NatMEG

DONDERS INSTITUTE

Radboud University

Open science and good practices

making your MEG/EEG research future proof

Robert Oostenveld

Donders Institute, Radboud University, Nijmegen, NL

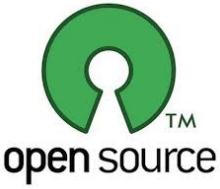
NatMEG, Karolinska Institute, Stockholm, SE

Outline

- Issues that we are facing
  - Reproducibility crisis
  - Complexity / efficiency
- Solutions that are being proposed
  - Open science
  - Improved research data management
  - Skills to learn

Open Science

- Open access publications
  - Open peer review
- Open educational resources
- Open methodology
  - Open source
  - Open hardware
- Open data



Markus Neuschäfer; https://www.slideshare.net/mneuschaefer/1504-open-knowledgefolien

Why do Open Science?

Democratic – science should be accessible for all

Pragmatic – it is more efficient to collaborate

Infrastructure – it results in better tools

Public – the public deserves to be well informed

Measurement – results are better quantified

But also some other motivations… lack of trust and of reproducibility

Lack of trust - in society

http://harrieverbon.blogspot.nl/2012/11/diederik-stapel-werd-ook-betaald-door.html

Lack of trust - reproducibility

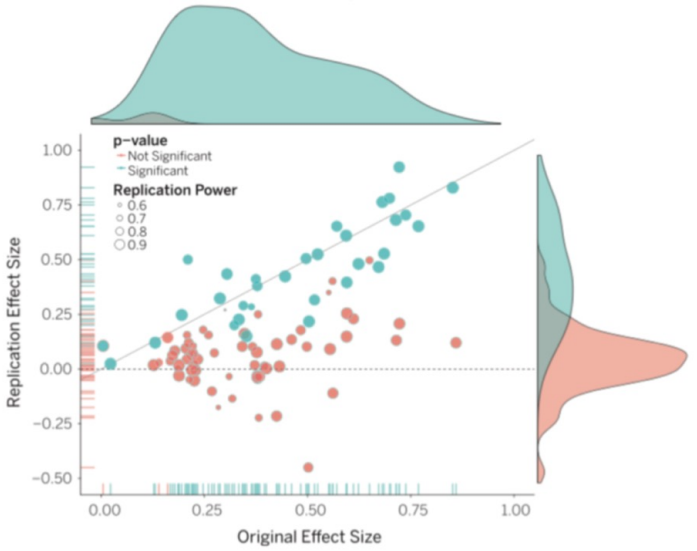
RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.



Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

Open Science Collaboration, Science (2015). DOI: 10.1126/science.aac4716

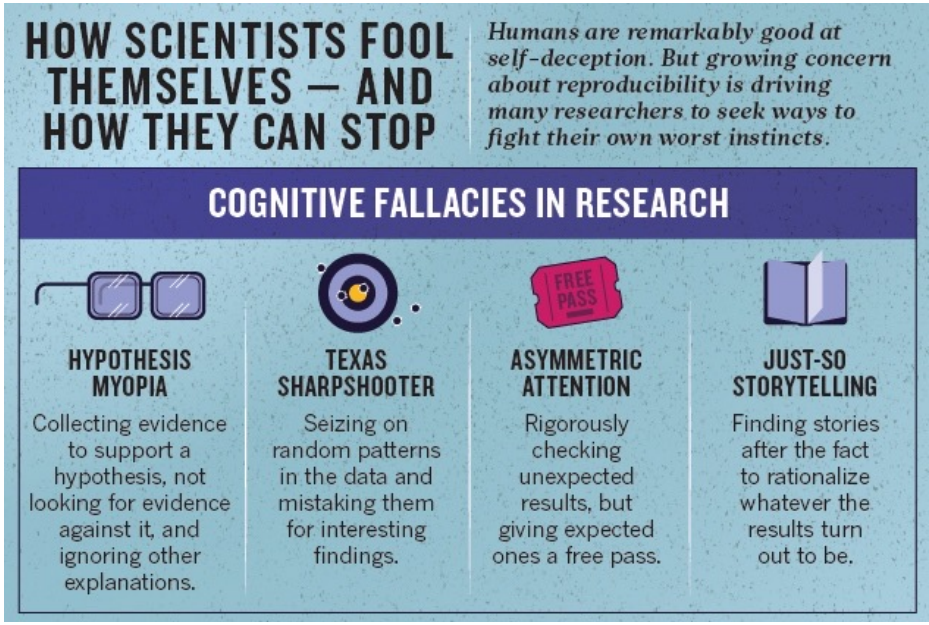
Incentives

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

Outline

- Issues that we are facing
  - Reproducibility crisis
  - Complexity / efficiency
- Solutions that are being proposed
  - Open science
  - Improved research data management
  - Skills to learn

