

Open Science and good practices

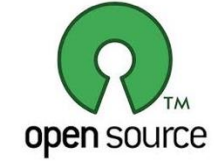
Robert Oostenveld

Donders Institute, Radboud University, Nijmegen, NL
Karolinska Institutet, Stockholm, SE

robert.oostenveld@donders.ru.nl

My personal contributions to Open Science

FieldTrip toolbox for MEG/EEG analysis
The EEGsynth project for artistic neuro- and biofeedback



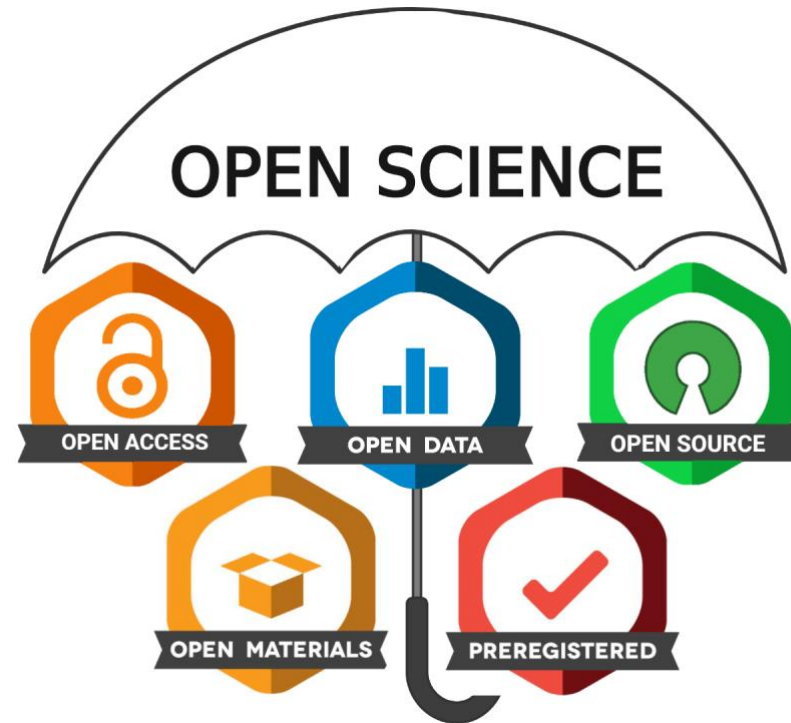
Human Connectome Project (HCP)
The Donders and Radboud Data Repository
Brain Imaging Data Structure (BIDS)



“DIY” projects with with open hardware



OHMB Open Science Room
BrainHack meetings
Hackathons at the Donders



Open Science

Open educational resources

Open access publications

Open peer review

Pre-registration

Registered reports

Open methodology

Open source

Open hardware

Open materials

Open data

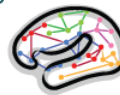
...



coursera



PeerJ



HUMAN
Connectome
PROJECT



open source



open hardware



OpenNEURO



Better

Why do ~~Open~~ Science?

Democratic

everyone should be able to
contribute

Pragmatic

it is more efficient to collaborate

Infrastructure

it results in better tools

Public

results should be accessible to all

But also ...

to tackle the lack of trust
to address poor reproducibility
to improve quality
because it is the right thing to do

Research cycle



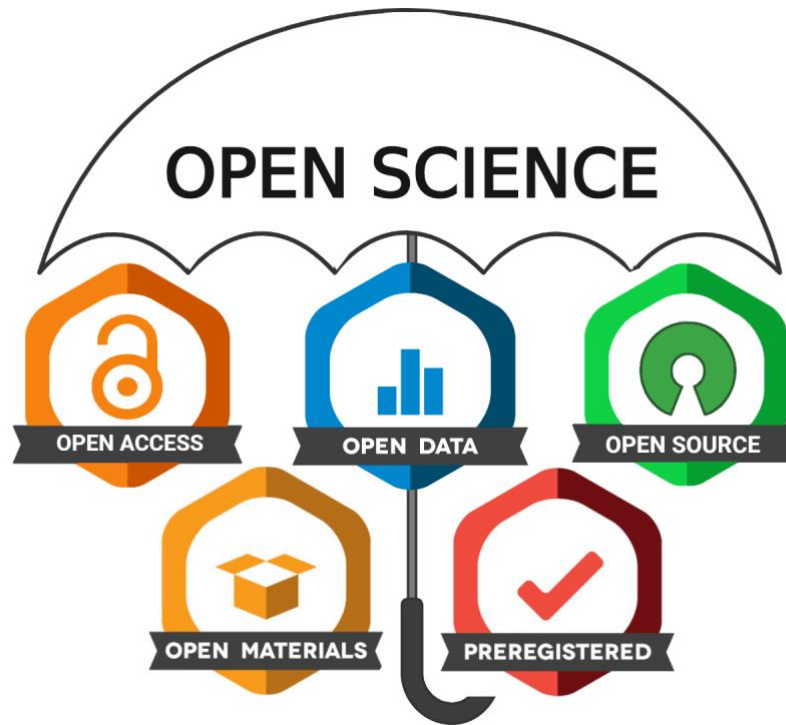
You can make this cycle more open by ...



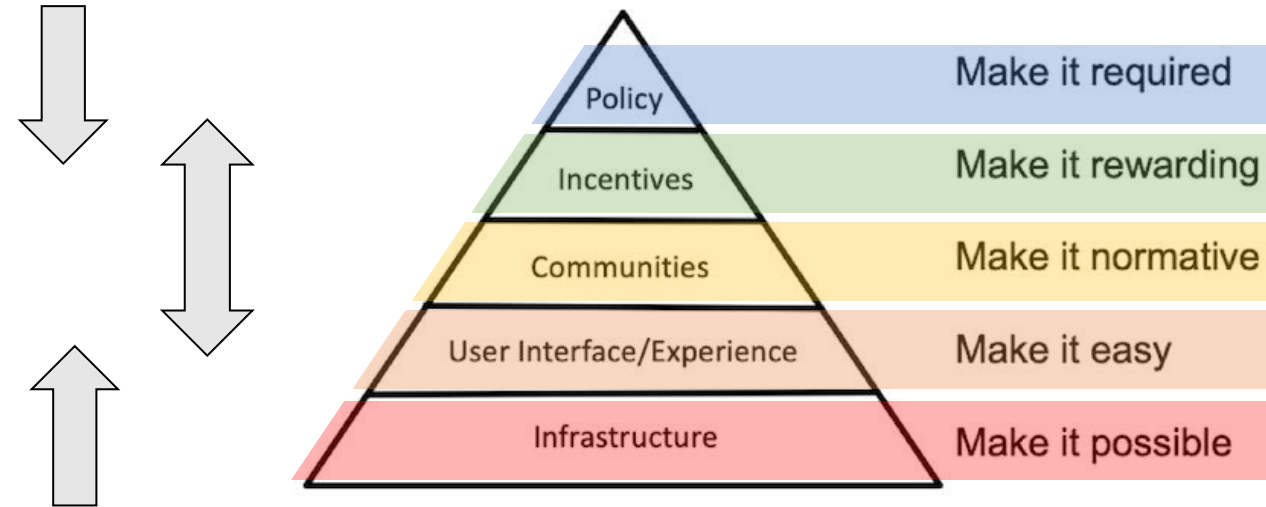
- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO



