Open Access: what, how and why?

Paola Masuzzo

@pcmasuzzo *paola.masuzzo@ugent.be*

Computational omics and systems biology group, VIB

Ghent University, Ghent, Belgium

Data Sharing in Neuroimaging

Daniele Marinazzo

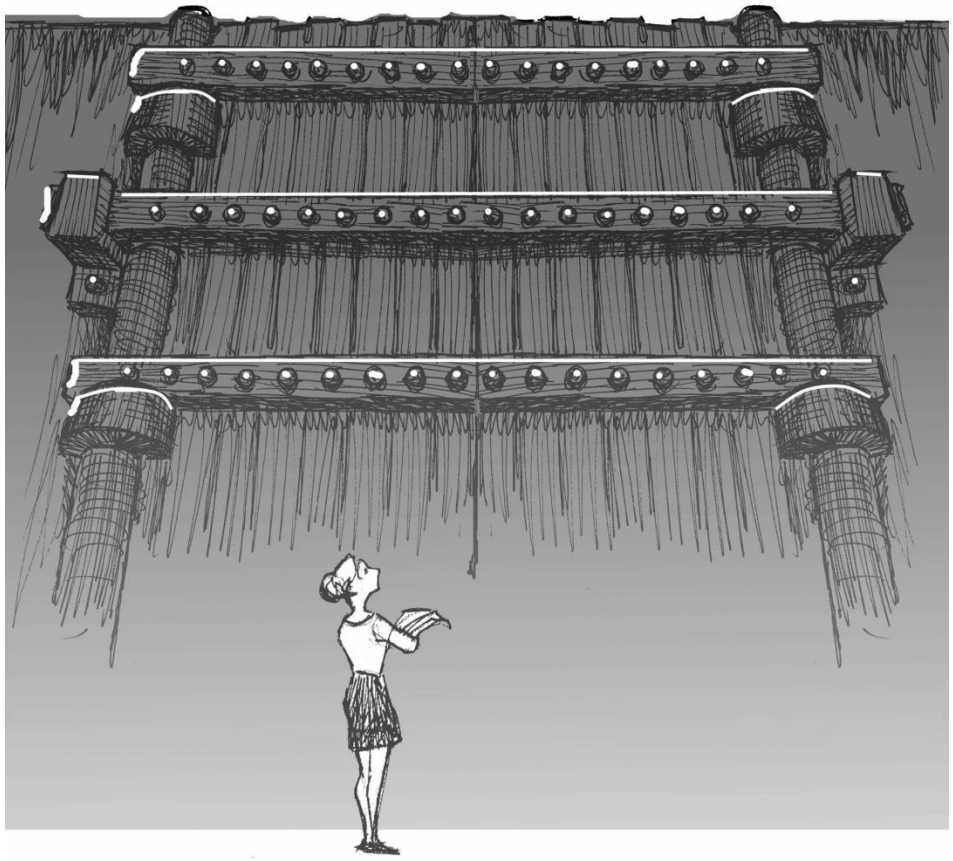
@dan_marinazzo *daniele.marinazzo@ugent.be*

Department of data analysis, faculty of psychology and pedagogical sciences

Ghent University, Ghent, Belgium

OpenAIRE Workshop, 31 May 2016

CC BY-SA 4.0



Cartoon by John R. McKiernan

http://whyopenresearch.org

CC BY-SA 4.0

Open Access explained!

https://www.youtube.com/watch?v=L5rVH1KGBCY

CC BY-SA 4.0

What?

How?

Why?

CC BY-SA 4.0

What?

How?

Why?

CC BY-SA 4.0

What is Open Access?

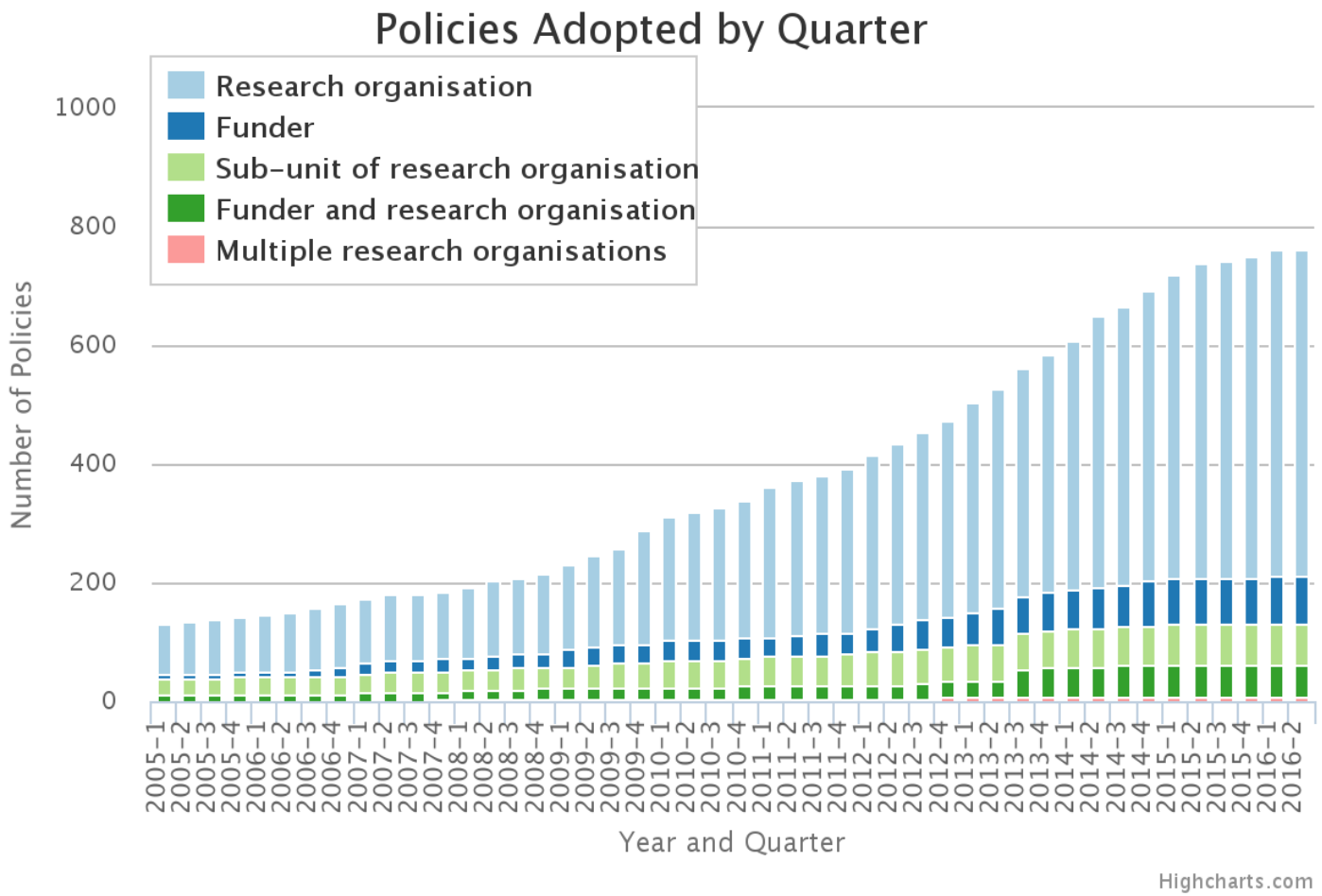
- Open Access (OA) refers to the removal of major obstacles to accessing, sharing and re-using the outputs of scholarly research.

