

amazon web services Public Data Sets

HCP Data now mirrored on Amazon Web Services

WU-Minn HCP Data - 500 Subjects + MEG2

Open Dataset Explore Subjects Download Image Data

HCP public data releases include high-resolution MR scans from healthy adults and four imaging modalities: structural images (T1w and T2w), resting-state fMRI (rsfMRI), task-fMRI (tfMRI), and high angular resolution diffusion imaging (dMRI). Behavioral data is also largely available, with some restrictions. Furthermore, MEG data is available for some subjects. The Open Access Dataset includes imaging data and most behavioral data. To protect subject privacy, some of the data (e.g., which subjects are twins) are part of a Restricted Access dataset.

Last Updated:

ACCESS: ☒ Restricted Access Terms Accepted Data Available on Amazon S3

KEYWORDS: HCP, MRI, CONNECTOME, MEG, RESTING STATE, DIFFUSION, RfMRI, DMRI, FMRI

526 SUBJECTS WITH MRI DATA
67 SUBJECTS WITH MEG DATA
542 SUBJECTS WITH BEHAVIORAL DATA

WU-Minn HCP Lifespan Pilot Data

Open Dataset Explore Subjects Download Image Data

The WU-Minn HCP consortium is acquiring and sharing pilot multimodal imaging data acquired across the lifespan, in 6 age groups (4-8, 8-9, 14-15, 25-35, 45-55, 65-75) and using scanners that differ in field strength (3T, 7T) and maximum gradient strength (70-100 mT/m). The scanning protocols are similar to those for the WU-Minn Young Adult HCP except shorter in duration. The objectives are (i) to enable estimates of effect sizes for identifying group differences across the lifespan and (ii) to enable comparisons across scanner platforms, including data from the MGH Lifespan Pilot. The initial data releases in August, 2014 include unprocessed Phase1a image data and will be regularly updated with Phase1B data; minimally preprocessed data will be released in the fall of 2014.

Last Updated:

ACCESS: ☒ Open Access Terms Accepted

KEYWORDS: LIFESPAN, HCP, FMRI

27 SUBJECTS WITH MRI DATA

MGH HCP Adult Diffusion

Open Dataset Explore Subjects Download Image Data

The MGH HCP team has released diffusion imaging and structural imaging data acquired from 35 young adults using the customized MGH Siemens 3T Connectome scanner, which has 300 mT/m maximum gradient strength for diffusion imaging.

Last Updated:

35 SUBJECTS WITH MRI DATA

Accessing ConnectomeDB

- Registration required
- Data use terms: open and restricted.
- Aspera plugin required for bulk downloads (i.e. packages)
- Browser for easy navigation
- Command line for scripted access

The screenshot shows a web browser window with the URL <https://db.humanconnectome.org/app/template/Login.vm>. The page header includes the 'CONNECTOME db' logo. A yellow warning banner at the top states: 'Plugin Required! A browser check indicates that you do not have the Aspera Connect plugin installed. The Human Connectome Project is using an Aspera server to dramatically boost data transfer speeds. Please install this plugin before using any of ConnectomeDB's services, including downloading. (Need help with Aspera install?)'. The main content area features a 'Log In' section with fields for 'USERNAME' and 'PASSWORD', and a 'Register' button. A modal window titled 'Register an account with the Human Connectome Project' is open, containing the following fields and options:

- CREATE A USERNAME:** Text input with 'dan3' entered.
- CREATE A PASSWORD:** Password input field.
- CONFIRM YOUR PASSWORD:** Password input field.
- Account Information:**
 - FIRST NAME:** Text input with 'Dan' entered.
 - LAST NAME:** Text input with 'Marcus' entered.
 - EMAIL:** Text input with 'marcusd@mir.wustl.edu' entered.
 - INSTITUTION:** Text input with 'Washington University' entered.
- Mailing List Subscriptions (Optional):**
 - ☒ Send me announcements of future data releases
 - ☒ Subscribe me to the HCP Data Users email forum

At the bottom of the modal, it states 'All fields required for registration.' with 'Cancel' and 'Register' buttons. The footer of the page reads 'ConnectomeDB is a product of the Human Connectome Project' and includes links for 'ConnectomeDB Tutorial' and 'Contact Support'.