Improving scientific *procedures*

- Design and implement your anlaysis
  - how to start with new (pilot) analysis pipelines
  - how to scale these to publication-quality group analysis
- Handling of scripts, data, and results
- FAIR data management
  - BIDS for organizing your data
  - Repositories for sharing your raw data
  - Publication of your analyses details
- Practical issues of sharing data and analysis details
- Legal issues and privacy of your subjects

Single-subject versus group-analysis

https://humanconnectome.org/study/hcp-young-adult

https://github.com/Washington-University/megconnectome

Frontiers in Neuroscience - From raw MEG/EEG to publication: how to perform MEG/EEG group analysis with free academic software

https://github.com/robertoostenveld/Wakeman-and-Henson-2015

Small or large data
Small or large computers

Note: “big data” is complex data, “large data” is large in size but not per see complex

Sharing of analysis details (code)

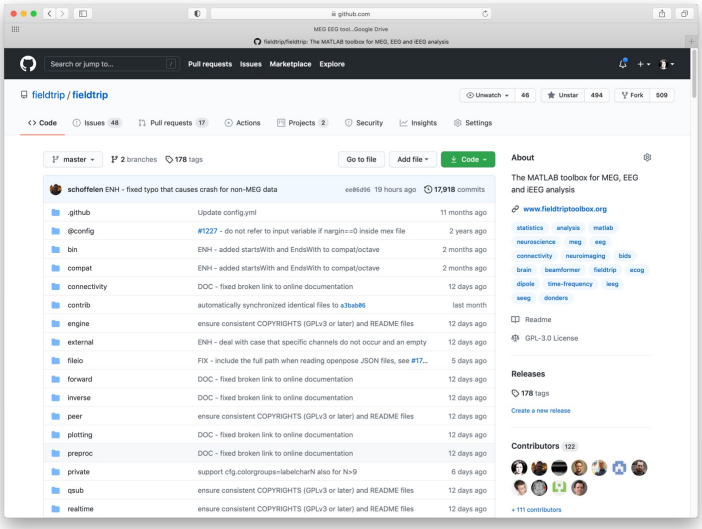
GitHub

Manage versions of your analysis scripts

- Github, Gitlab, Bitbucket
- Backup and share between computers
- Collaborate and review

