

Open and Big Data for Life Imaging
March 30, 2015

Technologies for Data Sharing in Neuroimaging

David N. Kennedy, PhD.

Child and Adolescent NeuroDevelopment Initiative

Department of Psychiatry

University of Massachusetts Medical School

Overview

- **Why Share?**
- **Framework for Sharing**
- **Existing Solutions**
- **Main Problems**

Data Sharing is Key to Scientific Reproducibility

Reproducibility, in general, has a number of manifestations:

- **Publication-level Replication.** Take any given publication and cast it in a reproducible fashion. This is important due to the impact academia places on the publication as the principle currency of scientific productivity.
 - Eliminate errors
 - Make it complete (data, workflow, execution, results)
 - 3rd party certifiably re-executable
- **Generalizable Reproducibility** across publications. Huge problem since we dis-incentivize publication of replication studies as 'not novel'
 - Rerun same analysis on different 'similar' data (subjects and acquisition variables)
 - Rerun 'comparable' analysis on same data (methods variables)
 - Rerun same analysis on expanded 'similar' data (increase nominal power)

The Cost of Irreproducibility

- Irreproducibility costs money. Because of variability of results, more grants get funded to keep attempting to ‘finally’ resolve the mounting conundrum.
- Irreproducibility costs trust. Variability of results is part of why progress in neuropsychiatric disorders has been excruciatingly slow. Akin to NIMH’s Research Domain Criteria (RDoC), we advocate RAnC (Reliable Analysis Criteria).
- Irreproducibility costs lives. Any delay in turning findings into treatment or prevention has a tangible impact on the lives of millions of people worldwide.

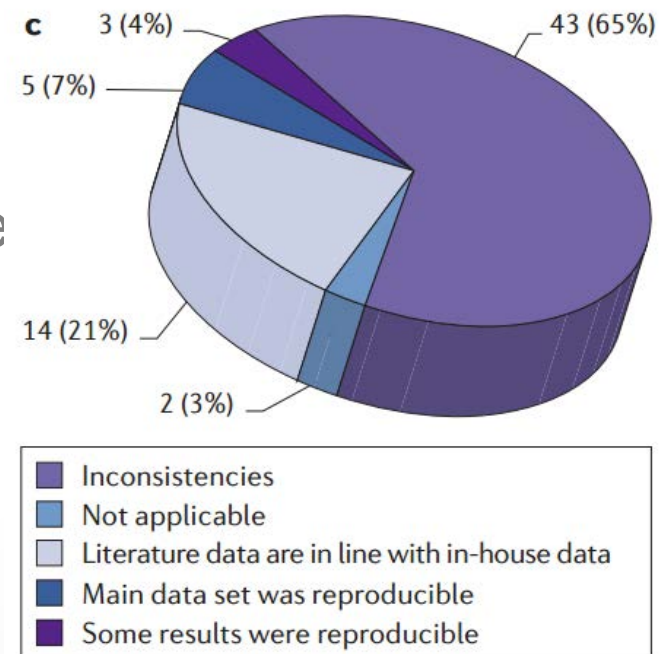


Chart from Prinz, et al. *Nature Reviews Drug Discovery* 10, 712 (September 2011)
Bayer Healthcare

Causes of Reproducibility/ Generalizability Issues

- Low power
- Mistakes
- Ineffective Data Sharing
- Methodological Variance
- Sampling
- P – hacking (P-phishing)
- Publication Bias

- Only reporting potentially significant results, and not the rest

P-Curve: A Key to the File-Drawer

Uri Simonsohn
University of Pennsylvania

Leif D. Nelson
University of California, Berkeley

Joseph P. Simmons
University of Pennsylvania

Because scientists tend to report only studies (publication bias) or analyses (*p-hacking*) that “work,” readers must ask, “Are these effects true, or do they merely reflect selective reporting?” We introduce *p-curve* as a way to answer this question. *P-curve* is the distribution of statistically significant *p* values for a set of studies ($ps < .05$). Because only true effects are expected to generate right-skewed *p-curves*—containing more low (.01s) than high (.04s) significant *p* values—only right-skewed *p-curves* are diagnostic of evidential value. By telling us whether we can rule out selective reporting as the sole explanation for a set of findings, *p-curve* offers a solution to the age-old inferential problems caused by file-drawers of failed studies and analyses.

Journal of Experimental Psychology: 2014, Vol. 143, No. 2, 534–5

NATURE REVIEWS | NEUROSCIENCE
VOLUME 14 | MAY 2013 | 365

Power failure: why small sample size undermines the reliability of neuroscience

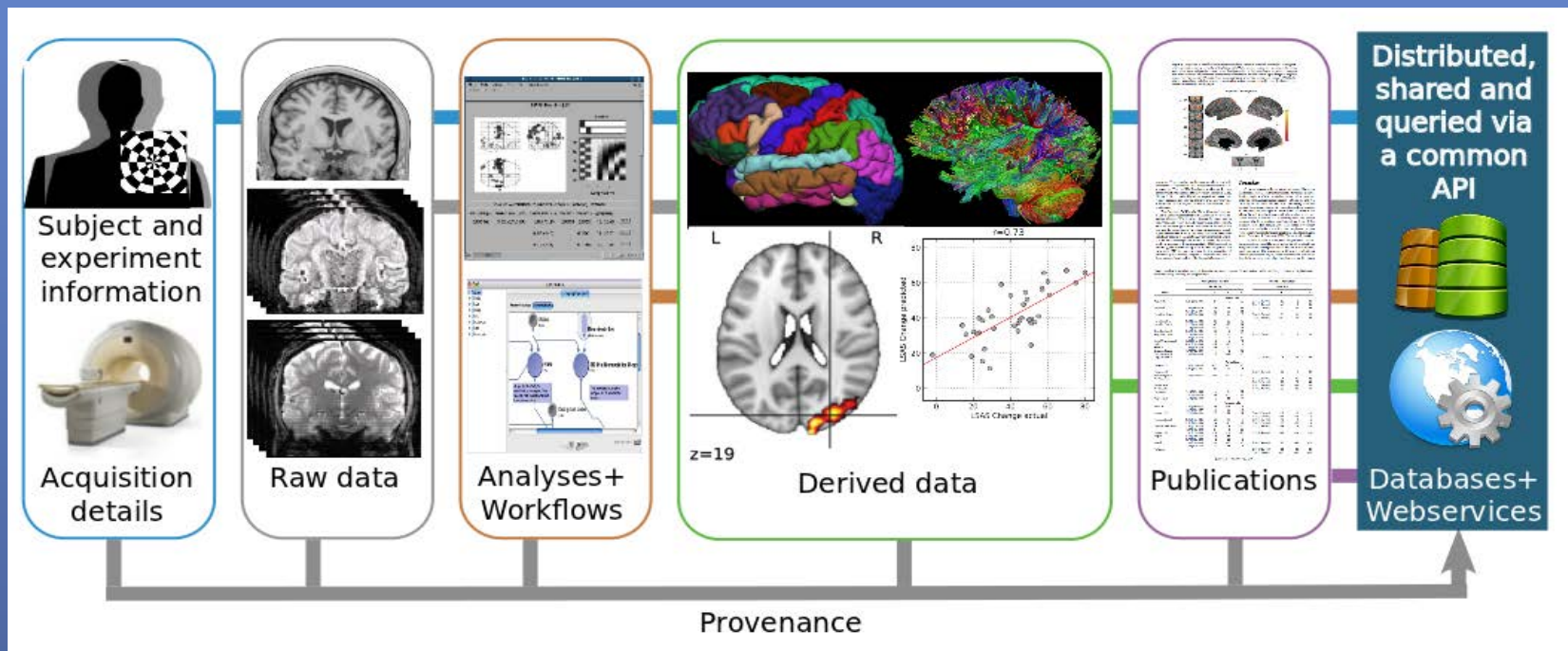
Katherine S. Button^{1,2}, John P. A. Ioannidis³, Claire Mokrysz¹, Brian A. Nosek⁴, Jonathan Flint⁵, Emma S. J. Robinson⁶ and Marcus R. Munafò¹

FAIR Data

- Data FAIRport (<http://www.datafairport.org/>): valuable scientific data should be 'FAIR':
Findable, Accessible, Interoperable and Re-usable.
- Provenance: provenance-aware data capture, analysis workflows, and results markup.
- Publication should include the necessary supplemental access to the requisite data, workflows and results to maintain maximum reproducibility of the scientific literature,
- Why? These principles are critical to the protection of the public investment in the scientific enterprise and to maximize the potential impact of this research in return for our participants' involvement in this project.
- How?
-

Framework

Data Flow and Stages of Data Sharing Opportunities



Stages of Information Sharing

(towards scientific reproducibility)

Data



Analysis



Results



Interpretation

**Data
Repositories**



**Standardized
Workflow
Description**



**Meta-Data
Repositories**



**Literature
Repositories**



**Mission: to develop metadata and data standards for reproducible research;
to develop standards for archiving, storing, sharing, and re-using neuroscience data and databases**



Electrophysiology Task Force

Focuses on electrophysiological data and databases

Use cases for handling and converting electrophysiology metadata

Web page on tools for converting between data formats

Further development of Neuroshare API

Continued engagement of industry vendors



Neuroimaging Task Force

Focuses on neuroimaging data and databases of neurological/psychiatric disorders and cognitive function