Alternative sources of HCP data

- Connectome in a Box
- Amazon Public Data Sets
- All options require first signing the data use terms in ConnectomeDB



Data organization

- The 4 standard imaging sessions have been merged into 1 “mega” session.
- The scans in the mega session are ordered in a standard sequence (not necessarily reflecting the temporal order).
- Key data files have been assembled into “packages” to streamline distribution.
- Some interesting processed data sits outside the package structure.

Packages

- Specific packages for structural, resting, diffusion, and each task.
- File organization is well documented here:
http://www.humanconnectome.org/documentation/S500/HCP_S500+MEG2_Release_Appendix_III.pdf
- Same file organization on ConnectomeDB, CINABox, and Amazon.
- All files can be retrieved individually from ConnectomeDB.

CONNECTOMEDB ACTIVITIES FROM THE BROWSER

ConnectomeDB Browser Demo -- Main themes

- Basic navigation
- Some special fields (metadata)
- Filtering data
- Saving subject groups
- Digging deeper into individual subjects

CONNECTOMEDB ACTIVITIES FROM COMMAND LINE

Why use the command line?

- ConnectomeDB has a rich programming interface (API)
 - Use the XNAT REST API
 - <https://wiki.xnat.org/display/XNAT16/Using+the+XNAT+REST+API>
- Build ConnectomeDB access directly into scripts.
- Batch download specific files.
- WARNING: The API does NOT use Aspera.

Some handy command line tools

- CURL
 - Access specific URLs
- XNAT Data Client
 - Access data using XNAT tags
- Python, bash, or your favorite scripting language
 - Access data programmatically

ConnectomeDB Command Line Demo --

Main themes

- Basic REST API patterns
- Accessing subject groups
- Accessing specific files follows standard package structure
- Querying on subject metadata
- Batch retrieval of files

HCP ON THE CLOUD



What is the “cloud”?

Amazon Web Services (AWS)

- “‘Cloud Computing’, by definition, refers to the on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing.”
- Simple Storage Service (S3): File storage
- Elastic Compute Cloud (EC2): Computers
- Many other services (databases, block storage, workflow, etc.) not used by HCP

Why use the cloud for HCP projects?

- HCP data is hosted on S3
 - You don't need to download and store it locally!
 - You don't need to buy Connectome in a Box!
- HCP pipelines are preconfigured on EC2
 - You don't need to install and configure the pipelines locally.
 - You don't need to have local computing resources to process large quantities of data.
- The cloud ain't free (but it can be pretty cheap).

Simple Storage Service (S3)

- Files are stored on S3 in “buckets”, which are used to partition and manage the data.
- Each file is accessible at a specific URL.
 - Example: <http://bucket1.s3.amazonaws.com/file1>
- But working with URLs is complicated – file systems are easier.
- S3fs can be used to “mount” an S3 bucket as a file system on a computer – they look just like regular files.

Elastic Compute Cloud (EC2)

- Computers on EC2 are called “instances”.
- “Amazon Machine Images” (AMIs) are preconfigured computer templates with specific applications preinstalled.
- Each instance runs a selected AMI.
- Each instance runs on selected “hardware”.
- Run lots of instances to build your own high performance compute cluster.
- Use Starcluster to manage jobs on your cluster



