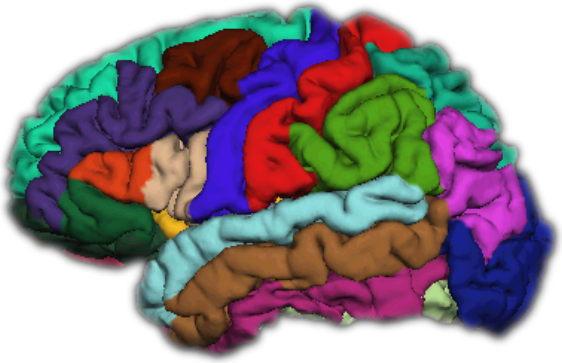
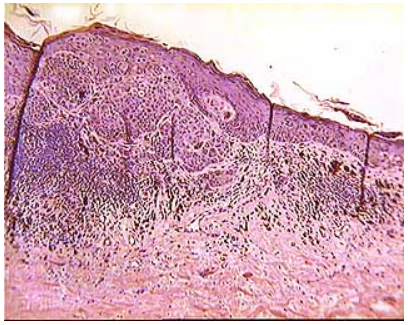
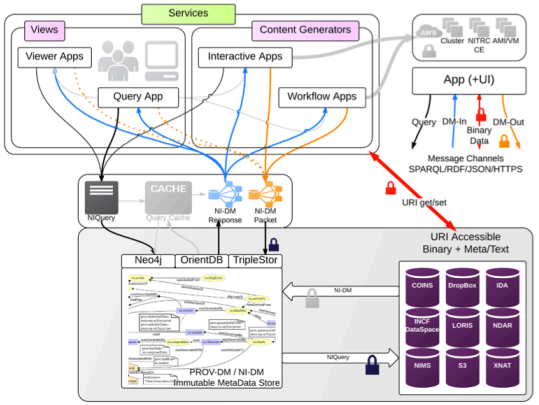
Sage: Week 2