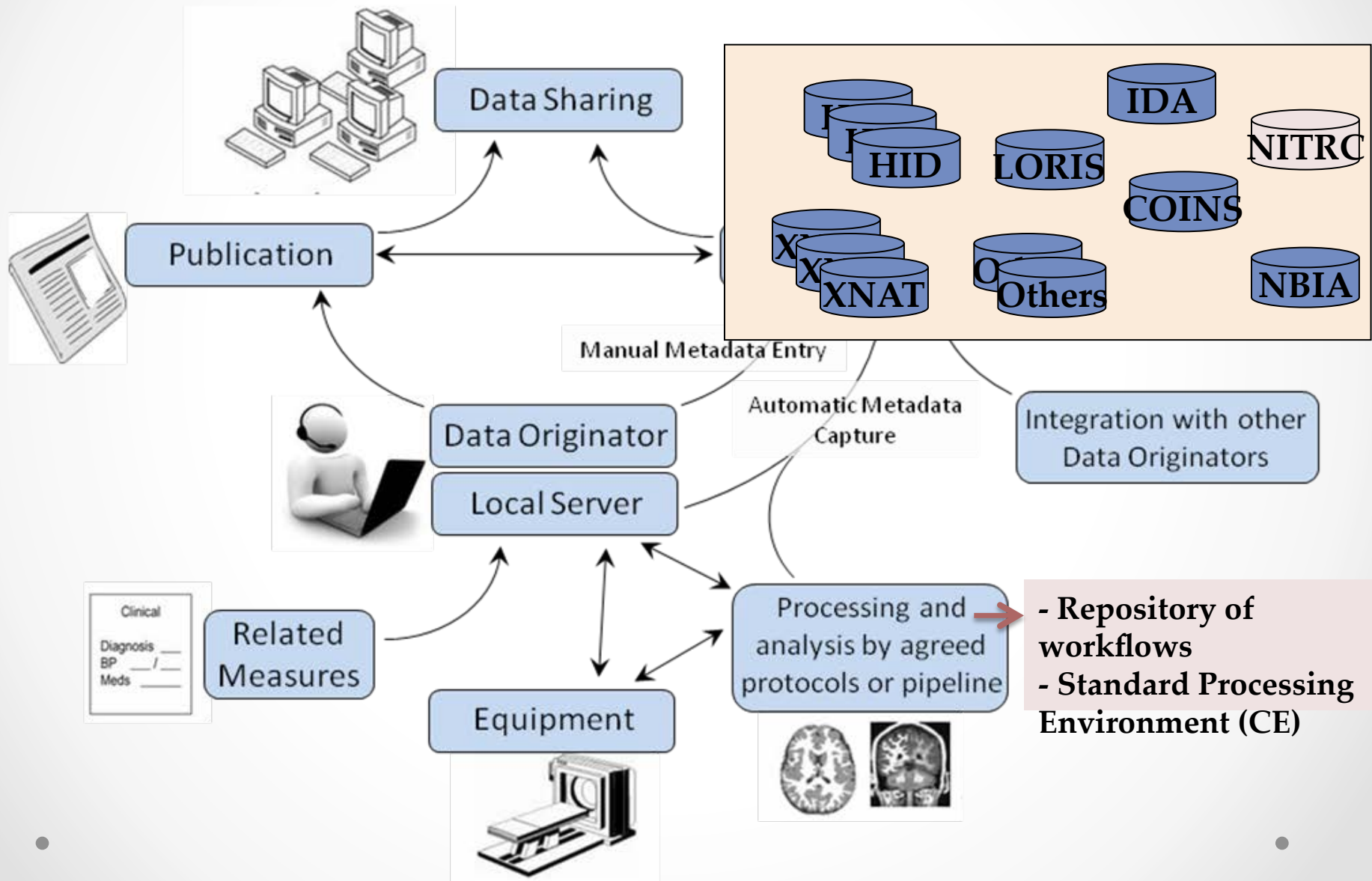
One-Click share tool

Standard description of neuroimaging data/metadata & common API

Standard data formats

Standard workflows and provenance tracking

Data-sharing infrastructure



Many Existing Resources

- Image (Raw Data) Level
 - Required
 - ADNI (IDA), NDAR (NIMH), HCP (XNAT), etc.
 - Optional
 - NITRC (XNAT), COINS, LORIS, OpenfMRI
- Derived Data
 - Statistical Maps
 - NeuroVault
 - Activation Foci
 - BrainMap, SuMSDB, etc.

XNAT has become the world. The following list out there.

Clinical and Translational CTSA are funded by the medical discovery to improve

- Brain-CODE at the
- Harvard Catalyst
- ICTS at the Univ
- ICTS at Washing
- Laboratory of Cl

The Human Connectome Project (HCP) is the foundation from 1,200 subjects at the backbone of the public

Institutional Operations XNAT is in use in neuro

- Biomedical Imag
- The Brain Imagin
- Centre of Excell
- Valencia
- Centre of Medic
- Cornell Universi
- Emory Universit
- The Genes_Cog
- Hoglund Brain Jr
- Institute of Canc
- Mayo Clinic Neu
- Medical-Image A
- Neuroimaging R
- New York State I
- Northwestern Un
- Psychiatric Neur
- The Rotman Res
- University of low
- University of Wa
- Vanderbilt Unive
- Wake Forest Uni

Multi-Site Studies

- Cardiovascular F
- IMAGEN Study
- Dominantly Inhe
- The Jackson He
- PREDICT-HD st
- The Injury and T
- International Ne
- GENetic Frontot
- Prospective Urb
- The Virtual Brain
- CENTER-TBI: la
- National Consor

Clinical / Pharmacology

- PharmaCog – Pr
- early clinical dev

Data Sharing

- OASIS Brains
- mBIRN Morpho
- XNAT Central
- Minimal Interval
- SchizConnect: S

The Qualitative Image

The Quantitative Image methods for the measurement facilitating clinical decisions

- PET-MRI for Ass
- Quantitative MR
- Quantitative MR
- Quantitative Image

- Technology
- Methodology
- Application

As neuroimaging research continues to grow, dynamic neuroinformatics systems are necessary to store, retrieve, mine, and share the massive amounts of data. The Collaborative Informatics and Neuroimaging Suite (COINS) has been created to facilitate communication

.....

LOGIN



LONI Image Data Archive

IDA HOME

ABOUT

NEWS

DOCUMENTATION

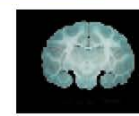
SOFTWARE

DATA

The LONI Image Data Archive (IDA) is a user-friendly environment for archiving, searching, sharing, tracking and disseminating neuroimaging clinical data. The IDA is utilized for dozens of neuroimaging research projects across North America and Europe and accommodates MRI, PET, and other imaging modalities. A flexible data de-identification engine and encrypted file transmission help ensure compliance with patient-privacy regulations. Data are stored on redundant servers with daily and weekly on- and off-site backups.

Archiving data in the IDA is simple, secure and requires no specialized hardware, software or personnel. All that is required is a computer with internet access and web browser software. The IDA automatically extracts relevant metadata from the de-identified image files allowing data to be searched and browsed. Once archived, data may be downloaded and/or streamed into the LONI Pipeline processing environment. Integration of the LONI Data format translation engine allows users to download image data in a number of file formats in addition to the original file format.

Image Data Storage, Protection & Sharing



ABOUT

- Overview

NEWS

- New Features
- Announcements

DOCUMENTATION

- Instructions
- Collaboration
- FAQ

SOFTWARE

- Debabeler
- De-identification Debabeler
- Inspector

DATA

- Available

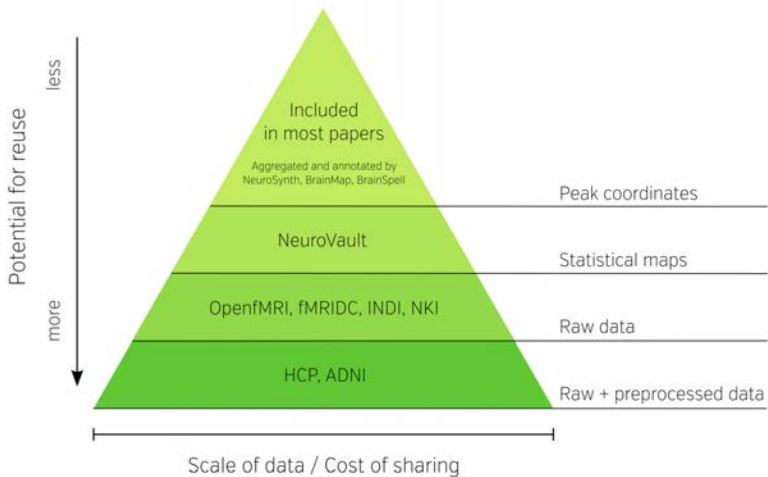


Download

© 2015 LONI. All rights reserved.

Transferring data from ida.loni.usc.edu...

Home Research



NeuroVault (beta)
Add new collection
See all collections
FAQ
Give feedback
Log in

NeuroVault

A public repository of unthresholded brain activation maps



What is it?

A place where researchers can publicly store and share unthresholded statistical maps produced by MRI and PET studies.

Why use it?

- Interactive visualization
- A permanent URL
- Publicly shareable
- Improves meta-analyses

Supported by

Get started and upload an image!

Latest collections of images

Name	Number of images
The integration of negative affect, pain and cognitive control in the cingulate cortex	3
The WU-Minn Human Connectome Project: An overview	47
The Default Mode of Human Brain Function Primes the Intentional Stance	6
Validating the Why/How contrast for functional MRI studies of Theory of Mind	5
Multiple brain networks contribute to the acquisition of bias in perceptual decision-making	6

Like 86
Tweet 282
+1 26
Share 26

Integrated Views

PING Data Portal - DataExploration

Multimodal Imaging Laboratory, UCSD
This project is funded by the [NIH / NIDA](#)



Projects ▾ Applications ▾ Documents Forum ▾ Help ▾ My account Log out

[Get Permalink To This Page](#)



Figure 1: The coefficient of determination R^2 (adjusted) is 0.3962025 ($n = 1099$). The AIC score of the full model is 24022.45.

R^2 (adj.): 0.3962025

Discovery and Action

The Three NITRC's

Neuroimaging Informatics Tools and Resources Clearinghouse
nitrc.org

- NITRC Resource Registry (NITRC-R)
- NITRC Image Repository (NITRC-IR)
- NITRC Computational Environment (NITRC-CE)



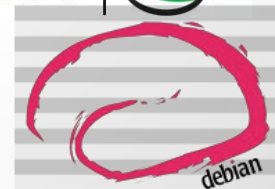
NITRC



NITRC



NITRC



NITRC Computational Environment (NITRC-CE)

<https://aws.amazon.com/marketplace>

- Powered by NeuroDebian
- Dynamic, 'summonable', cloud-based (Amazon, Azure, local VM) computational platform
- Scalable performance (1-16+ cores)
- FSL, AFNI, FreeSurfer, and many more



Amazon Web Services Home


Sign in or Create a new account | Your Account | Help | Sell on AWS Marketplace


CE Your Software

NITRC

[Visit the NITRC Website](#)

NITRC Products (2)

 **NITRC** **NITRC Computational Environment**
★★★★★ (3) | Version v0.30-all regions | Sold by [NITRC](#)
Free Tier Eligible
\$0.00/hr for software + AWS usage fees
NITRC-CE is a virtual computing platform pre-configured with many neuroimaging data analysis applications. NITRC-CE joins the family of successful NITRC services starting ...
Linux/Unix, Ubuntu 12.04 | 64-bit Amazon Machine Image (AMI)

 **NITRC** **NITRC Computational Environment for Cluster Compute Instances**
Version v0.30 CC | Sold by [NITRC](#)
\$0.00/hr for software + AWS usage fees
NITRC-CE for Cluster Compute Instances provides the same virtual computing platform for neuroimaging data analysis as NITRC-CE but on high performance computing machines ...
Linux/Unix, Ubuntu 12.04 | 64-bit Amazon Machine Image (AMI)

