

Toward an Open Science Ecosystem in Neuroimaging

Russ Poldrack
Stanford University

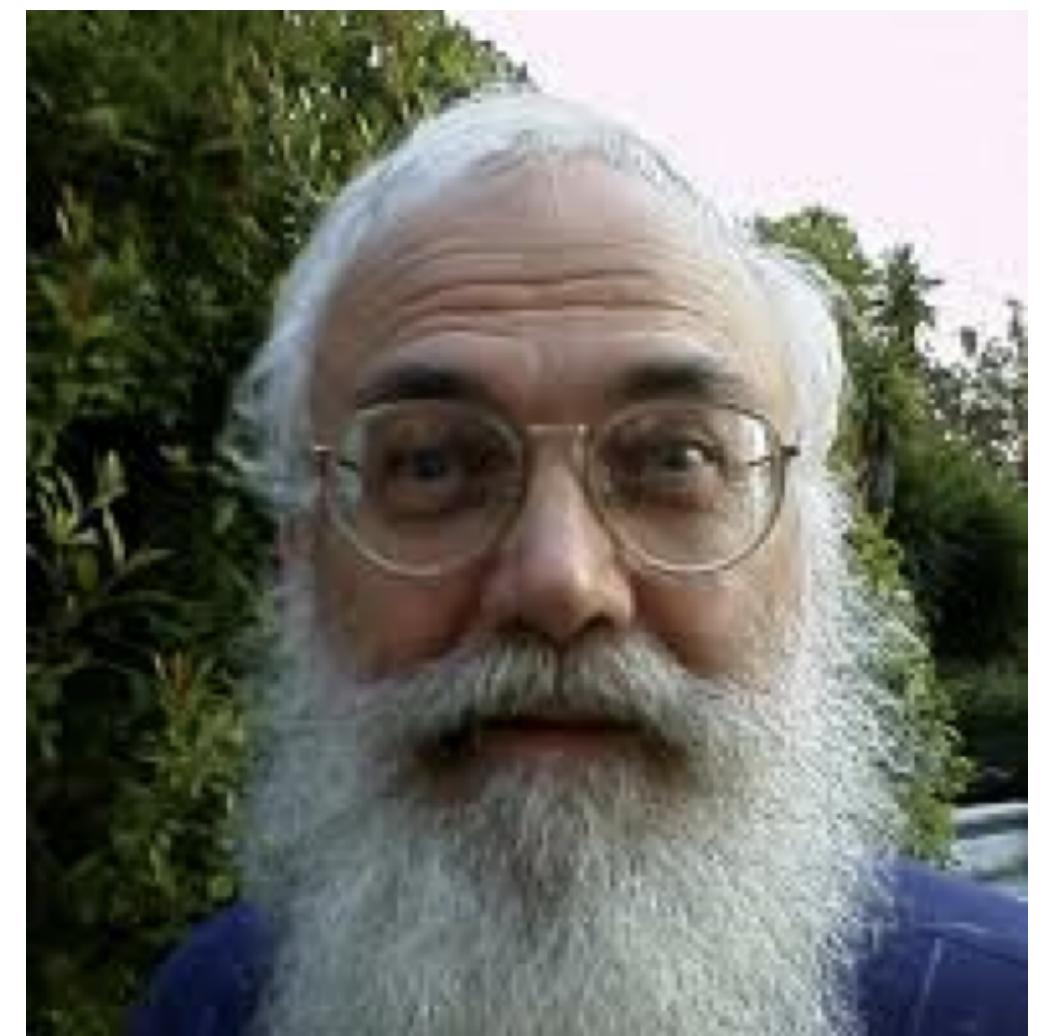
Transparency is essential for reproducibility

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

<https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html>

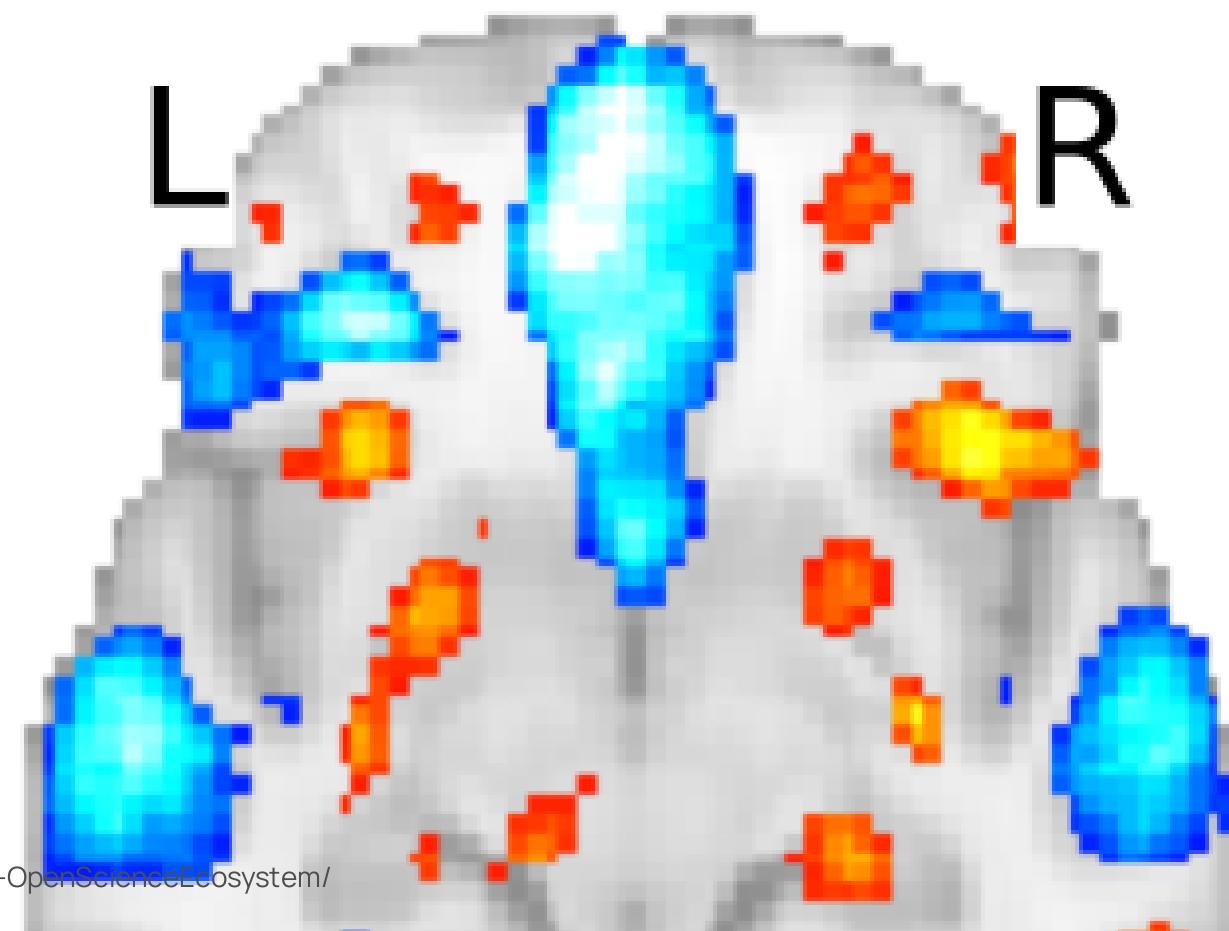
“we can distill Claerbout’s insight into a slogan:

An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development



Why neuroimaging is a best-case scenario for open science

- Magnetic resonance imaging (MRI) is the primary tool for studying human brain structure and function
- MRI data are digital end-to-end
 - From MRI scanner to automated analysis
 - Usually zero/few manual analysis



A false start for fMRI data sharing

The fMRI Data Center

fMRI IDC

SEARCH fMRIIDC Database FOR SUBMIT

[My Account](#) [Request List \(Empty\)](#)

► HOME

DATABASE

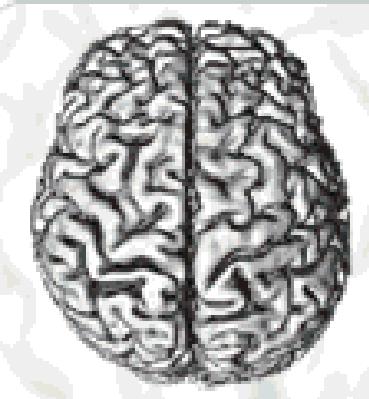
SUBMISSIONS

RESOURCES

DATA MANAGEMENT TOOL

HELP

ABOUT US



A public repository of peer-reviewed fMRI studies and their underlying data.

Funded By
The National Science Foundation
The W. M. Keck Foundation
The National Institutes of Mental Health
A Sun Center of Excellence for Neuroscience



INFORMATION

[How do I get started?](#)
Answers to questions commonly posed by first-time visitors.

[Q&A about fMRIIDC](#)
A comprehensive list of frequently asked questions about the fMRIIDC.

[Available Datasets](#)

PROJECT STATISTICS

Registered users: 1912
Datasets available: 110
Dataset requests: 1789
[More database statistics...](#)

Updated November 18, 2005



Special Collections
Data from special or rare populations of subjects.



Summer Workshops

5 / 65

A false start for fMRI data sharing

nature
neuroscience

A debate over fMRI data sharing

nature

3 August 2000 Volume 406 Issue no 6795

Whose scans are they, anyway?

This letter comes from a group of scientists who are publishing papers using fMRI to understand the links between brain and behavior. We are writing in reaction to the recent announcement of the creation of the National fMRI Data Center (www.fmridc.org). In the letter announcing the creation of the center, it was also implied that leading journals in our field may require authors of all fMRI related papers accepted for publication to submit all experimental data pertaining to their paper to the Data Center. ...

2010: The year data sharing broke in neuroimaging

Toward discovery science of human brain function

Bharat B. Biswal^a, Maarten Mennes^b, Xi-Nian Zuo^b, Suril Gohel^a, Clare Kelly^b, Steve M. Smith^c, Christian F. Beckmann^c, Jonathan S. Adelstein^b, Randy L. Buckner^d, Stan Colcombe^e, Anne-Marie Dogonowski^f, Monique Ernst^g, Damien Fair^h, Michelle Hampsonⁱ, Matthew J. Hoptman^j, James S. Hyde^k, Vesa J. Kiviniemi^l, Rolf Kötter^m, Shi-Jiang Liⁿ, Ching-Po Lin^o, Mark J. Lowe^p, Clare Mackay^c, David J. Madden^q, Kristoffer H. Madsen^f, Daniel S. Margulies^r, Helen S. Mayberg^s, Katie McMahon^t, Christopher S. Monk^u, Stewart H. Mostofsky^v, Bonnie J. Nagel^w, James J. Pekar^x, Scott J. Peltier^y, Steven E. Petersen^z, Valentin Riedl^{aa}, Serge A. R. B. Rombouts^{bb}, Bart Rypma^{cc}, Bradley L. Schlaggar^{dd}, Sein Schmidt^{ee}, Rachael D. Seidler^{ff,u}, Greg J. Siegle^{gg}, Christian Sorg^{hh}, Gao-Jun Tengⁱⁱ, Juha Veijola^{jj}, Arno Villringer^{ee,kk}, Martin Walter^{ll}, Lihong Wang^q, Xu-Chu Weng^{mm}, Susan Whitfield-Gabrieliⁿⁿ, Peter Williamson^{oo}, Christian Windischberger^{pp}, Yu-Feng Zang^{qq}, Hong-Ying Zhangⁱⁱ, F. Xavier Castellanos^{b,j}, and Michael P. Milham^{b,1}

4734–4739 | PNAS | March 9, 2010 | vol. 107 | no. 10

- “Comprehensive mapping of the functional connectome, and its subsequent exploitation to discern genetic influences and brain–behavior relationships, will require multicenter collaborative datasets. Here we initiate this endeavor by gathering R-fMRI data from 1,414 volunteers collected independently at 35 international centers. We demonstrate a universal architecture of positive and negative functional connections, as well as consistent loci of inter-individual variability. ...”

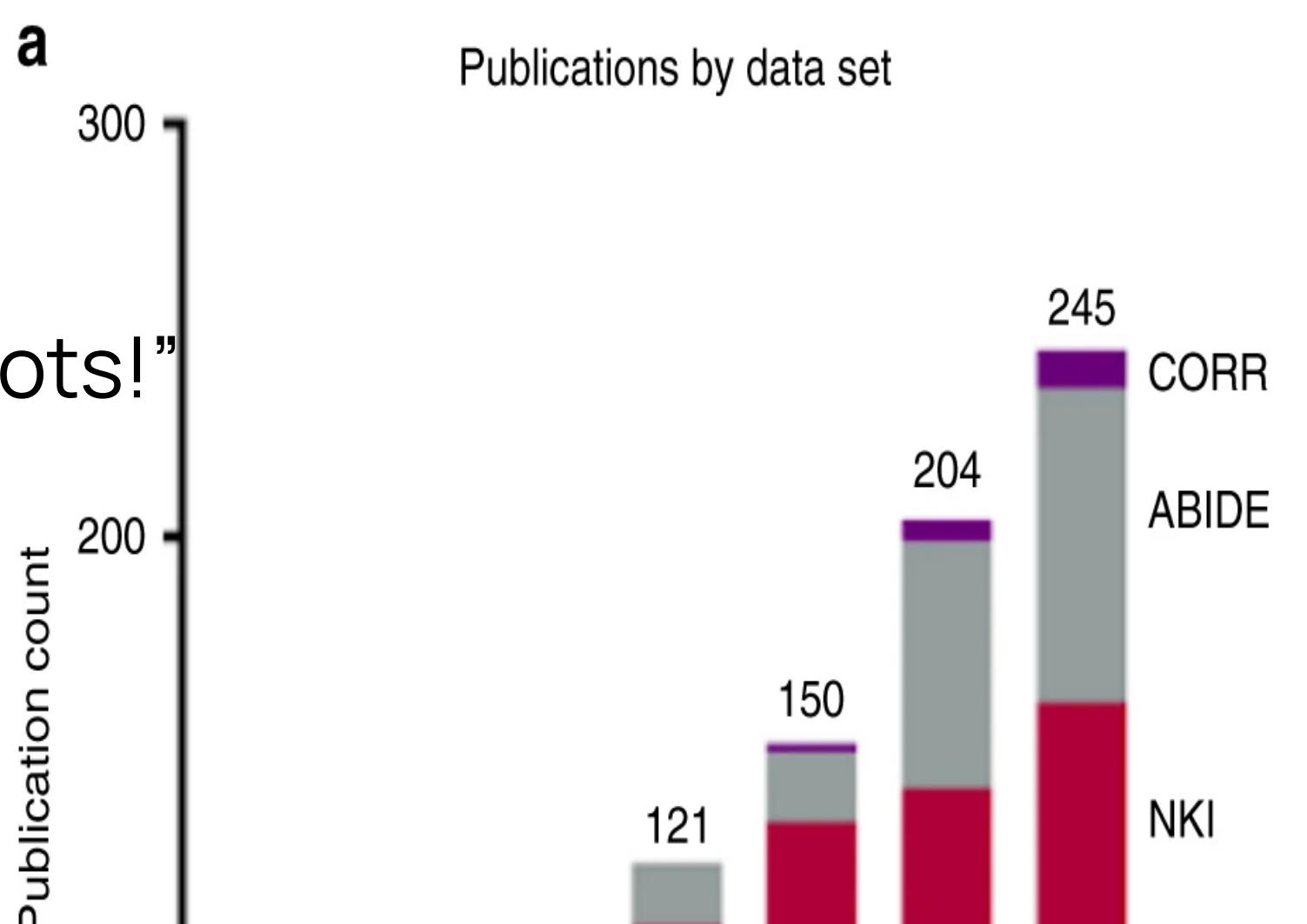
Data sharing is becoming the norm in neuroimaging

Anonymous senior researcher:
“OHBM has been taken over by the open science zealots!”