Put HCP Data in
an S3 Bucket

Mount HCP Data
from S3 on EC2
Instances



Run EC2 Instances
preconfigured w/
HCP Pipelines



Manage batch
processing with
StarCluster

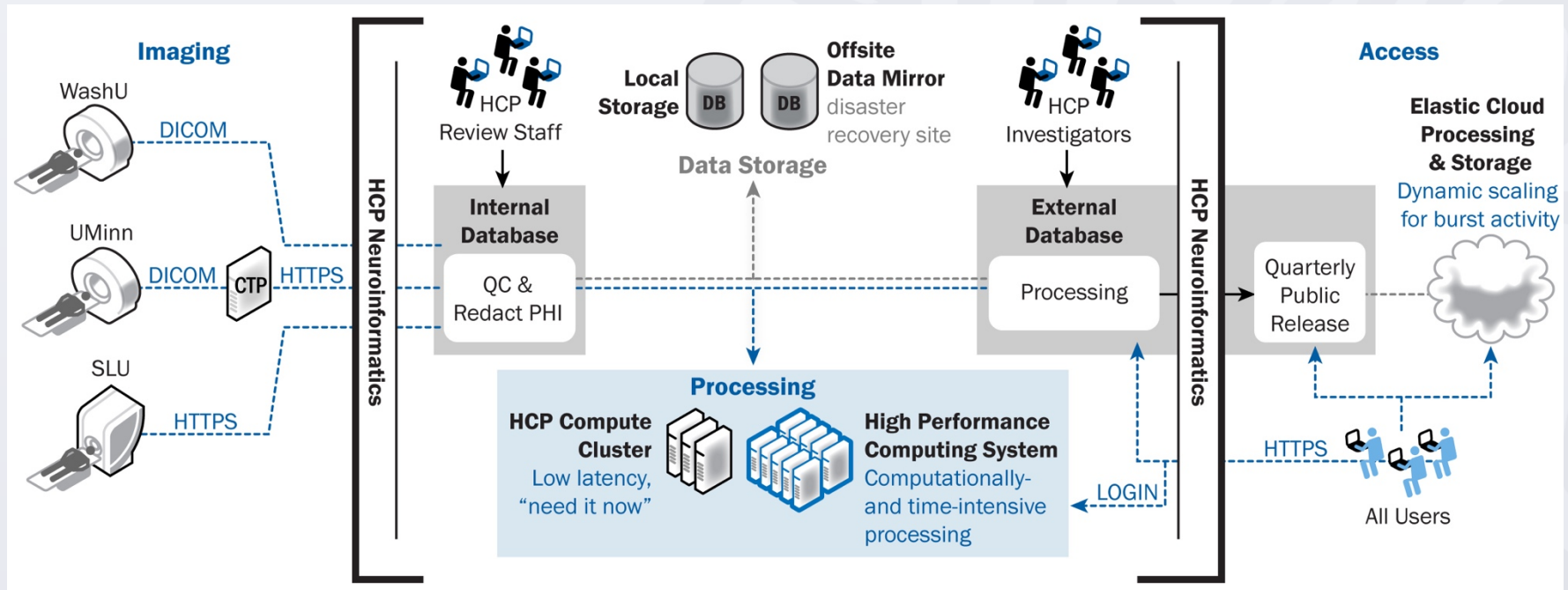
Strategies for using AWS

- Use Elastic Block Storage (EBS) for your own data.
- Use EC2 Spot Instances.
- Fork the NITRC HCP AMI to customize pipelines.