NITRC Computational Environment (NITRC-CE)

- Powered by NeuroDebian
- Dynamic, ‘summonable’, cloud-based (Amazon, Azure, local VM) computational platform
- Scalable performance (1-16+ cores)
- FSL, AFNI, FreeSurfer, and many more



aws marketplace

Amazon Web Services Home


Sign in or Create a new account

Your Account | Help | Sell on AWS Marketplace

Shop All Categories Search AWS Marketplace GO Your Software

NITRC Computational Environment

Sold by: NITRC

 NITRC-CE is a virtual computing platform pre-configured with many neuroimaging data analysis applications. NITRC-CE joins the family of successful NITRC services starting with the flagship, NITRC-Resources, the “go to” place for neuroimaging tools and resources. NITRC Image Repository offers a select set of community-generated neuroimaging data sets, while this service, NITRC Computational Environment, offers the convenience of cloud-based computing against NITRC-IR data sets or your data sets. We welcome any suggestions on how to improve this service to make it a user friendly tool for ... [Read more](#)

Customer Rating	★★★★★ (3 Customer Reviews)
Latest Version	v0.30-all regions (Other available versions)
Base Operating System	Linux/Unix, Ubuntu 12.04
Delivery Method	64-bit Amazon Machine Image (AMI) (Learn more)
Support	See details below
AWS Services Required	Amazon EC2, Amazon EBS

Highlights

- Need resources on demand to compute against your neuroimaging data? Tired of fighting for institutional compute resources and just need to get the compute done? Use NITRC-CE!
- Need access to the most popular neuroimaging analysis tools? Each release has more of the most popular neuroimaging tools, check our User Guide for a complete listing of installed packages. Use these resources separately, or pipeline them; we're agnostic!
- Need access to the most popular community-generated and curated neuroimaging analysis data sets? Access

Continue

You will have an opportunity to review your order before launching or being charged.

Pricing Details

For region **US East (Virginia)**

Free Tier Eligible

This product can be used for free on a Micro instance for up to 750 hours per month if you qualify. See details.

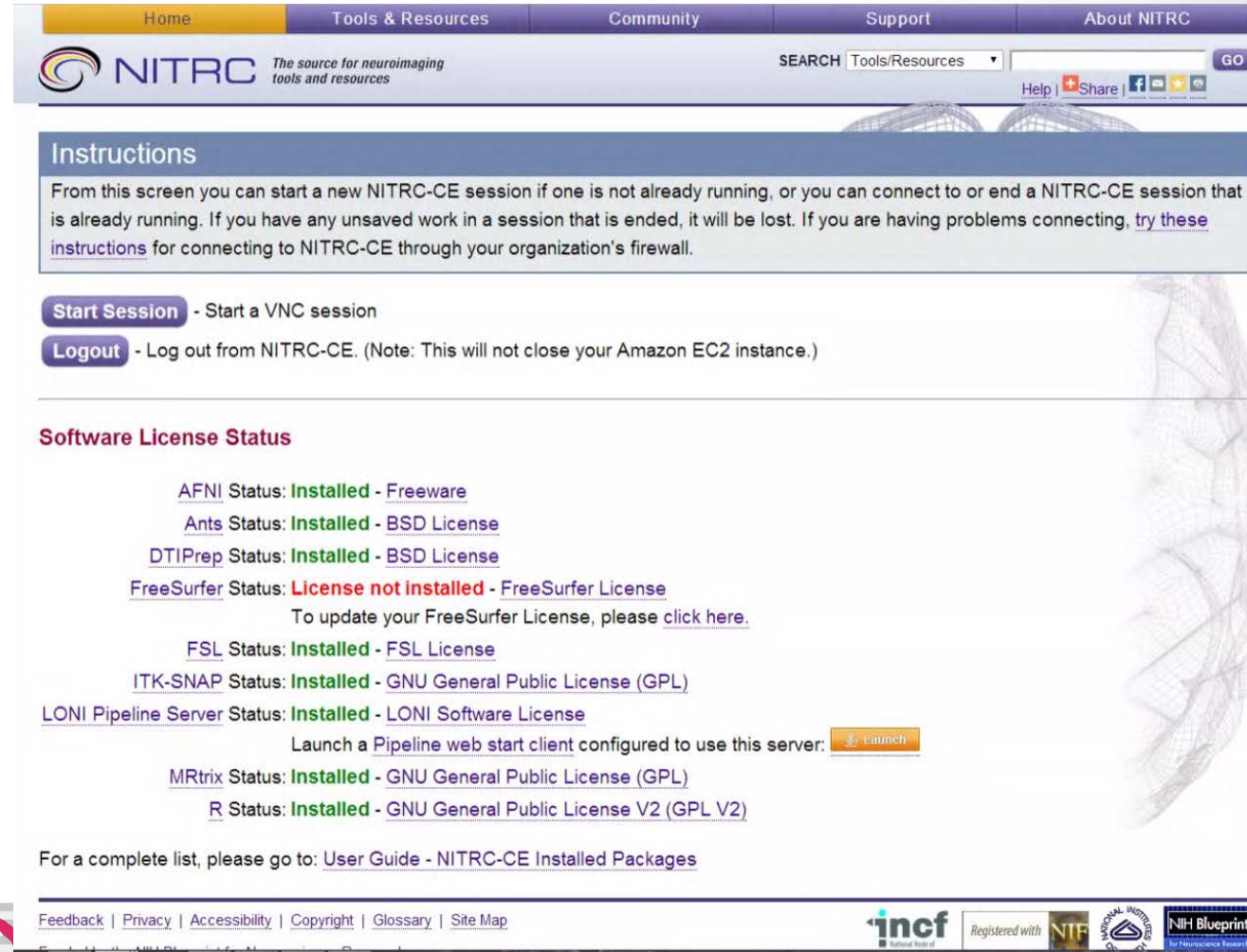
Hourly Fees

Total hourly fees will vary by instance type and EC2 region.

EC2 Instance Type	Software	EC2	Total
Standard Micro (t1.micro)	\$0.00/hr	\$0.02/hr	\$0.02/hr
Standard Small (m1.small)	\$0.00/hr	\$0.06/hr	\$0.06/hr
Standard Medium (m1.medium)	\$0.00/hr	\$0.12/hr	\$0.12/hr
Standard Large (m1.large)	\$0.00/hr	\$0.24/hr	\$0.24/hr
Standard XL (m1.xlarge)	\$0.00/hr	\$0.48/hr	\$0.48/hr
High-Memory XL (m2.xlarge)	\$0.00/hr	\$0.41/hr	\$0.41/hr
High-Memory 2XL (m2.2xlarge)	\$0.00/hr	\$0.82/hr	\$0.82/hr
High-Memory 4XL (m2.4xlarge)	\$0.00/hr	\$1.64/hr	\$1.64/hr

NITRC Computational Environment (NITRC-CE)

- Powered by NeuroDebian
- Dynamic, ‘summonable’, cloud-based (Amazon, Azure, local VM) computational platform
- Scalable performance (1-16+ cores)
- FSL, AFNI, FreeSurfer, and many more



The screenshot shows the NITRC-CE web interface. At the top is a navigation bar with links: Home, Tools & Resources, Community, Support, and About NITRC. Below this is a header section with the NITRC logo and tagline 'The source for neuroimaging tools and resources'. A search bar is located on the right. The main content area is titled 'Instructions' and provides information on starting or connecting to a NITRC-CE session. Below the instructions are two buttons: 'Start Session' (to start a VNC session) and 'Logout' (to log out from NITRC-CE). The 'Software License Status' section lists various software packages and their license status: AFNI (Installed - Freeware), Ants (Installed - BSD License), DTIPrep (Installed - BSD License), FreeSurfer (License not installed - FreeSurfer License), FSL (Installed - FSL License), ITK-SNAP (Installed - GNU General Public License (GPL)), LONI Pipeline Server (Installed - LONI Software License), MRtrix (Installed - GNU General Public License (GPL)), and R (Installed - GNU General Public License V2 (GPL V2)). A 'Launch' button is also present. At the bottom, there is a footer with links for Feedback, Privacy, Accessibility, Copyright, Glossary, and Site Map, along with logos for incf, NITF, and NIH Blueprint.

Home Tools & Resources Community Support About NITRC

NITRC The source for neuroimaging tools and resources

SEARCH Tools/Resources GO

Help | Share |

Instructions

From this screen you can start a new NITRC-CE session if one is not already running, or you can connect to or end a NITRC-CE session that is already running. If you have any unsaved work in a session that is ended, it will be lost. If you are having problems connecting, [try these instructions](#) for connecting to NITRC-CE through your organization's firewall.

Start Session - Start a VNC session

Logout - Log out from NITRC-CE. (Note: This will not close your Amazon EC2 instance.)

Software License Status

AFNI Status: **Installed** - [Freeware](#)

Ants Status: **Installed** - [BSD License](#)

DTIPrep Status: **Installed** - [BSD License](#)

FreeSurfer Status: **License not installed** - [FreeSurfer License](#)
To update your FreeSurfer License, please [click here](#).

FSL Status: **Installed** - [FSL License](#)

ITK-SNAP Status: **Installed** - [GNU General Public License \(GPL\)](#)

LONI Pipeline Server Status: **Installed** - [LONI Software License](#)
Launch a [Pipeline web start client](#) configured to use this server: [Launch](#)