Arno Klein

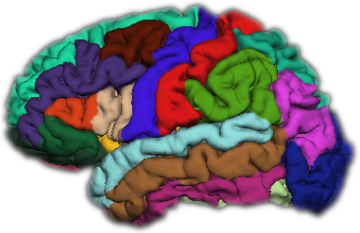






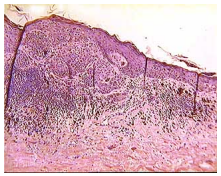
Brain images

Mindboggle and brain imaging-related challenges



Histopathology images

Melanoma image classification challenge



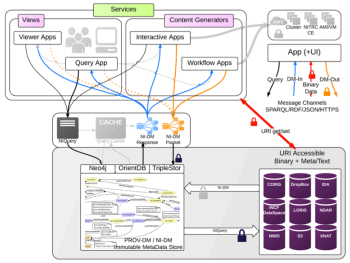
Clinical audio recordings

Crowdsourced Parkinson's voice validation

Proposals for open science endeavors

Semantic-web-based science

Unending challenges



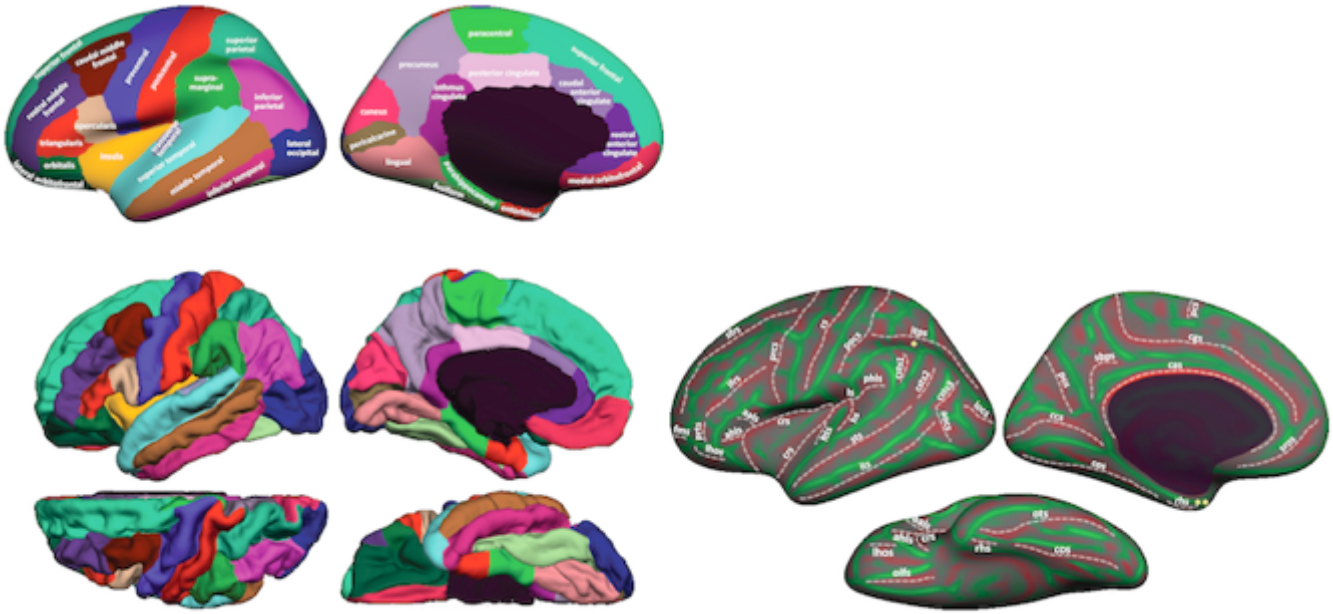
Mindboggle

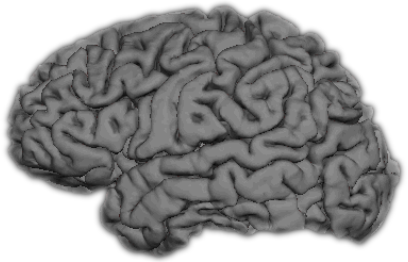
Software Data Papers People

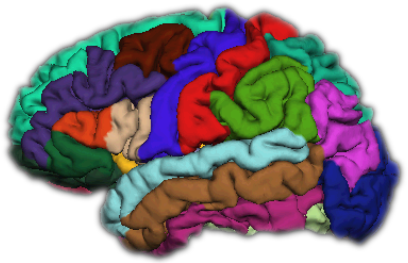
Data

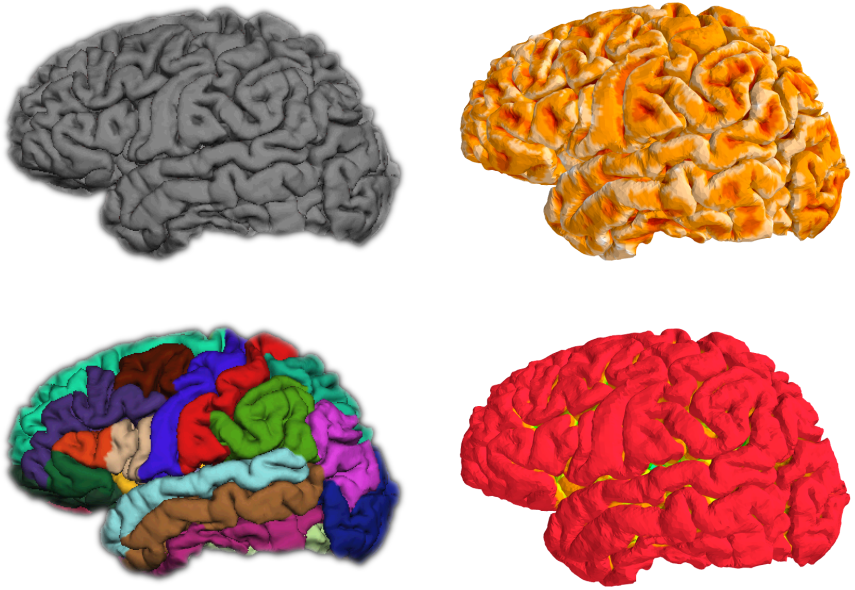
Welcome to the largest collection of publicly available, manually labeled human brain image data in the world!
Please cite the following article and this website when making use of Mindboggle-101 data:
101 labeled brain images and a consistent human cortical labeling protocol
Arno Klein, Jason Tourville. Frontiers in Brain Imaging Methods. 6:171. DOI: 10.3389/fnins.2012.00171