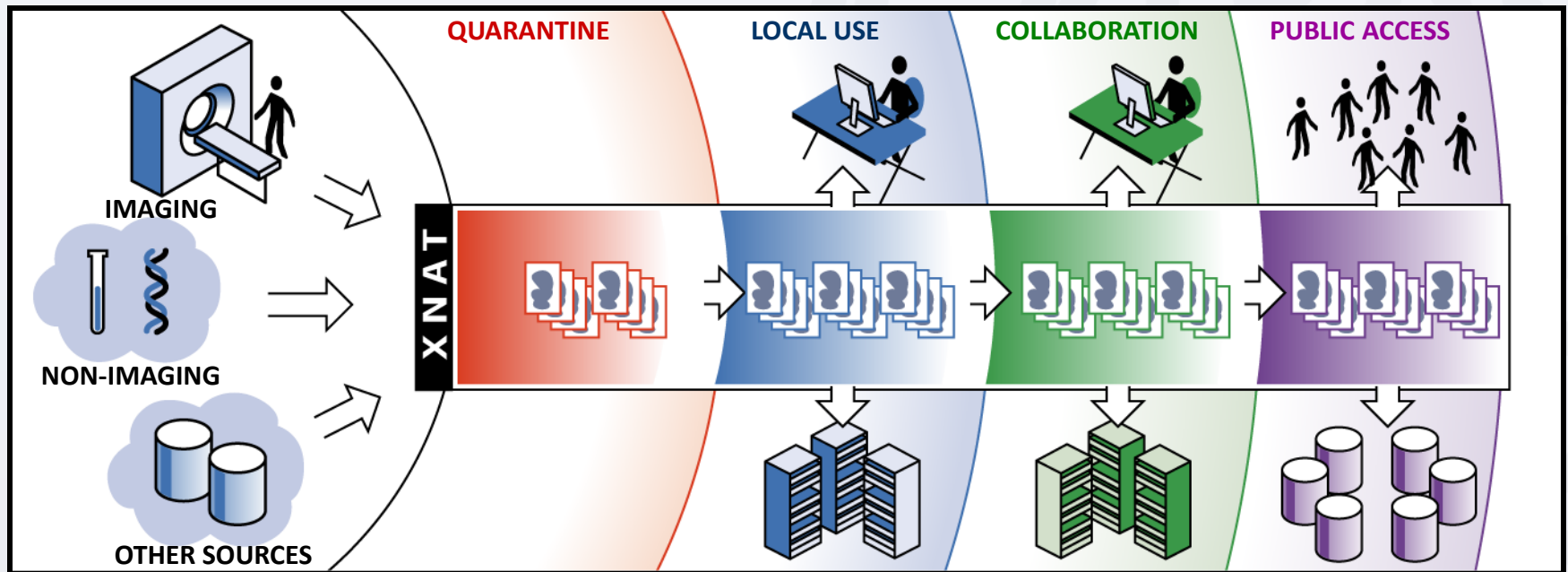
RUNNING YOUR OWN HCP INFORMATICS SYSTEM

Why run your own HCP Informatics System?

- You are running an HCP-like study.
- You enjoy pain.



The XNAT workflow



- Data organization
- Quality control
- Data access
- Security
- Visualization
- Automation
- Integration
- Data sharing

XNAT is...

Feature rich

Archive, manage, process, view,
and share imaging and related
data.

A platform

Clinical/translational research
Institutional repositories
Multi-center studies
Data sharing

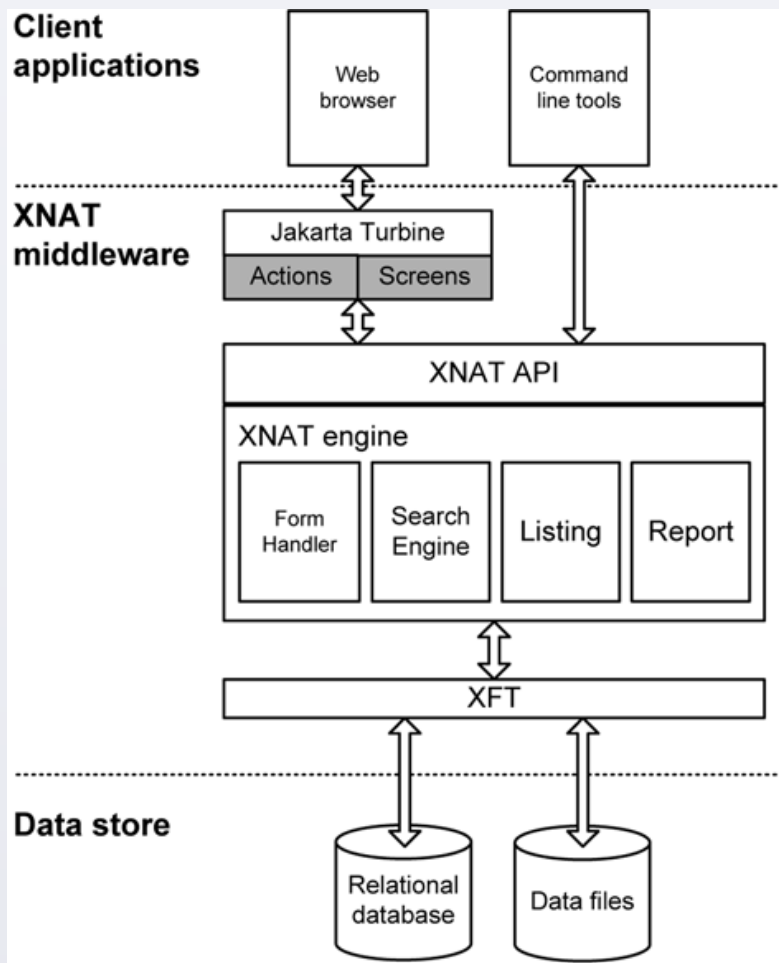
Open

Open source
Open API
Free (though commercial
support is available)
Used by organizations around
the world.