We can all benefit of it ...
and we can all contribute to it



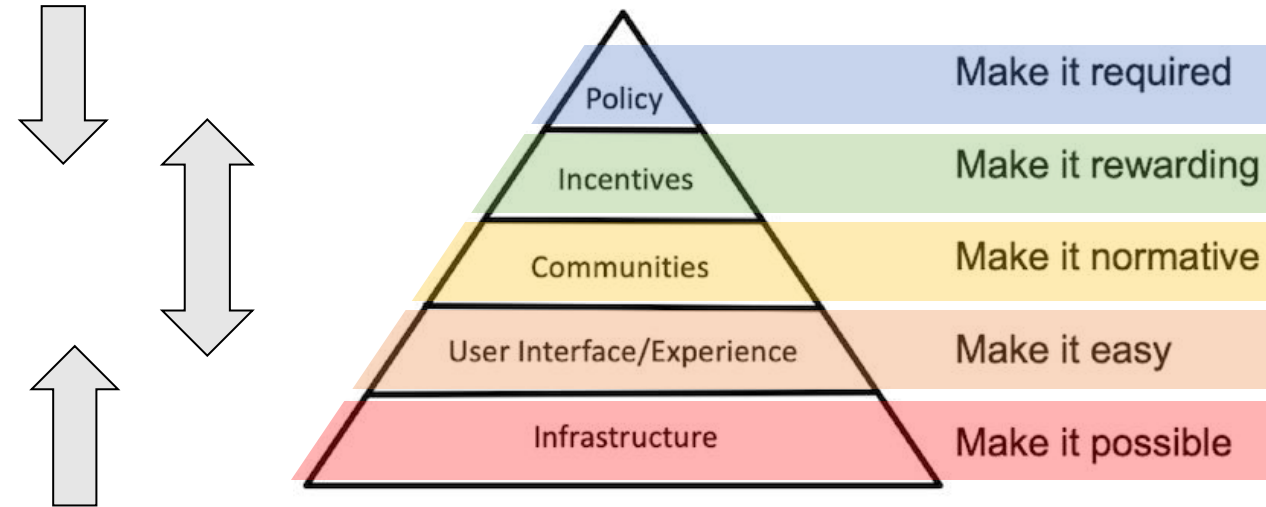
Contributing to a Culture Change



Not everyone can do everything.

Change requires personal *and* organisational investments.

Contributing to a Culture Change



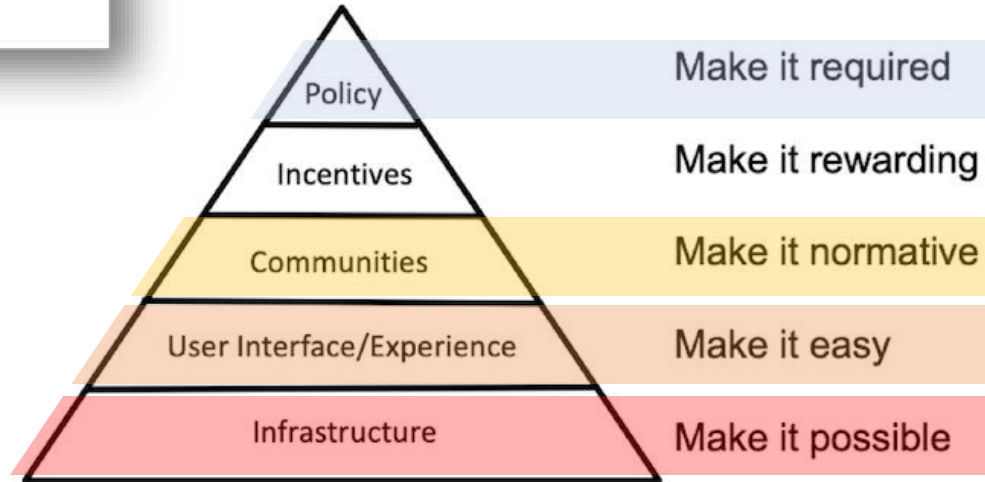
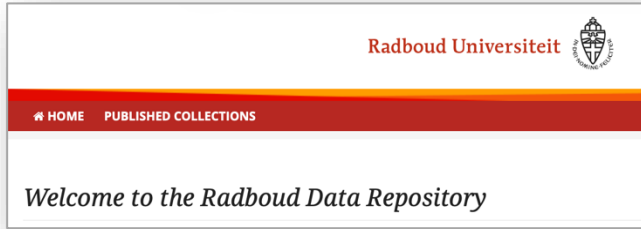
Each of us is - from our position - only capable of limited influence:

policy makers from the top downwards

research engineers and support staff from bottom-up

many of us in between

Personal experiences with Open Science

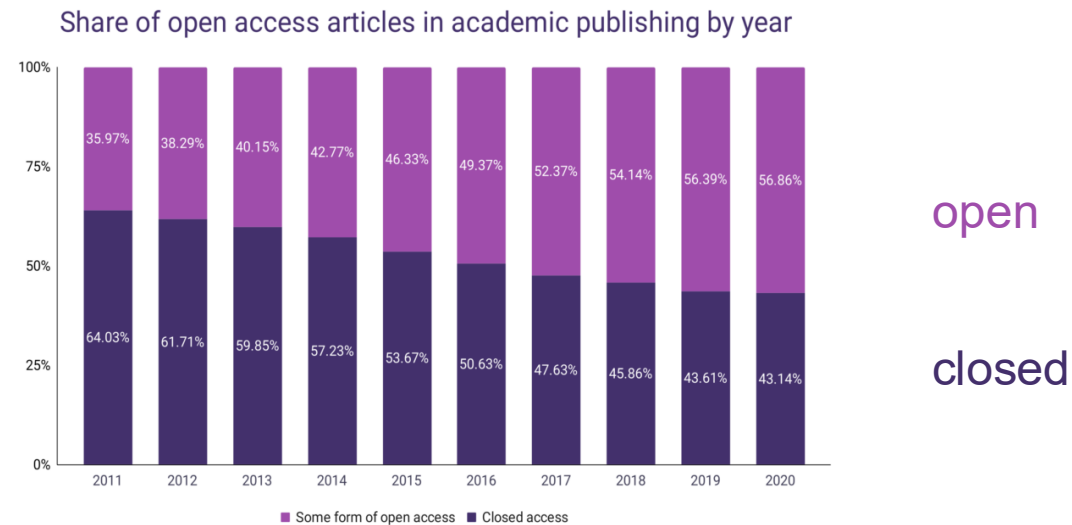


Where is Open Science successful?

Open Access publications

More than 18.000 open access publishers, including Elsevier, Springer, etc.
20% - 54% of all papers are now open access (Simard et al. 2022).

However, article processing charges make it hard for some researchers to publish in high-impact (=expensive) journals.



Where is Open Science successful?

Open Access publications

More than 18.000 open access publishers, including Elsevier, Springer, etc.

20% - 54% of all papers are now open access (Simard et al. 2022).

However, article processing charges (APCs) make it hard for some researchers to publish in high-impact (=expensive) journals.

Open Source analysis software

Widely used in many research fields (R, Python, Julia, JASP, FieldTrip).

However, challenging relation between open source and sustainability.

Open Data

Standard for many national and EU-level funding schemes.

However, privacy considerations and inconsistent interpretation of GDPR.

Tension between Open Source/Open Data and other interests

Academic software development, support and maintenance requires long-term efforts, but Research Software Engineers (RSEs) are rare.

FieldTrip is an Open Source toolbox, but MATLAB is commercial, so FieldTrip is Free-as-in-speech (Libre ) but not Free-as-in-beer (Gratis )

Note: the English language does not distinguish “free ” from “free .

Companies like Google (Colab) and Microsoft (Github) are actively trying to bind Open Source software projects into their ecosystems on the premise of their services being free ()...

Companies like FaceBook and Google are funding Open Source development conditional on it being licensed under MIT/BSD() , not GPL ()...

Data platforms like OpenNeuro and DataDryad use CC0 (“anything goes”) rather than CC-SA (“share alike”)...

Tension between stakeholders

Not as simple as
free-as-in-speech (libre ) versus free-as-in-beer (gratis ) .