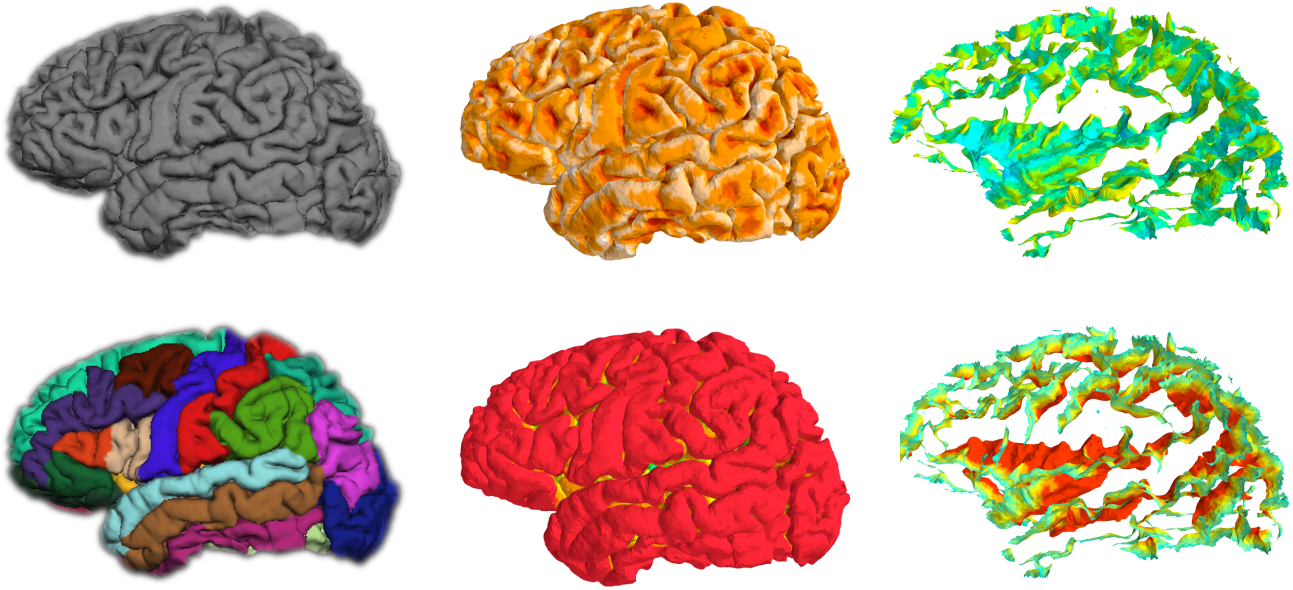
See the README, labeling protocol, the CHANGELOG, and MD5SUMS,
which describe the labeled nifti volumes (nii), vtk surfaces (vtk), and FreeSurfer files (mgh, etc.).
Except where noted, all data are licensed under a Creative Commons License: (cc) BY-NC-SA