BrainMap
NeuroSynth

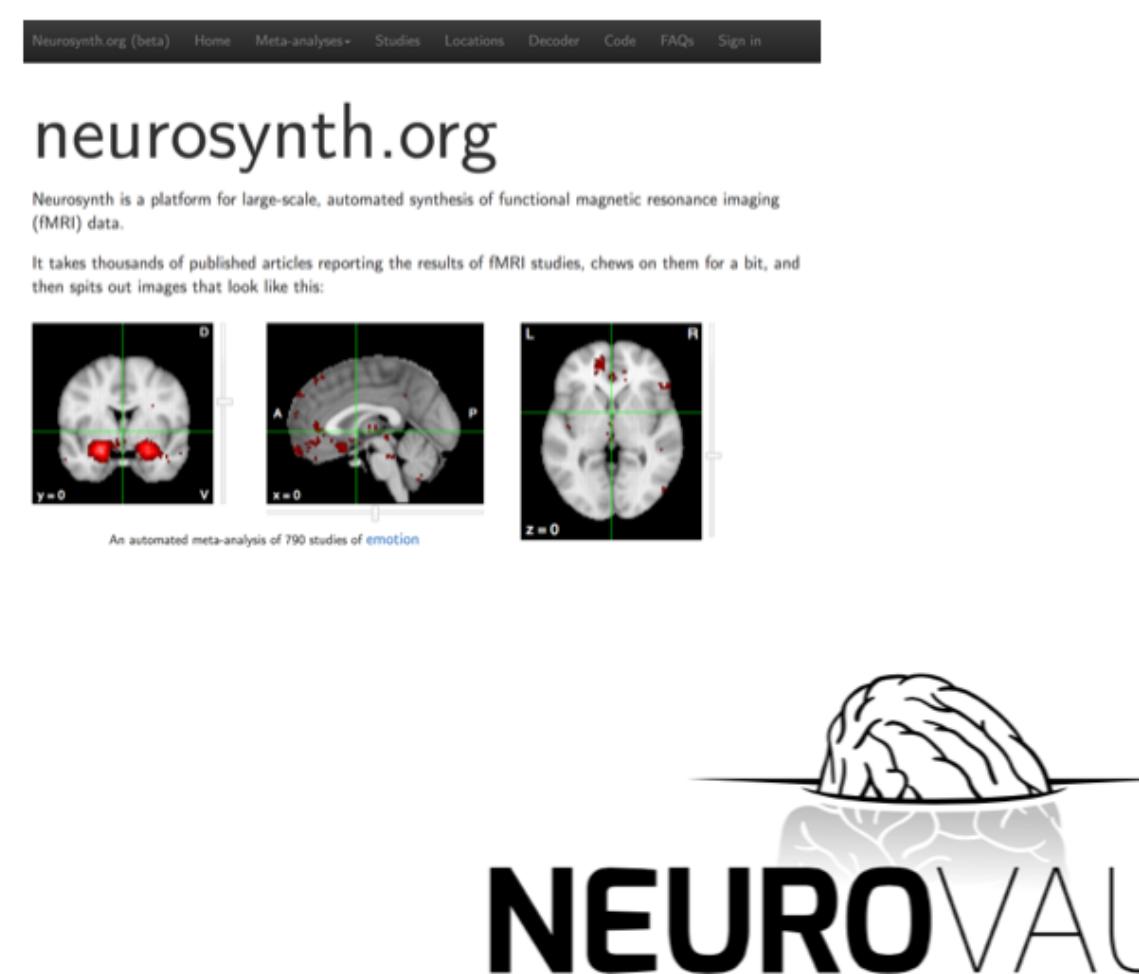
Peak coordinates only

<https://boldrake.github.io/talks/OpenScienceEcosystem/>



An open ecosystem for retrospective data sharing

Breadth



- Neurosynth.org: Open database of published neuroimaging coordinates
- Neurovault.org: Open archive for neuroimaging results
- OpenNeuro.org: Open

Maximally open sharing

- Data shared under maximally permissive data use agreements:
 - Neurosynth: Open Data Commons Open Database License v1.0
 - Neurovault: CC0
 - OpenNeuro: CC0
- All data available programmatically via web API



- CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in

Neurosynth: Sharing activation coordinates

- Brain activity is reported in a (somewhat) standardized coordinate system

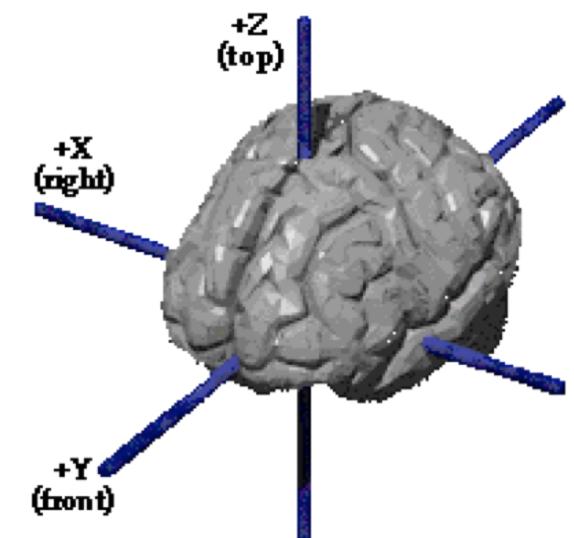


Table 1
Regions that showed a condition × time interaction in the ANOVA analysis

No.	Region	Hemisphere	BA	x	y	z	mm ³
1	Middle/superior temporal gyrus	L	21/22/37	-52	-54	9	13257
2	Inferior frontal gyrus	L	45/46/9	-49	26	6	2781
3	Posterior cerebellum	L		-19	-79	-38	2214
4	Dorsomedial PFC	L	9/8	-11	42	47	3051
5	Left anterior PFC	L	10	-37	49	15	2025
6	Inferior parietal cortex	L	40/7	-42	-58	47	3132
7	Dorsal premotor cortex	L	6	-43	0	50	1485
8	Lingual gyrus	L	17	-10	-95	-2	378
9	Middle /superior temporal gyrus	R	21/22/37	52	-40	5	16470
10	Inferior frontal gyrus	R	45/46	51	28	6	2241
11	Posterior cerebellum	R		22	79	24	2808

Creating meta-analytic maps

- Automated Coordinate Extraction
 - Automatically extracts activation tables from fMRI papers for 17 journals
 - Current database has 14,371 papers (with full text)
 - 84% sensitivity, 97% specificity against manual database (SumsDB)

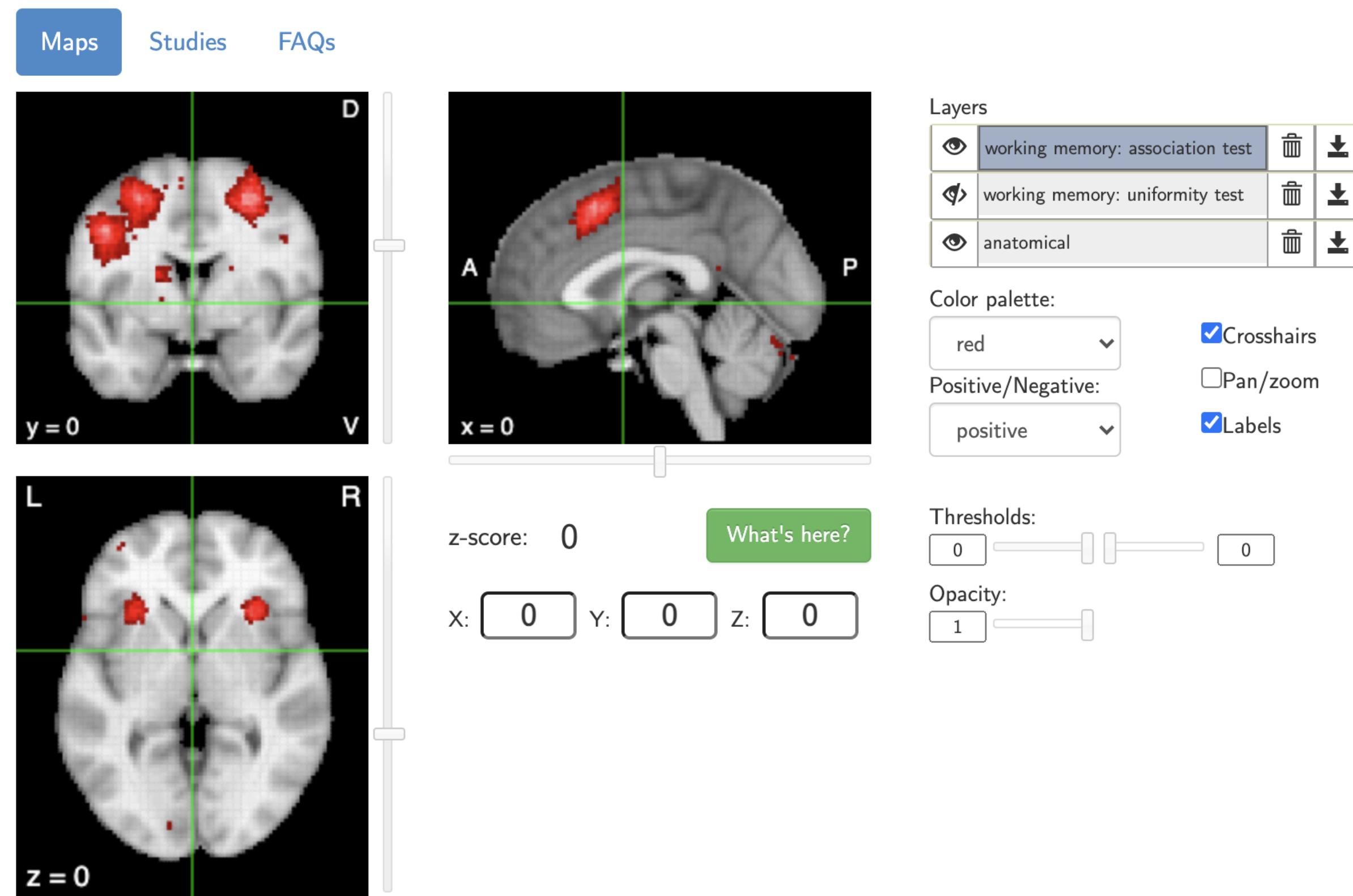
Yarkoni et al, 2011, *Nature Methods*

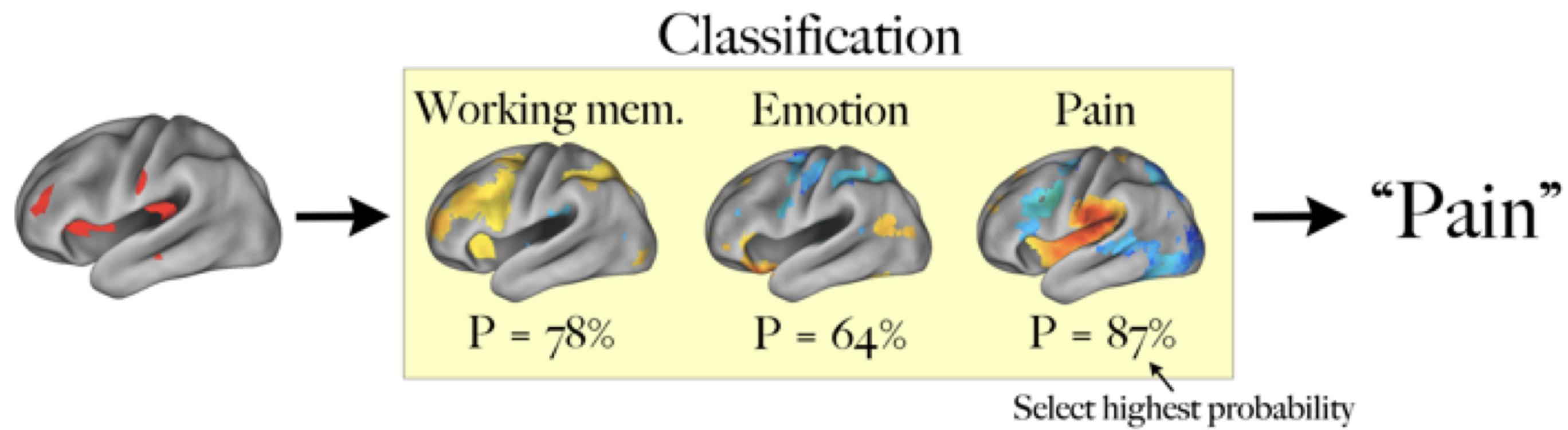
X	Y	Z
12	57	-6
33	21	15
24	-6	51
28	10	18

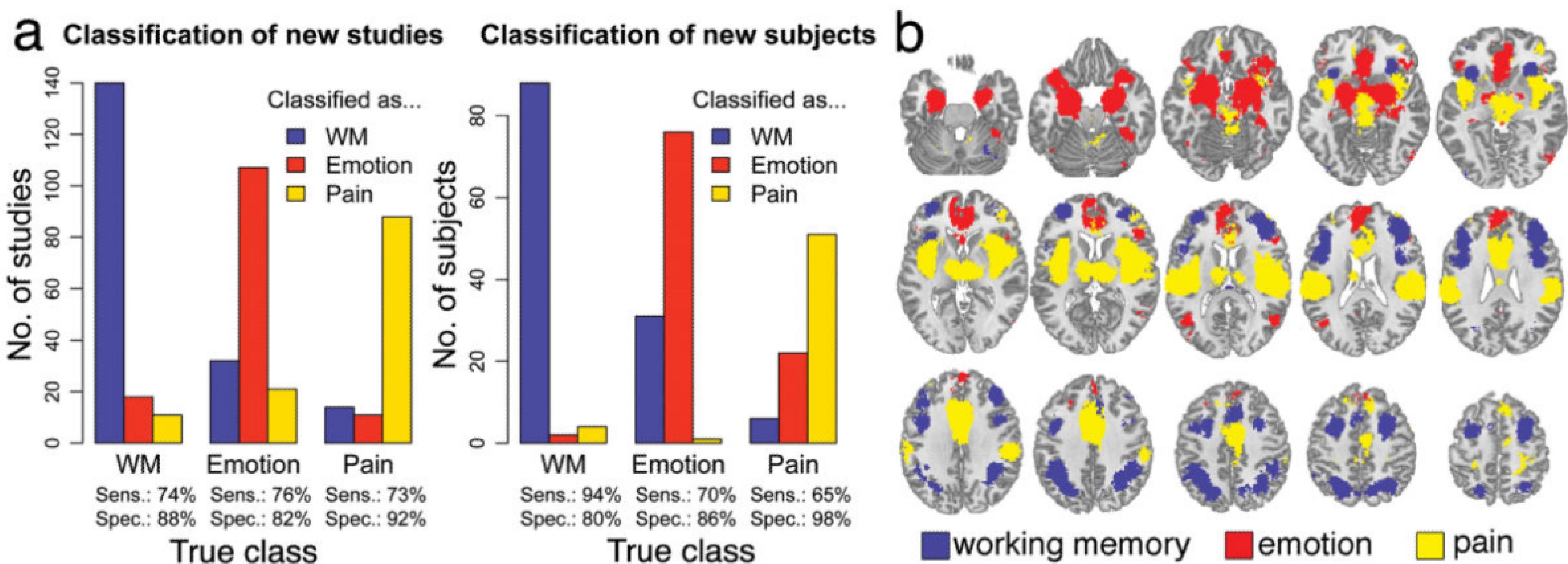
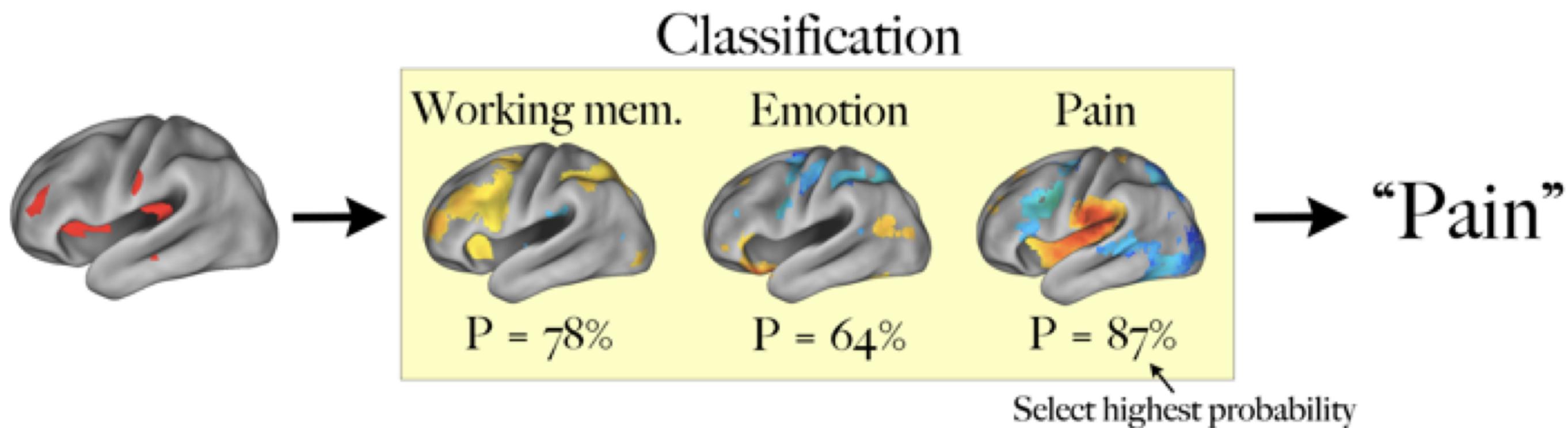
working memory

An automated meta-analysis of 1091 studies

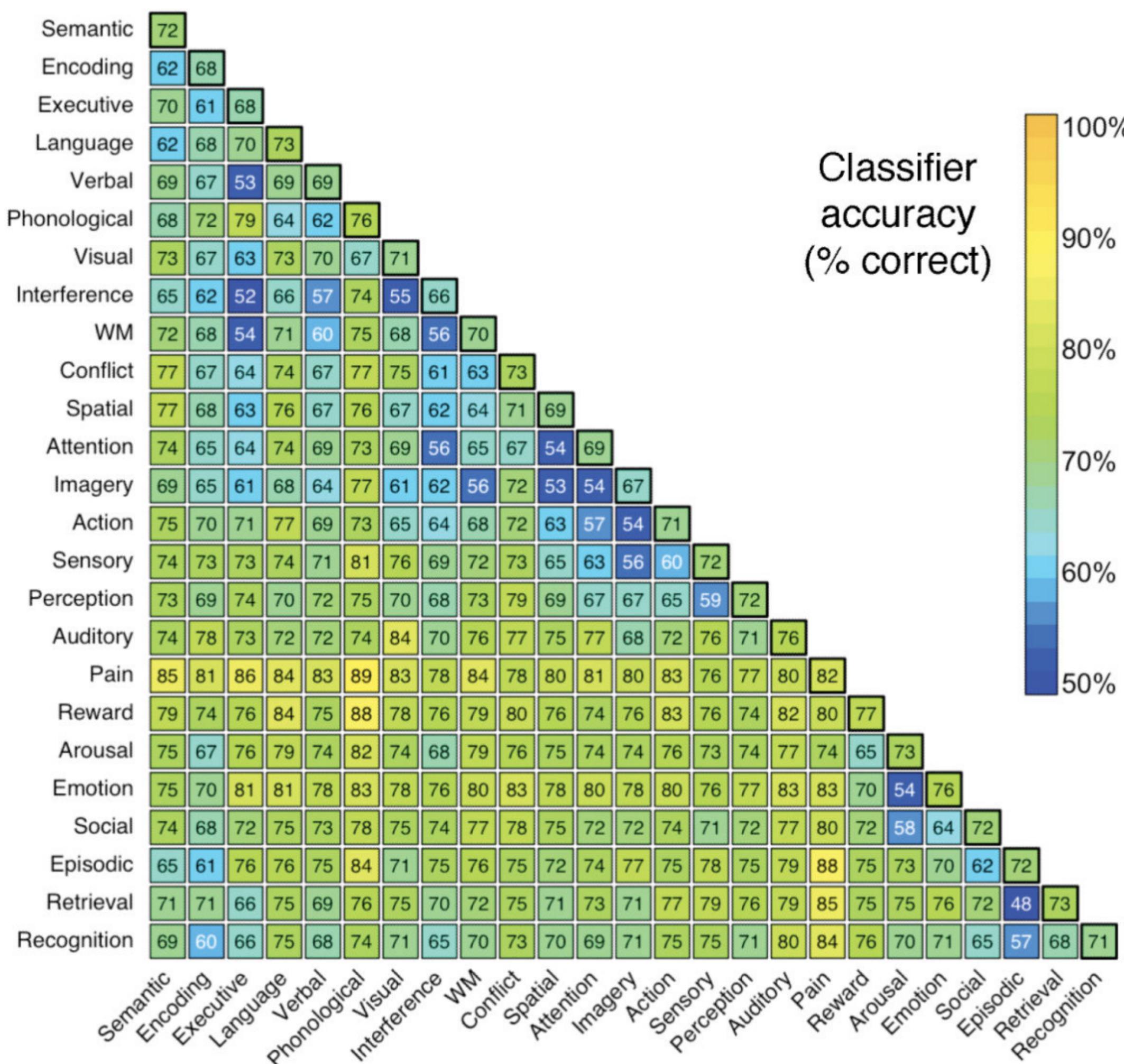
Search for another term:







Decoding brain activity patterns using Neurosynth

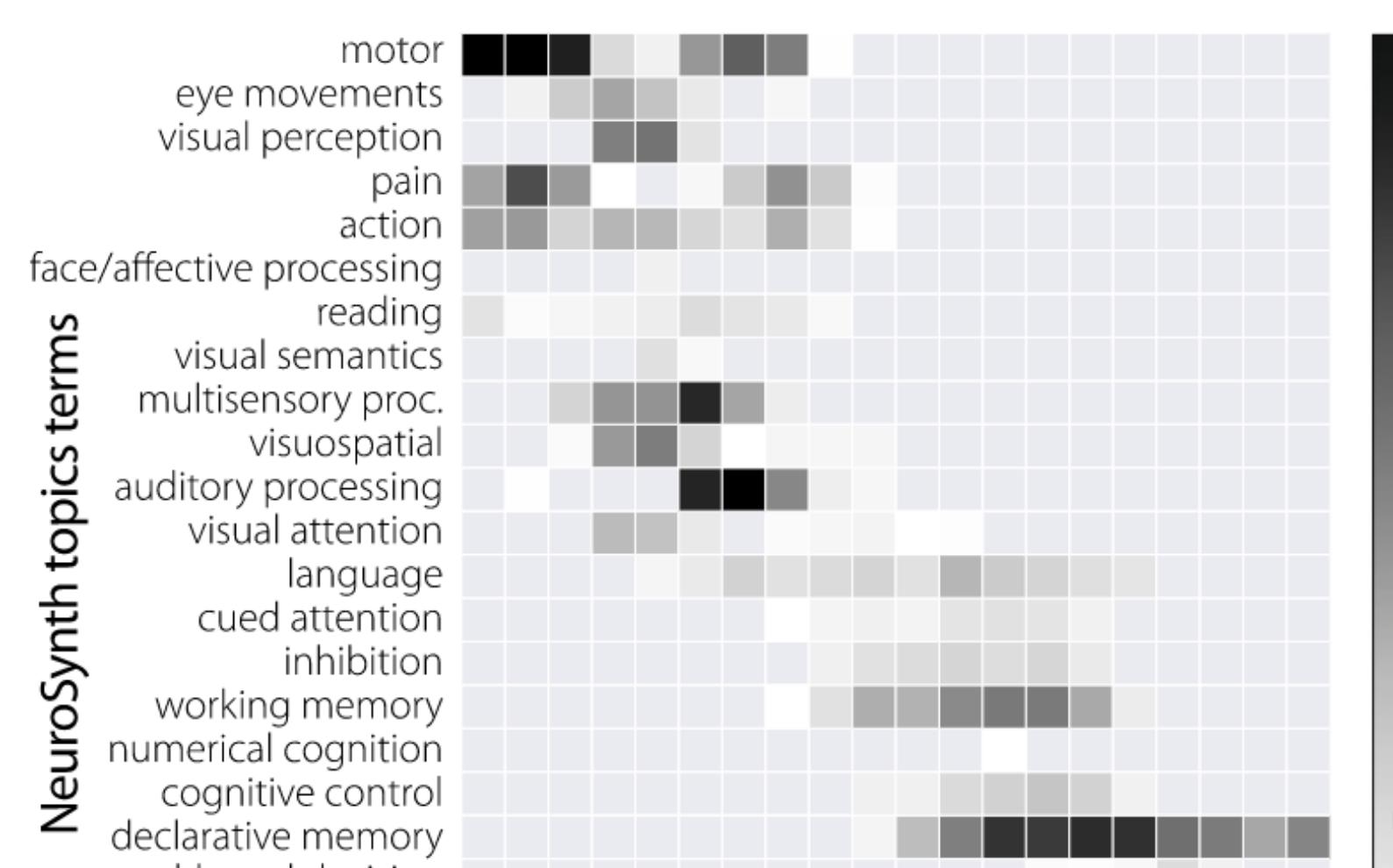


Example of Neurosynth usage

Situating the default-mode network along a principal gradient of macroscale cortical organization

Daniel S. Margulies^{a,1}, Satrajit S. Ghosh^{b,c}, Alexandros Goulas^d, Marcel Falkiewicz^a, Julia M. Huntenburg^{a,e}, Georg Langs^{f,g}, Gleb Bezgin^h, Simon B. Eickhoff^{i,j}, F. Xavier Castellanos^{k,l}, Michael Petrides^m, Elizabeth Jefferies^{n,o}, and Jonathan Smallwood^{n,o}

12574–12579 | PNAS | November 1, 2016 | vol. 113 | no. 44



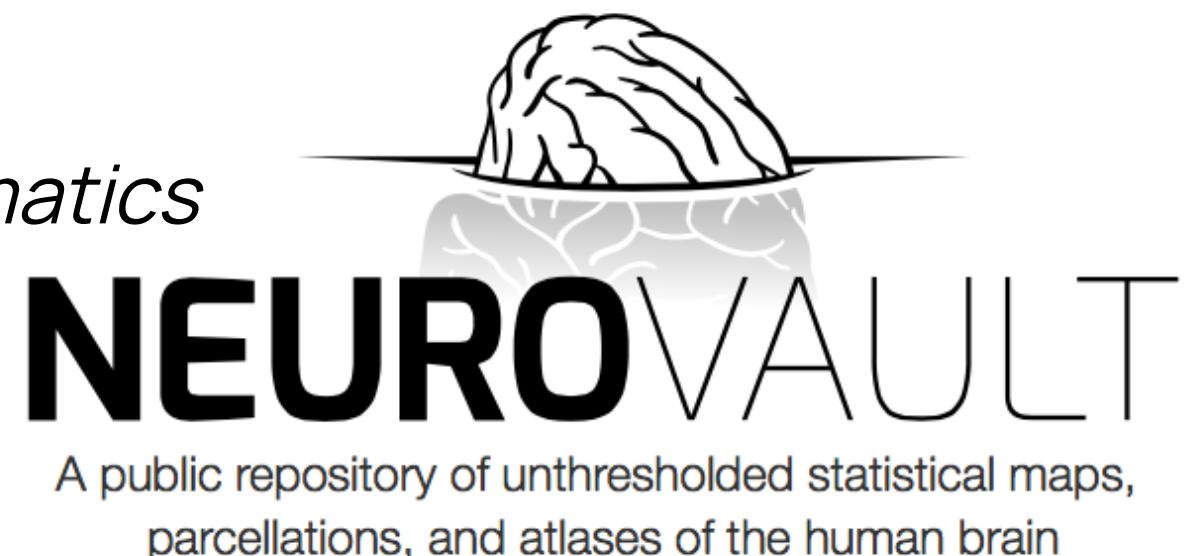
- Identified gradients of functional organization across the cortex
- Used Neurosynth to identify the most common terms associated with each gradient

Neurovault: Sharing neuroimaging results

- The results of most neuroimaging studies are images with statistical estimates at each voxel

Gorgolewski et al., 2015, *Frontiers in Neuroinformatics*

- Neurovault.org is an open archive for these results



- **Collections**

- A set of images (such as all images from a particular paper) can be uploaded as a collection
- Each collection receives a persistent identifier

NeuroVault Collections ▾ Metaanalyses ▾ About ▾ RussPoldrack ▾ Search Search

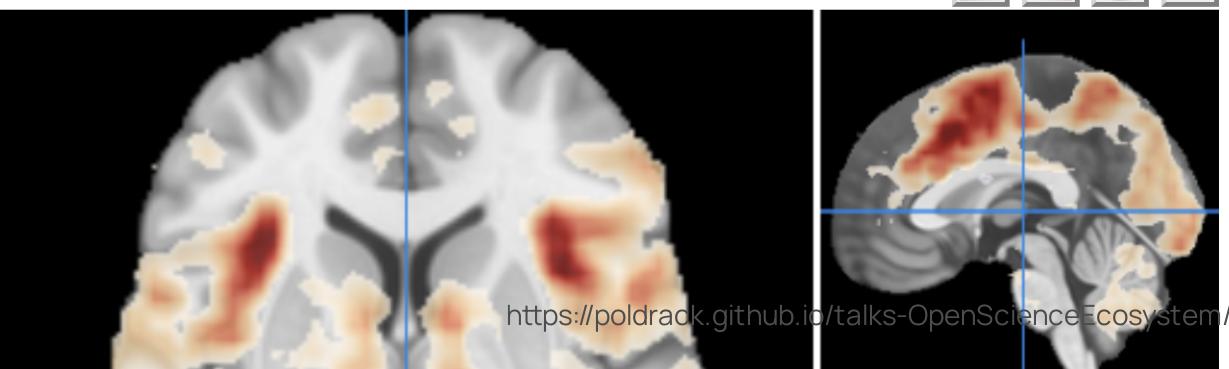
Preprocessed Consortium for Neuropsychiatric Phenomics dataset

Related article: <http://doi.org/10.12688/f1000research.11964.2>

Source data:

3D View Download

File View Settings Help



Group Metadata

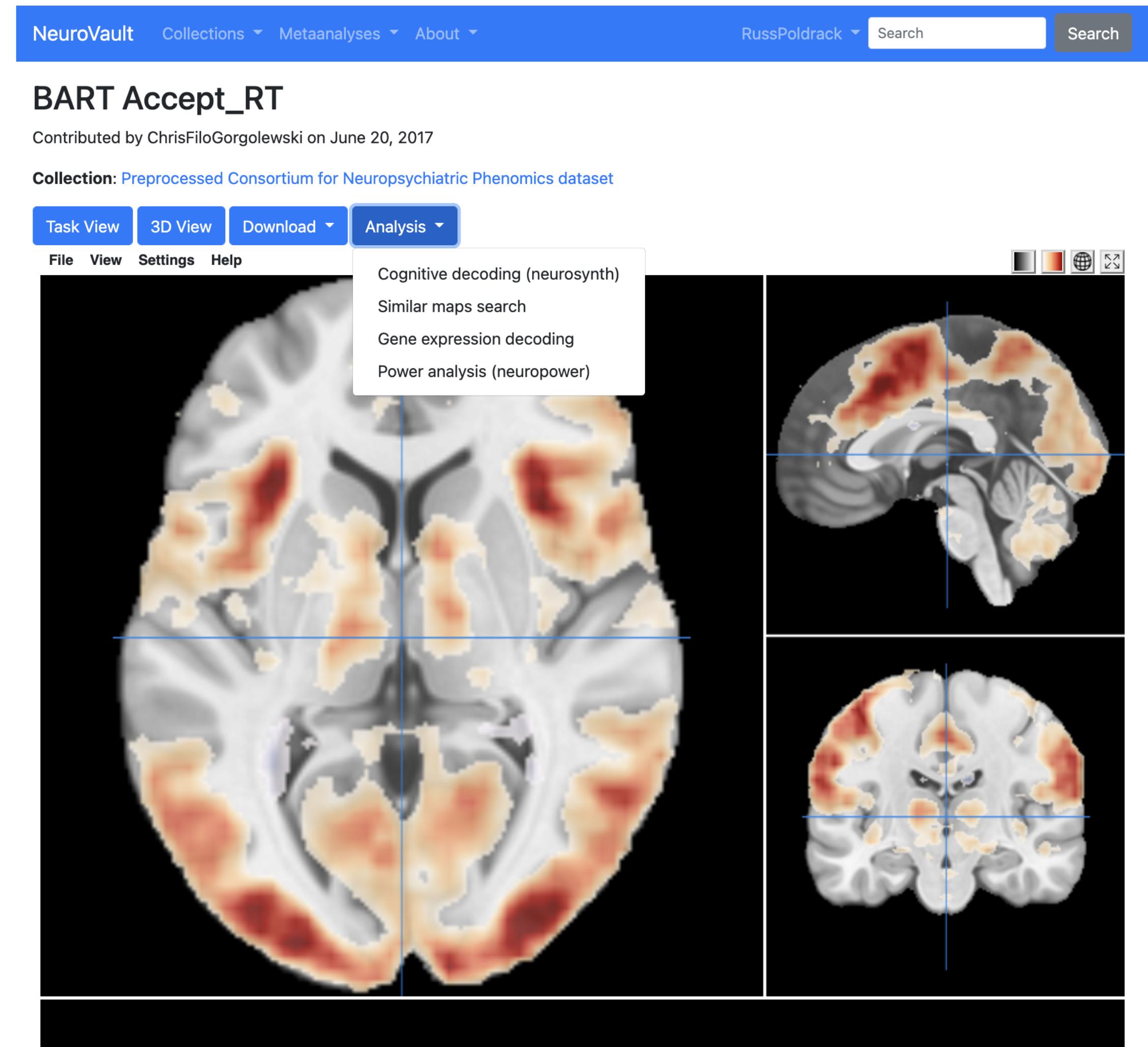
Show 7 entries Search:

View	ID	Name	Type
	49974	BART Accept	T map

<https://poldrack.github.io/talks-OpenScienceEcosystem/>

- **Image browser**

- Individual images can be browsed and downloaded
- A number of analysis tools can also be applied
- Each image also receives a persistent identifier

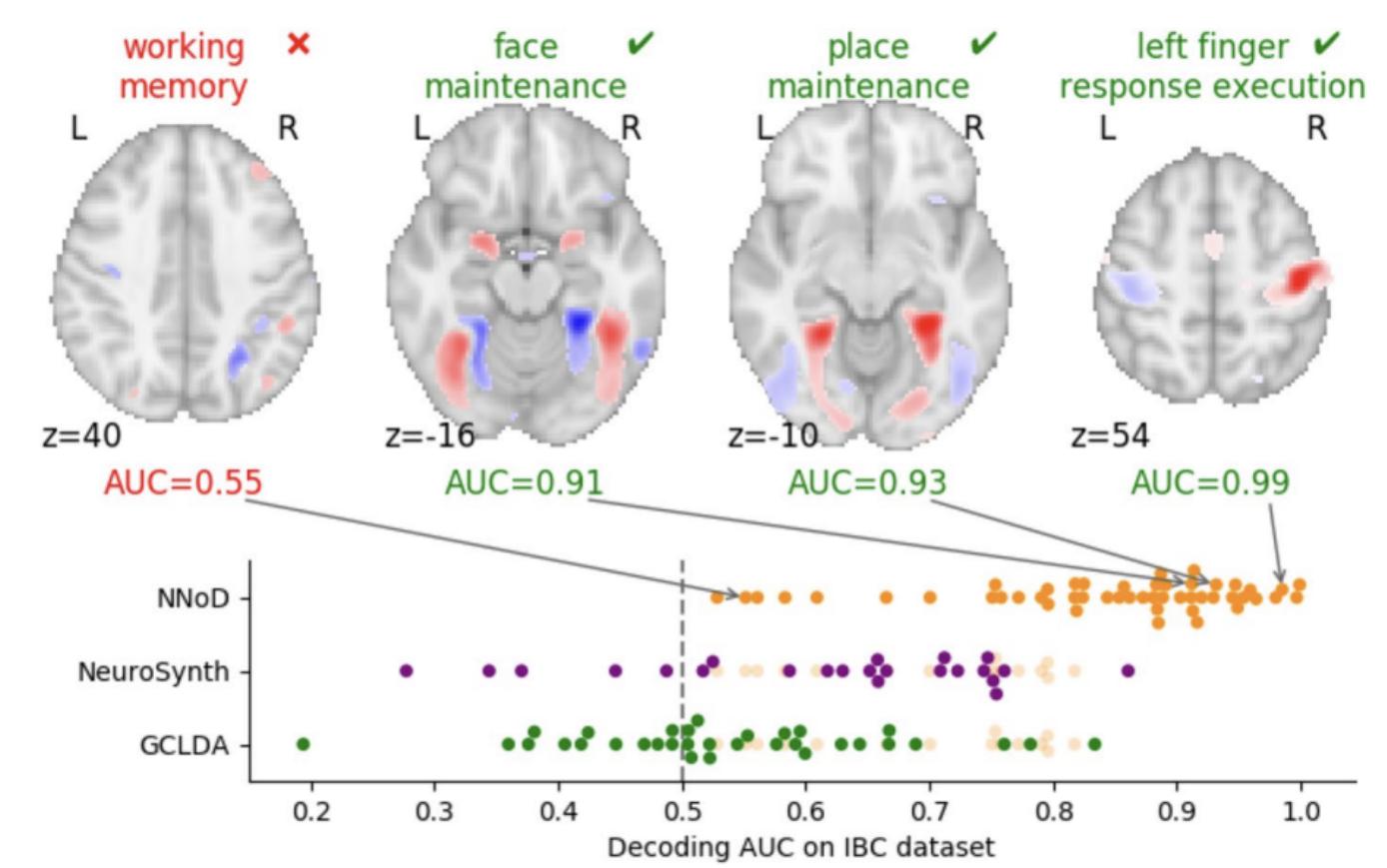
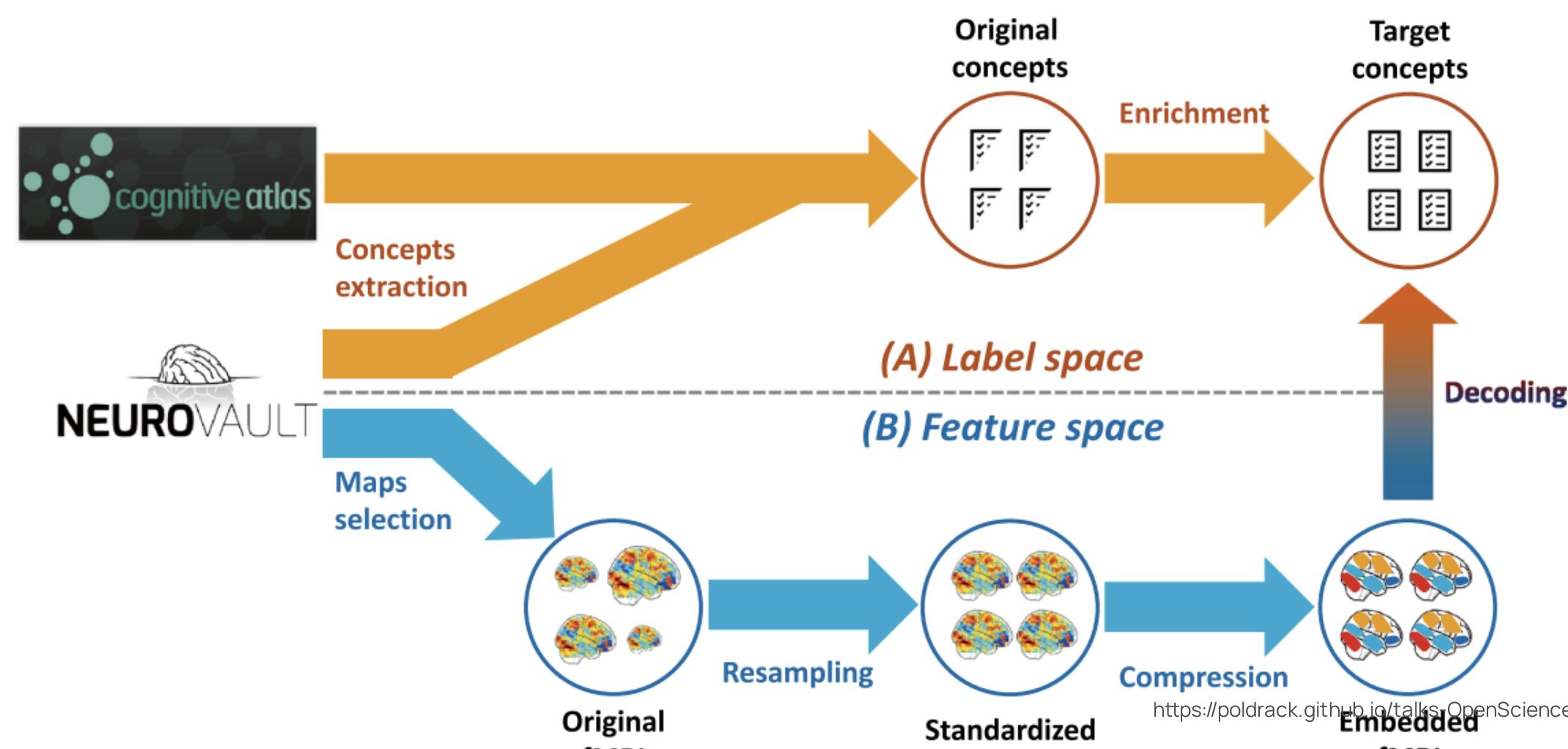