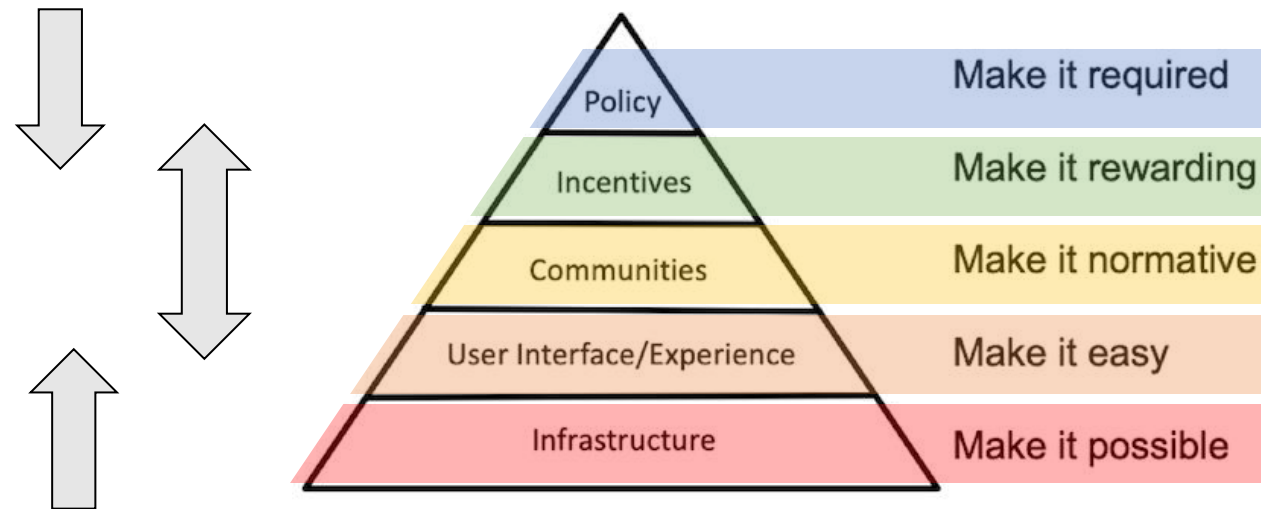
Academic researcher needs career incentives and reward.

Interest of the **society and community** in developing actual products based on academic research, and bringing those to market.

Many Open-Science-minded **companies** would not mind becoming the next Elsevier or Novo Nordisk, and companies want to leverage Open Science to their advantage.

Very different responsibilities and (sometimes personal) stakes of academic researchers, research engineers, librarians, data stewards, privacy officers, institutional management, ...

Changing towards an Open Science culture *inside academia*



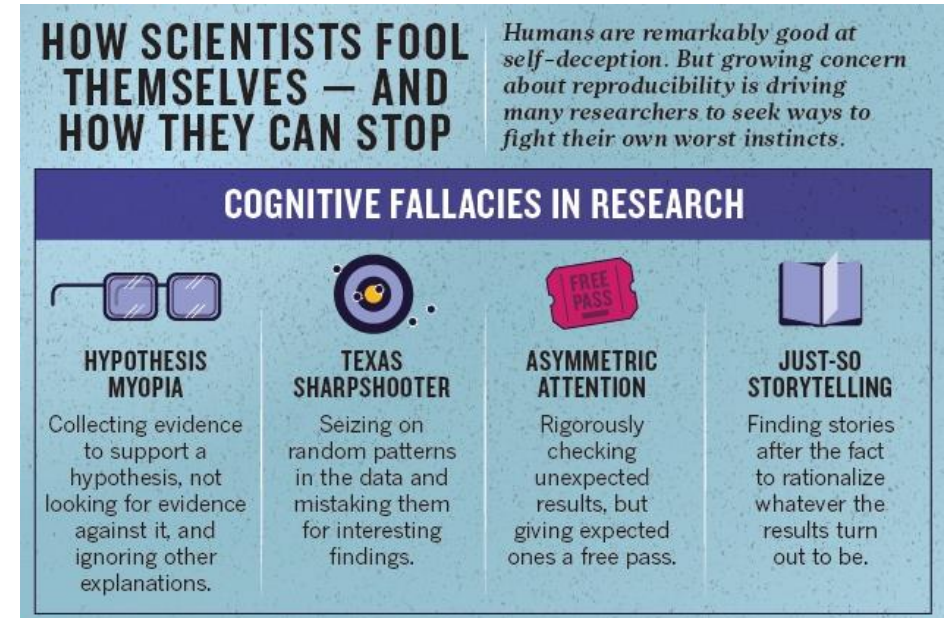
Incentives

Your career will benefit from

- Many publications
- High-impact publications
- Spectacular results

This may result in undesired behavior

- P-hacking
- Harking



Research article

The natural selection of bad science

Paul E. Smaldino✉ and Richard McElreath

Published: 01 September 2016 | <https://doi.org/10.1098/rsos.160384>

Improving scientific procedures

Design and implement your analysis

- how to start with new (pilot) analysis pipelines, preregister
- how to scale these to publication-quality group analysis

Handling of scripts, data, and results

- collaboration and version control

FAIR data management

- [BIDS](#) for organizing your data
- repositories for sharing your raw data
- publication of all analysis details, including the ugly ones

Practical issues of sharing analysis details and data

Legal issues and privacy of your research participants

Sharing of analysis details (code)



Manage versions of your analysis scripts

Github, GitLab, Bitbucket

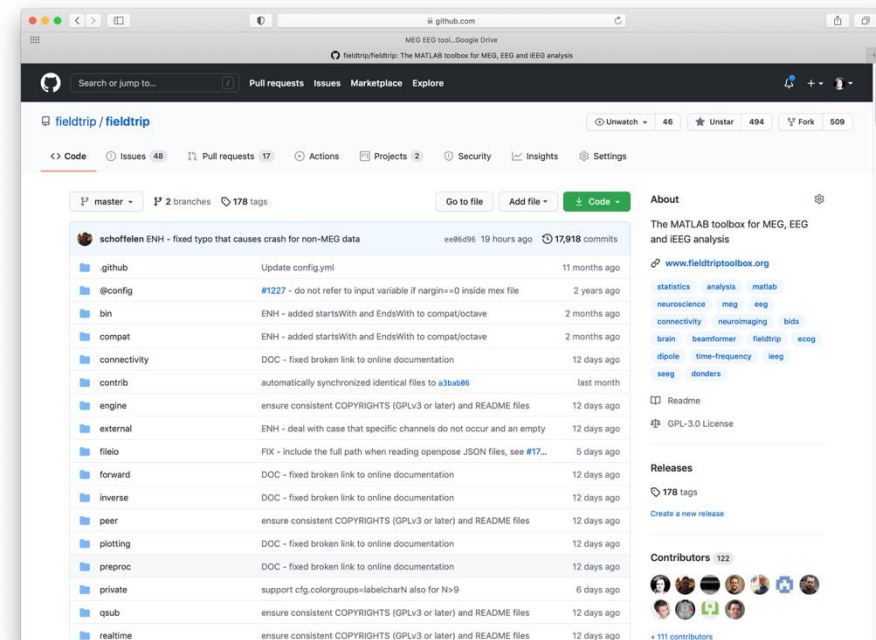
Backup and share with yourself between computers