- The research process is then facilitated by ensuring rapid and widespread access to research findings such that all communities have the opportunity to build upon them.

- Read full definition @Budapest OA Initiative: http://www.budapestopenaccessinitiative.org/read

CC BY-SA 4.0

There are currently over 700 OA policies and mandates recorded worldwide



http://roarmap.eprints.org

CC BY-SA 4.0

What?

How?

Why?

CC BY-SA 4.0

There are basically two routes to OA

- The **Gold** route: **freely accessible** research articles at the point of publication (sometimes accompanied by APCs).

- The **Green** route: author self-archiving; a version of the peer-reviewed article is posted online to a repository or website.

CC BY-SA 4.0

What?

How?

Why?

CC BY-SA 4.0

There are several reasons to go for OA!

CC BY-SA 4.0

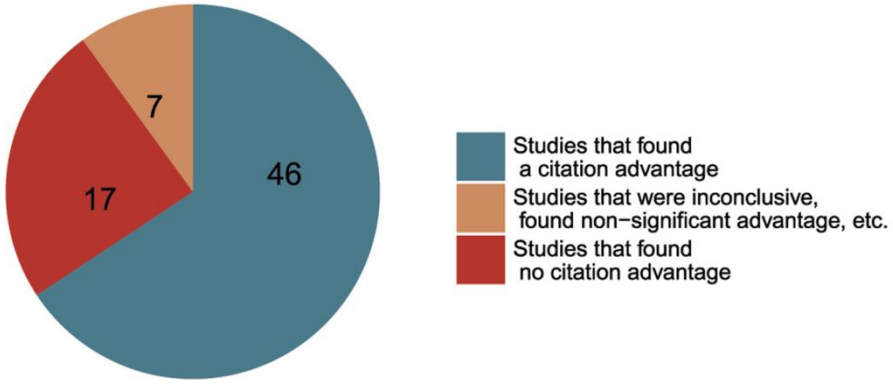
There are several reasons to go for OA!

CC-BY Danny Kingsley & Sarah Brown

CC BY-SA 4.0

The academic case for OA

- OA enables higher documented impact of scholarly articles through availability and re-use

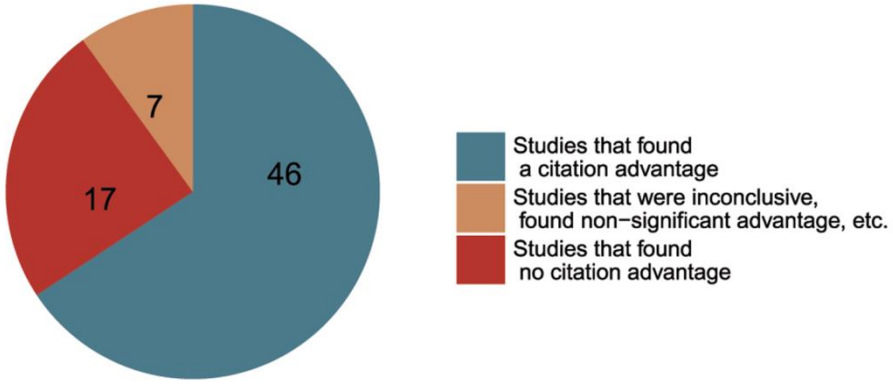


Tennant JP, Waldner F, Jacques DC *et al.* The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research 2016, 5:632 (doi: 10.12688/f1000research.8460.1)

CC BY-SA 4.0

The academic case for OA

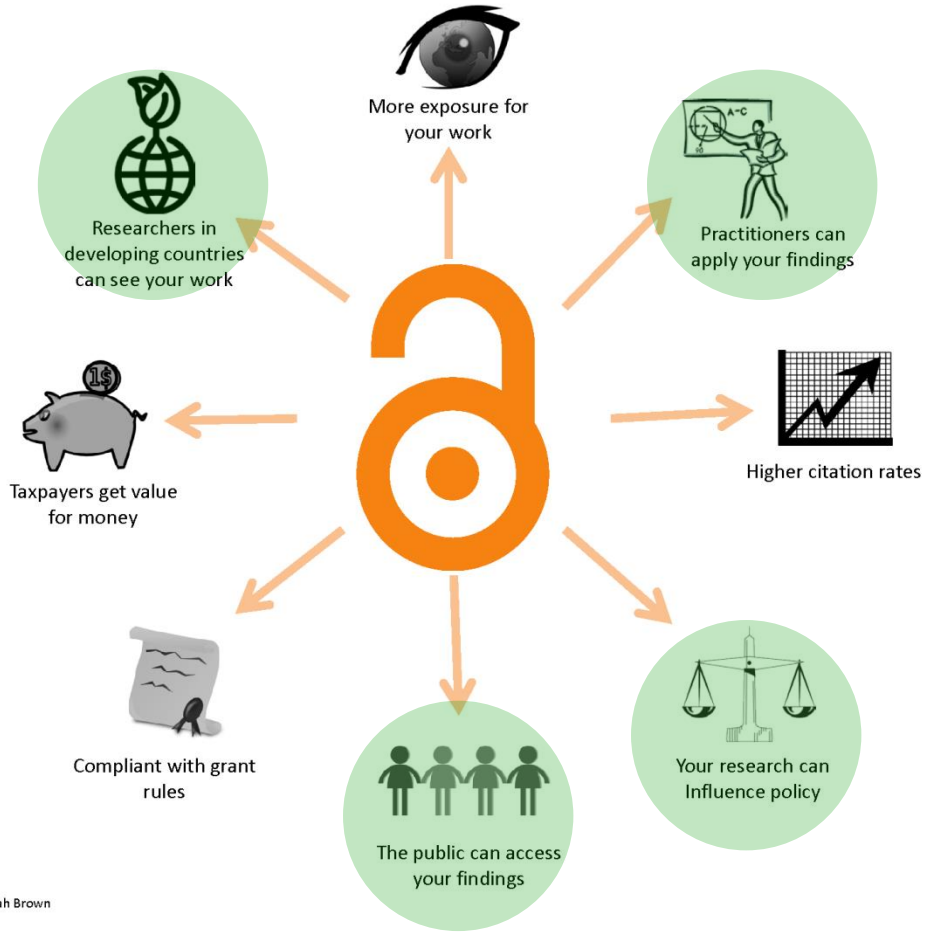
- OA enables higher documented impact of scholarly articles through availability and re-use



- OA non-restrictively allows researchers to use automated tools to mine the scholarly literature (TDM)

CC BY-SA 4.0

And even more reasons...!