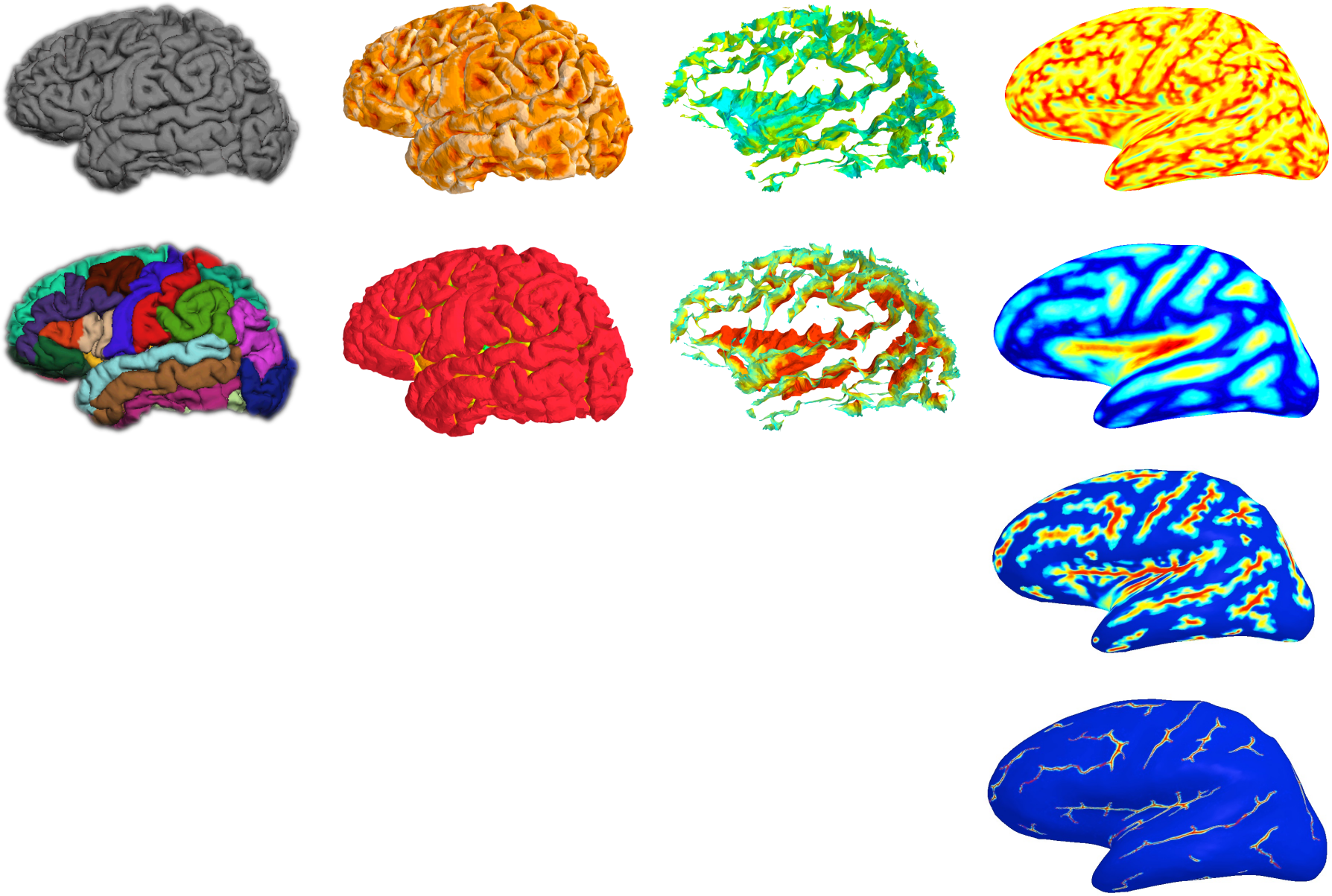












International Neuroimaging
Data-Sharing Initiative

Autism Brain Imaging
Data Exchange

FUNCTIONAL
CONNECTOMES
PROJECT

NIH NDAR National Database for Autism Research

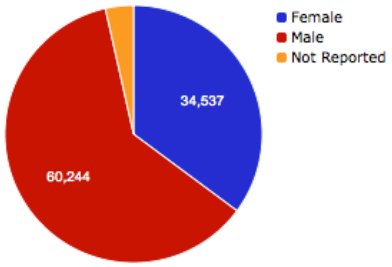
Serving the autism research community

Home | Query | Harmonization Tools | Cloud | Contribute | Request Access | Policy | Tutorials | About | FAQ | login

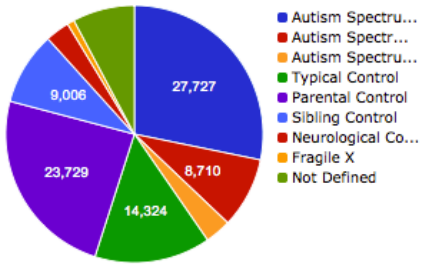
Query Data | Data from Labs | Data from Papers | Query By Data Dictionary | Query Instructions

Use "Select Data" below to query the data available in NDAR. Then, select download to create a package and download your results. Use the Data tab above to search in other ways. For more information on search see our Methods.