Collaborate and review

Also used for FieldTrip development

Toolbox code

Website



Managing, versioning and sharing your code

Start with version control

```
> git init
```

Write the pipeline for a single subject

```
> git commit
```

Manage subject differences

```
> git commit
```

Run for all subjects

```
> git commit
```

Do group analysis

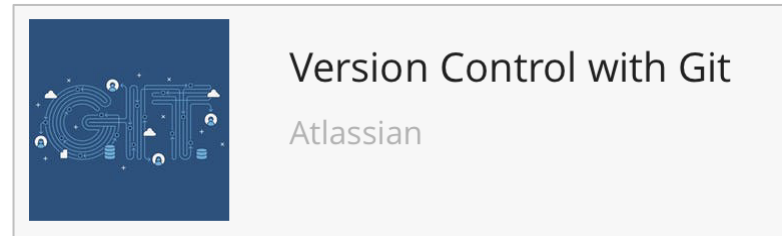
```
> git commit
```

Share your pipeline along with the paper and data

```
> git push
```



<https://www.fieldtriptoolbox.org/development/git/>



<https://www.coursera.org/learn/version-control-with-git>



<https://software-carpentry.org/lessons/>

A full example

Open Access **publication**

Data on Donders repository (data.ru.nl)

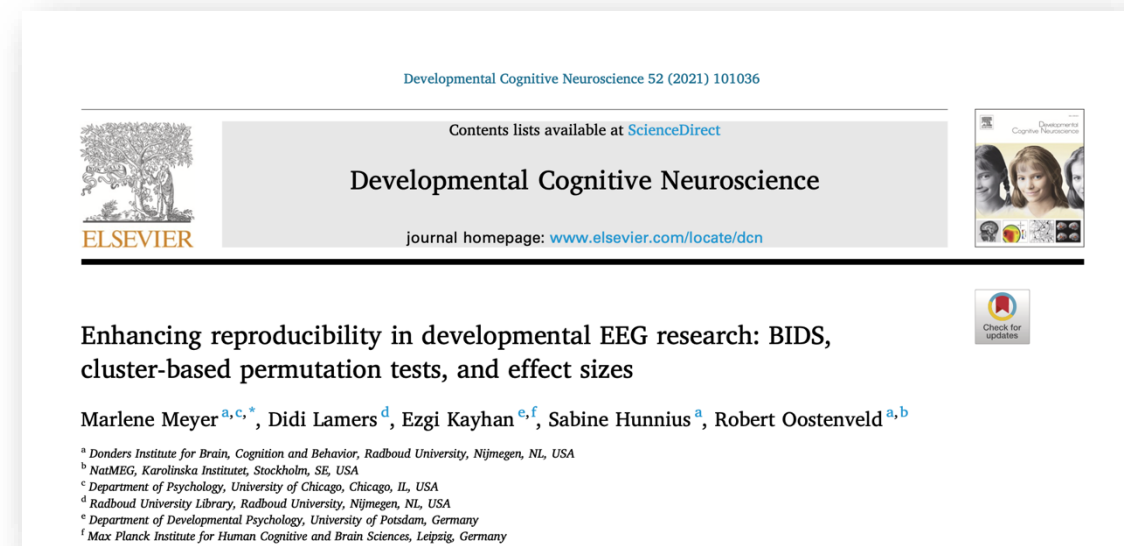
[doi: 10.34973/gvr3-6g88](https://data.ru.nl/doi/10.34973/gvr3-6g88)

[doi: 10.34973/g4we-5v66](https://data.ru.nl/doi/10.34973/g4we-5v66)

Code on GitHub

<https://github.com/Donders-Institute/infant-cluster-effectsizes>

Transparent reporting, not only p-values, also including **effect sizes**



Managing and sharing research data

Why manage research data?

- Improve efficiency and quality of research

- You and your colleagues can use existing data to jump-start new projects

- Research findings can be re-visited upon new insights

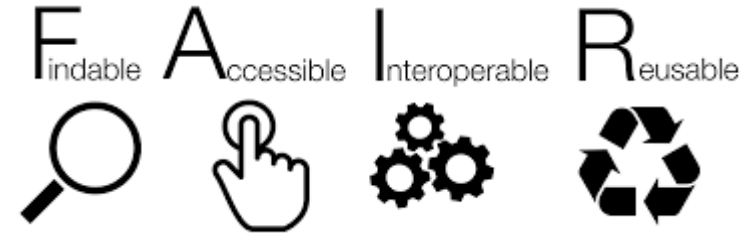
Why share data?

- Publishers require it

- Funders require it

- It is just the “right thing to do”

Open Data



Findable

Make your data available on repository with a persistent identifier (DOI, handle) and metadata

Accessible

Be explicit about data usage terms (agreement with downloader)

Interoperable

Make your data human and machine readable, e.g. BIDS

Reusable

Make sure you document enough details, e.g. “data descriptor” paper
this can be cited, along with citing our data -> measurable impact!

Brain Imaging Data Structure



What is is?



BIDS is a way to organize your existing raw data

To improve consistent and complete documentation

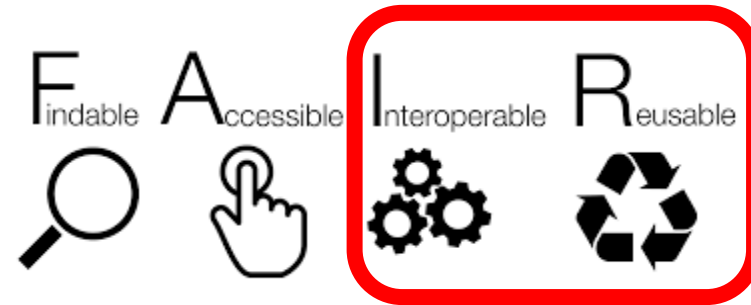
To facilitate re-use by your future self and others