CC-BY Danny Kingsley & Sarah Brown

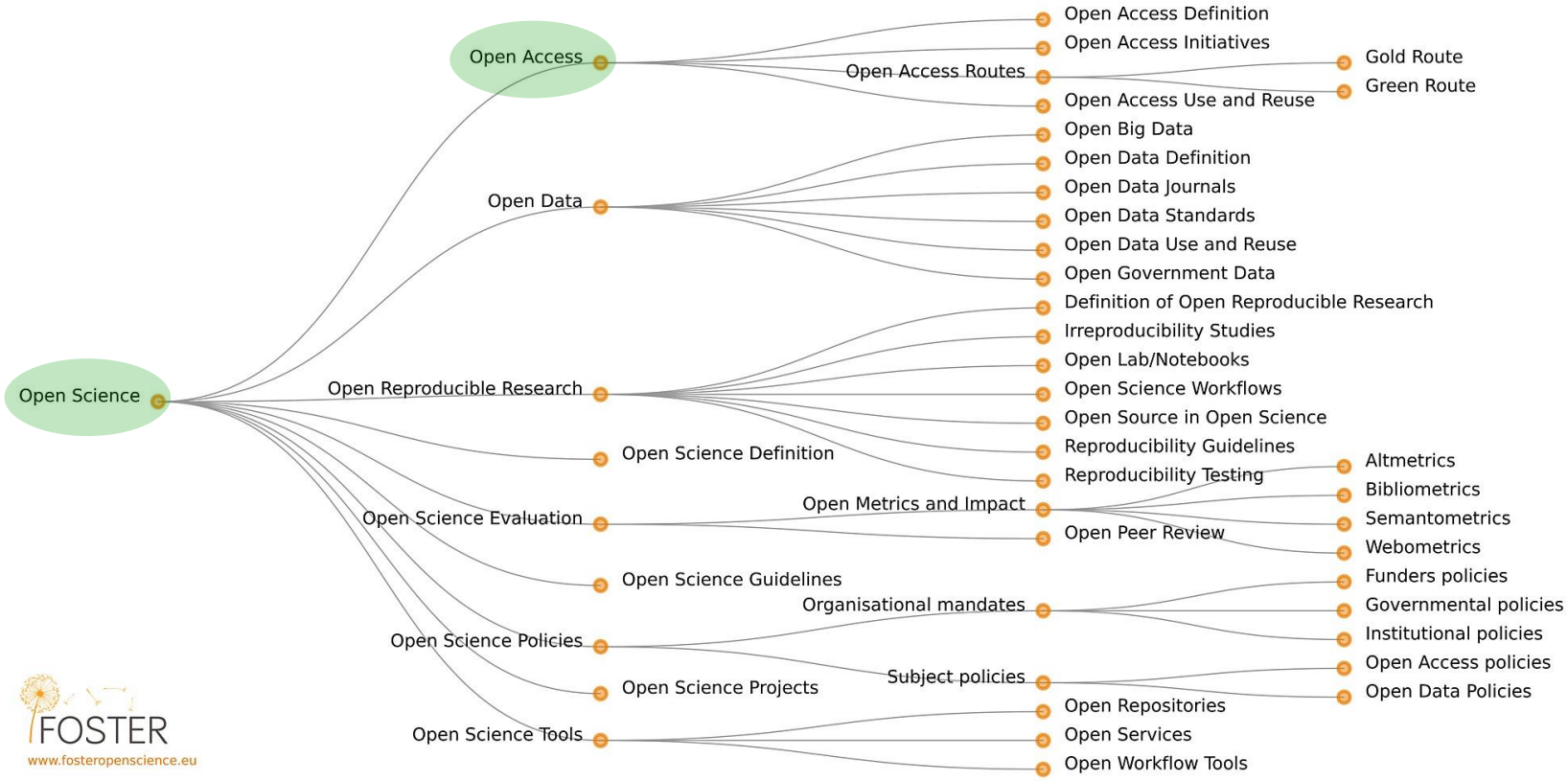
CC BY-SA 4.0

If you want to know more about OA (and you should!)

- The academic, economic and societal impacts of Open Access: an evidence-based review (http://f1000research.com/articles/5-632/v1)
- Budapest OA Initiative: http://www.budapestopenaccessinitiative.org
- FOSTER EU OA Resources: https://www.fosteropenscience.eu/foster-taxonomy/open-access
- SPARC OA: http://sparcopen.org/open-access
- OpenCon 2016: http://www.opencon2016.org