Example of Neurovault usage

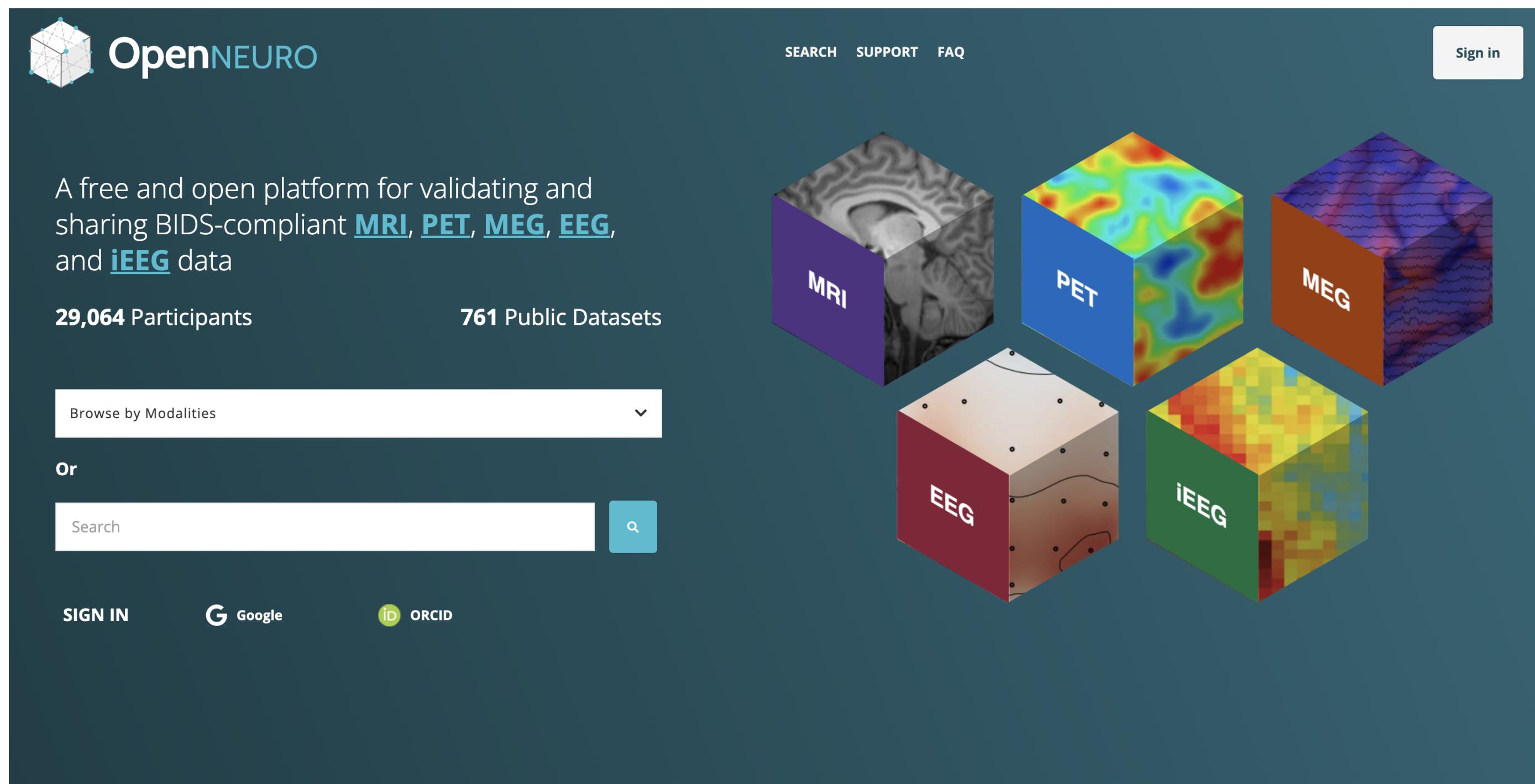
Comprehensive decoding mental processes from Web repositories of functional brain images

Romuald Menuet^{5,6}, Raphael Meudec^{1,2,3,6}, Jérôme Dockès⁴, Gael Varoquaux^{1,2,3} & Bertrand Thirion^{1,2,3✉}

Scientific Reports | (2022) 12:7050



OpenNeuro: Sharing raw and processed neuroimaging data



Validation Using BIDS

The [Brain Imaging Data Structure](#) (BIDS) is an emerging standard for the organization of neuroimaging data.

Want to contribute to BIDS?

Visit the [Google discussion group](#) to contribute.



OpenNeuro Runs on DataLad

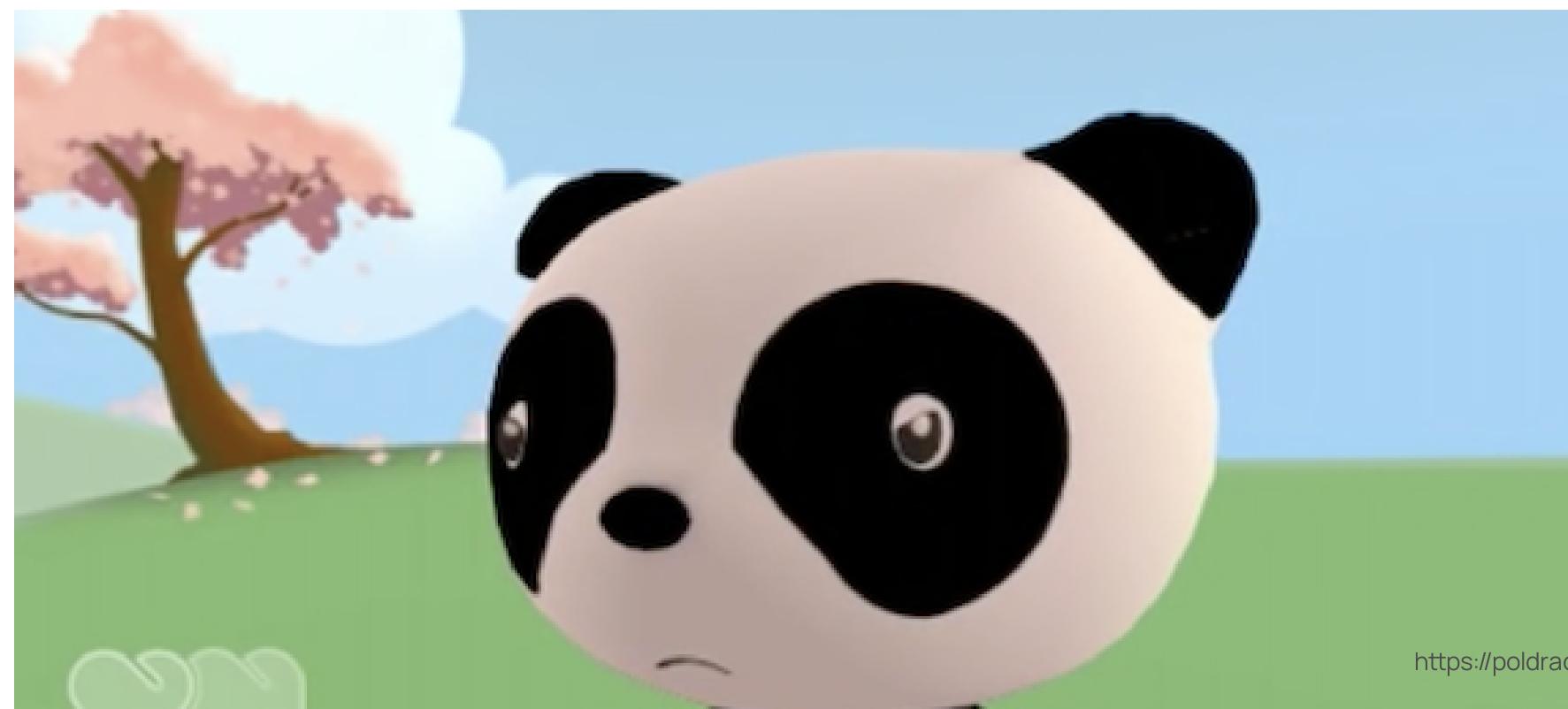
Want to access OpenNeuro datasets with DataLad? Visit the [dataset collection on GitHub](#).

A data management solution built on [Git](#) and [Git-annex](#). Read more about [DataLad](#).

Simply sharing data is not sufficient
It must be shared in a way that makes it useful!

It's easy to share data badly

Data Sharing and Management Shafu in 3 Short Acts
<https://www.youtube.com/watch?v=N2zK3sAtr-4>



- I received the data, but when I opened it up it was in hexadecimal
- Yes, that is right
- I cannot read hexadecimal
- You asked for my data and I gave it to you. I have done what you asked.

Brain Imaging Data Structure (BIDS)

- A community-based open standard for neuroimaging data
 - A file organization standard
 - A metadata standard



SCIENTIFIC DATA<sup>110110
0111101
11011110
011101101</sup>

OPEN

SUBJECT CATEGORIES

- » Data publication and archiving
- » Research data

Received: 18 December 2015

Accepted: 19 May 2016

Published: 21 June 2016

The brain imaging data structure,
a format for organizing and
describing outputs of neuroimaging
experiments

Krzysztof J. Gorgolewski¹, Tibor Auer², Vince D. Calhoun^{3,4}, R. Cameron Craddock^{5,6}, Samir Das⁷, Eugene P. Duff⁸, Guillaume Flandin⁹, Satrajit S. Ghosh^{10,11}, Tristan Glatard^{7,12}, Yaroslav O. Halchenko¹³, Daniel A. Handwerker¹⁴, Michael Hanke^{15,16}, David Keator¹⁷, Xiangrui Li¹⁸, Zachary Michael¹⁹, Camille Maumet²⁰, B. Nolan Nichols^{21,22}, Thomas E. Nichols^{20,23}, John Pellman⁶, Jean-Baptiste Poline²⁴, Ariel Rokem²⁵, Gunnar Schaefer^{1,26}, Vanessa Sochat²⁷, William Triplett¹, Jessica A. Turner^{3,28}, Gaël Varoquaux²⁹ & Russell A. Poldrack¹

The development of BIDS

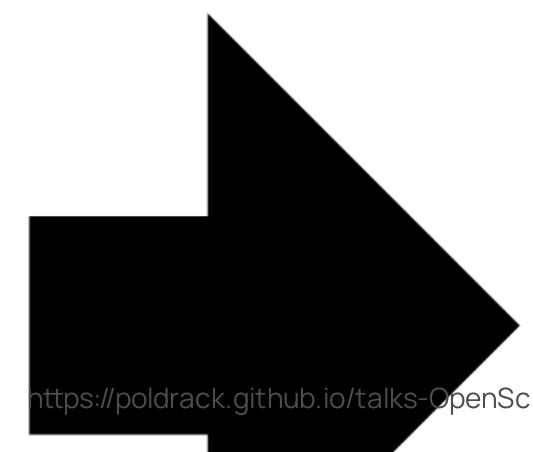
- January 2015
 - Initial stakeholder meeting at Stanford (funded by INCF)
 - Initiated development of a draft standard
- September 2015
 - Draft standard posted to BIDS web site with 22 example datasets
 - Solicited feedback from community

BIDS Principles

- *Adoption is crucial*
 - Keep it as similar to existing practices as possible
 - Don't let technology override usability!
 - Focus on engaging the community
- *Don't reinvent the wheel*
 - Use existing standards when possible

From DICOM to BIDS

```
dicomdir/
  1208200617178_22/
    1208200617178_22_8973.dcm
    1208200617178_22_8943.dcm
    1208200617178_22_2973.dcm
    1208200617178_22_8923.dcm
    1208200617178_22_4473.dcm
    1208200617178_22_8783.dcm
    1208200617178_22_7328.dcm
```



```
my_dataset/
  participants.tsv
  sub-01/
    anat/
      sub-01_T1w.nii.gz
    func/
      sub-01_task-rest_bold.nii.gz
      sub-01_task-rest_bold.json
    dwi/
```

The importance of automated validation

Summary

- 40 Files, 18.42kB
- 13 - Subjects
- 1 - Session

Available Tasks

- rhyme judgment

Available Modalities

- bold
- T1w

BIDS Extensions

- BIDS was originally focused on structural/functional MRI data
- BIDS extension process allows extension of the standard through BIDS Extension Proposals (BEPs) initiated by the community
 - Patterned after the Python Enhancement Proposal (PEP) process

11 Completed BEPs:

BEP #	Title
BEP001	Quantitative MRI (qMRI)
BEP003	Common Derivatives
BEP005	Arterial Spin Labeling (ASL)
BEP006	Electroencephalography (EEG)
BEP007	Hierarchical Event Descriptor (HED) Tags

The growing usage of BIDS: An example

- MRIQC Web API
 - Crowdsourced database of MR QC metrics
 - QC metrics from ~375K unique BOLD scans and ~280K T1w scans as of June 2022
 - Publicly available:

SCIENTIFIC DATA

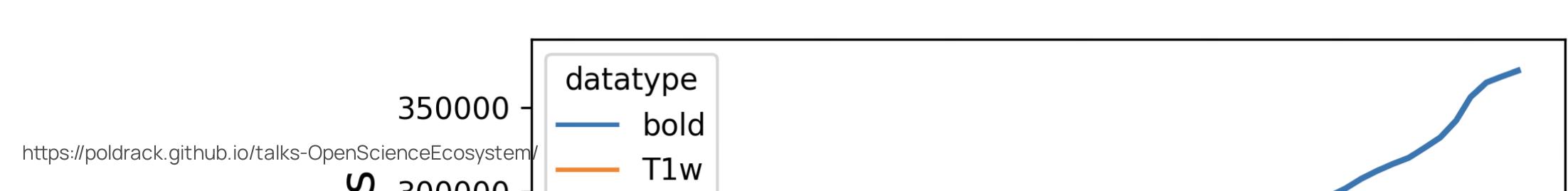
OPEN

DATA DESCRIPTOR

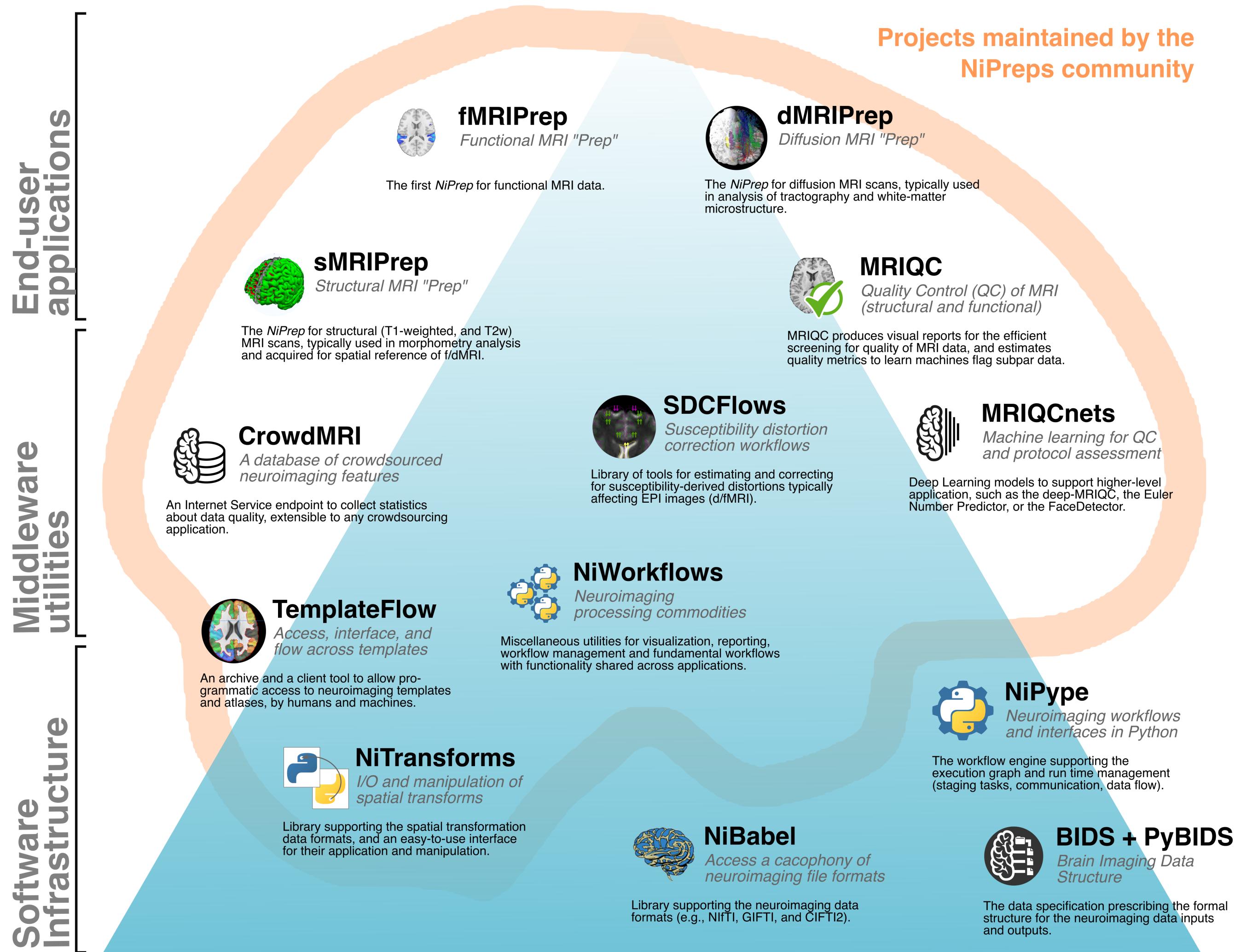
Received: 19 September 2018
Accepted: 12 March 2019