BIDS is not

A new file format

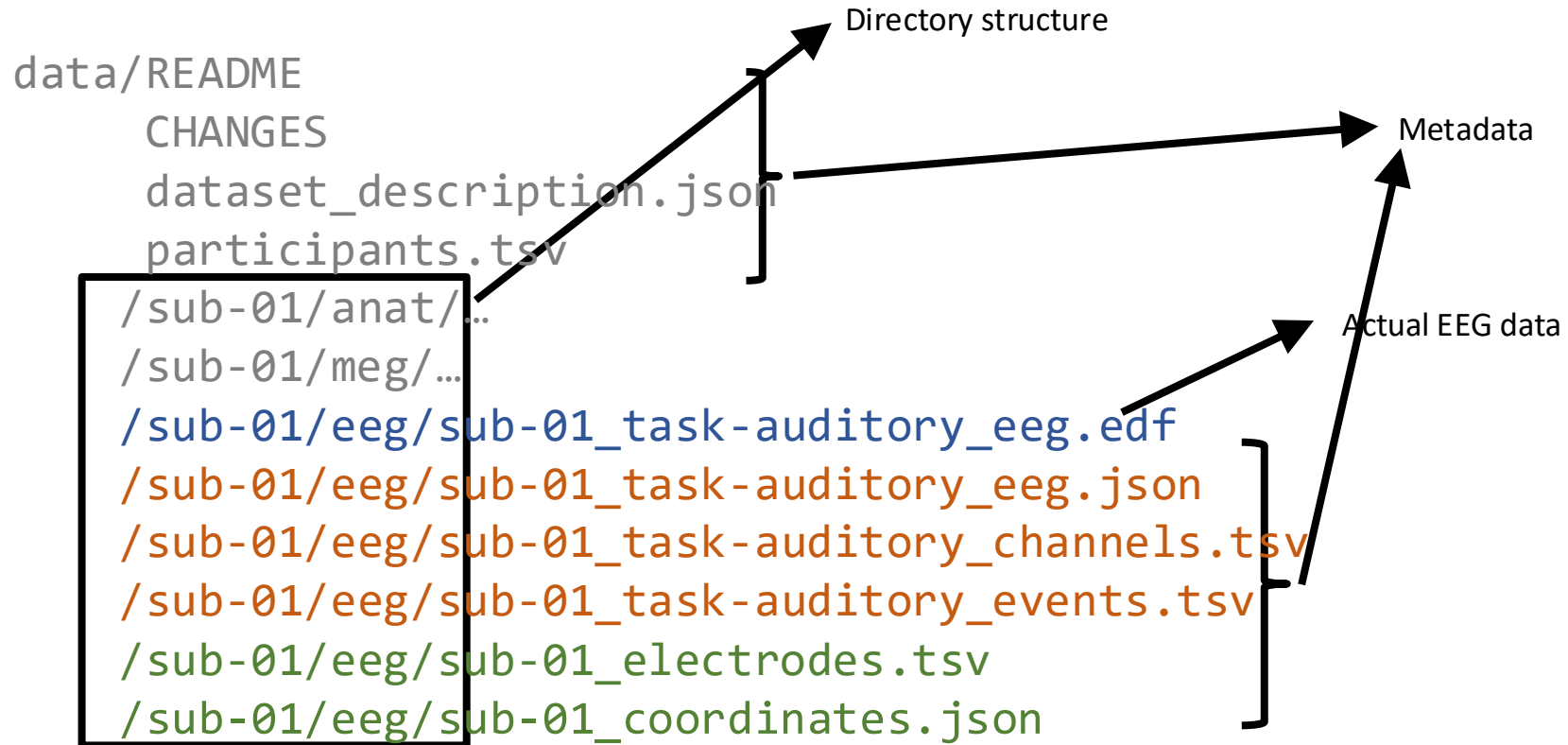
A search engine

A data sharing platform



BIDS for MRI, MEG, EEG, iEEG ...

also fNIRS, PET, motion, eye-tracker, genetics, etc.



BIDS sidecar files for metadata

see also <https://github.com/bids-standard/bids-examples>



- 1) represent otherwise missing data
- 2) make it easier to query/search

As example for EEG:

_participants.tsv and json
_sessions.tsv and json
_scans.tsv and json

_eeg.json
_channels.tsv and json
_electrodes.tsv and json
_coordinates.json
_photos.jpg

```
{  
  "TaskName": "matchingpennies",  
  "TaskDescription": "The task is emulating a game of 'matching pennies'",  
  "SamplingFrequency": 5000,  
  "Manufacturer": "Brain Products",  
  "ManufacturersModelName": "BrainAmp DC",  
  "CapManufacturer": "Brain Products",  
  "CapManufacturersModelName": "actiCAP 64Ch Standard-2",  
  "EEGChannelCount": 10,  
  "EOGChannelCount": 0,  
  "ECGChannelCount": 0,  
  "EMGChannelCount": 0,  
}
```

name	type	units	status	status_description
FC5	EEG	uV	bad	Contains high frequency noise
FC1	EEG	uV	good	n/a
C3	EEG	uV	good	n/a
CP5	EEG	uV	good	n/a
CP1	EEG	uV	good	n/a
FC2	EEG	uV	good	n/a
FC6	EEG	uV	bad	Low correlation with other channels
C4	EEG	uV	good	n/a
CP2	EEG	uV	good	n/a
CP6	EEG	uV	good	n/a

BIDS sidecar files for metadata

very convenient for the “_events.tsv”



onset	duration	trial_number	event_type	executed_movement	correct_movement	shown_stimulus	included_trial	reason_for_exclusion
12	23	1	stimulus	n/a	n/a	white X	0	test trial
35	9.50336620000598	1	stimulus	n/a	n/a	red X	0	test trial
35.35148169999959	0	1	finger press	2	2	n/a	0	test trial
35.847265999997	0	1	finger press	1	1	n/a	0	test trial
36.3019795	0	1	finger press	2	2	n/a	0	test trial
36.7276351000037	0	1	finger press	3	3	n/a	0	test trial
37.1842814999982	0	1	finger press	2	2	n/a	0	test trial
37.6747901000053	0	1	finger press	1	1	n/a	0	test trial
38.1721732000005	0	1	finger press	3	3	n/a	0	test trial
38.615335800001	0	1	finger press	2	2	n/a	0	test trial
39.1269308000046	0	1	finger press	4	4	n/a	0	test trial
39.5899466999981	0	1	finger press	2	2	n/a	0	test trial
40.0537673000072	0	1	finger press	4	4	n/a	0	test trial
40.5511086000042	0	1	finger press	1	1	n/a	0	test trial
44.5	5	1	stimulus	n/a	n/a	white X	0	test trial
61.7599999999948	22.02	2	stimulus	n/a	n/a	white X	1	n/a
83.7709132000018	9.50284050000482	2	stimulus	n/a	n/a	red X	1	n/a
84.3465505000058	0	2	finger press	2	2	n/a	1	n/a
84.8345718000055	0	2	finger press	1	1	n/a	1	n/a
85.2867590999958	0	2	finger press	2	2	n/a	1	n/a



Data and metadata should be organized in a human- and machine-readable format

Reuse existing data and metadata standards where possible

NIFTI, EDF, NWB

DICOM, CogPo, SI units

Metadata should be human-readable, hence TSV and JSON

One BIDS specification, not separate ones, hence consistent definitions and use of terms across modalities

Semantic versioning (version x.y.z), hence backwards incompatible changes must wait until version 2.0

BIDS technical principles



File naming with some redundance

Inheritance of metadata

“Source” versus “raw” versus “derived”

Making a BIDS dataset



```
sub = {'01', '02', '03', '04', '05', '06', '07', '08', '09', '10'};

% for subject 3 the age is unknown, for subject 2 the gender is not specified
age = [11 96 nan 77 82 87 18 40 26 80];
gender = {'f' [] 'f' 'f' 'm' 'm' 'm' 'm' 'm'};

for subidx=1:numel(sub)
    cfg = [];
    cfg.method = 'copy';
    cfg.datatype = 'eeg';

    % specify the input file name, here we are using the same file for every subject
    cfg.dataset = sprintf('raw/eeg/sub_%d.vhdr', i);

    % specify the output directory
    cfg.bidsroot = 'bids';
    cfg.sub = sub{subidx};

    % specify the information for the participants.tsv file
    cfg.participants.age = age(subidx);
    cfg.participants.gender = gender{subidx};

    % specify some general information that will be added to the eeg.json file
    cfg.InstitutionName = 'Donders Institute, Radboud University';
    % provide the mnemonic and long description of the task
    cfg.TaskName = 'changedetection';

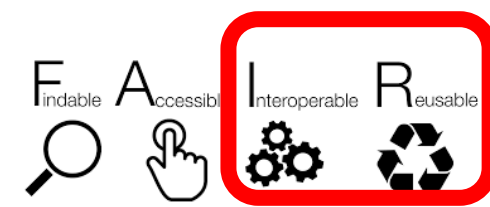
    % these are EEG specific
    cfg.eeg.PowerLineFrequency = 50; % since recorded in the EU
    cfg.eeg.EEGReference = 'M1'; % the left mastoid

    data2bids(cfg);
end
```

<https://www.fieldtriptoolbox.org/example/bids/>

BIDS is not a new file format

So which file formats are used?



MRI and PET

NIFTI, not DICOM or Analyze or MINC

MEG

Original manufacturers file formats (*.ds, *.fif)

EEG

BrainVision Core format (*.vhdr/vmrk/eeg)

[European Data Format](#) (*.edf)

EEGLAB (HDF5 *.mat file that is renamed to *.set)

Biosemi

iEEG

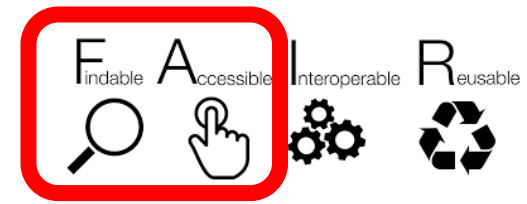
BrainVision, EDF, EEGLAB

[Neurodata Without Borders](#) (*.nwb files)

[MEF3](#) (*.mefd directory)

BIDS is not a search engine

But it standardizes the metadata



Generic search engines (i.e., Google and Bing) will not use BIDS metadata and structure

Domain specific search engines can use it

<https://commons.datacite.org>

<https://datasetsearch.research.google.com>

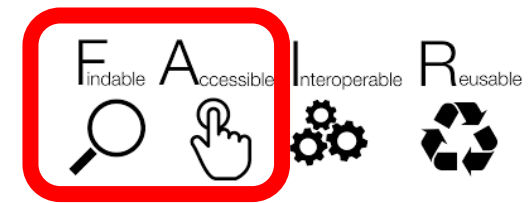
<https://www.datalad.org>



Tools based on Large language models (LLMs) like ChatGPT might use it in the future...