CC BY-SA 4.0

Open Access in the Open Science taxonomy

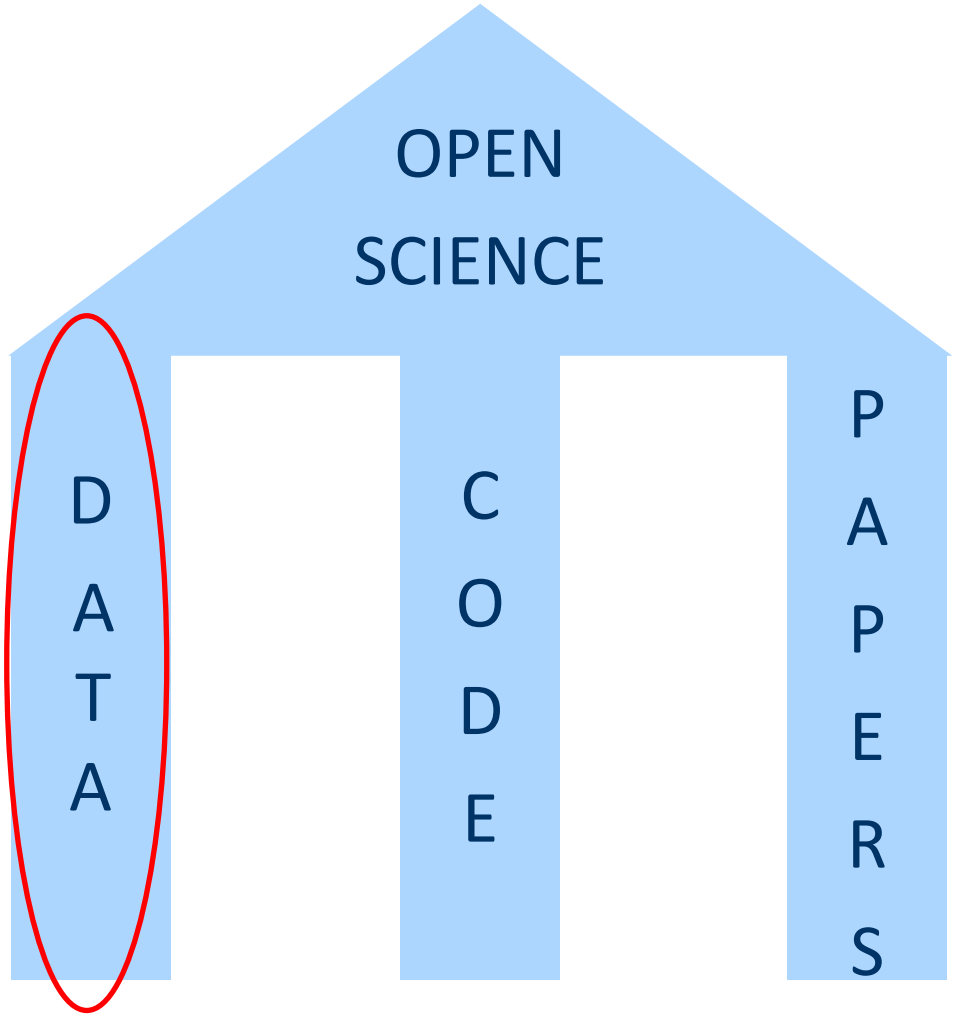


https://www.fosteropenscience.eu/taxonomy/term/5

CC BY-SA 4.0

CC BY-SA 4.0

The pillars of open science



Gorgolewski and Poldrack, BiorXiv 2016

CC BY-SA 4.0

Chris Gorgolewski

Stanford Center For
Reproducible Neuroscience

A practical guide for improving transparency and reproducibility in neuroimaging research

Krzysztof J. Gorgolewski, Russell Poldrack
doi: http://dx.doi.org/10.1101/039354

bioRχiv beta
THE PREPRINT SERVER FOR BIOLOGY

http://www.slideshare.net/chrisfilo1/share-and-reuse-how-data-sharing-can-take-your-research-to-the-next-level

CC BY-SA 4.0

An overview of the main public fMRI datasets

CC BY-SA 4.0

NKI-Rockland

Table Of Contents

Studies
Recruitment
Participant Schedule
MRI Protocol
Assessments
Sample Characteristics
- Age
- Sex
- DSM-IV Diagnoses

Data Sharing
NKI-RS Lite Releases
NKI-RS Team

Quick search

Go

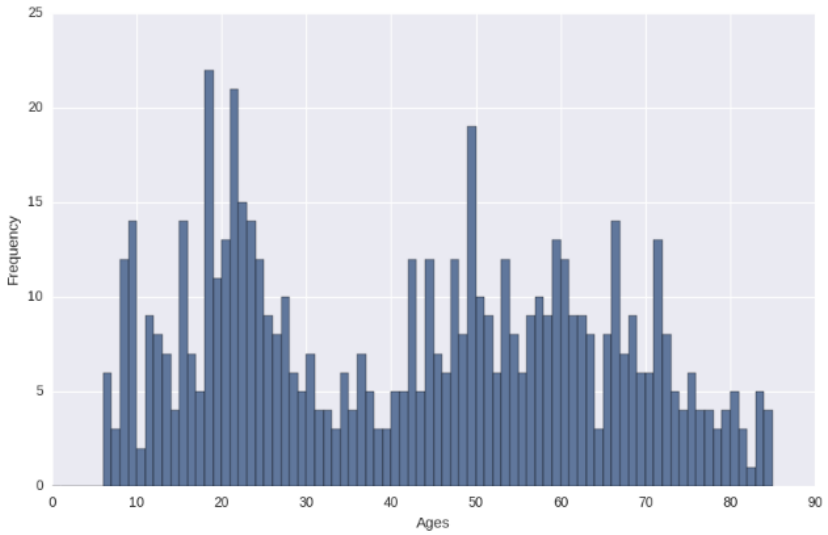
Enter search terms or a module, class or function name.

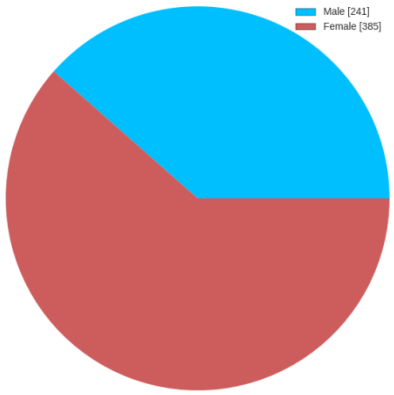
An open neuroscience project brought to you by:

Current Sample Characteristics (As of July 2015)

The NKI-RS is designed to be a representitive, heterogeneous community sample. Exclusion criteria are minimal, and subjects vary widely in many respects. The following figures provide basic demographic information for the NKI-RS as of the 689th subject. Of these 689 participants, 657 have imaging data available.