Also used for FieldTrip development

- Toolbox code improvements
- Website



Managing 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
```

FieldTrip

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

Version Control with Git

Atlassian

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

software carpentry

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

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

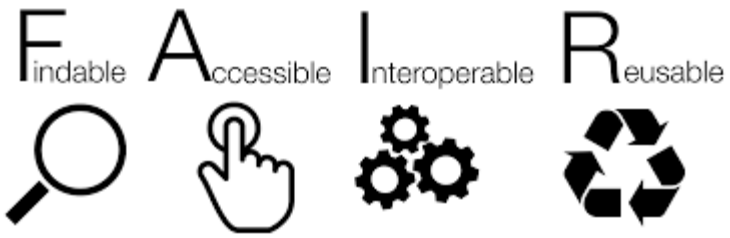
Why *share* data?

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

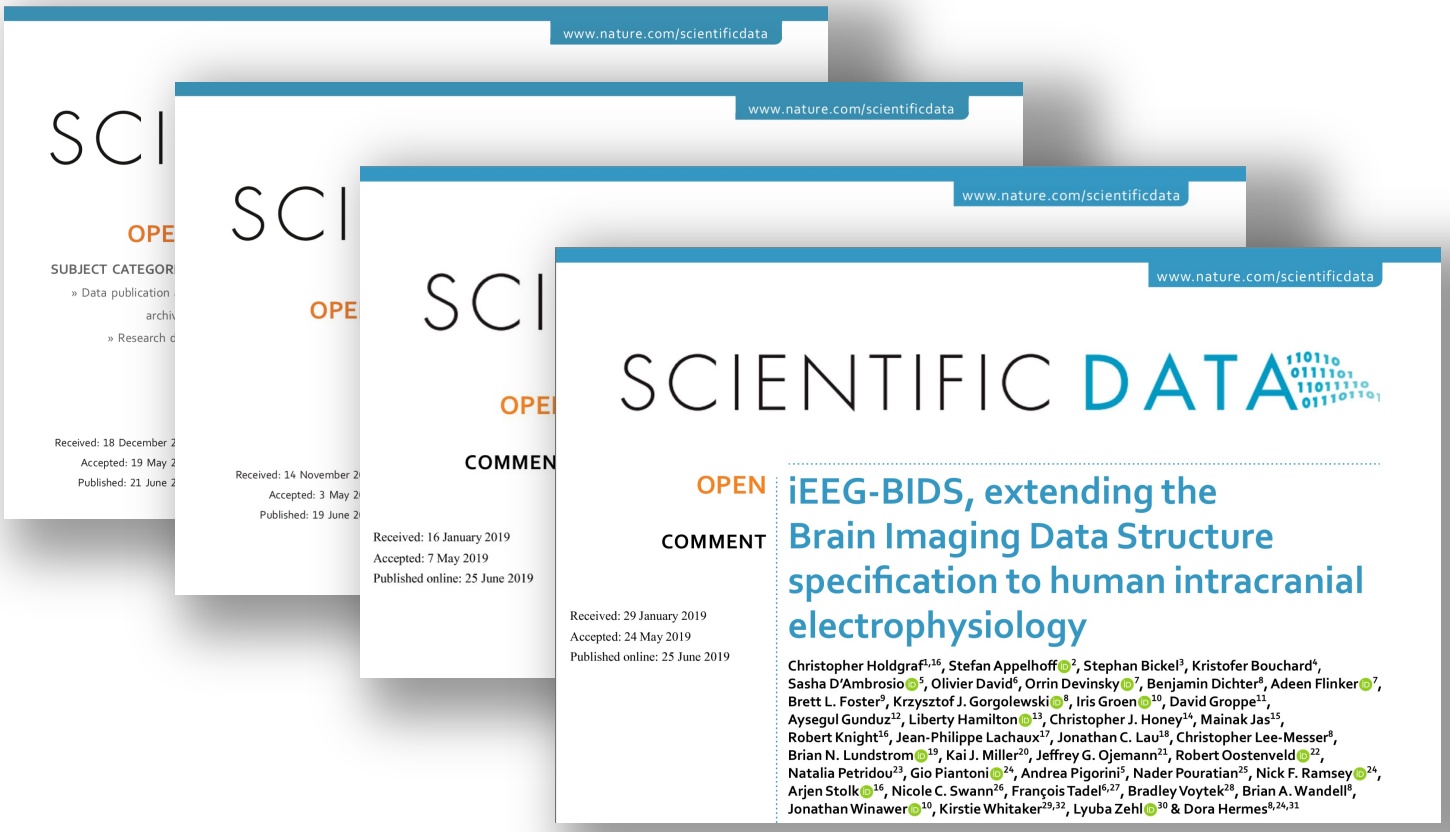
Reusable

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

BIDS
BRAIN IMAGING DATA STRUCTURE



http://bids.neuroimaging.io

What is is?

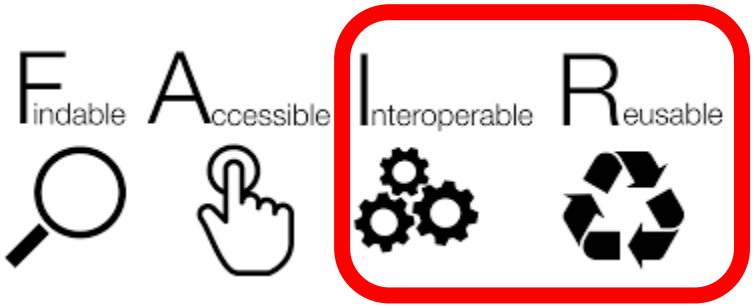
BIDS
BRAIN IMAGING DATA STRUCTURE