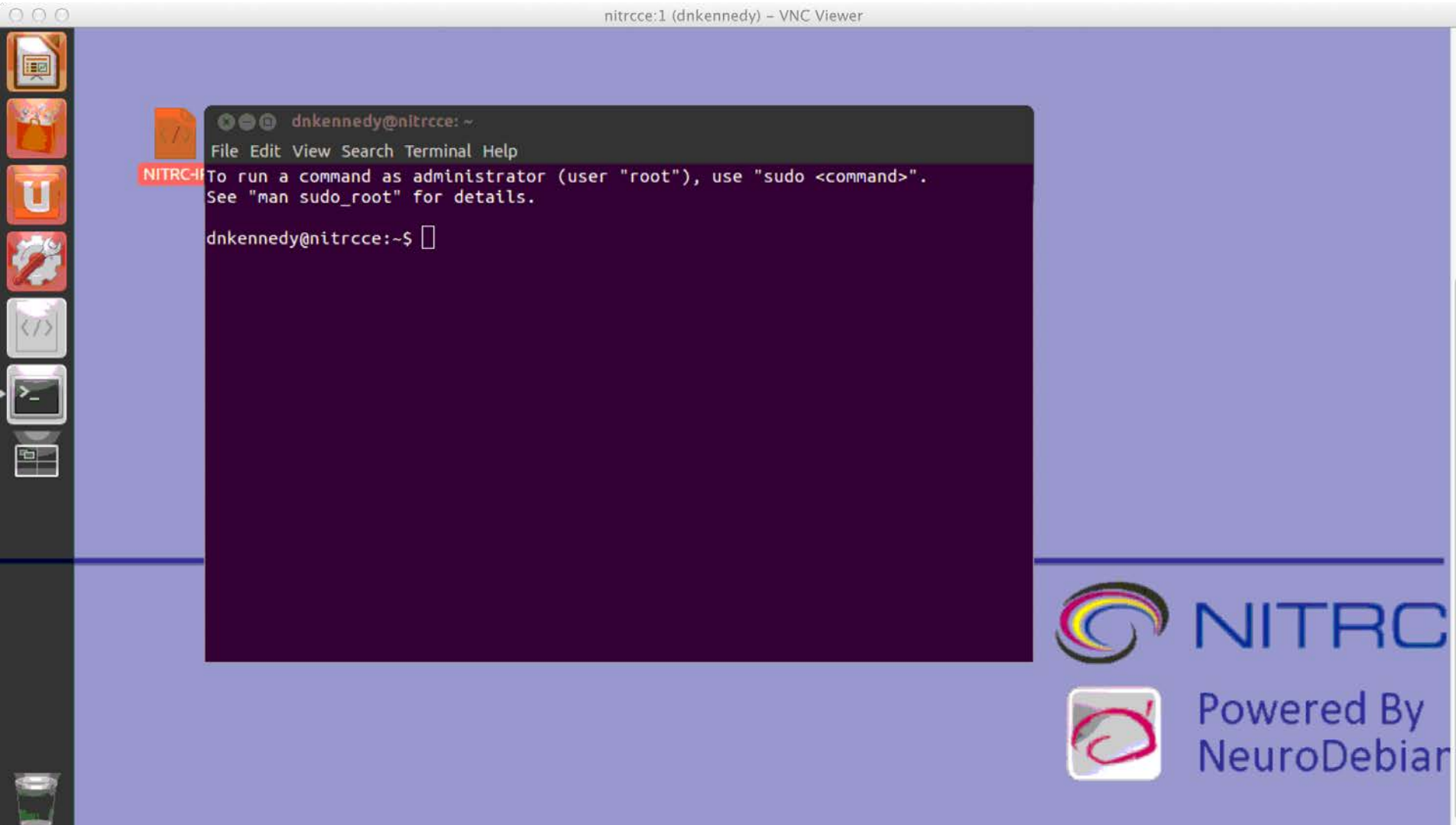
MRtrix Status: **Installed** - [GNU General Public License \(GPL\)](#)

R Status: **Installed** - [GNU General Public License V2 \(GPL V2\)](#)

For a complete list, please go to: [User Guide - NITRC-CE Installed Packages](#)

[Feedback](#) | [Privacy](#) | [Accessibility](#) | [Copyright](#) | [Glossary](#) | [Site Map](#)

incf National Institute of Neuroinformatics
Registered with **NITF**
National Institutes of Health
NIH Blueprint for Neuroscience Research

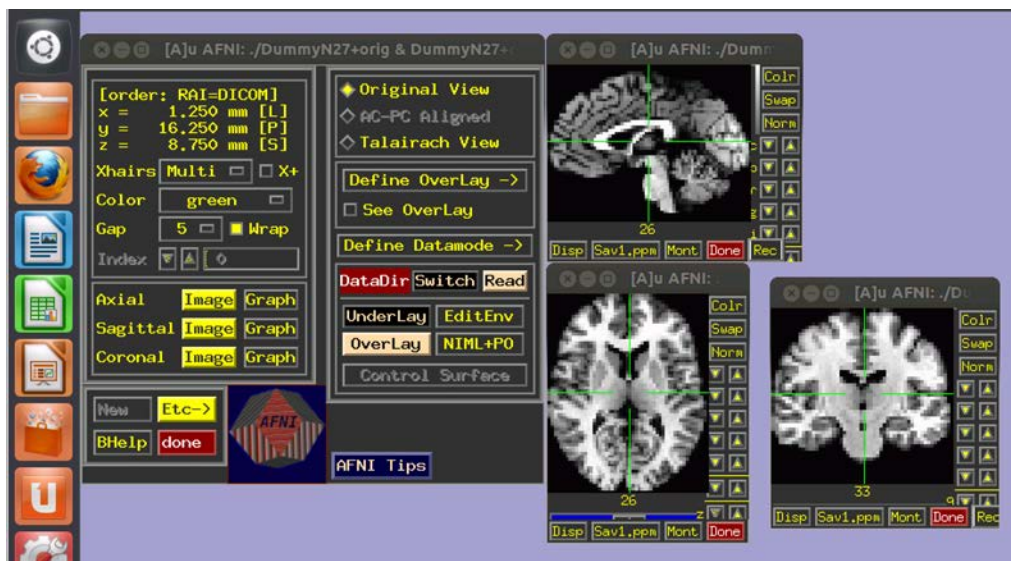
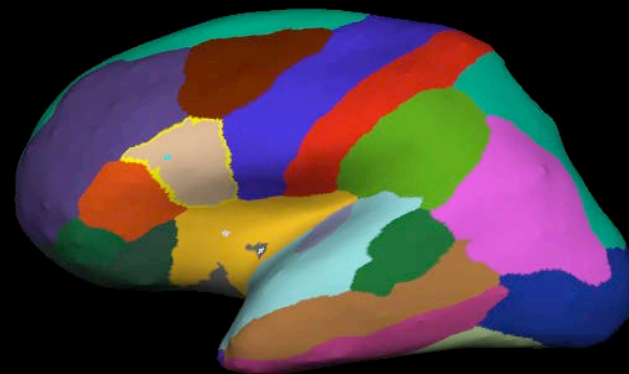
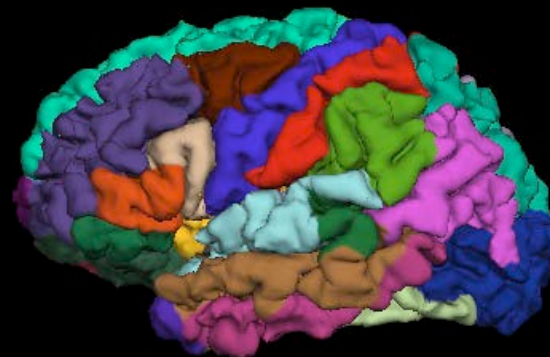
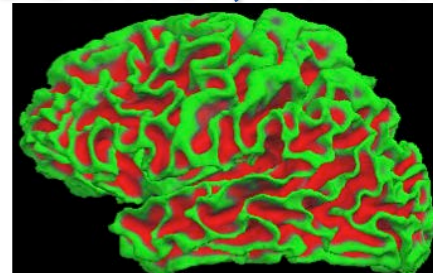
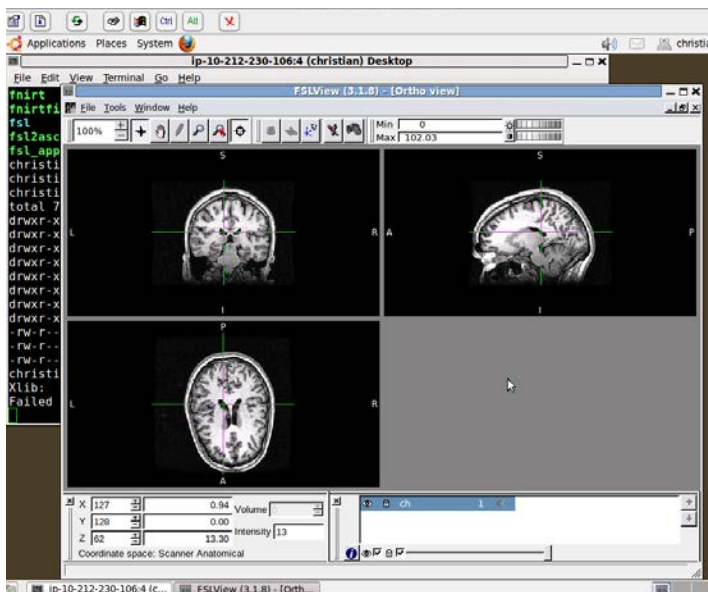




NITRC



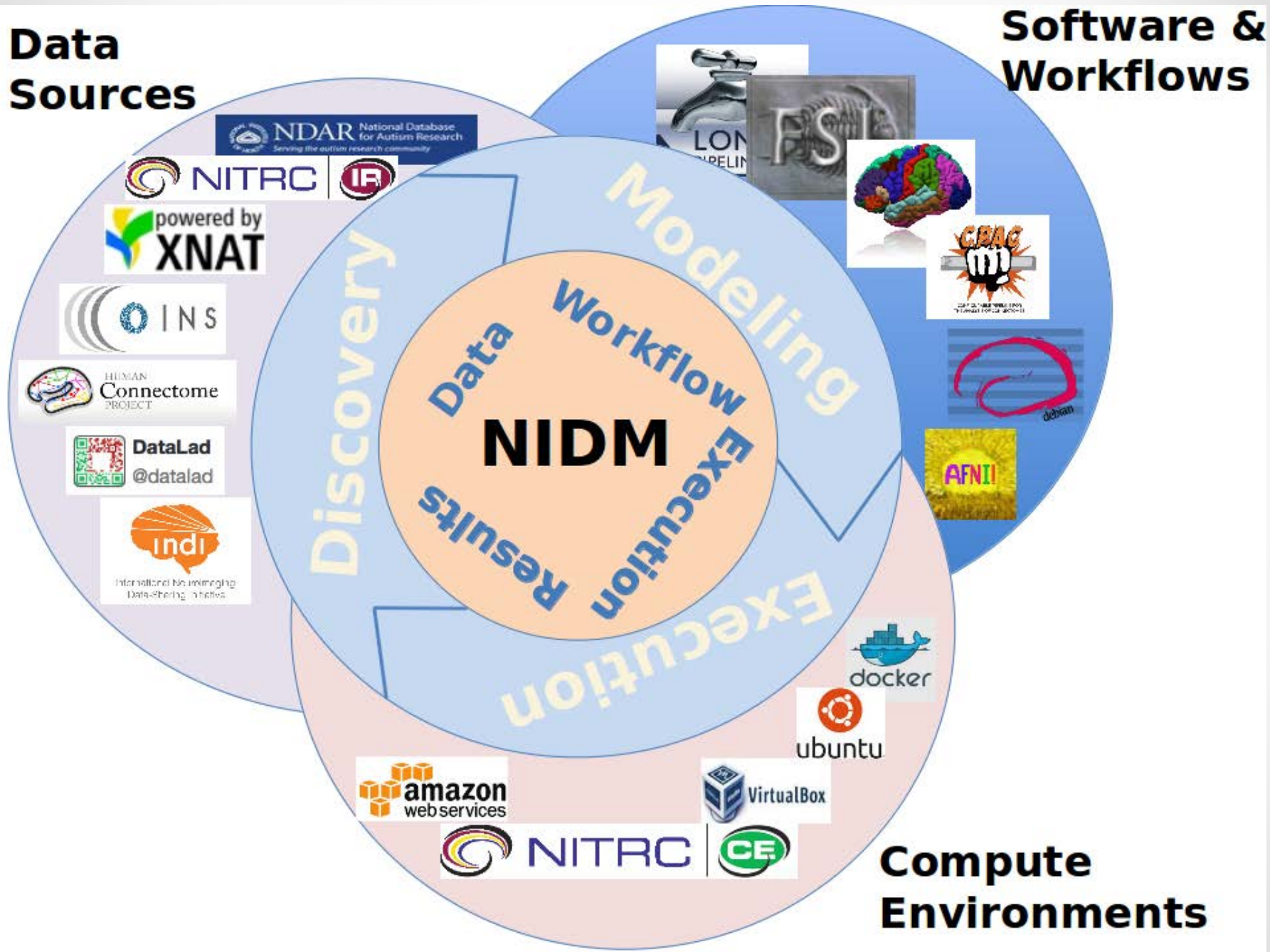
Sample Views: FSL, FreeSurfer, AFNI



NITRC ...The Source for Neuroimaging Tools and Resources

Data Sources

Software & Workflows





Provenance and NI-DM:

- The NI-DM data model
- Capture content from data and workflows (provenance)
- Apps for storing and querying NI-DM stores

Provenance and NI-DM:

- **The NI-DM data model**

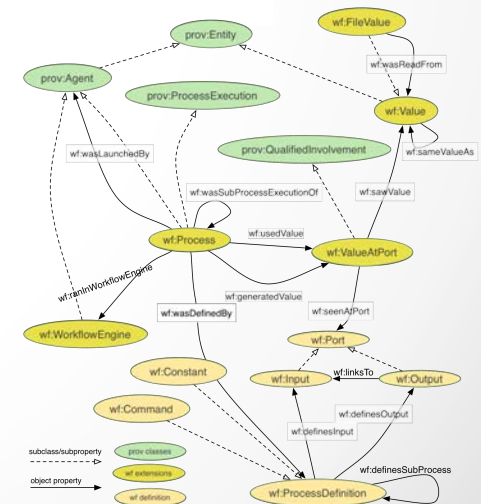
Neuroimaging Data Model (NI-DM) (Keator et al., 2013)

- Based on PROV-DM hence borrows PROV ontology (PROV-O)
 - Structured information encoding
 - Consistent vocabulary
 - Metadata standards via domain specific object models
-
- **Terms**
A lexicon of all things brain imaging. (e.g., DICOM terms, software specific terms, statistic terms, paradigm terms)
 - **Object Models**
Structured information in brain imaging (e.g., directory structures, CSV/Tab delimited files, brain imaging file formats)
 - **Integrated provenance**
How are entities generated or derived and by what or who?

The diagram illustrates the NIDM stack architecture, organized into five levels:

- Level 4:** NIDM Dataset Descriptor (orange bar)
- Level 3:** NIDM Experiment (Data Acquisition), NIDM Workflow (Image Processing), and NIDM Results (Statistical Model) (colored boxes with logos and arrows).
 - NIDM Experiment:** Includes logos for OpenfMRI, XNAT, OINS, and HUMAN CONNECTOME PROJECT.
 - NIDM Workflow:** Includes logos for FSL, SPM, NITRC, and GE.
 - NIDM Results:** Includes a flowchart showing steps like Model fitting, Inference, and a graph showing a statistical model.
- Level 2:** NIDM Core Vocabulary (blue bar)
- Level 1:** PROV Family of Specifications (purple bar)
- Level 0:** Semantic Web Technologies (SPARQL, RDF) (grey bar)

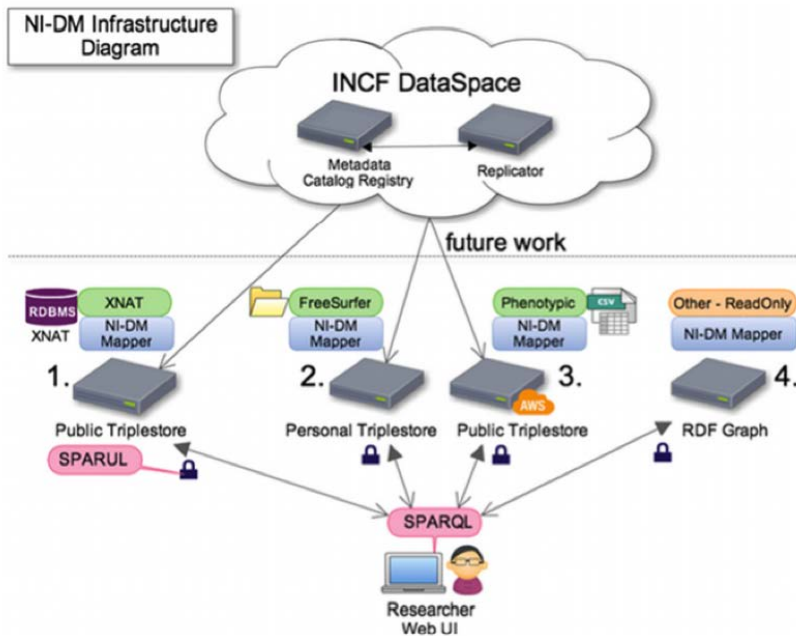
A vertical pink bar on the right is labeled **NIQuery API**. A small graph on the right shows relationships between Agent, Entity, and Activity.



Workflow Ontology (WFO) extensions for PROV-DM

Provenance and NI-DM:

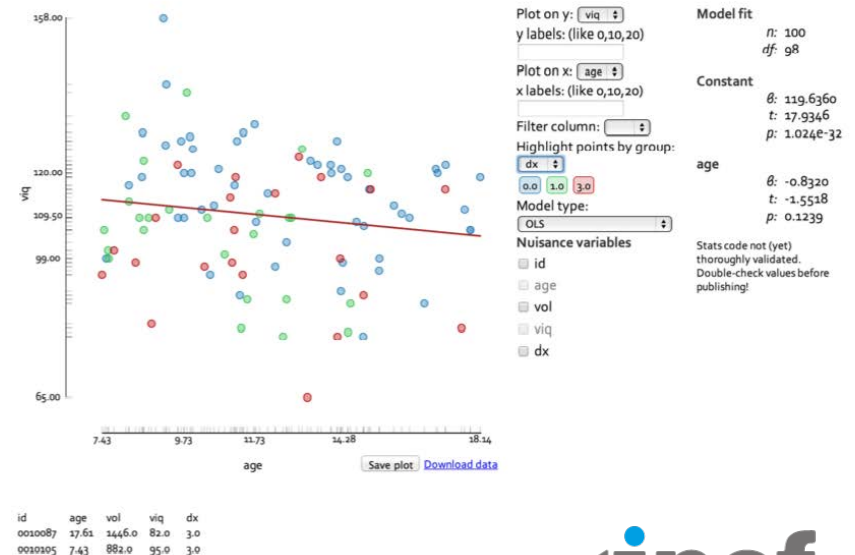
- Capture content from data and workflows (provenance)
- Apps for storing and querying NI-DM stores



Federated Query

App call: <http://localhost:5000/?url=http://bit.ly/1atAL00>

Scatterize: <https://github.com/njvack/scatterize>



Problems & Barriers

- **Incentive**
- **Cost**
- **Credit**
- **Ethics**
- **Harmonization**
- **Reproducibility**

Incentive Barrier

- Funders & Publishers beginning to require sharing
- PubMed Central 'success story' for open publications
- Education & Simplification needed

The 'Cost' Barrier

The Sharing Tax

(aka Data Persistence Insurance)

- Sharing costs money: someone pays for it
- Well, not just ‘someone’, but rather the researcher themselves, regardless of:
 - If centrally supported
 - Imposed on the researchers themselves
 - Mandated in the costs of a specific program
- Given a finite budget, the cost of the sharing tax comes out of the number of subjects from whom data is acquired

~~The Sharing Tax~~

Data Persistence Insurance, continued

- What is the magnitude of the ~~sharing tax~~ persistence insurance?
 - ~5-10% of the cost of acquisition
- What return do we need on this?
 - ~10-20%, at least, in improved scientific 'value'
- Now we have to document the real world benefit of sharing
- Or else...

ANGRY MOB FUN RUN



One modest proposal:

- Funding agencies could set a future ‘market’ for ‘commercial’ data archival (10 years) at a target price.
 - i.e. for neuroimaging, a target cost equal to or less than 5% of the acquisition cost: is a typical hour of MRI scanning costs \$500, then this would be \$25 for archival of an typical 400MB scanning session
- The magnitude of this ‘potential future market’ per year can be estimated by the number of funded MRI sessions by the funding agency multiplied by the target price
 - This is a pretty big market (~\$25-50M), should entice commercial or other vendors
- This ‘costs’ the funding agency nothing
 - i.e. they pay the same as they would otherwise; cost would be 5% reduction in subjects, in exchange for guaranteed future data accessibility

Neuroinform (2014) 12:361–363
DOI 10.1007/s12021-014-9239-0

EDITORIAL

Data Persistence Insurance

David N. Kennedy

Credit Barrier

Nested DOIs

- Publication DOI
 - DataSet DOI
 - Individual DataElement DOIs

New publications using shared data cite the original Individual DataElement DOIs in the new DataSet DOI as part of the New Publication DOI

Credit where credit is due: Through the citation of the original Individual DataElement DOIs in new publication datasets, original data providers and original publications are cited

Data DOI's

- For tracking pooled use credit
- Documentation of of novel cohorts
- Cohort DOIs
- Derived data DOIs (or handles)

Creator *

Name: *

David N. Kennedy

Name Identifier:

0000-0002-9377-0797

Identifier Scheme:

ORCID

Scheme URI:

<http://www.orcid.org>

Title *

Title: *

CANDISHARE_HC001_procmg

Type:

☒ Main title ☐ Alternative title ☐ Subtitle ☐ Translated title

Publisher: *

UMass Child and Adolescent NeuroDevelopment Initiative

Publication Year: *

2014

Resource Type:

Dataset

Resource Type Description:

MRI

Abstract

Descriptive information:

MRI scan of a 13.5 year old male healthy control subject.