Age





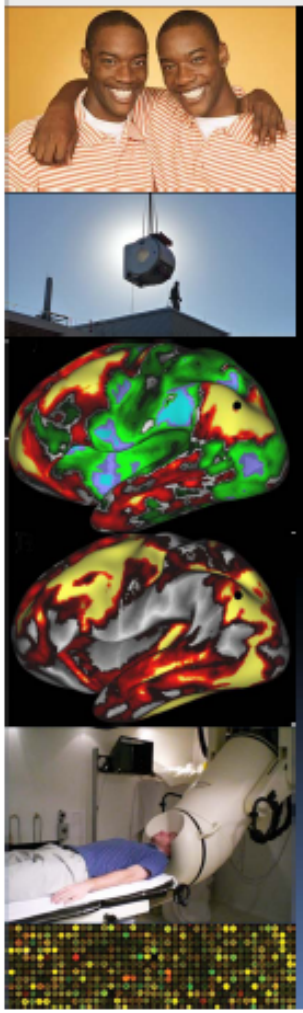
Diagnosis (Past and Current)	Total
No Diagnosis or Condition on Axis I	341
Alcohol Abuse	69
Cannabis Abuse	53
Major Depressive Disorder Single Episode In Full Remission	47
Alcohol Dependence	33
Cannabis Dependence	18
Panic Disorder Without Agoraphobia	18
Major Depressive Disorder Recurrent In Full Remission	17
Attention-Deficit/Hyperactivity Disorder NOS	16
Specific Phobia	16
Depressive Disorder NOS	15
Generalized Anxiety Disorder	15
Cocaine Abuse	13
Attention-Deficit/Hyperactivity Disorder Predominantly Inattentive Type	13
Attention-Deficit/Hyperactivity Disorder Combined Type	13
Social Phobia	13
Posttraumatic Stress Disorder	11
Cocaine Dependence	9
Enuresis (Not Due to a General Medical Condition)	9
Obsessive-Compulsive Disorder	8
Dysthymic Disorder	6
Amphetamine Abuse	6
Sedative Hypnotic or Anxiolytic Abuse	5
Major Depressive Disorder Recurrent In Partial Remission	5
Eating Disorder NOS	5
Oppositional Defiant Disorder	5
Separation Anxiety Disorder	4
Bereavement	4
Major Depressive Disorder Single Episode In Partial Remission	4
Panic Disorder With Agoraphobia	3
Major Depressive Disorder Recurrent Moderate	3
Psychotic Disorder NOS	2
Chronic Motor or Vocal Tic Disorder	2
Opioid Dependence	2
Attention-Deficit/Hyperactivity Disorder Predominantly Hyperactive-Impulsive Type	2
Hallucinogen Abuse	2
Major Depressive Disorder Recurrent Unspecified	2
Bipolar II Disorder	1
Schizophrenia Paranoid Type	1
Major Depressive Disorder Single Episode Unspecified	1
Major Depressive Disorder Recurrent Mild	1
Major Depressive Disorder Single Episode Moderate	1
Mood Disorder Due to General Medical Condition	1
Mood Disorder Due to Substance	1
Major Depressive Disorder Recurrent Severe Without Psychotic Features	1
Sedative Hypnotic or Anxiolytic Dependence	1
Hallucinogen Dependence	1
Opioid Abuse	1
Tic Disorder NOS	1
Anxiety Disorder Due to General Medical Condition	1
Inhalant Abuse	1
Transient Tic Disorder	1
Bulimia Nervosa	1
Encopresis Without Constipation and Overflow Incontinence	1
Body Dysmorphic Disorder	1
Bipolar Disorder NOS	1
Anxiety Disorder NOS	1

CC BY-SA 4.0

Human Connectome Project

HUMAN Connectome PROJECT

Mapping structural and functional connections in the human brain



Project Goals I

- Study a large population:
  - 1,200 healthy adults
  - 300 twin pairs and their non-twin siblings
- Cutting-edge neuroimaging methods
  - 3T Skyra MRI, customized gradient (UMinn -> Wash U)
  - 7T MRI (UMinn, 200 subjects); perhaps also 10.5T
  - dMRI/tractography; R-fMRI; Task-fMRI
  - MEG/EEG (100 subjects)
- Extensive behavioral testing
- Blood samples for genotyping

CC BY-SA 4.0

SchizConnect.org

Query terms	Number of images & distinct subjects returned				
	Total	fBIRN @ HID	COBRE @ COINS	MCICShare @ COINS	NUSDAST @ XNAT
T1 SCZ + CON	3178	1048	484	264	1382
	737	185	173	95	284
3T T1 20–60 Yr	1348	750	492	106	–
	345	112	180	53	
T1 healthy controls	1846	655	243	264	684
	439	113	94	95	137
Sternberg SCZ 3 T <= 30 Yr	210	210	–	–	–
	18	18			
Resting BP + CON	103	–	103	–	–
	101		101		
Sensory motor + T1, SCZ + CON, 30–50 Yr (multimodal)	1310	1140	–	170	–
	119	83		36	
All MRI	21,309	13,552	1596	4347	1814
	1029	251	198	212	368

CC BY-SA 4.0

ElectroEncephalography (EEG) data

IEEG.org

Network | Data

Welcome!

http://ieeg.org

Sharing, collaborating, and analyzing data on the cloud

IEEG.ORG is a collaborative initiative funded by the National Institutes of Neurological Disorders and Stroke. This initiative seeks to advance research towards the understanding of epilepsy by providing a platform for sharing data, tools and expertise between researchers. The portal includes a large database of scientific data and tools to analyze these datasets.

Sign up | Download Toolbox | Read Documentation

You can also: Penn, Mayo Clinic, NIH National Institute of Neurological Disorders and Stroke

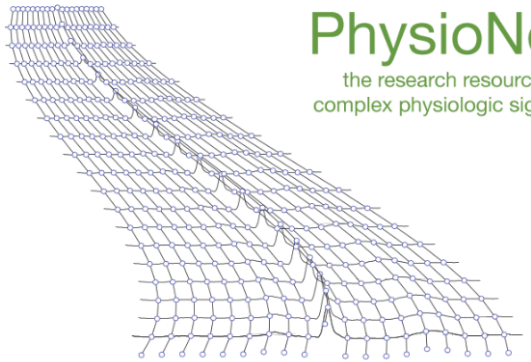
Overview | Downloads | Publications

Downloads: To request access to the TUH EEG Corpus, please fill out **this form.** You will receive an automatically-generated username and password via email.

- TUH EEG Epilepsy Corpus (v0.0.0): This is a subset of the TUH EEG Corpus that contains 100 subjects with and without epilepsy. We plan to expand this to 1,000 patients of each type. To request access to this data, please fill out the attached form.
- TUH EEG Corpus (v1.0.0): This will be the final official release of all data collected from 2002-2014. Contact us for further details.
- TUH EEG Corpus (v0.6.0): A beta release used to collect feedback from the community.
- TUH EEG Corpus (v0.2): This is our first public release of the TUH EEG Corpus. This is a beta release intended to allow users to give us feedback on the data. There are 247 sessions, 615 EDF files, and over 150 hours of EEG data. The uncompressed data occupies about 8.3G of disk space.

The TUH EEG Corpus is freely available. The only reason we require registration is that we need to track who downloads the data. We also want to be able to inform you of any updates to the releases.

Once you have obtained the username and password, you can selectively download portions of the corpus using your browser. Due to the size of the data, the best way to transmit this data is via a hard disk. Email us for details.



PhysioNet

the research resource for complex physiologic signals

PhysioNet offers free web access to large collections of recorded physiologic signals (PhysioBank) and related open-source software (PhysioToolkit). PhysioBank databases are made available under the ODC Public Domain Dedication and License v1.0.