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

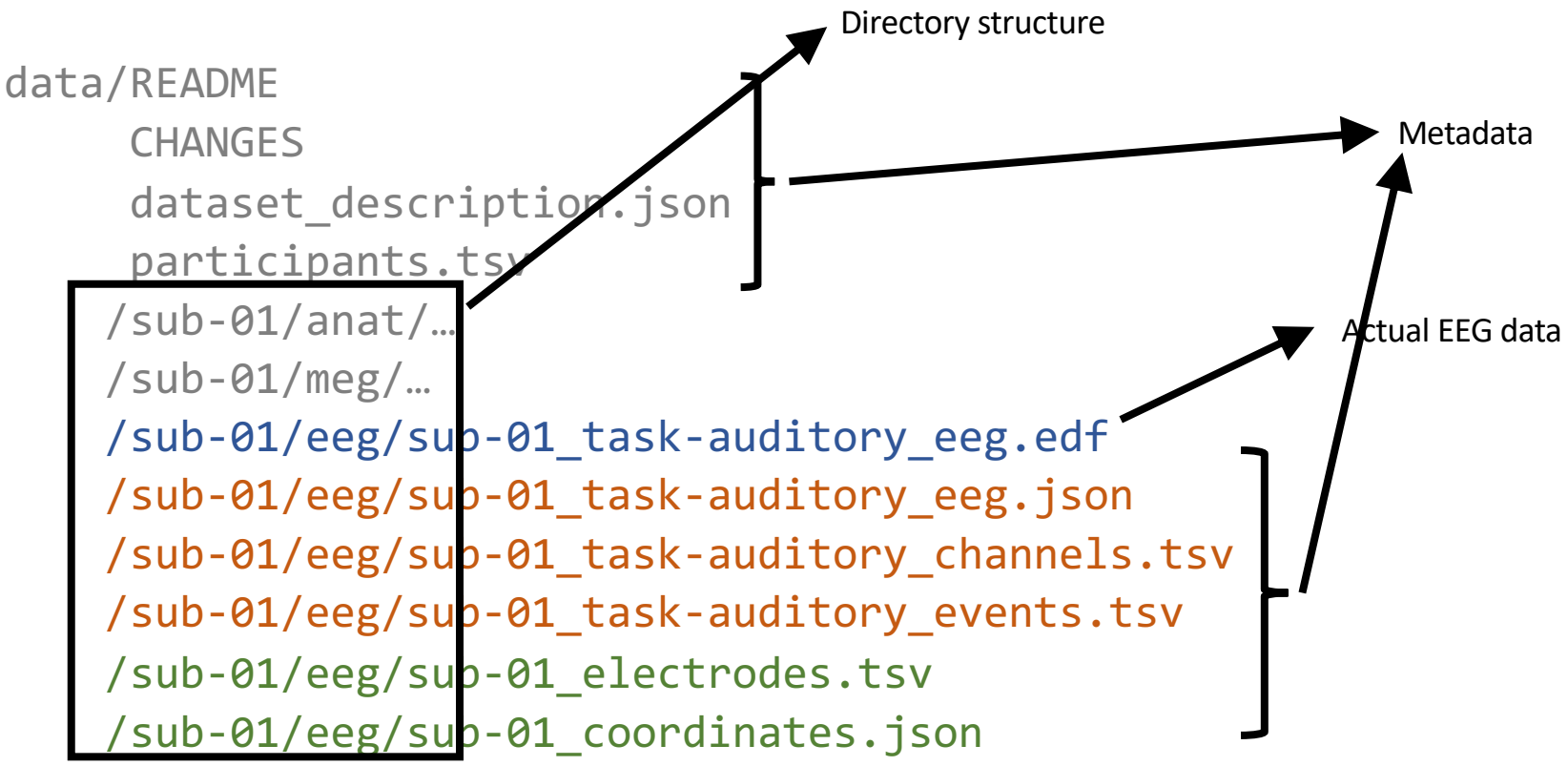
BIDS is not

- A new file format
- A search engine
- A data sharing platform



BIDS for MRI, MEG, EEG, iEEG …

in future also PET, 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 penn
    "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 conceptual principles

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

Pareto principle: 80% of the value from 20% of the use-cases

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

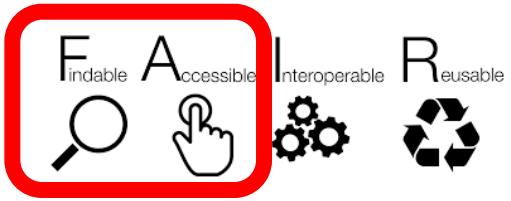
Semantic versioning (version x.y.z), hence backwards incompatible changes must wait until version 2.0

BIDS technical principles

File naming with some redundance

Inheritence of metadata

“Source” versus “raw” versus “derived”



BIDS is not a search engine