Type:

Abstract

Subject

Subject:

Magnetic Resonance Imaging

Scheme:

MeSH

Scheme URI:

<http://www.ncbi.nlm.nih.gov/mesh>

Date

Date:

2014-07-18

Date Type:

Submitted

Alternate Identifier

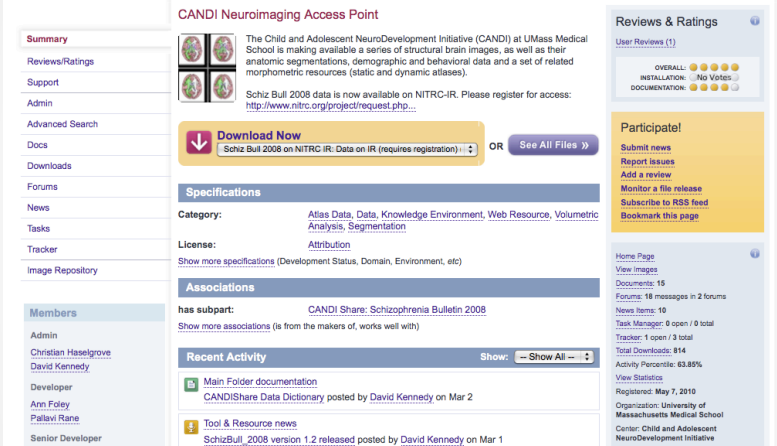
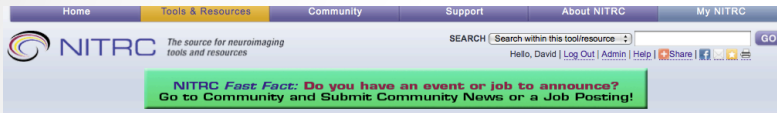
Identifier:

HC_001_V0.5

Identifier Type:

<http://www.nitrc.org>

Integrated Publication



CANDIShare - Downloads

Release	Date	Filename	Size	DL	Arch	Type
Dynamic Atlasses	2012-02-28 16:44					
		Dynamic Probability Server (url)		14	Any	URL
SchizBul_Atlas_V1.1	2012-02-14 18:19					
		Atlasses_V1.1.tar.gz	8.26 MB	5	Any	.gz
SchizBul_Atlas_V1.8	2011-12-05 03:33					
		Atlasses_V1.0.tar.gz	3.29 MB	11	Any	.gz

Package: [SchizBul_2008g2](#) 2 Subscribers

Release	Date	Filename	Size	DL	Arch	Type
Schiz Bul 2008 on NITRC IR	2012-02-28 19:41					
		Data on IR (requires registration) (url)		22	Any	URL
ImprSegReg_V1.2	2012-02-14 21:59					
		SchizBul_2008_BFDwithPay_segimgreg_V1.2.tar.gz	113.56 MB	5	Any	.gz
		SchizBul_2008_BFDwithPay_segimgreg_V1.2.tar.gz	61.01 MB	4	Any	.gz
		SchizBul_2008_HG_segimgreg_V1.2.tar.gz	90.21 MB	7	Any	.gz
		SchizBul_2008_SS_segimgreg_V1.2.tar.gz	66.14 MB	4	Any	.gz
ImprSeg_V1.1	2011-07-08 16:11					
		SchizBul_2008_BFDwithPay_segimgtar.gz	89.48 MB	14	Any	.gz
		SchizBul_2008_BFDwithPay_segimgtar.gz	48.02 MB	11	Any	.gz
		SchizBul_2008_HG_segimgtar.gz	70.32 MB	25	Any	.gz
		SchizBul_2008_SS_segimgtar.gz	51.68 MB	11	Any	.gz
SS_1.0	2010-07-27 23:11					
		Link to this group in IBVO (url)		174	Any	URL
		SchizBul_2008_SS_Procimgtar.gz	49.02 MB	22	Any	.gz
BFDwithPay_1.8	2010-07-27 17:56					
		Link to this group in IBVO (url)		109	Any	URL
		SchizBul_2008_BFDwithPay_procimgtar.gz	45.52 MB	17	Any	.gz
BFDwithPay_1.0	2010-07-08 00:36					
		Link to this group in IBVO (url)		68	Any	URL
		SchizBul_2008_BFDwithPay_procimgtar.gz	54.87 MB	16	Any	.gz
HG_1.0	2010-07-01 16:00					
		Link to this group in IBVO (url)		83	Any	URL
		SchizBul_2008_HG_procimgtar.gz	66.38 MB	24	Any	.gz

[Link to Dynamic Atlases](#)

[Link to Static Atlases](#)

Link to NITRC-IR

Images & Segmentation

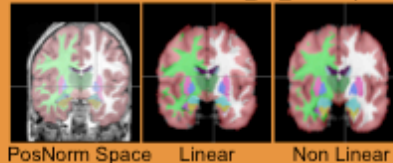
Image Downloads
and Links to IBVD



NITRC-IR

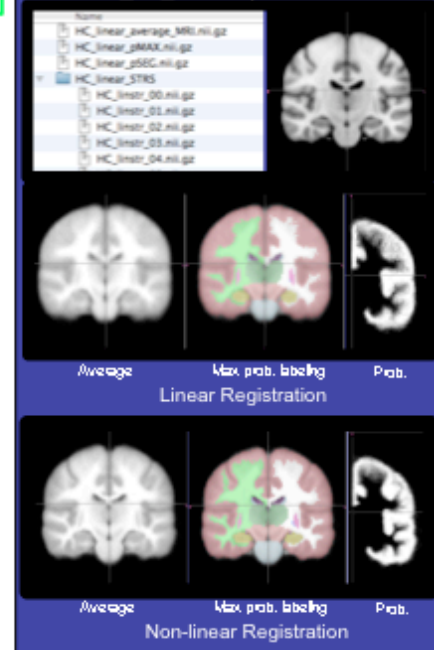


Individual Data MNI152_T1_2mm Space



PosNorm Space	Linear	Non Linear
---------------	--------	------------

'Static' Probability Atlases
MNI152 T1 2mm Space



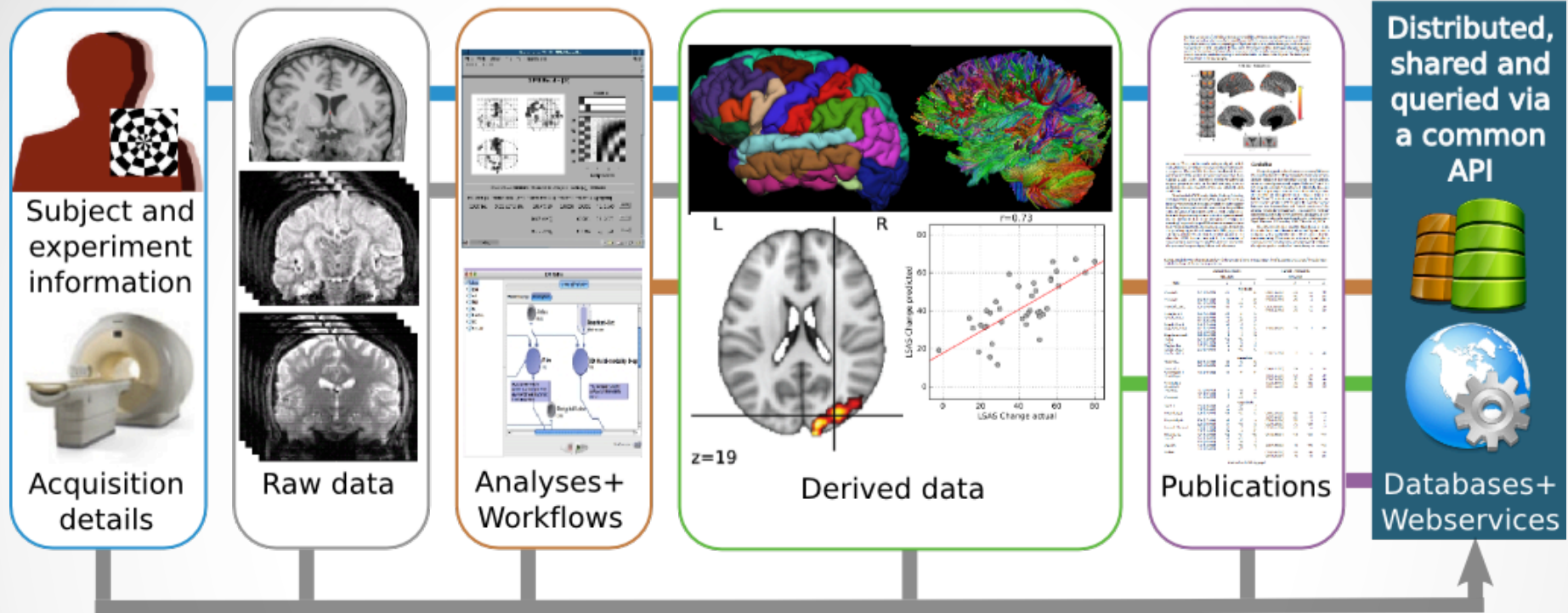
http://www.nitrc.org/projects/candi_share/

The 'Reproducibility' Barrier

[illegible]

In case you didn't think there was a problem with the current publication system... Multiple small N studies that themselves are not repeatable will not likely lead to a consistent view of a complex disease process...