PhysioNetWorks workspaces are available to members of the PhysioNet community for works in progress that will be made publicly available in PhysioBank and PhysioToolkit

Useful Links [site map]

Recent news

PhysioNet/CinC Challenge 2016

Looking for data?

Physionet Google Search: Use the keyword search at the top right of every page.

Physiobank Database Index: All of our databases sorted by category (ECG, EEG, etc).

Physiobank Record Search: Record search with key input information. Advanced user guide.

MIMIC-III: A massive healthcare dataset collected from over 40000 critical care patients.

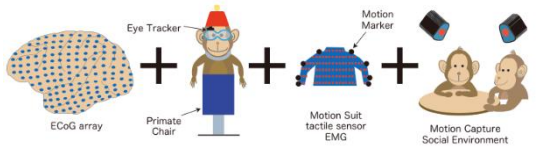
Looking for software?

PhysioToolkit Software Index: All the useful software contributed to our website.

WFDB Software Package: Our own large collection of software for signal reading, writing, processing, and automated analysis.

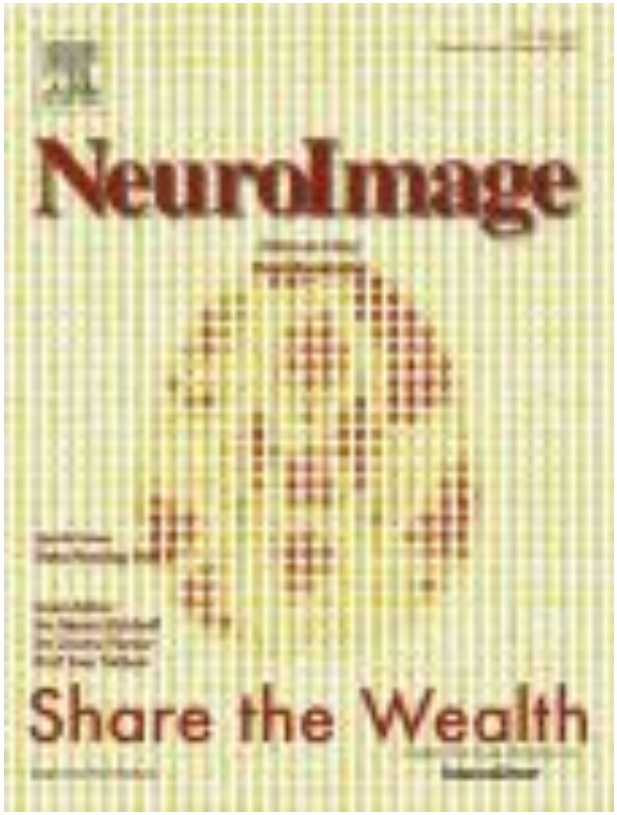
WFDB Matlab Toolbox: A Matlab implementation of the WFDB

NEUROTYCHO *beta*



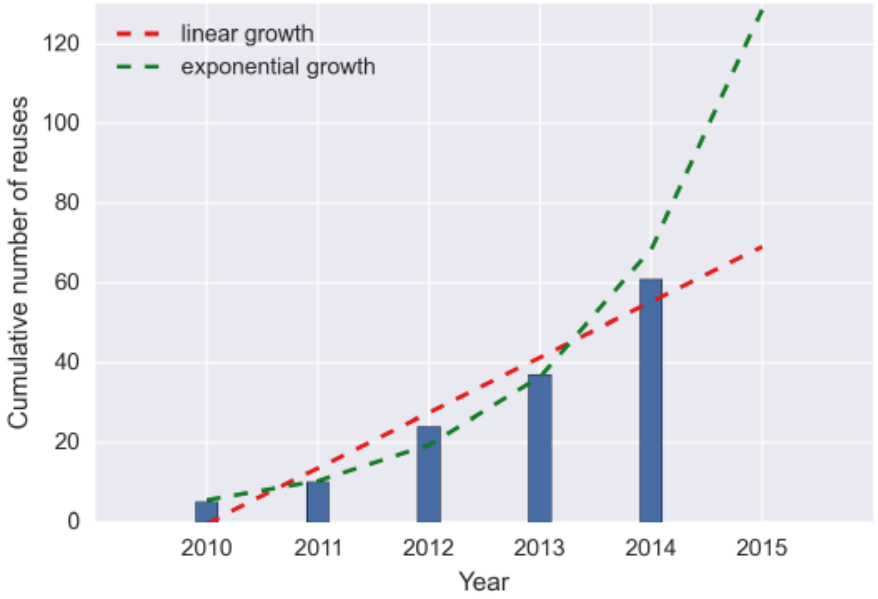
CC BY-SA 4.0

NeuroImage Special Issue, January 2016



CC BY-SA 4.0

Data Sharing Saves Money



Results cont.

63 number of papers and courses using OpenfMRI data

60% of papers used more than one dataset

39 number of reuses of the most popular dataset [7]

$878,400

how much it would cost to perform those studies without shared data

Gorgolewski et al. OHBM 2015

CC BY-SA 4.0

Data Sharing Concerns

- Fear of being scooped
- Fear that someone finds a mistake in your analysis
- Misconceptions about data ownership

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIAL

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

N Engl J Med 2016; 374:276-277 | January 21, 2016 | DOI: 10.1056/NEJMe1516564

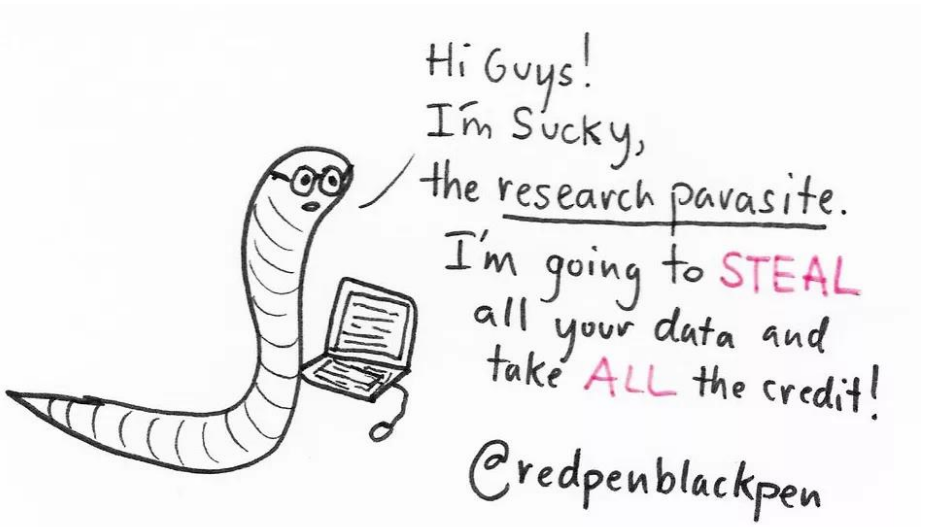
2 comments on PubPeer

Share:

Article | References | Citing Articles (8) | Metrics

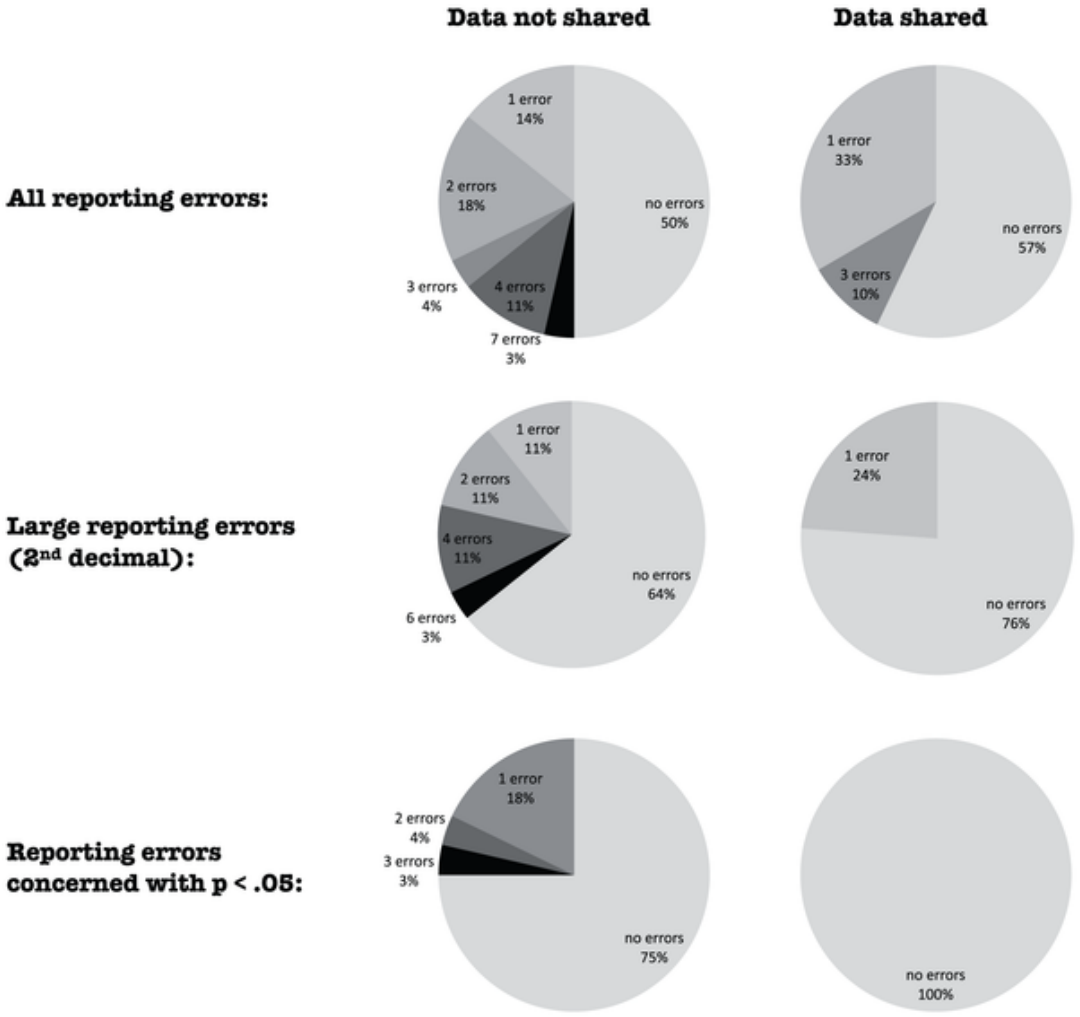
The aerial view of the concept of data sharing is beautiful. What could be better than having high-quality information carefully reexamined for the possibility that new nuggets of useful data are lying there, previously unseen? The potential for leveraging existing results for even more benefit pays appropriate increased tribute to the patients who put themselves at risk to generate the data. The moral imperative to honor their collective sacrifice is the trump card that takes this trick.

However, many of us who have actually conducted clinical research, managed clinical studies and data collection and analysis, and curated data sets have concerns about the details. The first concern is that someone not involved in the generation and collection of the data may not understand the choices made in defining the parameters. Special problems arise if data are to be combined from independent studies and considered comparable. How heterogeneous were the study populations? Were the eligibility criteria the same? Can it be assumed that the differences in study populations, data collection and analysis, and treatments, both protocol-specified and unspecified, can be ignored?



CC BY-SA 4.0

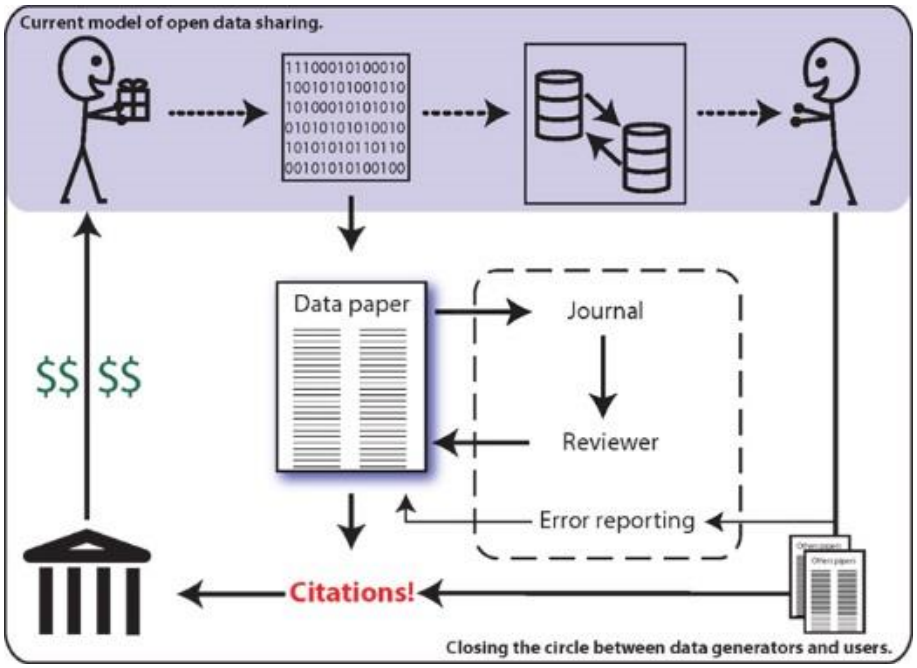
Data Sharing and Quality of Results



Wicherts JM, Bakker M, Molenaar D (2011) Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. PLoS ONE 6(11): e26828. doi:10.1371/journal.pone.0026828
http://journals.plos.org/plosone/article?id=info:doi/10.1371/journal.pone.0026828