The screenshot displays the ConnectomeDB HCP Dashboard for HCP Q1 Release Data. The interface includes a header with the ConnectomeDB logo and a navigation bar. The main content area shows the project description and a table of filtered subjects. The table has columns for Subject, Gender, Age, Full Imaging Compl., T1 Count, T2 Count, Non-Toolbox Compl., and Visual I. The table is filtered to show subjects with Gender 'M' and Age '26-30'. The table is paginated to show 1 of 1 Pgs (19 Rows).

Subject	Gender	Age	Full Imaging Compl.	T1 Count	T2 Count	Non-Toolbox Compl.	Visual I
114924	M	26-30	false	2	1	true	true
118932	M	26-30	true	1	1	true	true
130013	M	26-30	true	2	2	true	true
142828	M	31-35	false	1	1	true	true
144428	M	22-25	false	0	0	true	true
149337	M	31-35	true	1	2	true	true
153429	M	31-35	true	2	2	true	true
156637	M	31-35	true	2	1	true	true
201111	M	26-30	true	1	1	true	true
304020	M	31-35	false	2	2	true	true
329440	M	26-30	false	1	1	true	true
530635	M	26-30	true	2	2	true	true
672756	M	31-35	true	2	2	true	true
685058	M	31-35	false	2	2	true	true
865363	M	22-25	true	2	2	true	true
889579	M	31-35	true	2	2	true	true
896879	M	26-30	true	2	2	false	true
917255	M	31-35	true	1	1	true	true
937160	M	26-30	true	1	1	true	true

XNAT Architecture



In a Nutshell, XNAT...

- ...has had **2,489 commits** made by **53 contributors** representing **643,792 lines of code**
- ...is **mostly written in Java** with **an average number of source code comments**
- ...has **a codebase with a long source history** maintained by **a very large development team** with **increasing Y-O-Y commits**
- ...took an estimated **176 years of effort** (COCOMO model) starting with its **first commit in April, 2010** ending with its **most recent commit 1 day ago**

Source: <https://www.ohloh.net/p/XNAT>

IntraDB Demo – Main Themes

- Receives data directly from scanners.
- Includes extensive QC pipelines and annotations.
- Accessible only to HCP personnel.
- Data that passes QC is exported to ConnectomeDB

The screenshot displays the IntraDB web application interface for MR Session 100307_strc. The interface includes a navigation bar with 'Details' and 'Projects' tabs, and an 'Actions' menu on the right. The main content area shows session details and a QC status summary.

MR Session: 100307_strc

Details

Accession #:	HCPIntradb_E04465	Subject:	100307
Date Added:	08/23/2012 18:39:57 (admin)	Gender:	Female
Date:	08/23/2012	Handedness:	
Time:	17:23:21	Age:	26
Operator:	OF EB		
Scanner:	HCP3T SIEMENS ConnectomS		
Acquisition Site:	WASH U		

Actions

- Edit
- Process
- Upload
- Download
- View XML
- View Images
- Email
- Add Rad Read
- Manage Files
- View in XimgView
- Delete
- Additional Processing

Notes: We decided to run the second set of T1/T2 scans with PMC off after the first two PMC scans didn't look great. At beginning of second T1 scan, scanner rebooted itself and gave Savelog. (OF). Camera calibration was corrected immediately prior to this session. -MPH

Session QC Status

Summary of QC

Scan	SIMPLESNR	SNR	SIMPLESNR SigMean	SIMPLESNR SigSTD
9(T1w_MPR1)	13.1	593.5		150.2
10(T2w_SPC1)	13.7	251.6		138.1
19(T1w_MPR2)	13.0	593.8		149.6
20(T2w_SPC2)	13.6	246.0		135.6
21(T1w_MPR2)	12.9	588.5		152.9

Scans

Scan	Type	Series Desc	Usability(Override)	Files	Note
1	Localizer	Localizer	unusable	809.2 KB in 6 files	
2	AAHScout	AAHScout	unusable	20.4 MB in 129 files	
3	AAHScout	AAHScout_MPR_sag	unusable	836.9 KB in 10 files	

Summary

- Get your data on ConnectomeDB, Connectome in a Box, or Amazon Public Data Sets.
- Don't be afraid to get to know the command line.
- Get ready for some cloud computing!

Thanks to...

- Amazon Web Services
 - Hosting of HCP data thru Public Data Sets project
 - \$100 credits to all course attendees
- Neuroimaging Informatics & Tools Clearinghouse (NITRC)
 - NITRC Computing Environment w/ HCP pipelines