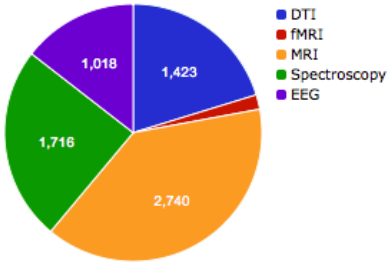
Data Distribution by Gender[1]



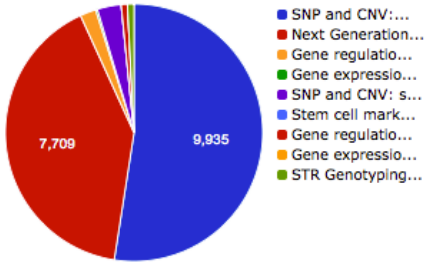
Phenotypic Data Distribution[1]



Neuroimaging Data Distribution



Genomic Data Distribution



[1] Numbers reported are subjects by age

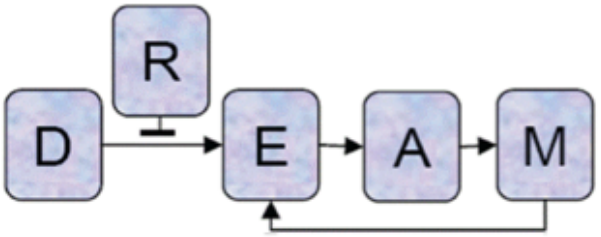
Select Data: All | Basic | Phenotype | Neuroimaging | ***omicSEARCH:*** Experiment | Results | **Show Results** | **Reset All** | **Download Data**

ADNI

ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

DREAM Challenges

Announcing the 2013 DREAM8.5 Challenges
Announcing the final results of the 2013 DREAM8 Challenges
More information about previous DREAM Challenges
About DREAM Challenges
About Synapse



Announcing the 2013 DREAM8.5 Challenges

We are pleased to announce three new DREAM8.5 challenges. Best performers in all DREAM8.5 Challenges will be invited to present at the 2014 DREAM conference (date and location to be determined). We are also working to establish publishing partners for each of these challenges. The DREAM8.5 Challenges are now open for registration, and will begin active problem-solving in late 2013 or early 2014.

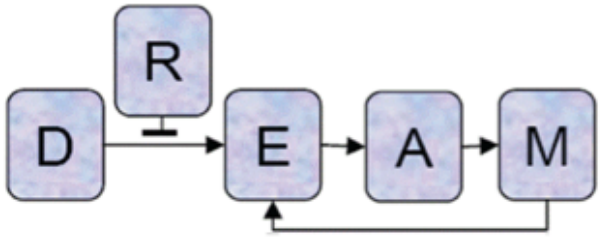
Click on a link below to read the Challenge detail and register for a DREAM8.5 Challenge.

Alzheimer's Disease Big Data DREAM Challenge #1

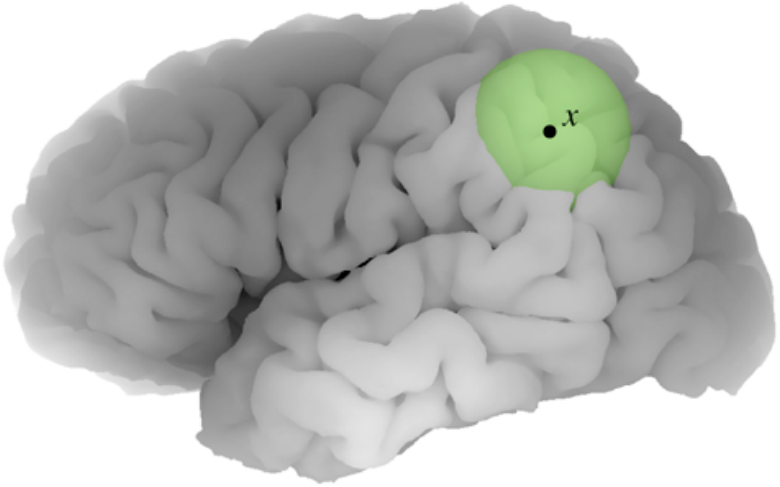
In the first of what will be a series of Alzheimer's Disease (AD) Big Data Challenges, participants will utilize data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Data will consist of cognitive, imaging, biological, and whole genome sequencing data on cohorts of volunteers, who range from cognitively normal, mild cognitive impairment and dementia. Participants will analyze the data to solve two sub-challenges: (1) Build a model that best predicts change over time in AD cognitive scores using all available test and adjacent data, and (2) Build a model that best predicts discordance between biomarkers suggestive of amyloid perturbations and lack of cognitive impairment. These models will be used to better understand the biomolecular mechanisms leading to Alzheimer's disease, and ultimately to develop new therapies. We expect to announce a publishing partner for this Challenge shortly.