but it standardizes the metadata

Generic search engines (i.e. web crawlers) will not use BIDS metadata and structure

Domain specific search engines might use it

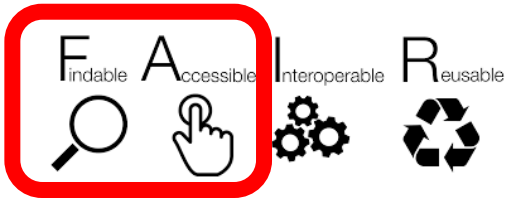
- https://search.datacite.org
- https://datasetsearch.research.google.com
- https://www.datalad.org

DataCite
FIND, ACCESS, AND REUSE DATA

Google Dataset Search Beta

EBRAINS

BIDS is not a data sharing platform



So where to share?

Institutional repository

- Donders https://data.donders.ru.nl
- Radboud University http://data.ru.nl
- In the UK Oxford, Cambridge, Edinburg
- …

National repository (in NL)

- https://easy.dans.knaw.nl
- https://dataverse.nl
- https://data.4tu.nl

Domain specific repository

- https://openneuro.org
- https://ebrains.eu

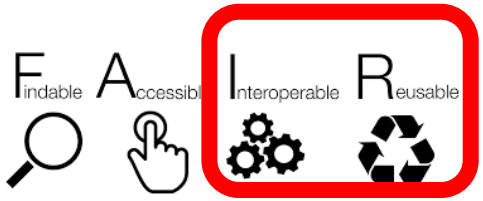