CC BY-SA 4.0

Even more credits for data sharing: data papers



- NeuroInformatics
- GigaScience
- Scientific Data
- F1000 Research
- Data in Brief
- Journal of Open Psychology Data

Gorgolewski, Milham and Margulies 2013

CC BY-SA 4.0

Repositories and formats

- OpenfMRI.org
- FCP/INDI
- COINS
- OSF.io
- figshare
- DataVerse
- DataDryad

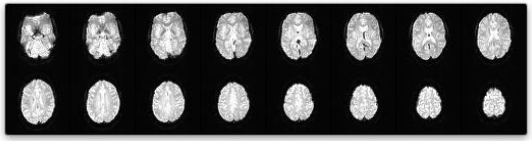
Quality assessment

mriqc: image quality metrics for quality assessment of MRI

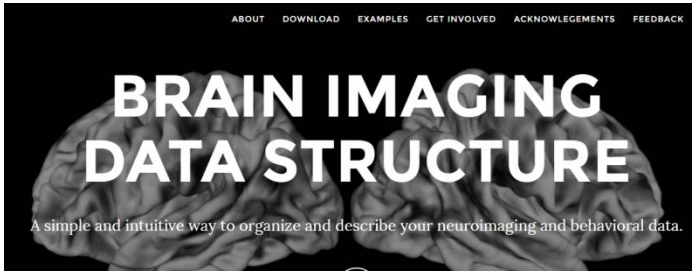
This pipeline is developed by the Poldrack Lab at Stanford University for use at the Center for Reproducible Neuroscience (CRN), as well as for open-source software distribution.

Preprocessed Connectomes Project

Overview Datasets Quality Assessment Publications Forum View on Github



Data Format, Guidelines, Metadata



http://bids.neuroimaging.io/

COBIDAS

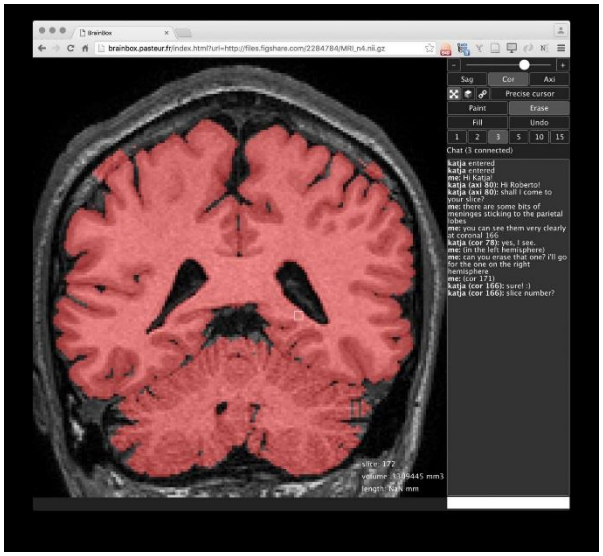
The report from the OHBM Committee on Best Practices in Data Analysis and Sharing (COBIDAS), Best Practices in Data Analysis and Sharing in Neuroimaging using MRI is ready. We received over 100 comments from the community, in the form of emails, blog posts and annotated Word/PDF files. The committee has carefully reviewed every comment and incorporated changes into the manuscript (see below). As a result, this unique best practice document really reflects views from across our field.

Organization for Human Brain Mapping

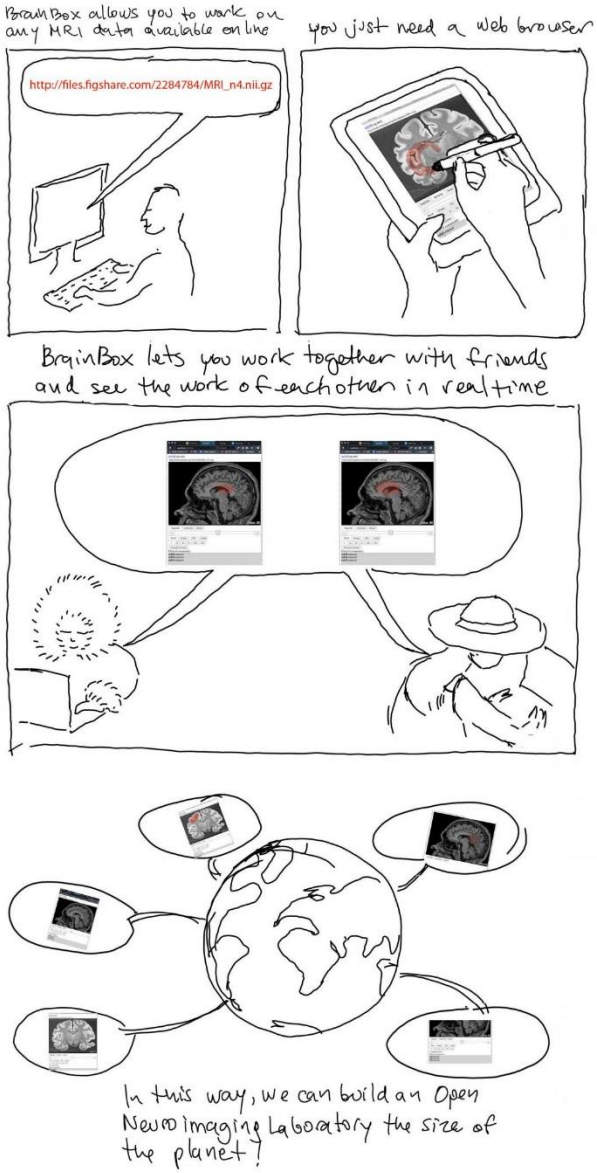
CC BY-SA 4.0

BrainBox, OpenScience Prize Finalist

The Open Science Prize



Collaboratively visualize and segment any brain image available online



CC BY-SA 4.0

Conclusions: why share data

- It's an ethical thing to do
- The journal (PLOS) or the funder (NIH and soon EU) requires it
- Increases exposure and impact of your research (and so your chances of getting the next grant)
- It's related to higher citation rate (Piwowar et al. 2007, 2013)
- Shows that you care

CC BY-SA 4.0

Thank you!

CC BY-SA 4.0