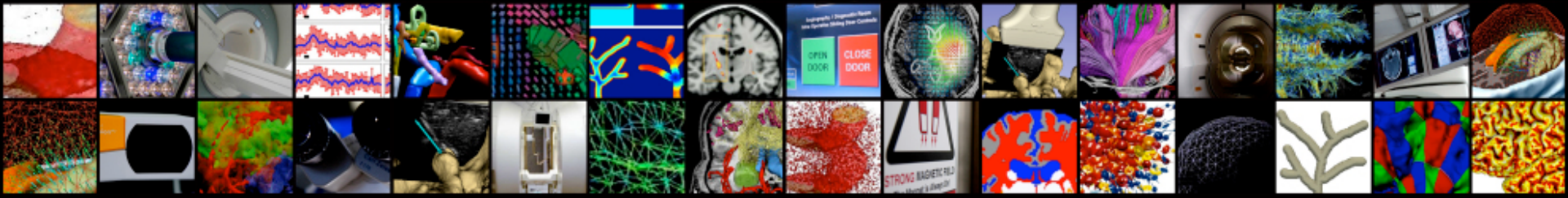
DREAM Challenges

Announcing the 2013 DREAM8.5 Challenges
Announcing the final results of the 2013 DREAM8 Challenges
More information about previous DREAM Challenges
About DREAM Challenges
About Synapse



Roberto Toro





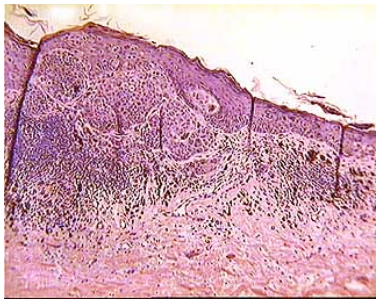
Melanoma image classification

Melanoma causes the majority of deaths related to skin cancer. However, four out of five patients with melanoma do not die of their disease and there is a significant over-diagnosis of lethal melanoma, leading to unnecessary suffering and cost to society. Providing patients with a more dependable prognosis than is generated by the subjective and unreliable visual evaluation of skin by a pathologist would have a profound impact on melanoma care.

In this challenge, participants would develop and test algorithms for automated scoring of malignancy from histopathology images. We will make available over 1,000 histopathology digital images from patients and malignancy scores that do not currently exist in the public domain, and would provide excellent training and testing data for the challenge. Sage Bionetworks would host all data, open source software, and results on their open source Synapse collaboration platform, to ensure that the entire process remains open and repeatable.

Johan Lundin
Institute for Molecular Medicine Finland (FIMM)
University of Helsinki, Finland

Paul J van Diest, MD, PhD
Professor and Head, Department of Pathology
University Medical Center Utrecht



foldit BETA

07:44:34 GMT

Solve Puzzles for Science

PUZZLES CATEGORIES GROUPS PLAYERS RECIPES CONTESTS
BLOG FEEDBACK FORUM WIKI FAQ ABOUT CREDITS

Pull Mode

Rank: 4 Score: 9587.911
Soloist Beginner Puzzle: Streptococcal Protein
Expires 3/27/2013 17:00 MZ (2 days, 17 hours)
▶ No bonuses or conditions

Cookbook

Shake Mutate Wiggle All Wiggle Backbone Wiggle Sidechains Help Glossary

Freeze Protein Remove Bands Disable Bands Reset Structures Reset Puzzle Align Guide

▲ Actions ▶ Undo ▶ Social ▶ Modes ▶ Behavior ▶ View ▶ Menu

▶ Chat - Group auto show
▶ Chat - Puzzle auto show
▶ Chat - Global auto show
▶ Notifications auto show

Crowdsourced Voice Validation Challenge

Validate health sensors through community engagement and collaborative data interpretation

patientslikeme Sage BIONETWORKS

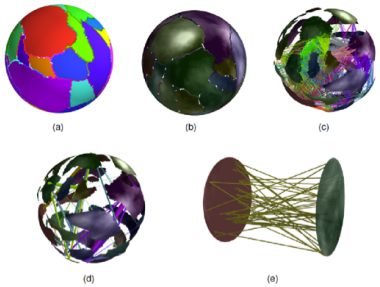
What is the problem or need for this research?