General repository

- https://zenodo.org
- https://dataverse.harvard.edu
- https://osf.io

Commercial publishers

- https://datadryad.org
- https://figshare.com

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



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

EEG

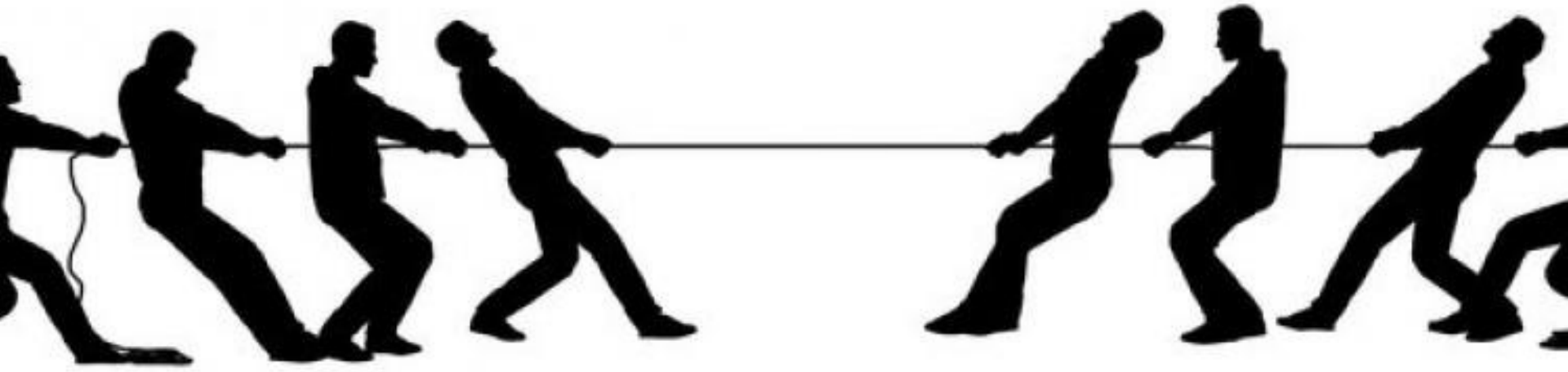
- BrainVision Core format
- European Data Format (*.edf)
- EEGLAB (HDF5 *.mat file that is renamed to *.set)
- Biosemi

iEEG

- BrainVision, EDF, EEGLAB
- Neurodata Without Borders (*.nwb files)
- MEF3 (*.mefd directory)

Upcoming: NIRS (SNIRF), motion capture (tbd)

Open data versus privacy



Personal data

name

address

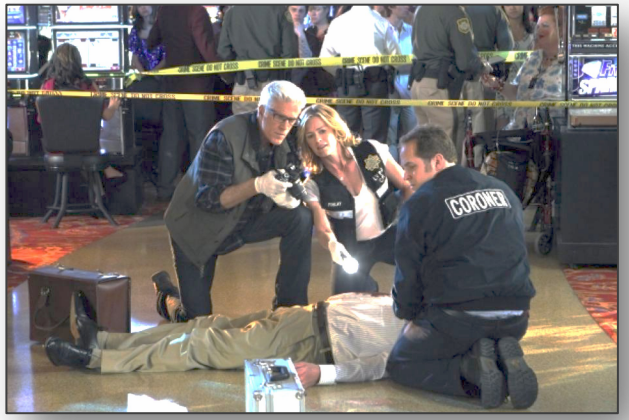
date of birth

phone number

license plate

IP address

...



Crime Scene Investigation
http://www.abc.net.au/news/2017-09-19/csi/8960590