Crowdsourced MRI quality metrics and expert quality annotations for training of humans and machines

Oscar Esteban  ¹, Ross W. Blair  ¹, Dylan M. Nielson  ², Jan C. Varada  ³, Sean Marrett  ³, Adam G. Thomas  ², Russell A. Poldrack  ¹ & Krzysztof J. Gorgolewski  ¹



BIDS enables a growing open-source software ecosystem



RESEARCH ARTICLE

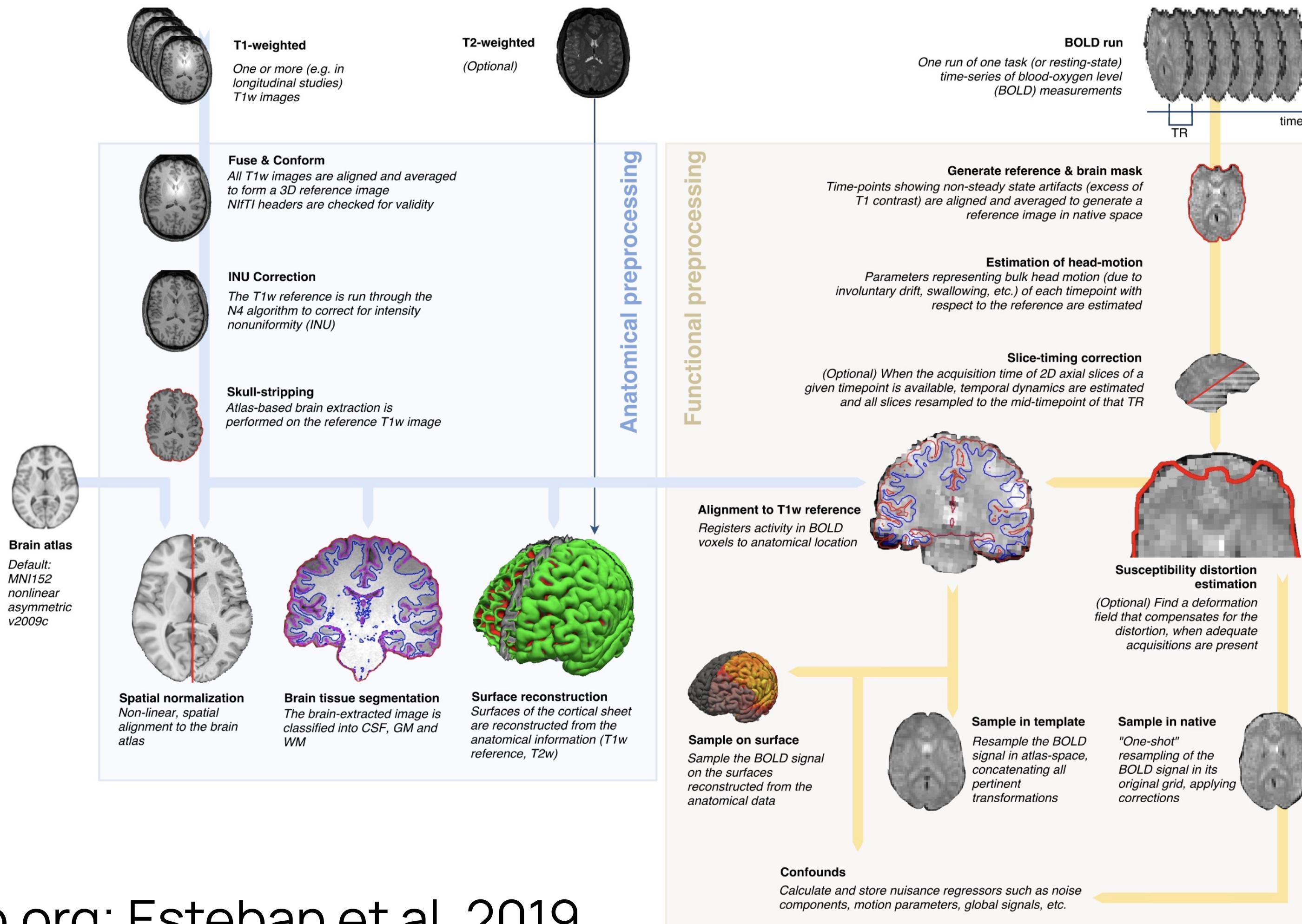
BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods

Krzysztof J. Gorgolewski^{1*}, Fidel Alfaro-Almagro², Tibor Auer³, Pierre Bellec^{4,5}, Mihai Capotă⁶, M. Mallar Chakravarty^{7,8}, Nathan W. Churchill⁹, Alexander Li Cohen¹⁰, R. Cameron Craddock^{11,12}, Gabriel A. Devenyi^{7,8}, Anders Eklund^{13,14,15}, Oscar Esteban¹, Guillaume Flandin¹⁶, Satrajit S. Ghosh^{17,18}, J. Swaroop Guntupalli¹⁹, Mark Jenkinson², Anisha Keshavan²⁰, Gregory Kiar^{21,22}, Franziskus Liem²³, Pradeep Reddy Raamana^{24,25}, David Raffelt²⁶, Christopher J. Steele^{7,8}, Pierre-Olivier Quirion¹⁵, Robert E. Smith²⁶, Stephen C. Strother^{24,25}, Gaël Varoquaux²⁷, Yida Wang⁶, Tal Yarkoni²⁸, Russell A. Poldrack¹

PLOS Computational Biology | <https://doi.org/10.1371/journal.pcbi.1005209> March 9, 2017

- Containerized applications that can be run on a BIDS dataset
 - Containers provide ease of use as well as better reproducibility

fMRIprep: Robust preprocessing of fMRI data



fmriprep.org; Esteban et al, 2019

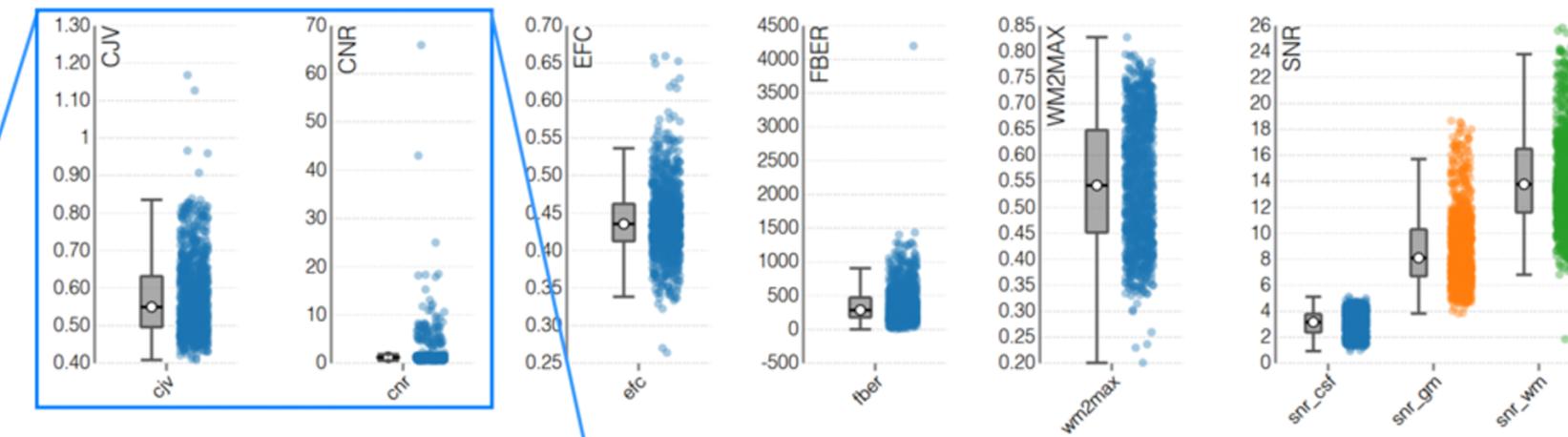
MRIQC: MRI quality control for BIDS data

①

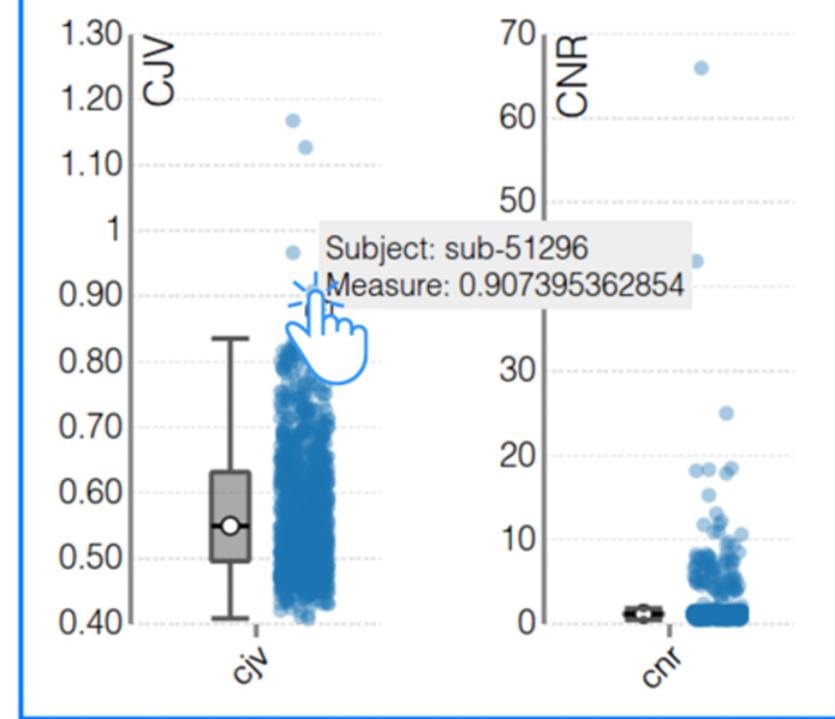
MRIQC: group anatomical report

Summary

- Date and time: 2017-02-05, 12:27.
- MRIQC version: 0.9.0-rc2.



②



Data points in the scatter plots of the group report can be clicked to open the corresponding individual report. This feature is particularly useful to identify low-quality datasets visually.

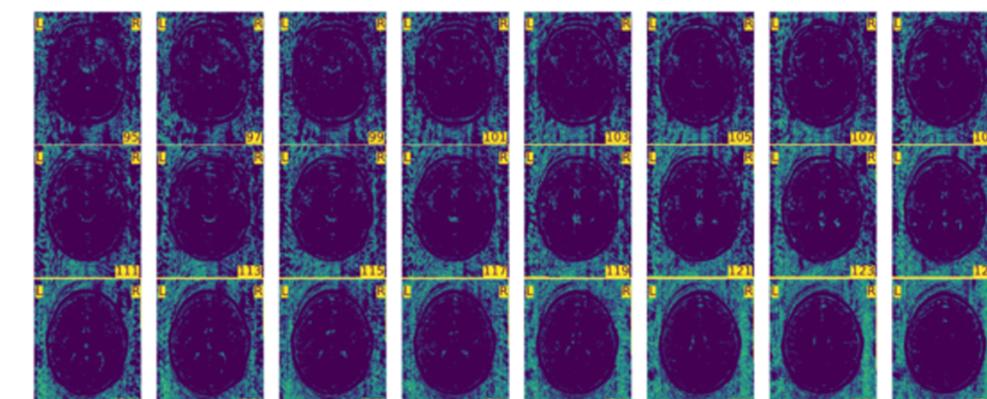
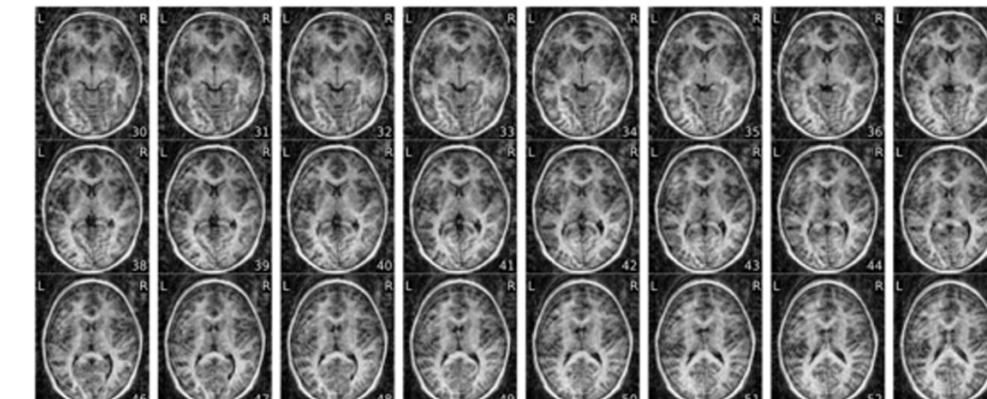
③

The individual reports show the calculated IQMs and metadata in the summary, and a series of image mosaics and plots designed for the visual assessment of images.

MRIQC: individual anatomical report

Summary

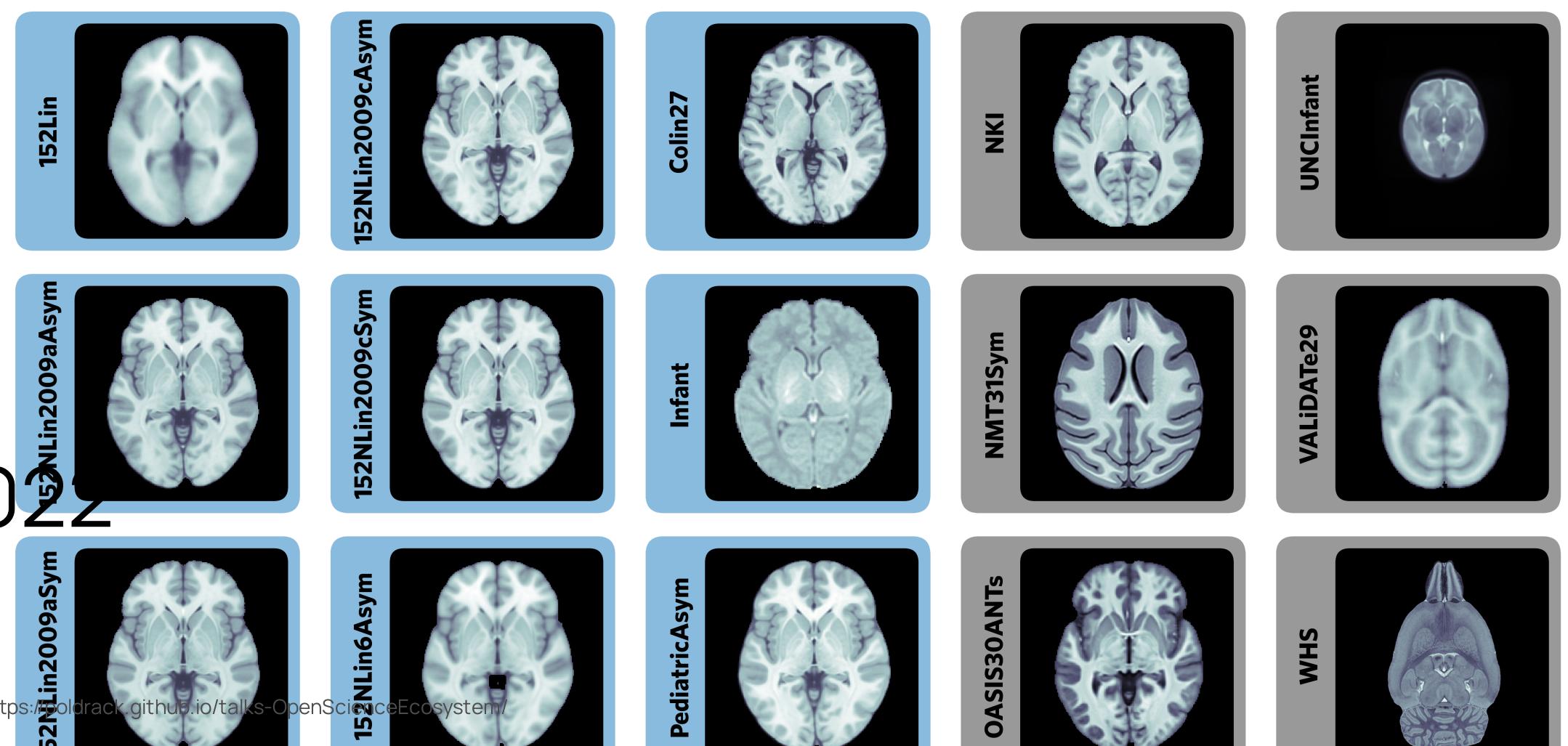
- Subject ID: 51296.
- Date and time: 2017-02-05, 03:44.
- MRIQC version: 0.9.0-rc2.



Tentemplateflow: FAIR Sharing of Neuroimaging Templates

- Templates and atlases are commonly used in neuroimaging
- There is a significant lack of clarity in the use of these templates

templateflow.org; Cricic et al., 2022



OpenNeuro: A BRAIN Initiative archive for BIDS data

The screenshot shows the OpenNeuro website homepage. At the top left is the OpenNeuro logo. To its right are links for SEARCH, SUPPORT, and FAQ, and a SIGN IN button. Below these are five large, colorful cubes representing different neuroimaging modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). To the left of the cubes, there is text about the platform being a free and open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, and iEEG data, and statistics showing 29,064 Participants and 761 Public Datasets. Below this is a search bar with a dropdown menu labeled "Browse by Modalities" and a text input field containing "reading". At the bottom are links for SIGN IN, Google, and ORCID.