Framework - Revisited



Provenance

Demographics
Stimuli
Scanner
Sequence

Assays
- Clinical
- Neuropsych
- Behavioral

MRI
- T1
- T2
dMRI
fMRI
- Task
- Resting
EEG, MEG,
PET, MRS

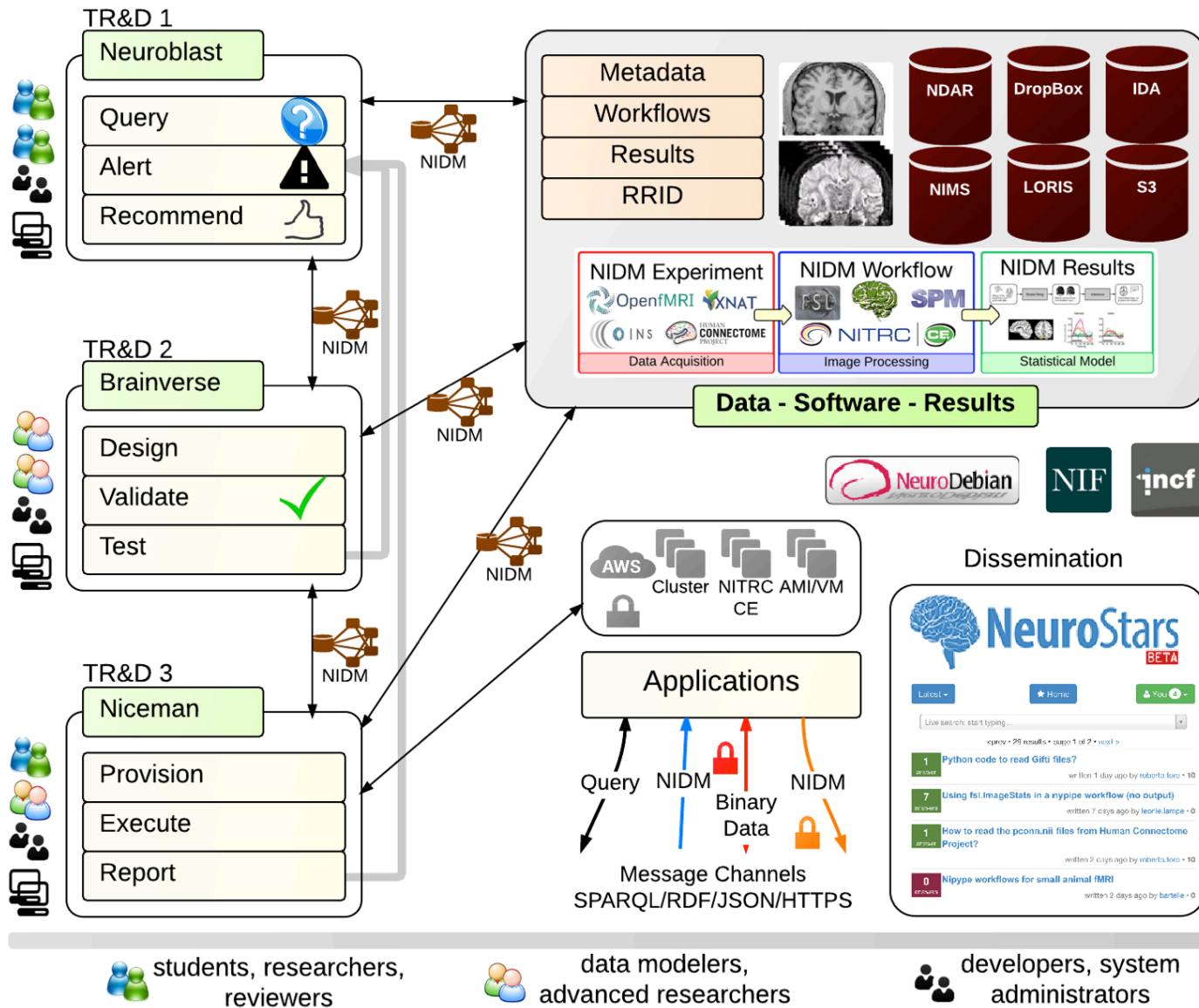
SPM
FSL
FreeSurfer
AFNI
...
Nipype
LONI
...
QA/QC

Contrasts
Correlation
Prediction

Journals
Preprint
archives
Websites
Blogs

XNAT
IDA
LORIS
COINS
Amazon S3
DropBox
Globus
Neurovault
Neurosynth

A User-centric Model



Part 1:

- User 1 writes a paper (structural MRI analysis) on 103 subjects (mixed gender, w/ and w/o a specific diagnosis, age range 5-18, etc.)
- Data is shared (as per publication requirement) in NITRC.
- 'Data' includes raw MRI scans and the results of specific segmentation workflow.
- Publication1 gets a doi (Pub1doi)

- Shared Data:

U1Sub1raw.nii

U1Sub1proc.nii

U1Sub2raw.nii

U1Sub2proc.nii

▪

▪

▪

U1Sub103raw.nii

U1Sub103proc.nii

Diagnostic and Sex Effects on Limbic Volumes in Early-Onset Bipolar Disorder and Schizophrenia

Jean A. Frazier^{1–4}, Steven M. Hodge⁵, Janis L. Breeze^{2,3}, Anthony J. Giuliano^{2,3,6}, Janine E. Terry³, Constance M. Moore^{2,3,7}, David N. Kennedy^{3,5,8}, Melissa P. Lopez-Larson^{2–4}, Verne S. Caviness^{5,8,9}, Larry J. Seidman^{2,6,10}, Benjamin Zablotzky³, and Nikos Makris^{5,11}

²Department of Psychiatry, Harvard Medical School; ³Child and Adolescent Neuropsychiatric Research Program, Cambridge Health Alliance; ⁴Department of Psychiatry, McLean Hospital; ⁵Center for Morphometric Analysis, Massachusetts General Hospital; ⁶Department of Psychiatry, Massachusetts Mental Health Center at the Beth Israel Medical Center; ⁷Brain Imaging Center, McLean Hospital; ⁸Department of Neurology, Harvard Medical School; ⁹Department of Pediatric Neurology, Massachusetts General Hospital; ¹⁰Department of Psychiatry, Massachusetts General Hospital; ¹¹Department of Neurology, Massachusetts General Hospital

Objective: The limbic structures in early-onset schizophrenia-spectrum illness (SZ) and bipolar disorder (BPD) were studied to discern patterns associated with diagnosis and sex. **Methods:** Thirty-five youths with DSM-IV BPD with-

hormone receptors. In addition, smaller thalamus was associated with SZ while larger right NA volumes were most related to BPD. This study underscores the importance of assessing diagnostic effects and sex effects on the brain in future studies and provides evidence that boys and girls with SZ and BPD may have differential patterns of neuropathology associated with disease expression.

Key words: mood disorders/psychosis/brain imaging technique/child psychiatry

Introduction

Schizophrenia (SZ) and bipolar disorder (BPD) both typically onset in late adolescence or early adulthood.^{1–4} Neurodevelopmental models of both disorders posit subtle disease processes that affect critical circuits in the brain early in development that then manifest in disease around the time of puberty or shortly thereafter.^{4–6} Not only is there increasing support for the neurodevelopmental model for both disorders but there is also emerg-

Description: Version 1.1 of the CANDI Share Schizophrenia Bulletin 2008 data.

Subjects ☐

SELECT

<< first < prev 1 2 3 4 5 next > last >>

20

1 of 6 Pgs (103 Rows)

Subject	M/F	Hand	YOB	MR Sessions
BPDwoPsy_030	F	R		1
BPDwoPsy_031	F	R		1
BPDwoPsy_032	M	R		1
BPDwoPsy_033	M	R		1
BPDwoPsy_034	M	R		1
BPDwoPsy_035	M	R		1
BPDwoPsy_036	M	R		1
BPDwoPsy_037	M	R		1
BPDwoPsy_038	M	R		1
BPDwoPsy_039	F	R		1

MR Session: BPDwoPsy_031_MR

Details

Projects

Accession # xnat_E02617

Date Added 2012-02-22 13:37:09.0 (dnkennedy)

Subject: BPDwoPsy_031

Gender: Female

Handedness: Right

Age: 15

Notes:

Scans

Scan

Type

Usability

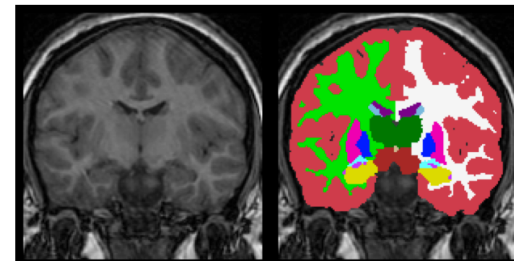
Files

☐ anat

anat

Show Counts

Image



Vox. Res. 0.9375, 0.9375, 1.5

Matrix 256 x 256

Part 2:

- Users 2-4 also write papers as above, share data. We have publication doi's for Pub2-4
- We have more shared data:

U2Sub1raw.nii

U2Sub1proc.nii

U2Sub2raw.nii

U2Sub2proc.nii

▪

▪

▪

U4Sub103raw.nii

U4Sub103proc.nii

Part 3:

- User 5 wants to use the shared data to do a new publication, looking at gender differences of 10 year olds in the healthy population. A query on the above shared data identifies 80 of the 270 subjects. The users obtains the raw data for these specific cases and performs another analysis workflow. A paper is written (Pub5doi)

- Data for the paper:

U1Sub2raw.nii	aka U5Sub1raw.nii	U5Sub1proc.nii
U1Sub7raw.nii	aka U5Sub2raw.nii	U5Sub2proc.nii
U1Sub27raw.nii	aka U5Sub3raw.nii	U5Sub3proc.nii
U1Sub31raw.nii	aka U5Sub4raw.nii	U5Sub4proc.nii
U1Sub34raw.nii	aka U5Sub5raw.nii	U5Sub5proc.nii
U2Sub7raw.nii	aka U5Sub6raw.nii	U5Sub6proc.nii
U2Sub14raw.nii	aka U5Sub7raw.nii	U5Sub7proc.nii
▪		
▪		
▪		
U4Sub37raw.nii	aka U5Sub270raw.nii	U5Sub80proc.nii