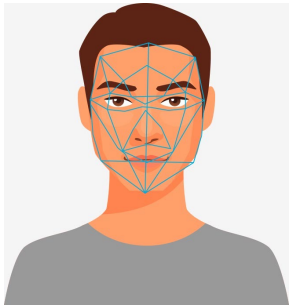
(Biometric) personal data

fingerprint

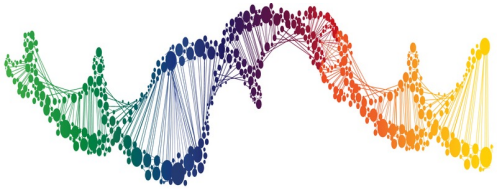
facial details

dental record

genetics



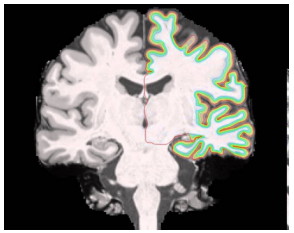
cortical folding pattern



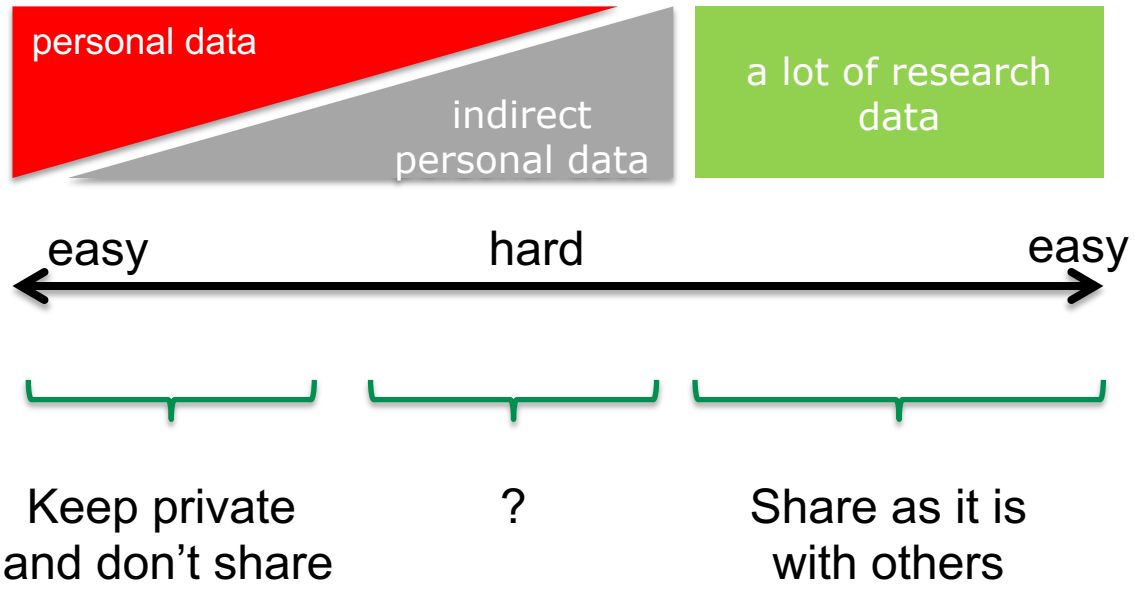
clinical data

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

- 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

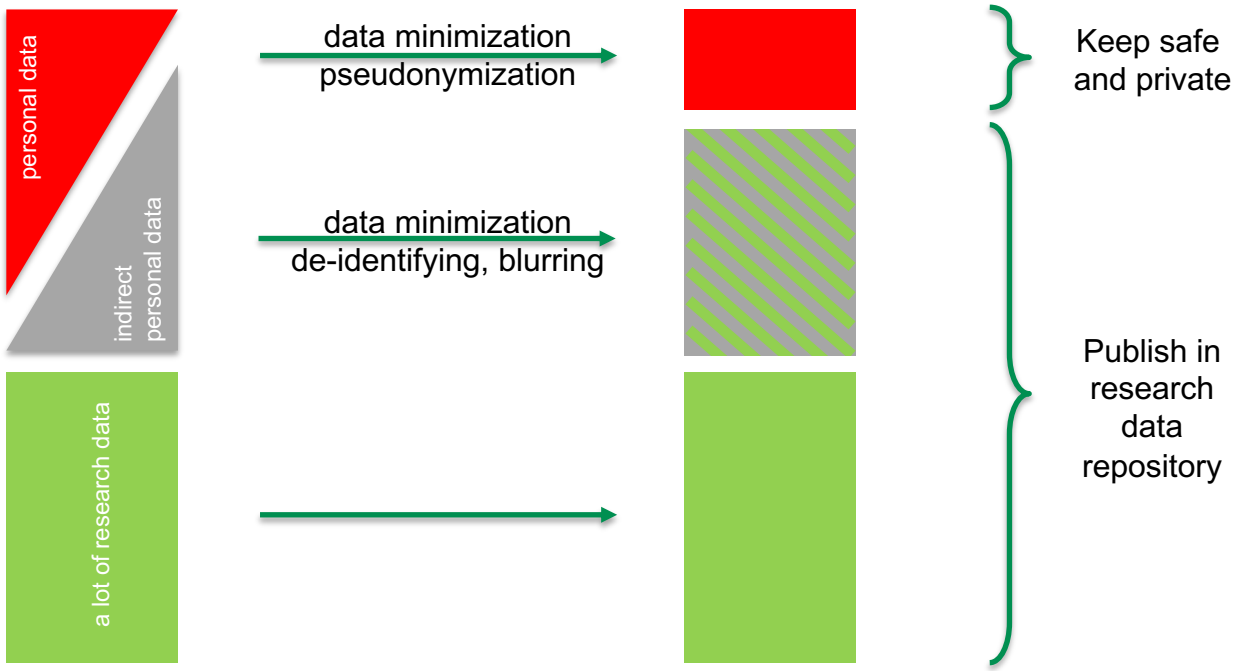
… for example blurring the age of the subject

1 month bins

5 or 10 year bins

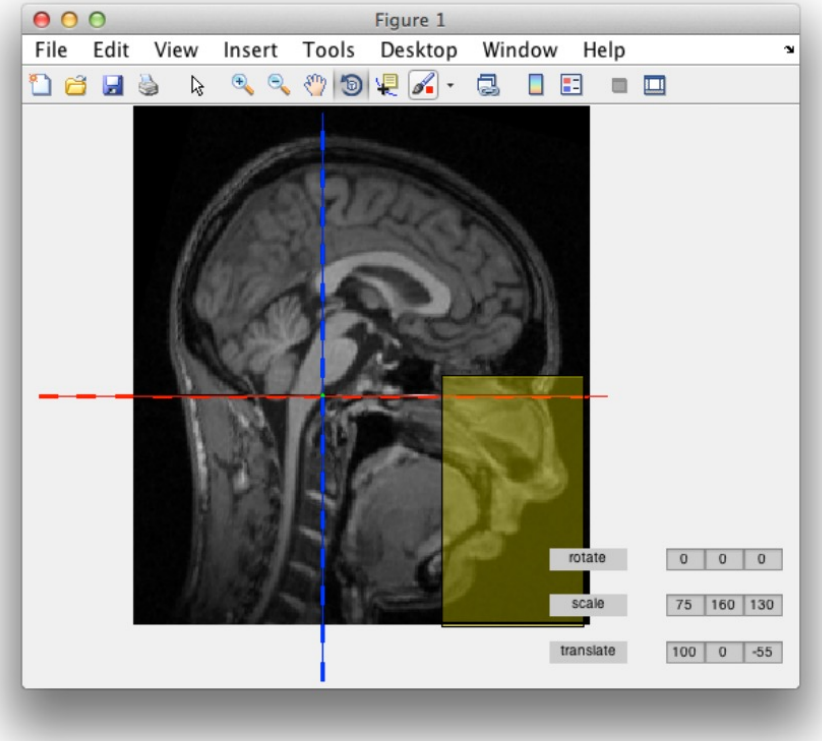
Personal and research data

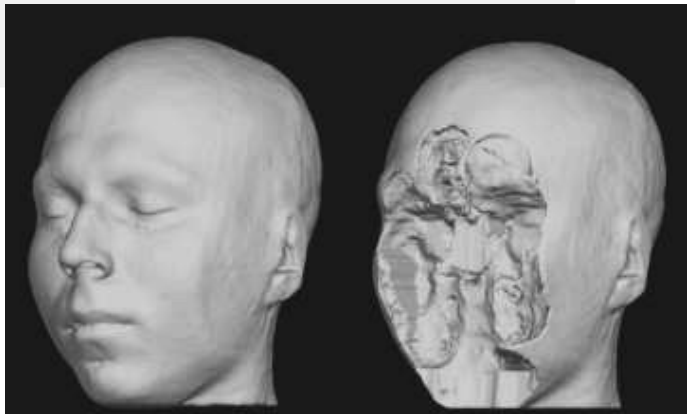
Personal and research data

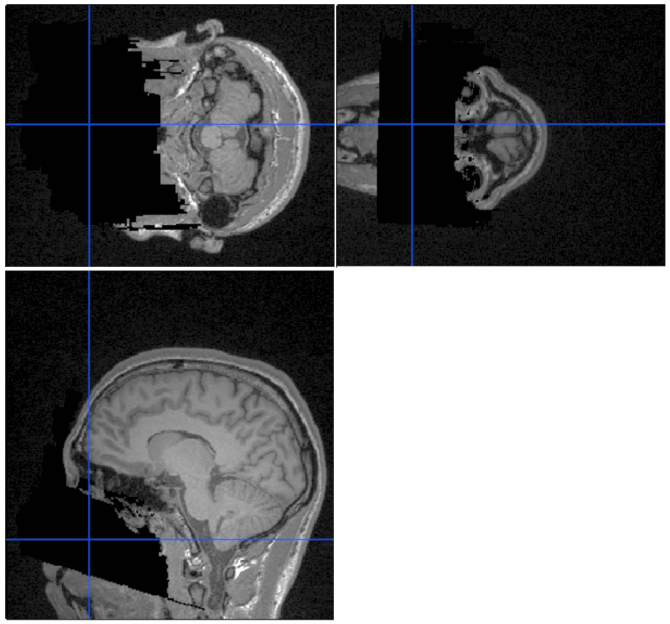


Sharing deidentified imaging data

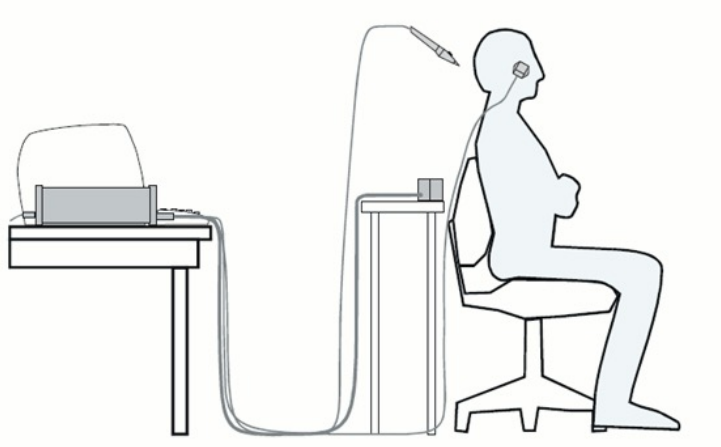
```
mri = ft_read_mri('oostenveld_r.mri');