The current siloed approach to biomedical research is ill-adapted to bring out value from the emerging tools and amazing data now available to query biological systems. Our proposal presents a community-based collaborative approach to doing research and monitoring diseases.

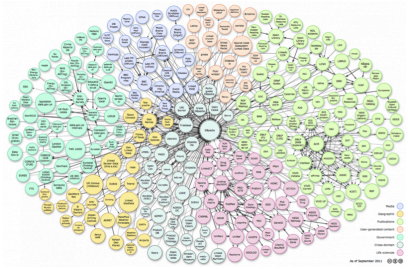
Topic Area 1: BOLD: Brain Organization using Linked Data

Sage Bionetworks and MIT: Stephen Friend

CONCEPT



Linked brain data



Linked Open Data cloud

Linked Data standards may help to provide an organizational framework for indeterminate, heterogenous brain data to facilitate their visualization, analysis and interpretation.

APPROACH

A Linked Data representation may be the most appropriate organizational framework for the vast amount of indeterminate, conflicting, and heterogeneous brain data. The current structured information landscape is extremely sparse and sparsely connected and continues to require significant human curation effort. An opportunity exists to develop an intuitive interface for semi-automatically entering linked brain data. The resulting framework would easily enable seamless integration and queries and exploration of relationships across datasets, and identification of gaps in these relationships to expose new, cross-disciplinary research directions.

Impact

The most successful organizing principle to date for brain data, and brain imaging data in particular, has been the map metaphor. Data of interest are examined and reported in their spatial context, with visual depictions of multiple modalities and connections as overlays and paths. While this representation has served well for integration of limited numbers of images and for pairwise relationships, it fails to accurately and intuitively accommodate multiple spatial scales and modalities, and data less amenable to spatial representation such as correlations across temporal signals, complex network behaviors, and developmental structure-function relationships. The map metaphor also has difficulty representing data from multiple sources where the data are inconsistent or imprecise.

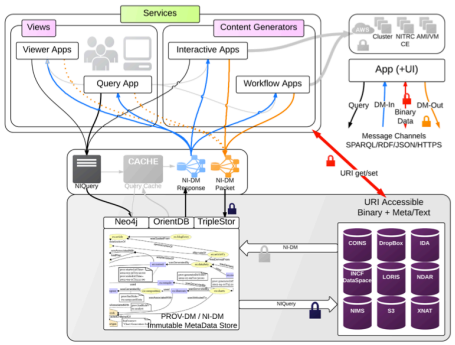
Context

The standard for communicating information in neuroscience is through unstructured prose in journal articles. The closest that human neuroimaging currently comes to providing a “reference architecture” in the sense of a reference space for brain data is in the form of Brodmann areas and Talairach coordinates. Attempts to link behavior to neurological state or brain location are hampered by the lack of structured information in publications and associated data sources. There does not exist a framework to differentiate, integrate, or reconcile multiple scales and modalities or to capture inconsistency and imprecision. This makes it incredibly difficult to integrate varied and possibly discordant data of even a single data type, let alone across heterogeneous data.

Topic Area 2: Semantically Accessible Reproducible Computation

Sage Bionetworks and MIT: Stephen Friend

CONCEPT



Semantic web tools that link structured databases and capture standardized provenance can aid reproducible scientific practices.

Example framework for decentralized computation and data storage mediated by standardized message passing

APPROACH

The figure shows a schematic diagram for a framework we are building for computation and data storage that operates using messages that conform to the W3C Provenance model. Different users can run the same workflows and many common tasks can be encapsulated in such containers. By focusing on standardizing communication and provenance capture rather than the computational environments, this has the added benefit of having the computation be carried out close to data when data are large and more distributed and decentralized when data are small.

Impact

An essential requirement for any scientific computational procedure is that it be repeatable. Reproducibility is the primary means for establishing validity in science. This is a nontrivial requirement, since scientists often require special-purpose algorithms or demanding computational environments. Indeed, recent high-profile articles have emphasized the limited reproducibility in life sciences. The methods section that most scientific publications rely on to describe data acquisition and analysis might provide a summary, but cannot capture the details necessary to reproduce all of the methods, nor can they provide the reviewer with sufficient information for verification. Therefore, most scientific claims are reviewed on the basis of perceived impact rather than veracity.