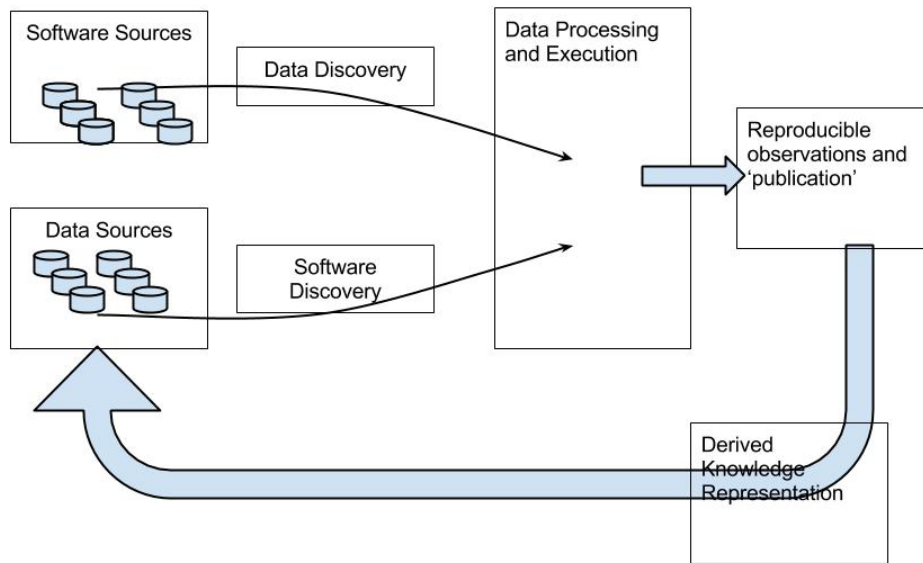
Filter(s): Age = 10

Label	Project	Subject	M/F	Age	Scans	Field Strength	Resting TR
BPDwoPsy_034_MR	cs_schizbull08	BPDwoPsy_034	M	10	anat(1)	1.5T	
BPDwoPsy_049_MR	cs_schizbull08	BPDwoPsy_049	M	10	anat(1)	1.5T	
BPDwoPsy_053_MR	cs_schizbull08	BPDwoPsy_053	F	10	anat(1)	1.5T	
BPDwoPsy_055_MR	cs_schizbull08	BPDwoPsy_055	M	10	anat(1)	1.5T	
BPDwPsy_075_MR	cs_schizbull08	BPDwPsy_075	M	10	anat(1)	1.5T	
BPDwPsy_081_MR	cs_schizbull08	BPDwPsy_081	F	10	anat(1)	1.5T	
HC_005_MR	cs_schizbull08	HC_005	F	10	anat(1)	1.5T	
HC_018_MR	cs_schizbull08	HC_018	F	10	anat(1)	1.5T	
HC_020_MR	cs_schizbull08	HC_020	M	10	anat(1)	1.5T	
HC_023_MR	cs_schizbull08	HC_023	M	10	anat(1)	1.5T	
IPCAS_26058_baseline	corr	IPCAS_26058	M	10	anat(1), rest(2)	3T	2500.0
IPCAS_26062_baseline	corr	IPCAS_26062	M	10	anat(1), rest(2)	3T	2500.0
IPCAS_26064_baseline	corr	IPCAS_26064	F	10	anat(1), rest(2)	3T	2500.0
IPCAS_26065_baseline	corr	IPCAS_26065	M	10	anat(1), rest(2)	3T	2500.0
IPCAS_26074_baseline	corr	IPCAS_26074	M	10	anat(1), rest(2)	3T	2500.0
KKI_1638334_1	adhd_200	KKI_1638334	F	10	anat(1), rest(1)		2500.0
KKI_1652369_1	adhd_200	KKI_1652369	F	10	anat(1), rest(1)		2500.0
KKI_1692275_1	adhd_200	KKI_1692275	F	10	anat(1), rest(1)		2500.0
KKI_1779922_1	adhd_200	KKI_1779922	M	10	anat(1), rest(1)		2500.0
KKI_1842810_1	adhd_200	KKI_1842810	M	10	anat(1), rest(1)		2500.0

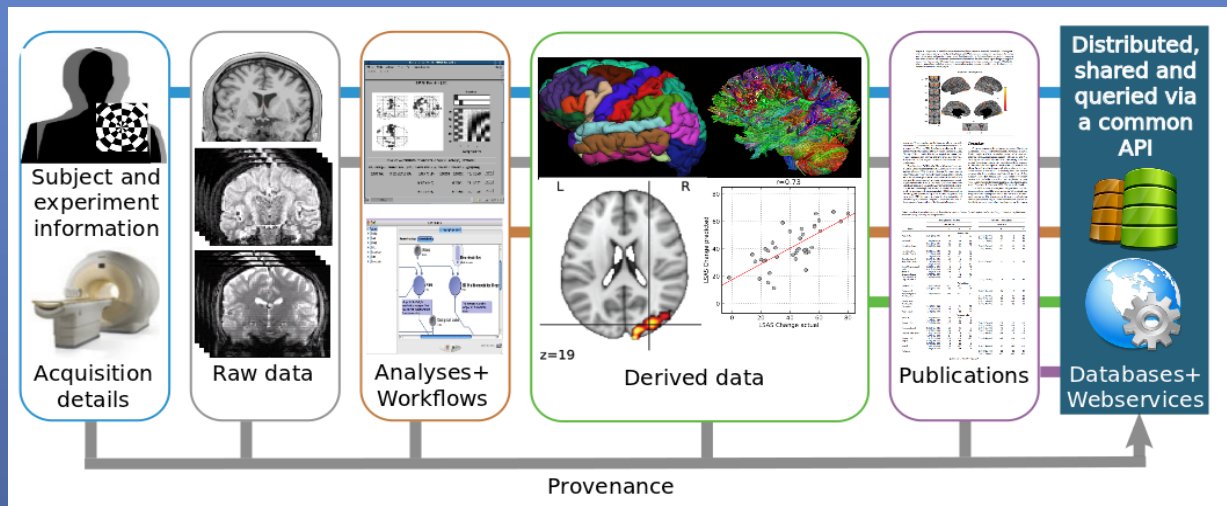
Overview

- **Framework for Sharing**
- **Existing Solutions**
 - **Mandated**
 - **ADNI (IDA)**
 - **Grassroots**
 - **NITRC (Fcon 1000, etc)**
- **Main Problems**
 - **Credit**
 - **Ethics**
 - **Cost**
 - **Harmonization**
 - **Reproducibility**

The Research Framework

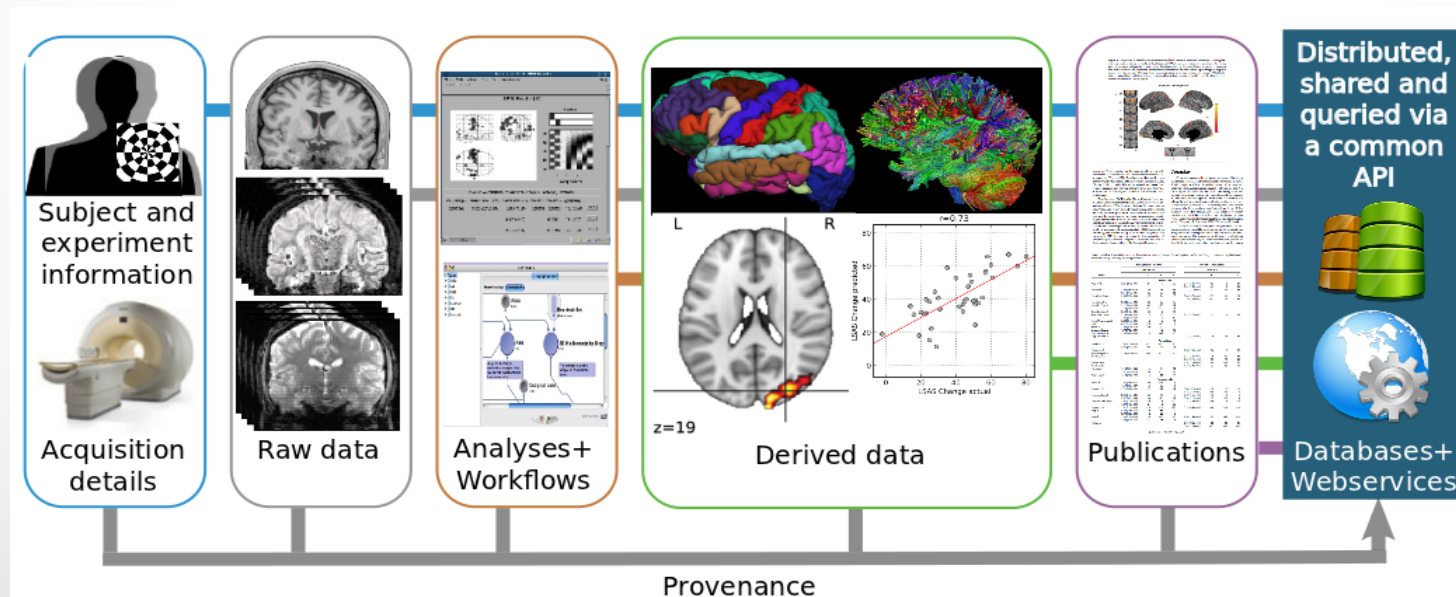


Data Flow and Stages of Data Sharing Opportunities



Resource Sharing: Beyond Data Sharing

- While we are huge proponents of resource sharing overall, in this Center we're not addressing the 'raw data sharing' topic; we are, rather, addressing the sharing of the 'public domain' aspects of publications that have been overlooked: the data descriptors, the workflow descriptors, the execution descriptors and the results descriptors. All of these elements are non-PHI completely sharable elements.



- All data will be marked up via NIDM (<http://nidm.nidash.org/>) and provided via a NIH-funded NITRC (Neuroimaging Informatics Tools and Resources Clearinghouse - nitrc.org) project. NIDM is the NeuroImage Data Model, an instance of the W3C PROV specification.

One-Click Data Share

Barrier: “Data sharing is too hard...”

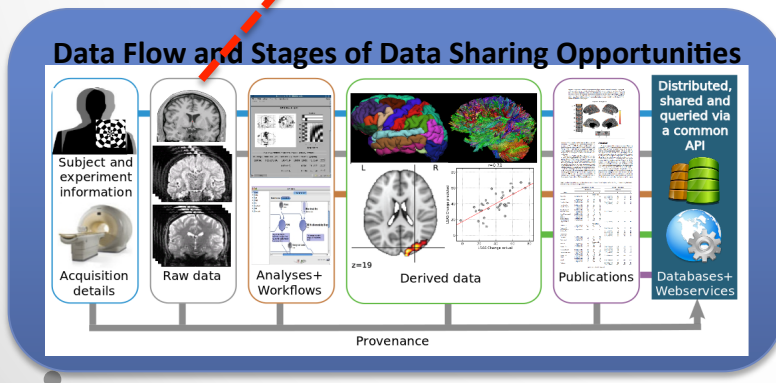
Solution: One-Click Data Sharing Tool

Data is anonymized and pushed to the INCF XNAT server. User is prompted for missing metadata. QA is performed and the results are available on line. The user is notified by e-mail.

Planned extensions:

- Promote to community
- Extend QC for structural & diffusion datasets
- Improvements to the user interface & ease of installation
- NIFTI support
- Process large databases to establish distribution of QC across scanners & populations

See <http://xnat.incf.org/> to access this resource.



One-Click Data Share

- Drag-n-Drop DICOM directory onto 'INCF Push' app

incf

Guest (Login) (Register)

Home Tools

PROJECT: demo > SUBJECT:CH01 > MR:CH01_S01 > BasicStructuralQA-301

Subject Detail

Details Projects

Accession # Incf_S
Date Added 2012-
Birth year --
Gender
Handedness

BasicStructuralQA Details

ImageSession_ID Incf_E00028
ID Incf_E00031
project demo
label BasicStructuralQA-301
source scan 301

SNR = 7.62654730211

This plot shows SNR for this scan on a histogram of all SNRs in this archive.

Experiments

Date	Experim
2010-12-20	MR Sess
	• Basic !
	• DTIPr

SNR Histogram

Count

SNR

min = 1.8, max = 306.2
value = 7.6, percentile = 51

Region	MinIntensity	MaxIntensity	RobustMinIntensity	RobustMaxIntensity	MeanIntensity	StdIntensity
external	0.0	49.03	0.0	898.0	4.09	27.15
brain	26.76	430.84	0.0	1338.0	207.09	101.29