BIDS is not a data sharing platform

So where to share?



Institutional repository

Donders/Radboud <http://data.ru.nl>

Utrecht Uni [Yoda](#)

In the UK [Oxford](#), [Cambridge](#), [Edinburg](#)

...

National repository (in NL)

<https://dataverse.nl>

<https://data.4tu.nl>

Domain specific repository

<https://openneuro.org>

<https://publicneuro.eu>

<https://ebrains.eu>

Generic repository

<https://zenodo.org>

<https://dataverse.harvard.edu>

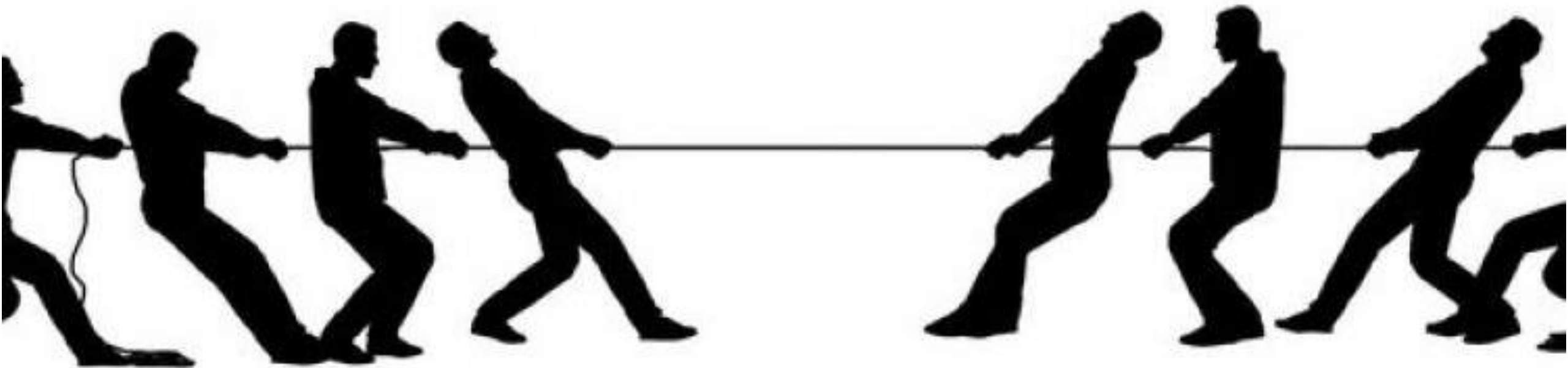
<https://osf.io>

Commercial publishers

<https://datadryad.org>

<https://figshare.com>

Open data versus privacy



Directly identifying personal data

name

address

date of birth

phone number

license plate

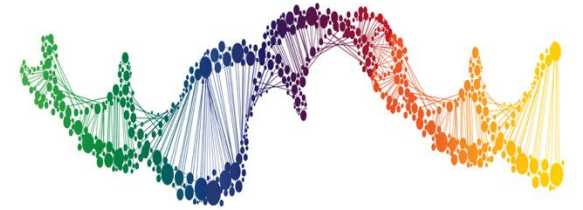
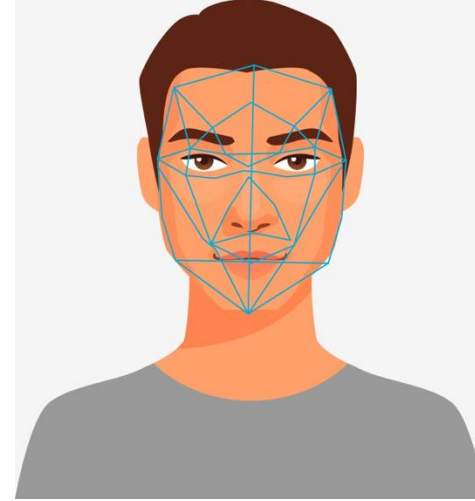
IP address

...

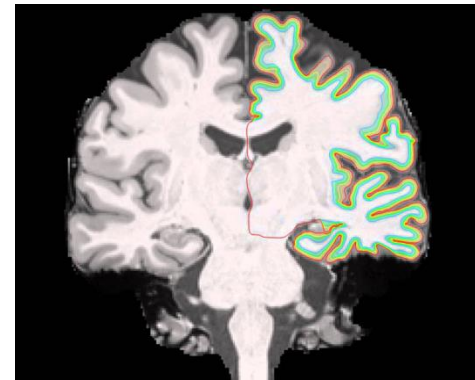


Indirectly identifying personal data

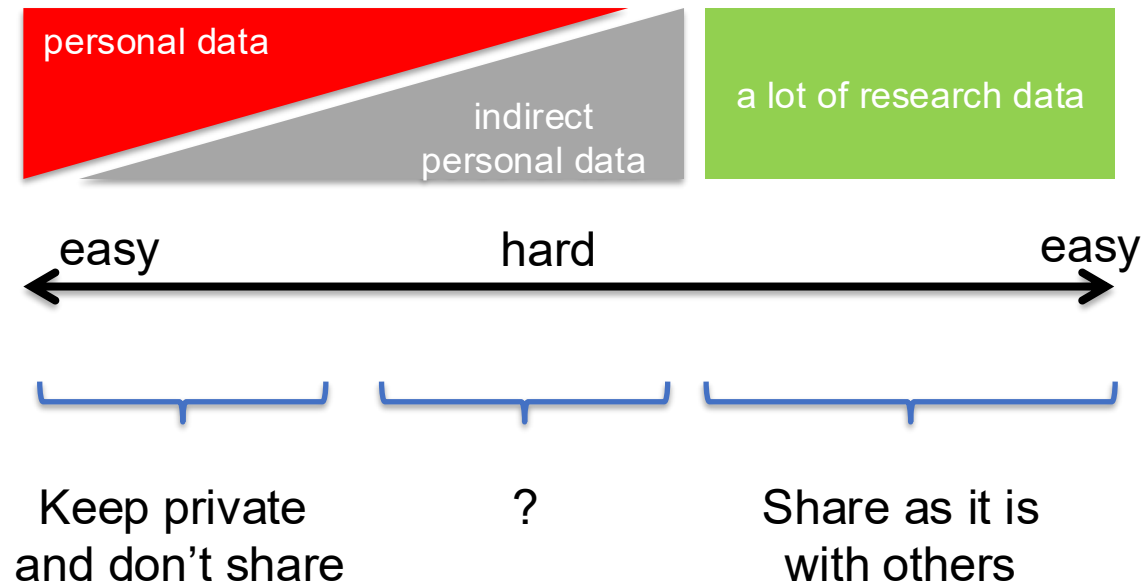
fingerprint
facial details
dental record
genetics



cortical folding pattern
clinical data
gait/movement pattern
responses on questionnaires



Gradient between personal and research data



Limit possible identification

Personal data

- restrict access to personal data

- protect the key that maps between the pseudonym and the identity

Biometric data

- data minimization only acquire, store and share data that is needed

- acquire ***anonymous*** data

- acquire data using a ***pseudonym***

- use ***de-identification*** techniques



Legal constraints

- collaboration: access only for specific authorized researchers

- sharing: access for everyone but only following data use agreement

Limit possible identification

Anonymous

You never knew the subjects identity to start with
This is not absolute and not guaranteed forever

Pseudonymization

Remove direct identifiers
Use a code instead of the subjects name

De-identification

Remove indirectly identifying features
Blur the indirect personal data
Deface anatomical MRI
Age at the time of acquisition instead of date of birth
Use age bins instead of years
Questionnaire outcomes rather than individual item scores
...