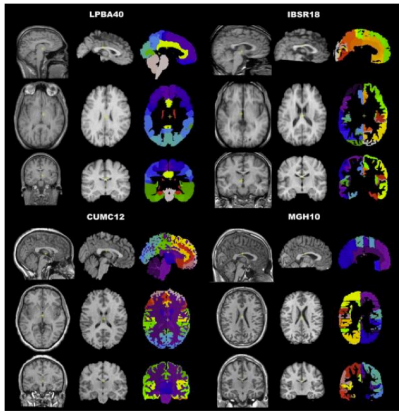
Context

Traditional databases do not store provenance and lack shared semantics (e.g., “age” in one database, “years” in another). The lack of shared semantics also negatively impacts how well a data management system can capture relationships across scales (nanometers to inches, microseconds to years), domains (autism vs. schizophrenia, inter-species), and methods (electron microscopy, magnetic resonance imaging). Another challenge to reproducibility in neuroscience is dealing with computational demands and infrastructure -- different sofware requirements, different hardware, and massive datasets.

Topic Area 3: Infrastructure for the Continuous Challenge: Evaluating Brain Registration

Sage Bionetworks: Stephen Friend

CONCEPT



Creating a challenge infrastructure that continues indefinitely for up-to-the-minute assessment of state-of-the-art of, for example, brain image registration. *Figure: Example data used in the largest brain image registration evaluation study.*

APPROACH

Like UCLA's online Segmentation Validation Engine, users would download source and target image pairs from a website, perform [registration] on the downloaded images, then upload their results. The server would contain well-curated, gold standard, manual anatomically labeled versions of the images, and would compute label overlap error measures between the transformed source labels and the target labels. These measures would be continually published to the website to provide an up-to-the-minute assessment of the state-of-the-art with regard to image registration. Additional manually labeled data could be generated to represent a cohort of interest.

Impact

All fields of neuroscience that employ brain imaging (and many that don't) need to communicate their results with reference to anatomical regions. In particular, comparative morphometry and group analysis of functional and physiological data require co-registration of brains to establish correspondences across brain structures. It is a rare occurrence for an objective evaluation to be performed on a broad sample of registration algorithms, even as new software algorithms appear and existing algorithms undergo changes, so the very basis for establishing correspondences in neuroscience is perpetually in question.

Context

The world's largest evaluation study of brain image registration methods took place almost five years ago, and has had a demonstrable impact on biomedical scientists' choice of registration algorithms inside and outside of neuroscience. However, many of the algorithms evaluated in that article have changed considerably and many new algorithms have appeared, and it is unclear how best to proceed short of conducting such a laborious evaluation study again and again. One solution to this problem lies in the form of the Segmentation Validation Engine (http://sve.bmap.ucla.edu/), which is an online resource to test and evaluate human brain tissue-class segmentation algorithms.

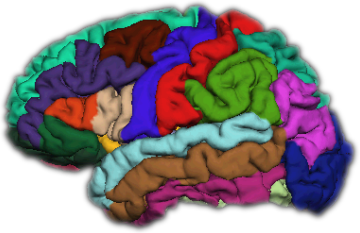
Mindboggle and brain imaging-related challenges

Run on ADHD-200, Abide, and other publicly available data

NDAR (National Database of Autism Research) Amazon instance

ADNI and AddNeuroMed data (**Alzheimer's challenge**)

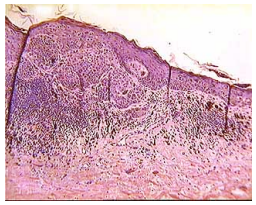
Brain variation **9 Challenge** with Roberto Toro



Melanoma image classification challenge

… MICCAI 2014 algorithm challenge with Andrew Trister

Fold.it-type human challenge with Zoran Popovic



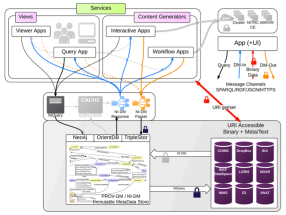
Crowdsourced Parkinson's voice validation

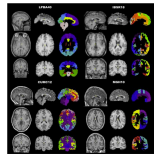
Prepare data with Max Little and Mike Kellen for a dry-run challenge

Semantic-web-based science (DARPA RFI submissions)

… Brain Organization through Linked Data

… Semantically Accessible Reproducible Computation



Unending challenges

… Registration evaluation resource