- Supports sharing of any validated BIDS dataset



OpenNEURO

SEARCH SUPPORT FAQ Sign in

Search All Datasets

Keywords ? Enter Keyword(s) to Search +

These filters return **194** datasets: CLEAR ALL

KEYWORD: reading X

reading X SORT BY: Relevance ↑

The Reading Brain Project L1 Adults
Uploaded by: Chanyuan Gu on 2022-01-07 - 10 months ago | Updated: 2022-01-05 - 10 months ago

MODALITY: MRI

TASKS: read task rest

OPENNEURO ACCESSION NUMBER: [ds003974](#) SESSIONS: 1 REPO: [black.github.io](#) RANK: 52 SCIENCE PAPERS: N/A PARTICIPANTS' AGES: N/A SIZE: 46.67GB FILES: 893

Modalities

- MRI
- PET
- EEG
- iEEG

☰ 38 / 65

Each shared dataset is versioned and receives a persistent identifier (DOI)

The screenshot shows the OpenNEURO dataset page for 'The Reading Brain Project L2 Adults'. The top navigation bar includes links for SEARCH, SUPPORT, FAQ, and SIGN IN. The main header features the dataset title 'The Reading Brain Project L2 Adults' with a 'MRI' icon. Below the header, there are sections for 'BIDS Validation' (4 WARNINGS, Valid), 'brainlife.io', and 'Clone'. The dataset's OpenNeuro Accession Number is listed as ds003988. The 'Authors' section lists Ping Li, Chun-Ting Hsu, Ben Schloss, Anya Yu, Lindsey Ma, Marissa Scotto, Friederike Seyfried, Chanyuan Gu. The 'Available Modalities' section shows 'MRI'. The 'Versions' section displays '1.0.0' (Created: 2022-02-01) with a 'Versions' dropdown arrow. The 'Files' tab is selected, showing a 'README' section with a note about privacy and a detailed description of the dataset's contents, including MRI scanning and behavioral tests. Other tabs include 'Download' and 'Metadata'.

SEARCH SUPPORT FAQ

SIGN IN

MRI The Reading Brain Project L2 Adults

BIDS Validation 4 WARNINGS Valid brainlife.io Clone

OpenNeuro Accession Number
ds003988

Authors
Ping Li, Chun-Ting Hsu, Ben Schloss, Anya Yu, Lindsey Ma, Marissa Scotto, Friederike Seyfried, Chanyuan Gu

Available Modalities
MRI

Versions
1.0.0 Created: 2022-02-01 Versions

Files README

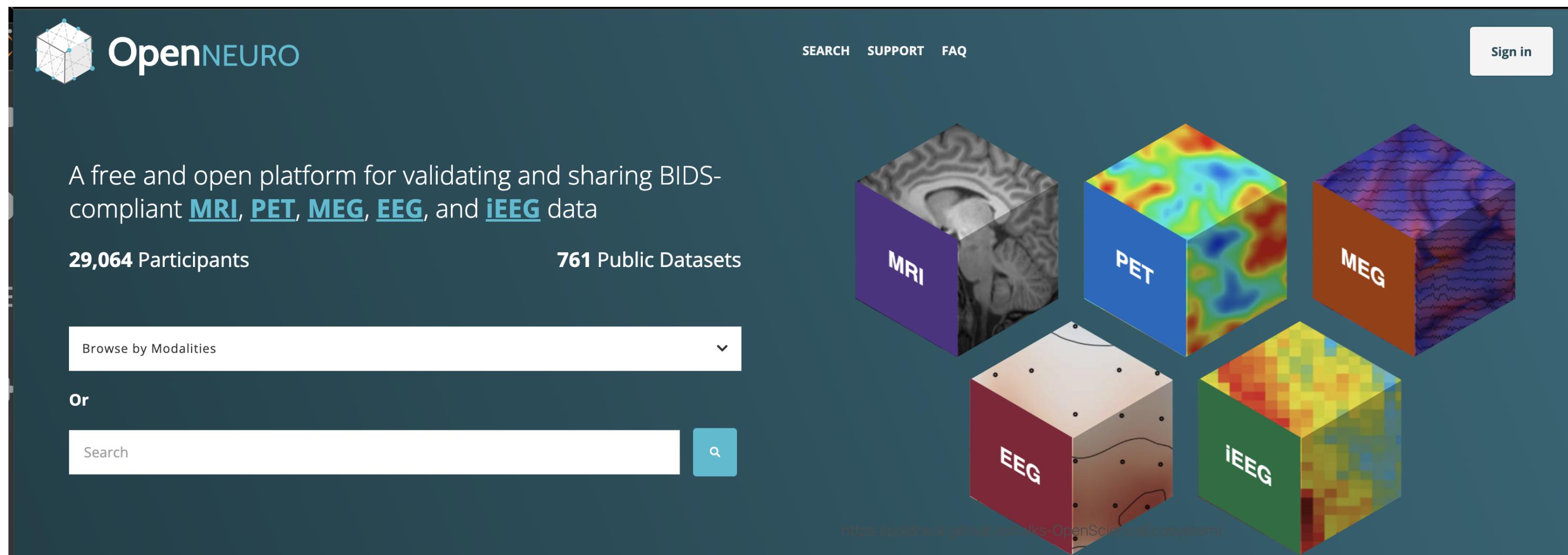
OpenNeuro curator note: This dataset was previously accessible at ds002317. The dataset was reuploaded due to privacy considerations.

This dataset contains the bilingual (L2) adult subset of the Reading Brain Project (RBP) data, focusing on 56 participants who underwent two sessions of testing: MRI scanning and behavioral tests tests; collected 2-3 days apart. During the first session, following the scanning of structural (T1-weighted) and resting-state data, participants performed a reading task with simultaneous eye-tracking and fMRI scanning, and the session ended with a diffusion tensor imaging (DTI) scan. The second session consisted of only behavioral tests, including five standardized tests: the Attention Network Test (ANT), Gray's Silent Reading Test (GSRT), Letter-Number Sequencing (LNS), Peabody Picture Vocabulary Test (PPVT-4), and Tower of Hanoi (ToH), followed by a survey: Reading Background Questionnaire (RBQ), which includes familiarity rating for the topics of our five reading texts. The raw data are all provided here. The first 28 subjects (sub-01 to sub-28) were native speakers of Mandarin Chinese living in the United States. Their data was collected in Hershey, PA. The second 28 subjects (sub-29 to sub-56) were native speakers of Mandarin Chinese who lived in China. Their data was collected in Beijing, China.

https://poldrack.github.io/TaskTalks-OpenScienceEcosystem/read_task_rest

An updated version of the methodology document will be available at http://blclab.org/reading_brain under L2 Adult dataset/

Any valid BIDS dataset can be shared via
OpenNeuro



The image shows the homepage of OpenNEURO, a platform for sharing neuroimaging data. The header features the OpenNEURO logo, a search bar, and links for "SEARCH", "SUPPORT", "FAQ", and "Sign in". Below the header, a main banner highlights "A free and open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, and iEEG data". It also displays statistics: "29,064 Participants" and "761 Public Datasets". A "Browse by Modalities" dropdown menu is shown, along with a search bar and a "Search" button. To the right, five 3D cube icons represent different data modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). The bottom of the page includes a navigation bar with three vertical bars and a URL: <https://poldrack.github.io/talks-OpenScienceEcosystem/>.

OpenNEURO

SEARCH SUPPORT FAQ Sign in

A free and open platform for validating and sharing BIDS-compliant MRI, PET, MEG, EEG, and iEEG data

29,064 Participants

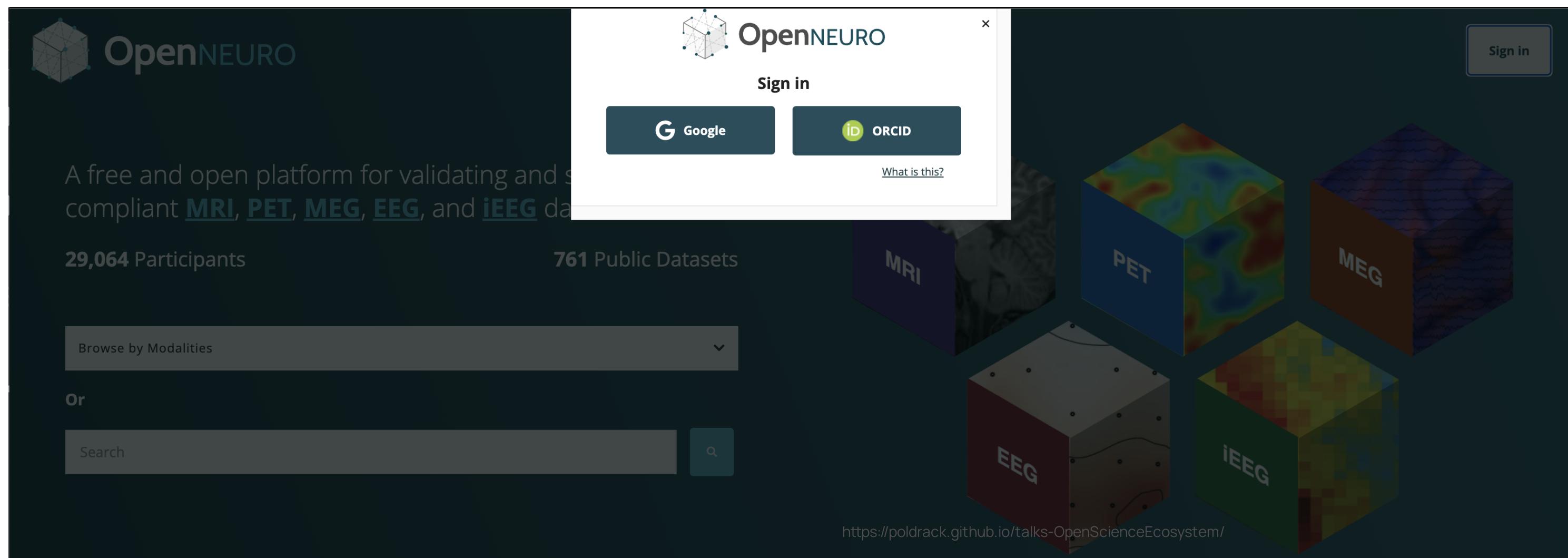
761 Public Datasets

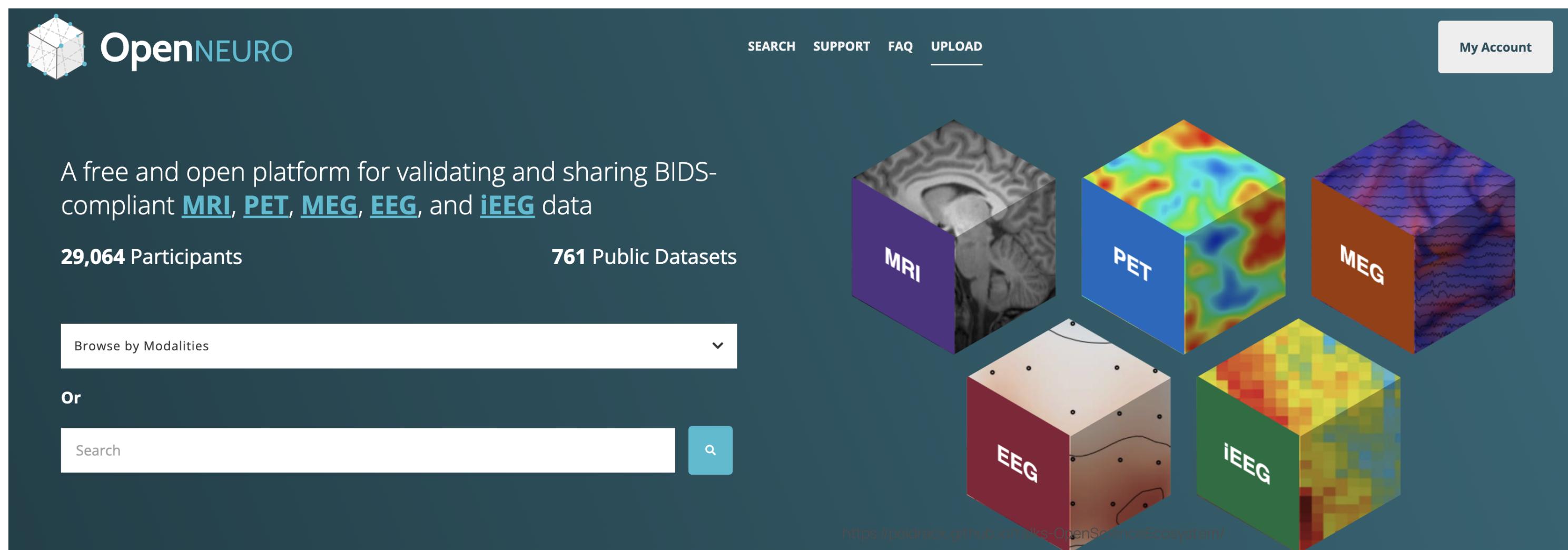
Browse by Modalities

Or

Search

https://poldrack.github.io/talks-OpenScienceEcosystem/





The image shows the homepage of the OpenNEURO platform. The header features the "OpenNEURO" logo with a brain cube icon. Navigation links include SEARCH, SUPPORT, FAQ, UPLOAD (underlined), and My Account. Key statistics are displayed: 29,064 Participants and 761 Public Datasets. A "Browse by Modalities" dropdown menu is open, showing options like MRI, PET, MEG, EEG, and iEEG. A search bar with a magnifying glass icon is also present. The main visual area displays five 3D cubes representing different data modalities: MRI (purple), PET (blue), MEG (orange), EEG (red), and iEEG (green). Below the cubes, there is a URL: <https://poldrack.github.io/talks-OpenScienceEcosystem/>.

OpenNEURO

SEARCH SUPPORT FAQ UPLOAD My Account

A free and open platform for validating and sharing BIDS-compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and [iEEG](#) data

29,064 Participants **761** Public Datasets

Browse by Modalities

Or

Search

<https://poldrack.github.io/talks-OpenScienceEcosystem/>



OpenNEURO

A free and open platform for validating compliant [MRI](#), [PET](#), [MEG](#), [EEG](#), and

29,064 Participants

Browse by Modalities

Or

Upload Dataset

Step 1: Select Files

Step 2: Validation

Step 3: Metadata

Step 4: Accept Terms

To protect the privacy of the individuals who have been scanned, we require that all scan data be defaced before publishing a dataset.

Select a [BIDS dataset](#) to upload

Select folder

close

My Account

The screenshot shows the OpenNEURO dataset upload process at Step 2: Validation. The main message indicates 3 warnings found in the dataset, which are not required to be fixed but will make it more BIDS compliant. A 'Continue' button is available to proceed. Below this, a yellow bar highlights 'VIEW 3 WARNINGS IN 15 FILES'. One warning is detailed: 'Warning: 1' about the 'SliceTiming' field being undefined, with a link to 'VIEW 13 FILES' for more details.

OpenNEURO

Rhyme judgment Edit

This dataset is public

BIDS Validation

Files Share Versioning Admin

How to Download

Upload Dataset

Step 1: Select Files Step 2: Validation Step 3: Metadata Step 4: Accept Terms

We found 3 warnings in your dataset. You are not required to fix warnings, but doing so will make your dataset more BIDS compliant. Continue or fix the issues and select folder again.

Continue

VIEW 3 WARNINGS IN 15 FILES

Warning: 1 VIEW 13 FILES

You should define 'SliceTiming' for this file. If you don't provide this information slice time correction will not be possible. 'Slice Timing' is the time at which each slice was acquired within each volume (frame) of the acquisition. Slice timing is not slice order -- rather, it is a list of times containing the time (in seconds) of each slice acquisition in relation to the beginning of volume acquisition.

<https://poldrack.github.io/talks-OpenScienceEcosystem/>

Follow 1 Bookmark 2

My Account

Russell A. Poldrack

45 / 65

The screenshot shows the OpenNEURO dataset upload interface. At the top, there's a navigation bar with the OpenNEURO logo, a search bar, and links for "My Account", "Follow 1", and "Bookmark 2". Below the navigation, a modal window titled "Upload Dataset" is open, divided into four steps: "Step 1: Select Files", "Step 2: Validation", "Step 3: Metadata" (which is active and highlighted in dark blue), and "Step 4: Accept Terms". A note below the steps says: "Incomplete fields in this form will make it more difficult for users to search for your dataset. We recommend completing the applicable fields to improve your search results." The "Step 3: Metadata" section contains several input fields: "DOI of papers from the source data lab" (with a dropdown menu showing "Papers that were published from the Lab that collected this dataset"), "Species" (dropdown menu), "Study Type" (dropdown menu), "Domain Studied" (input field), and "Number of Trials (if applicable)" (input field). On the left side of the modal, there's a sidebar with "BIDS Validation" dropdown, "Files", "Share", "Versioning", and "Admin" buttons, and sections for "How to Download" and "Download with your browser".

The screenshot shows the OpenNeuro dataset upload interface. At the top, there's a navigation bar with the OpenNeuro logo, a search bar, and links for "My Account", "Follow 1", and "Bookmark 2". Below the navigation is a dark sidebar with options like "MRI Rhyme judgment", "Edit", "BIDS Validation", "How to Download", and "Download with your browser". The main content area has a title "Upload Dataset" and a progress bar with four steps: "Step 1: Select Files", "Step 2: Validation", "Step 3: Metadata", and "Step 4: Accept Terms", where "Step 4" is highlighted in a dark blue box. A large text box contains the "Accept Terms" agreement, which states: "By uploading this dataset to OpenNeuro I agree to the following conditions: I am the owner of this dataset and have any necessary ethics permissions to share the data publicly. This dataset does not include any identifiable personal health information as defined by the [Health Insurance Portability and Accountability Act of 1996](#) (including names, zip codes, dates of birth, acquisition dates, etc). I agree to destroy any key linking the personal identity of research participants to the subject codes used in the dataset." It also includes a statement about the Creative Commons CC0 license and a link to the FAQ. At the bottom, it says "This dataset is not subject to GDPR protections" and provides a link: <https://poldracklab.org/talks-OpenScienceEcosystem/>.