Appropriate blurring depends
on the situation



1 month bins



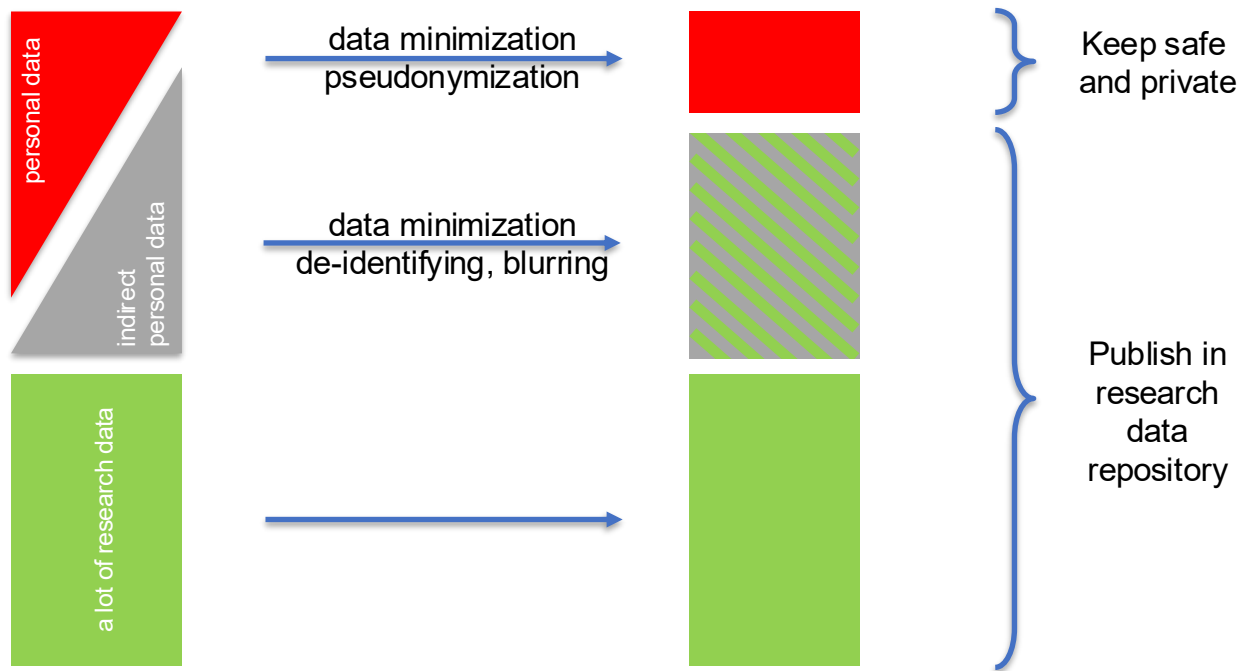
photo by Richard Ansett for IPA

5 or 10 year bins

Personal and research data

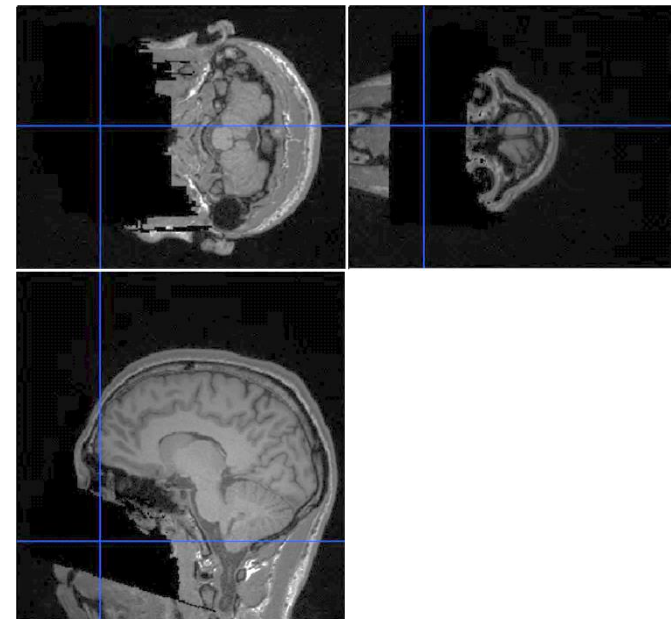
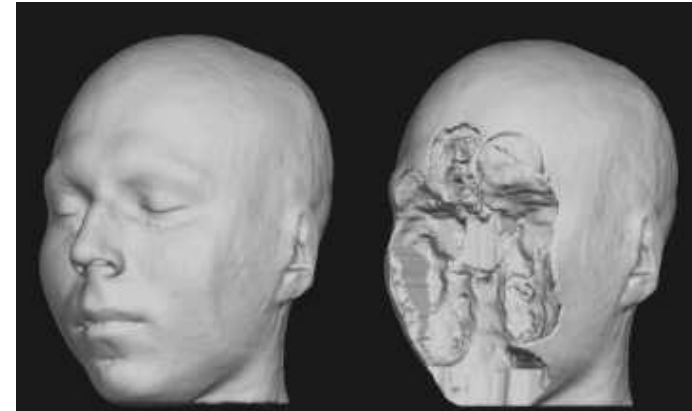
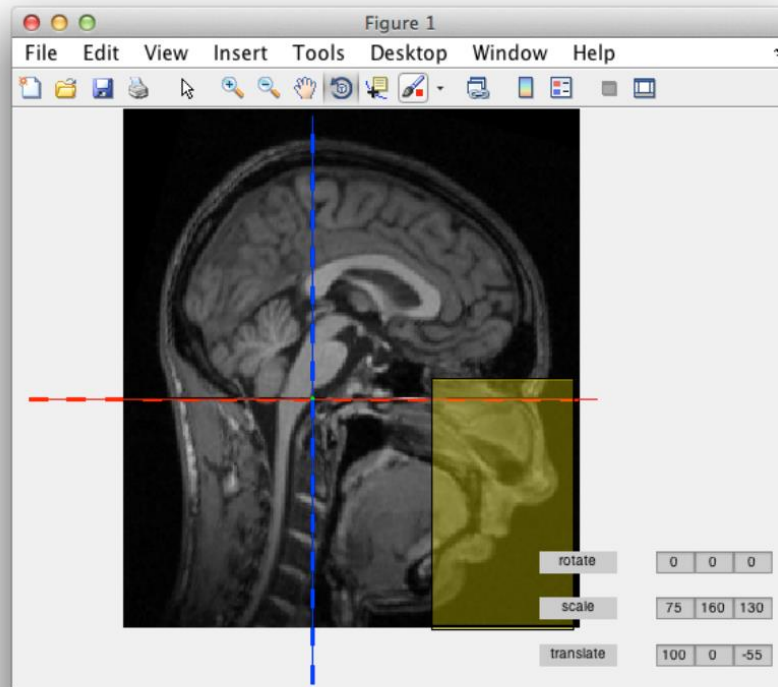