cfg = [];
mri_anon = ft_defacevolume(cfg, mri);
```







Coregistration between MEG/EEG and anatomy

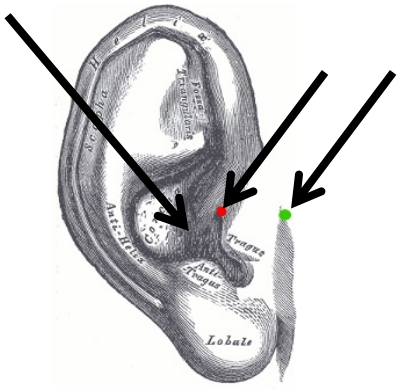


1) anatomical landmarks (lpa, rpa, nas)
2) HPI/HCL coil locations
3) scalp surface points

Lab specific conventions for landmarks and markers

Landmarks:
anatomically recognizable points on the head

Markers (or fiducials):
objects that are visible in multiple modalities, e.g. ehad localizer coils, reflective spheres, or vitamin E capsules



Solving the challenges of sharing potentially identifiable data

› 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 Vinding [1], Robert Oostenveld [2]

Affiliations + expand

PMID: 35378289 DOI: 10.1016/j.neuroimage.2022.119165

Free 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 White [i], Lyuba Zehl [j]

Show more

Share Cite

https://doi.org/10.1016/j.ynirp.2021.100053

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 Barker [3], Valentina Borghesani [4], Nils Broeckx [5], Patricia Clement [6], Kyrre E Emblem [7], Satrajit Ghosh [8, 9], Enrico Glerean [10, 11], Krzysztof J Gorgolewski [12], Marko Havu [13], Yaroslav O Halchenko [14], Peer Herholz [15], Anne Hespel [16], Stephan Heunis [17], Yue Hu [18], Chuan-Peng Hu [19], Dorien Huijser [20], María de la Iglesia Vayá [21], Radim Jancalek [22], Vasileios K Katsaros [23, 24], Marie-Luise Kieseler [25], Camille Maumet [26], Clara A Moreau [27], Henk-Jan Mutsaerts [28, 29], Robert Oostenveld [30], Esin Ozturk-Isik [31], Nicolas Pascual Leone Espinosa [32], John Pellman [33], Cyril R Pernet [34], Francesca Benedetta Pizzini [35], Amira Šerifović Trbalić [36], Paule-Joanne Toussaint [37], Matteo Visconti di Oleggio Castello [38], Fengjuan Wang [39], Cheng Wang [40], Hua Zhu [41]

Affiliations + expand

PMID: 33522661 PMCID: PMC8046140 DOI: 10.1002/hbm.25351

Free PMC article

Sharing of data

Institutional Repository

- Donders Repository

Generic repositories (note the DUA)

- Zenodo, Harvard DataVerse, DataDryad, …

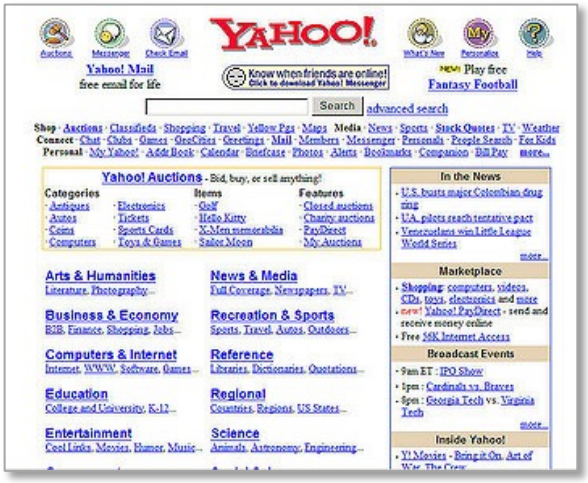
Specific repositories

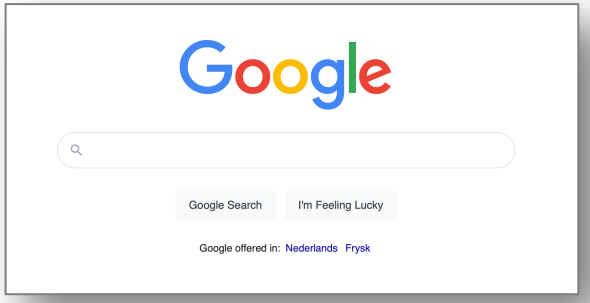
- Genetics, astromomy, openfmri, …

Re3data - repository of data repositories

Narcis – scholarly information (and data) in NL

Elsevier - datasearch





(Summary)

Things to look out for

Toolboxes and communities like FieldTrip 😁, but also MNE-Python, BrainStorm, EEGLAB, SPM, and others

Projects and services

- OSF
- COBIDAS
- #EEGManyLabs
- #ManyPipelines
- ARTEM-IS

(Summary)

General/transferable skills to acquire

Good research practices and a critical view

(MATLAB) coding and code management

GitHub or other code versioning/sharing

Data management, e.g. BIDS

Data sharing platforms, eg. OpenNeuro and OSF

Knowledge sharing, e.g. manuscripts but also using Wikipedia, the FieldTrip website, Youtube, etc