 OpenNEURO

SEARCH SUPPORT FAQ UPLOAD

 Rhyme judgment

 Following 1  Bookmark 0

This dataset has not been published! Before it can be published, please [create a version](#)

BIDS Validation 2 WARNINGS  Valid

 Files  Publish  Share  Versioning  Admin  Download  Metadata  Delete

New Version

Create a new version of this dataset for download and public access. This will begin an export of this dataset to GitHub and S3 if it has been made public.

1.0.0   

 <https://poldrack.github.io/talks-OpenScienceEcosystem/>

OpenNeuro Accession Number
ds04338

Authors
Xue, G., Russell A. Poldrack

Available Modalities


Google Calendar - Week of Jul 19, 2015
<https://www.google.com/calendar/render?pli=...>

OpenNEURO

SEARCH SUPPORT FAQ UPLOAD My Account

MRI Rhyme judgment

Follow 1 Bookmark 2

BIDS Validation ▾ 1 ERROR Invalid brainlife.io Clone ▾

Files View Draft Download Derivatives Metadata Deprecate Version

README

This dataset was obtained from the OpenfMRI project (<http://www.openfmri.org>). Accession #: ds003 Description: Rhyme judgment

Release history: 10/06/2011: initial release 3/21/2013: Updated release with QA information 2/18/2016: Updated orientation information in nifti headers for improved left-right determination

This dataset is made available under the Public Domain Dedication and License v1.0, whose full text can be found at <http://www.opendatacommons.org/licenses/pddl/1.0/>. We hope that all users will follow the ODC Attribution/Share-Alike Community

OpenNeuro Accession Number
ds000003

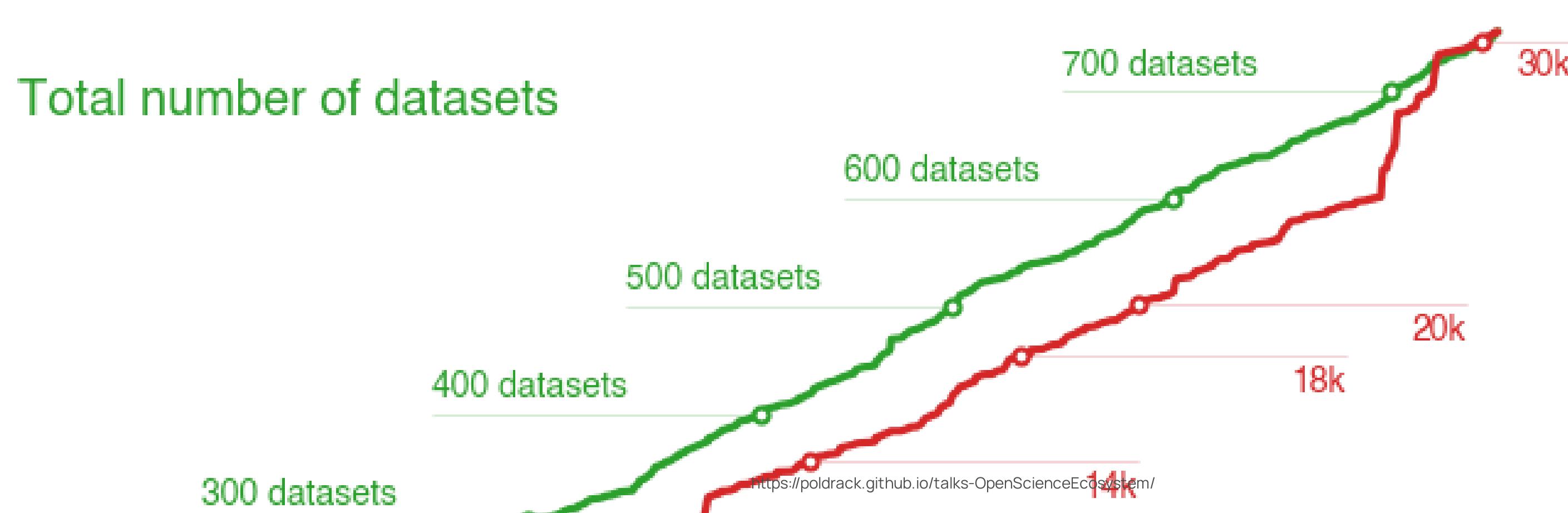
Authors
Xue, G., Russell A. Poldrack

Available Modalities
MRI

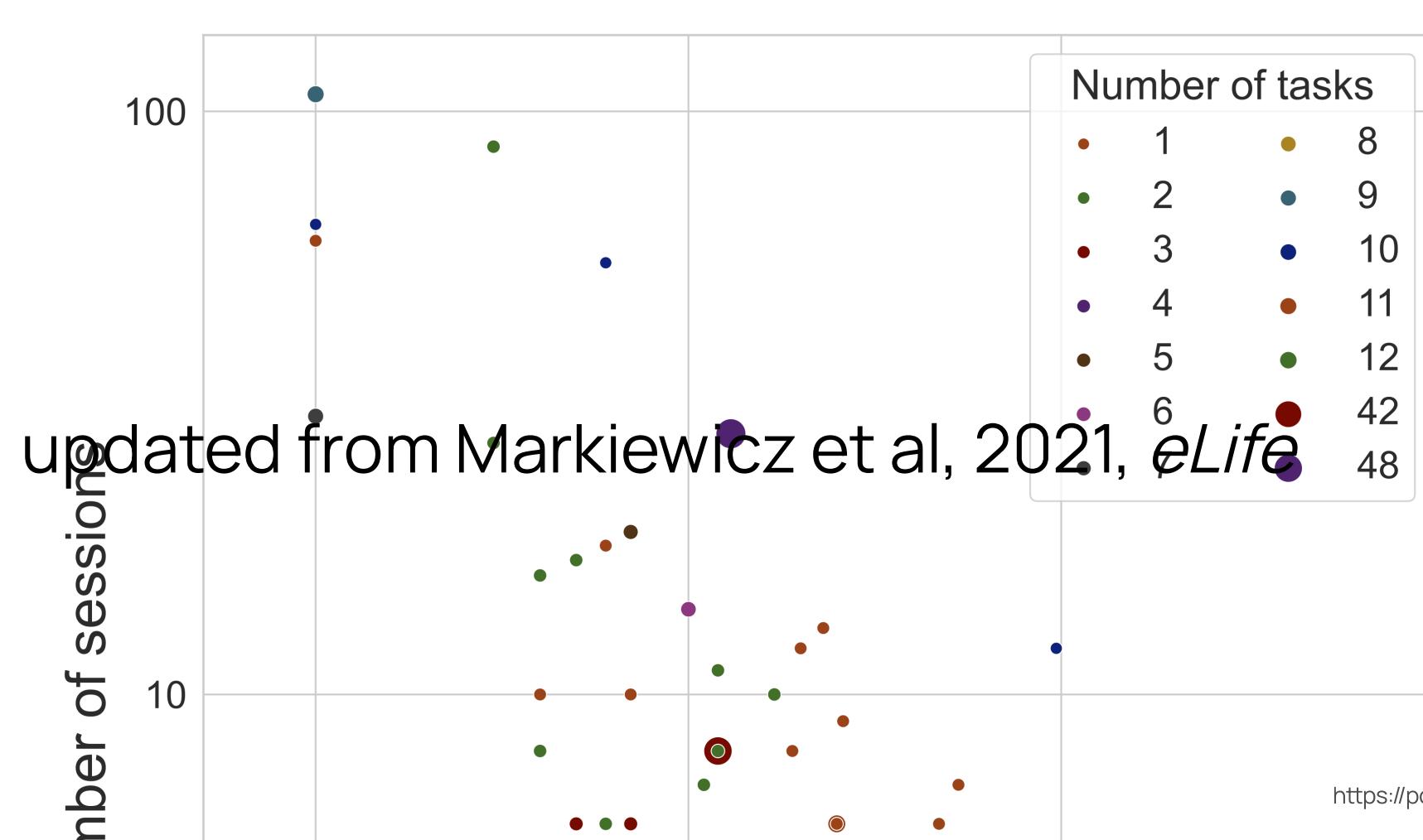
Versions

1.0.0 Versions ▾
Created: 2020-05-14

The growth of OpenNeuro



The diversity of OpenNeuro datasets



Datatype	#
mri - anat	597
mri - func	521
eeg	120
mri - dwi	67
meg	30
ieeg	17
beh	13

Species	#
Human	676
Mouse	20
Rat	12
NHP	2
phantoms	1
Juvenile pigs	1
Human, Mouse	1

Scholarly reuse of OpenNeuro datasets

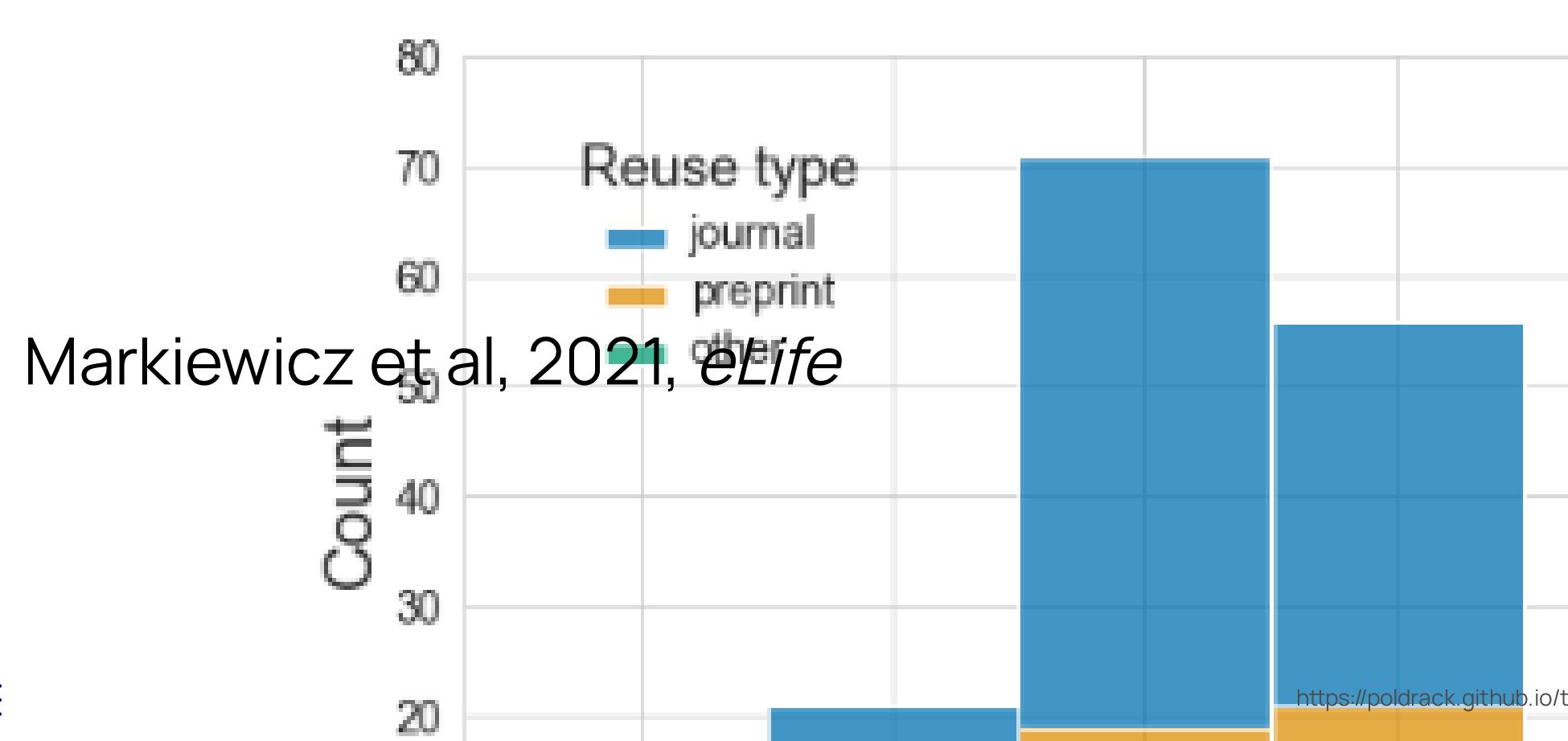


Figure 5. Published reuses of OpenNeuro datasets, split by the type of reuse. Note that the final bar includes only reuses identified through June 2021.

Processing of OpenNeuro data

brainlife.io: processing of MRI data

brainlife < Search Projects

DATASETS PROJECTS APPS PUBLICATIONS DATATYPES

PUBLIC/PROTECTED PROJECTS

HCP 3T / Diffusion
anat/t1w transform/nifti anat/t2w
hcp/freesurferpost raw dwi freesurfer
Human Connectome Project Datasets -
Diffusion MRI 3T (1200-subjects-data-
1112 sub | 6880 objs (4.51 TB)

HCP 7T / Diffusion
anat/t1w dwi
Human Connectome Project Datasets -
Diffusion MRI 3T (184 out of 1200-
150 sub | 300 objs (22.56 MB)

HCP 3T Retest / Diffusion
transform/nifti dwi hcp/freesurferpost
anat/t2w freesurfer anat/t1w
45 HCP 3T subjects retested

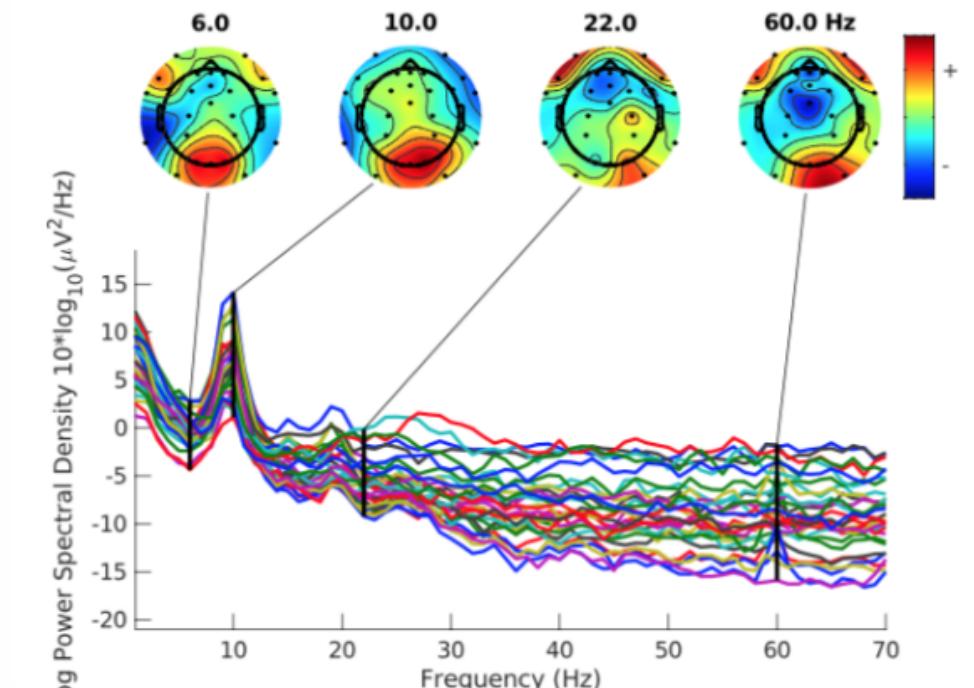
O3D
networkneuro freesurfer anat/t1w wmc-
deprecated dwi dtiinit recon track/tck
LIFE track/trk

O3D (Open Diffusion Data and Derivatives)
<https://poldrack.github.io/talks-OpenScienceEcosystem/>
is a reference repository for precision

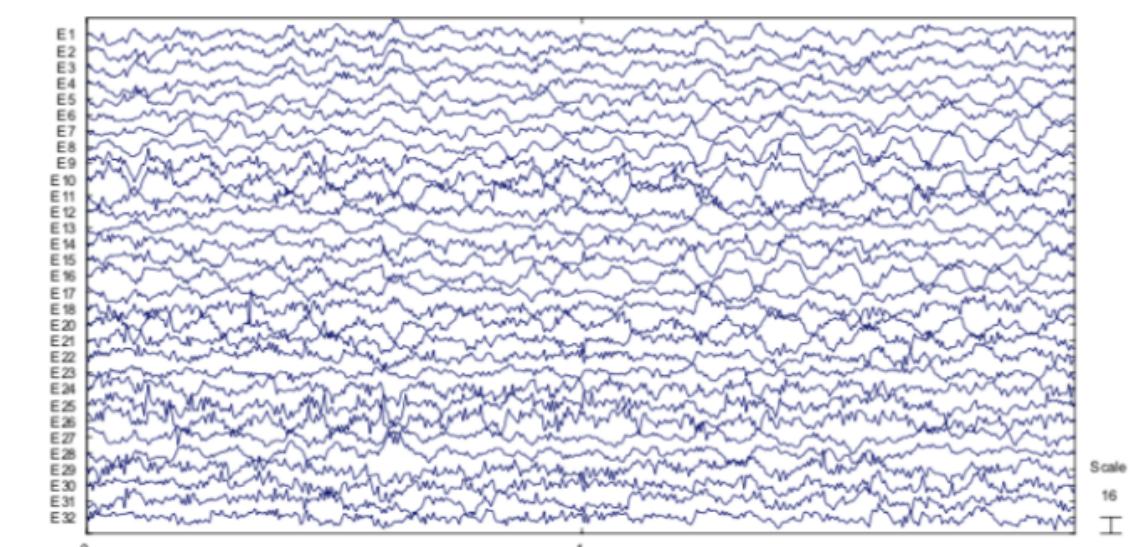
Data Summary

Subject: sub-001
Session(s): 1, Run(s): 1
Data size: 32 channels, 298k frames
Acceptable scalp channels: 100.0% (32 of 32) ⓘ
Acceptable data points channels: 90.9% (271k of 298k) ⓘ
Source quality metric based on independent component: 48.4% ⓘ