Personal and research data

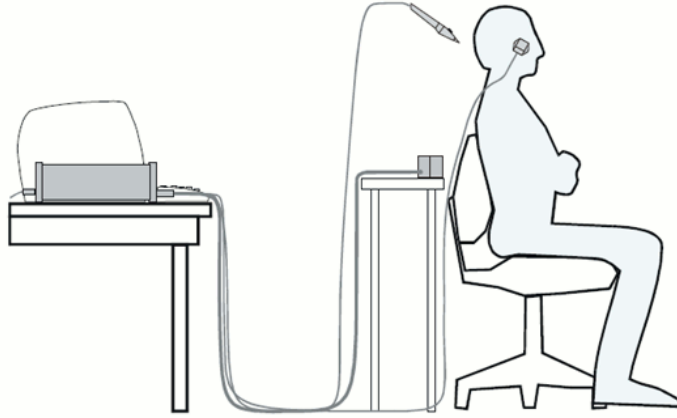


Sharing deidentified anatomical MRIs

```
mri = ft_read_mri('oostenveld_r.mri');  
  
cfg = [];  
mri_anon = ft_defacevolume(cfg, mri);
```



Coregistration between MEG/EEG and anatomy



- 1) anatomical landmarks (lpa, rpa, nas)
- 2) HPI/HCL coil locations
- 3) scalp surface points

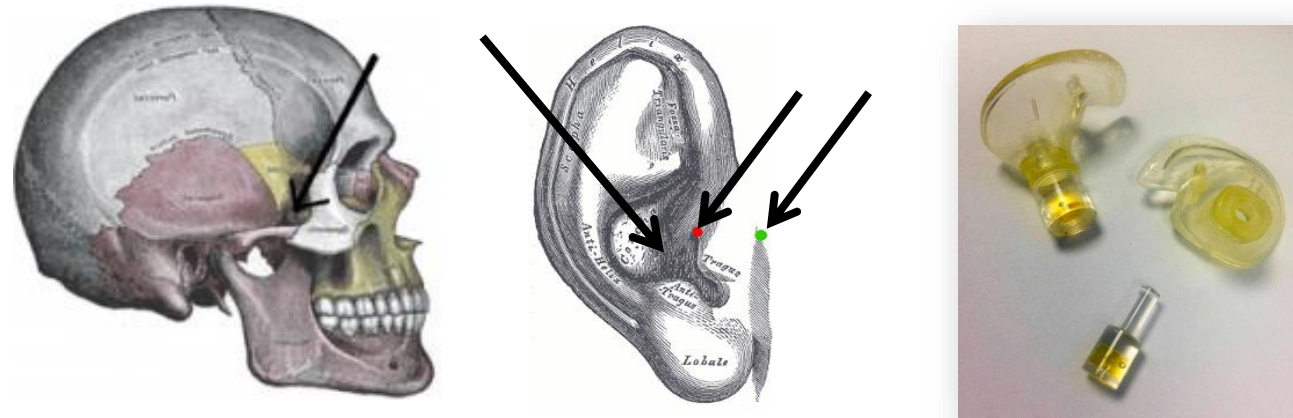
Lab specific conventions for landmarks and markers - > BIDS terminology

Landmarks:

anatomically recognizable points on the head

Markers or fiducials:

objects that are visible in multiple modalities, like head localizer coils, reflective spheres, or vitamin E capsules



Solving the challenges of sharing potentially identifiable data

Editorial > Hum Brain Mapp. 2021 May;42(7):1945-1951. doi: 10.1002/hbm.25351.

Epub 2021 Feb 1.

The Open Brain Consent: Informing research participants and obtaining consent to share brain imaging data

Elise Bannier^{1,2}, Gareth Barker³, Valentina Borghesani⁴, Nils Broeckx⁵, Patricia Clement⁶, Kyrre E Emblem⁷, Satrajit Ghosh^{8,9}, Enrico Glerean^{10,11}, Krzysztof J Gorgolewski¹², Marko Havu¹³, Yaroslav O Halchenko¹⁴, Peer Herholz¹⁵, Anne Hespel¹⁶, Stephan Heunis¹⁷, Yue Hu¹⁸, Chuan-Peng Hu¹⁹, Dorien Huijser²⁰, María de la Iglesia Vayá²¹, Radim Jancalek²², Vasileios K Katsaros^{23,24}, Marie-Luise Kieseler²⁵, Camille Maumet²⁶, Clara A Moreau²⁷, Henk-Jan Mutsaerts^{28,29}, Robert Oostenveld³⁰, Esin Ozturk-Isik³¹, Nicolas Pascual Leone Espinosa³², John Pellman³³, Cyril R Pernet³⁴, Francesca Benedetta Pizzini³⁵, Amira Šerifović Trbalić³⁶, Paule-Joanne Toussaint³⁷, Matteo Visconti di Oleggio Castello³⁸, Fengjuan Wang³⁹, Cheng Wang⁴⁰, Hua Zhu⁴¹

Affiliations + expand

PMID: 33522661 PMCID: [PMC8046140](#) DOI: [10.1002/hbm.25351](#)

[Free PMC article](#)

Pseudonymisation of neuroimages and data protection: *Increasing access to data while retaining scientific utility*

Damian Eke^{a,*,}, Ida E.J. Aasebø^b, Simisola Akintoye^c, William Knight^a, Alexandros Karakasidis^{d,e}, Ezequiel Mikulan^f, Paschal Ochang^a, George Ogoh^a, Robert Oostenveld^{g,h}, Andrea Pigorini^f, Bernd Carsten Stahl^a, Tonya Whiteⁱ, Lyuba Zehl^j

[Show more](#) v

[Share](#) [Cite](#)

<https://doi.org/10.1016/j.jnirp.2021.100053>

[Get rights and content](#)

Under a Creative Commons license

[Open access](#)

> Neuroimage. 2022 Jul 1;254:119165. doi: 10.1016/j.neuroimage.2022.119165. Epub 2022 Apr 1.

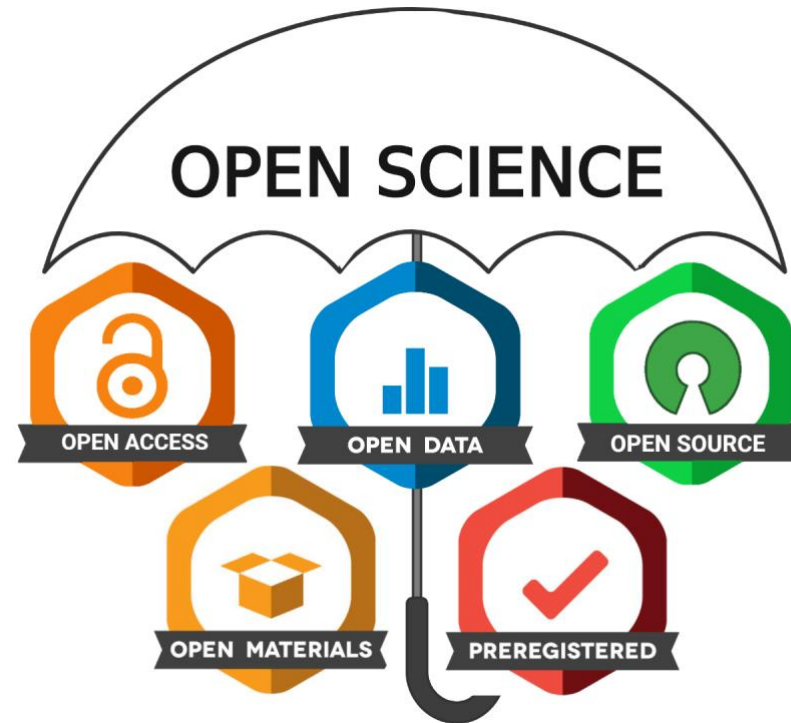
Sharing individualised template MRI data for MEG source reconstruction: A solution for open data while keeping subject confidentiality

Mikkel C Vinding¹, Robert Oostenveld²

Affiliations + expand

PMID: 35378289 DOI: [10.1016/j.neuroimage.2022.119165](#)

[Free article](#)



Open Science and good practices

Robert Oostenveld

Donders Institute, Radboud University, Nijmegen, NL
Karolinska Institutet, Stockholm, SE

robert.oostenveld@donders.ru.nl