Scalp channel log spectra



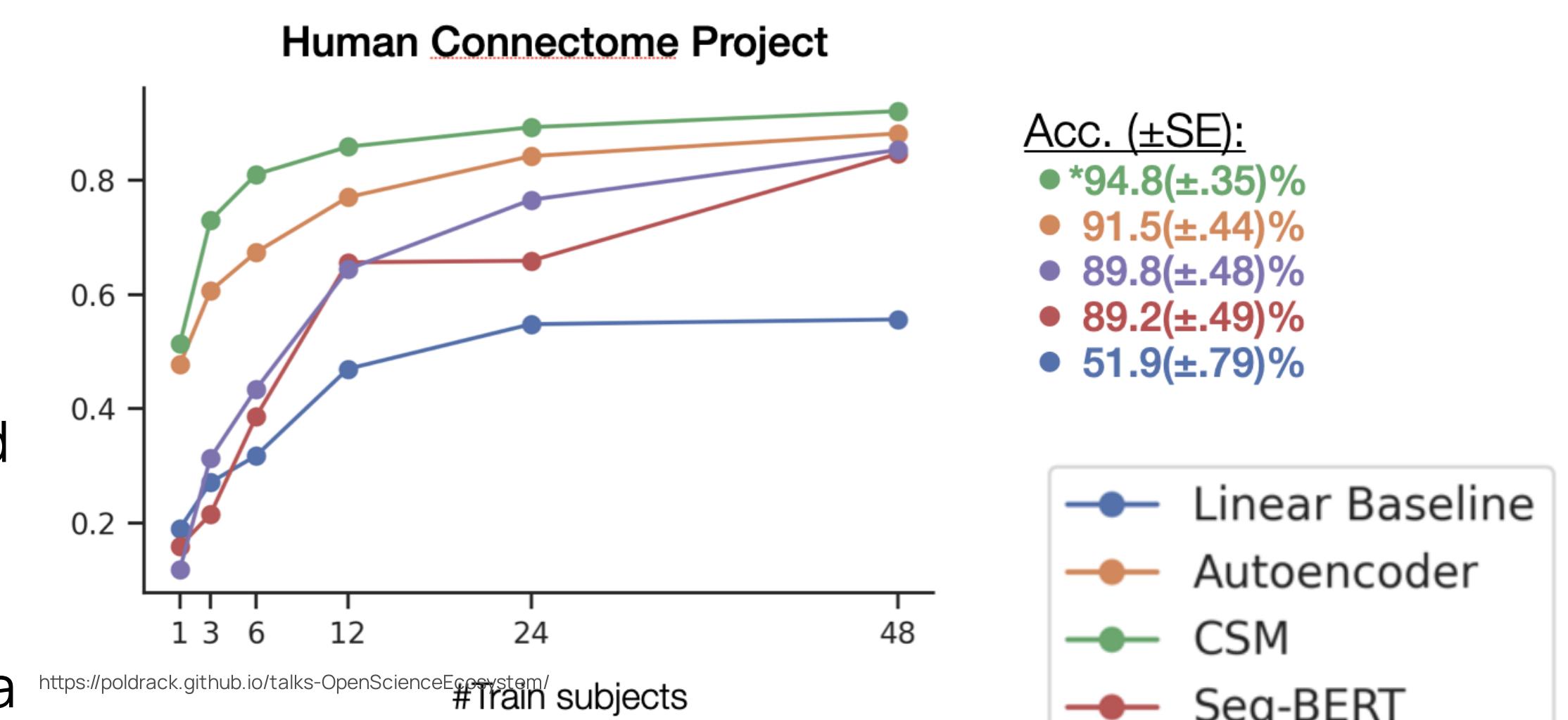
Sample scalp channel data (mid 2 seconds)



NEM

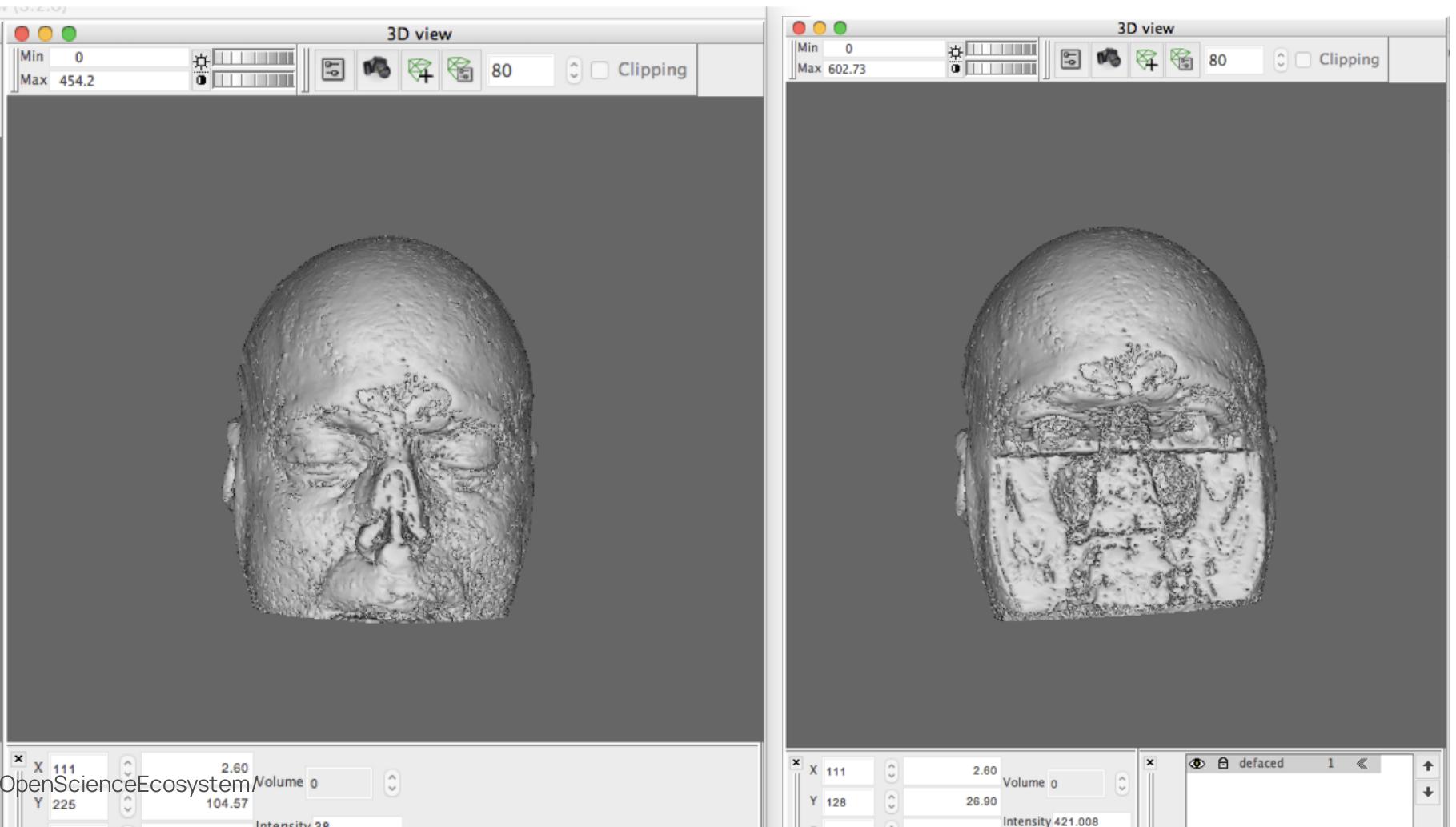
Example of OpenNeuro reuse

- A challenge for decoding brain activity from fMRI data is that most datasets are very small
- We used OpenNeuro to train a “foundation model”
Thomas, Re, & Poldrack, 2022, *NeurIPS*
 - A pre-trained model that can be used as a starting point for decoding models on smaller datasets
- We pre-train models on broad fMRI data



Challenges to open sharing

- All OpenNeuro MRI datasets must be *defaced*
 - To reduce risk of reidentification
- There is increasing risk that subjects might be reidentified even after defacing using advanced face recognition systems + face imputation tools (Schwartz et al., 2021)
- If the risk continues to rise, it may become necessary to move away from open sharing
 - This would be a huge loss for researchers



Keys to success in neuroimaging data sharing

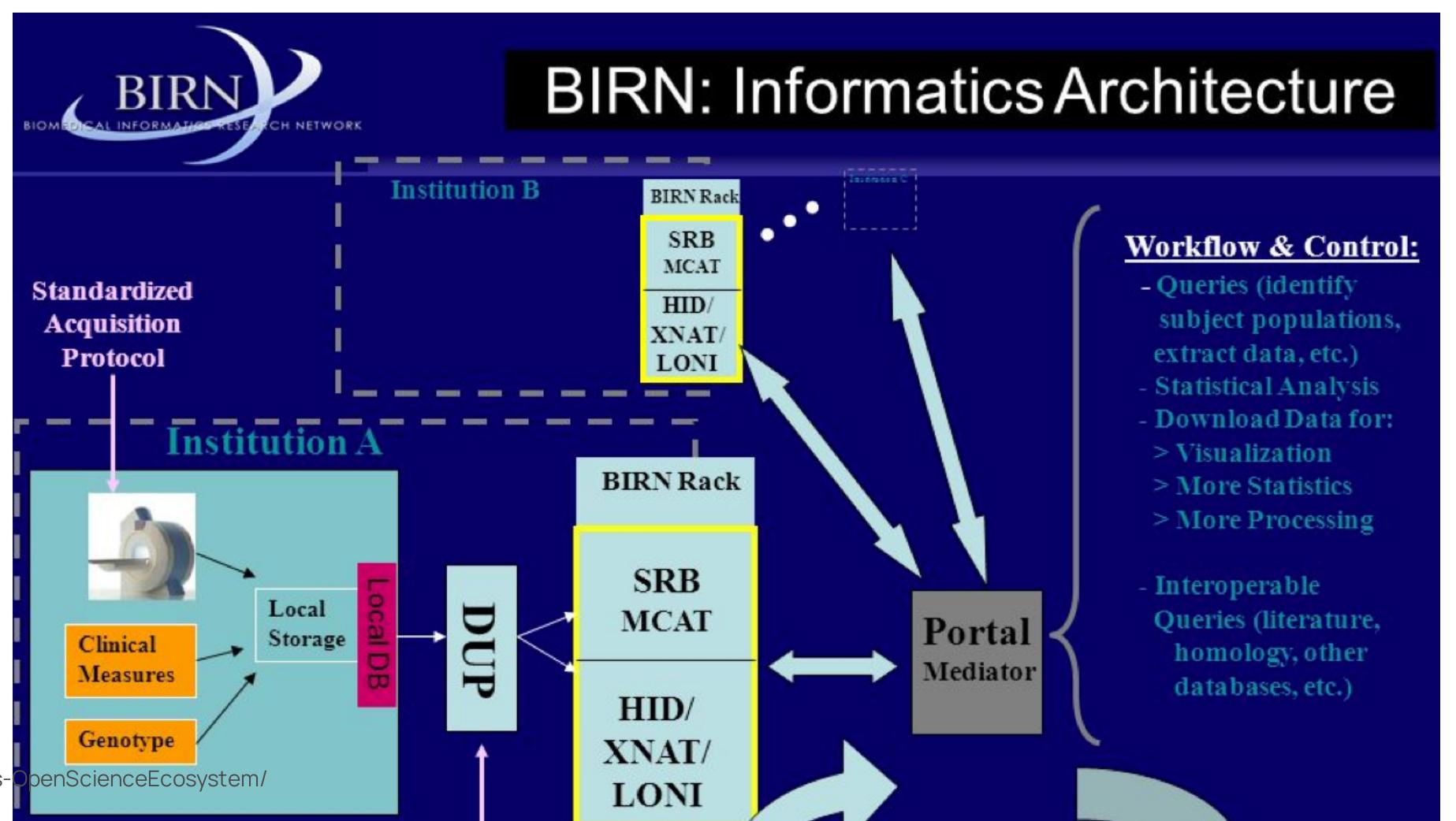
- *Data are digital end-to-end*
 - Minimizes manual steps in the process
- *Standardized file formats and data standards*
 - Makes data immediately usable by anyone
 - Reduces burden of curation and preparation
- *Demonstrated scientific utility*
- *Numerous success stories*

Lessons learned

- Community buy-in is essential
 - Mandates put in place before the community is ready can backfire
 - Unless they have overwhelmingly powerful advocates, as in genomics
 - Important that sharing advocates are members of community and eat their own dog food

Lessons learned

- Keep it simple and as close to standard practice as possible
 - Overengineered solutions have generally failed
 - If there are more than 2 acronyms...



Lessons learned

- Don't let the perfect be the enemy of the good
 - 20% of the effort will cover 80% of the datasets - focus on these!
 - There is a long tail of edge cases with loud advocates



Conclusions

- The field of neuroimaging has built an model ecosystem for open science and data sharing
- Infrastructure is critical to ease friction
- Community engagement has been key to adoption
- Need to keep the tools as close as possible to current practice

The Poldrack Lab



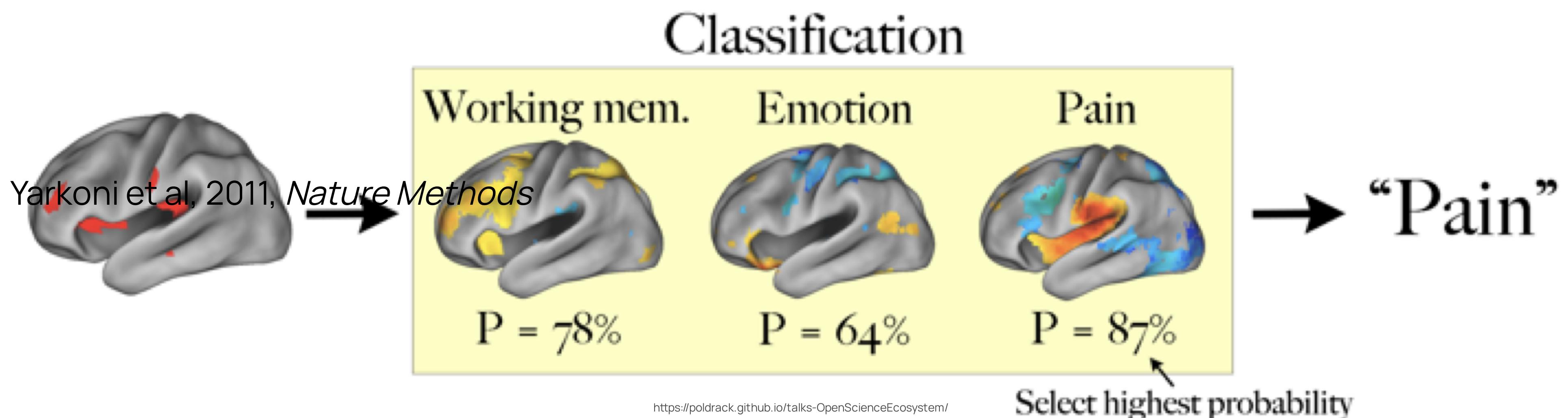
OpenNeuro Team



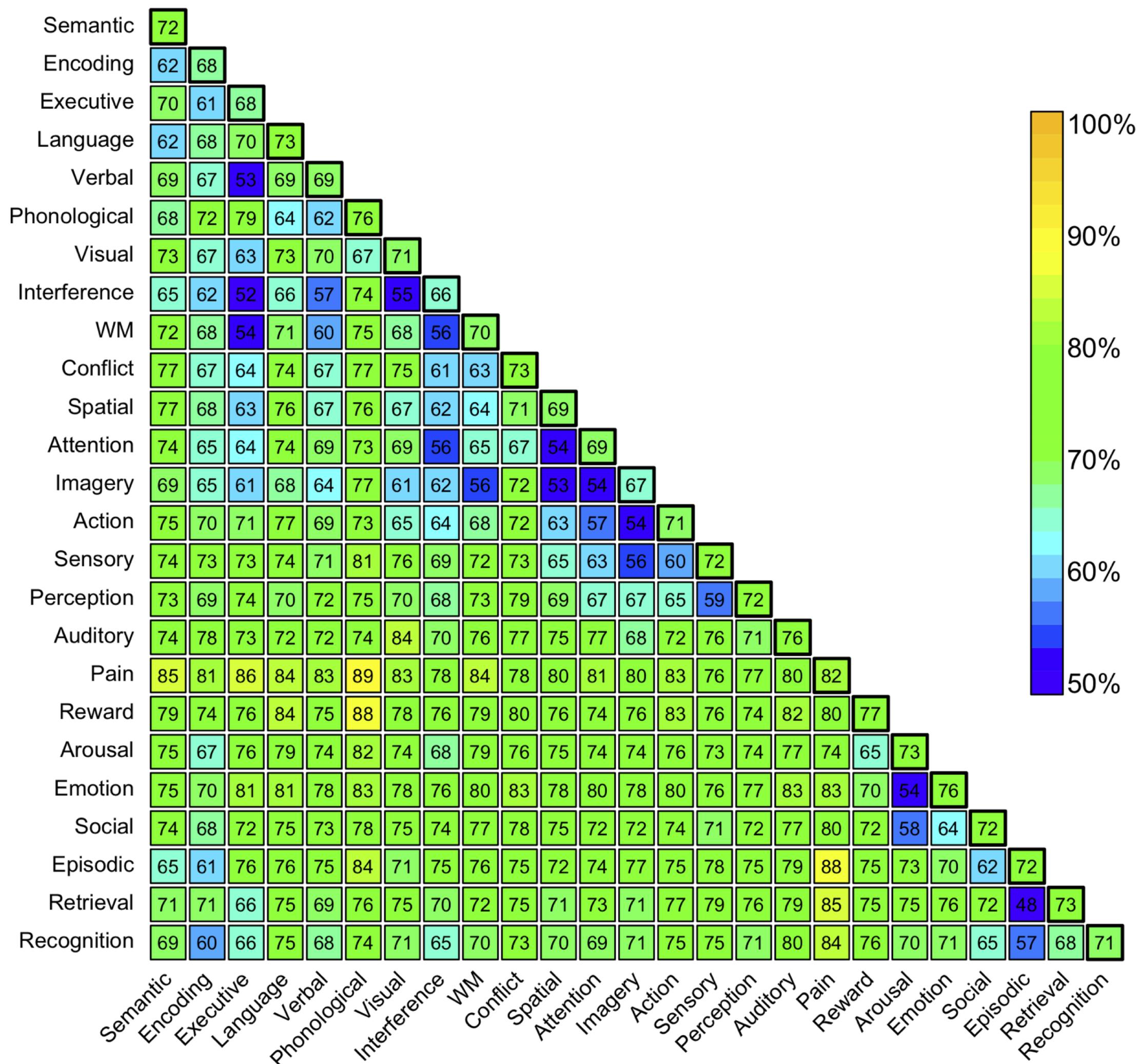
<https://poldrack.github.io/talks-OpenScienceEcosystem/>

Meta-analytic decoding using Neurosynth

- Given 2+ terms, can determine which is most likely given the data
- Naive Bayes classifier: assumes that all features (voxels) are independent; selects the most probable class
- Can apply this to any activation map—studies, individual subjects, etc.



- Cross-validated classification of all studies in database
- Select 25 high-frequency terms
- Pairwise classification: how well can we distinguish between the presence of each pair of terms?



Yarkoni et al, 2011, *Nature Methods*