Making MEG/EEG analysis future proof

Robert Oostenveld

MEG/EEG toolkit 2019

Outline

- Issues that we are facing
  - Reproducibility crisis
  - Complexity / efficiency
- Solutions that are being proposed
  - Open science
  - Big data(?)
  - Research data management

Open Science

Open educational resources

Open access publications

- Open peer review

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 – science should be disseminated to the wide public

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

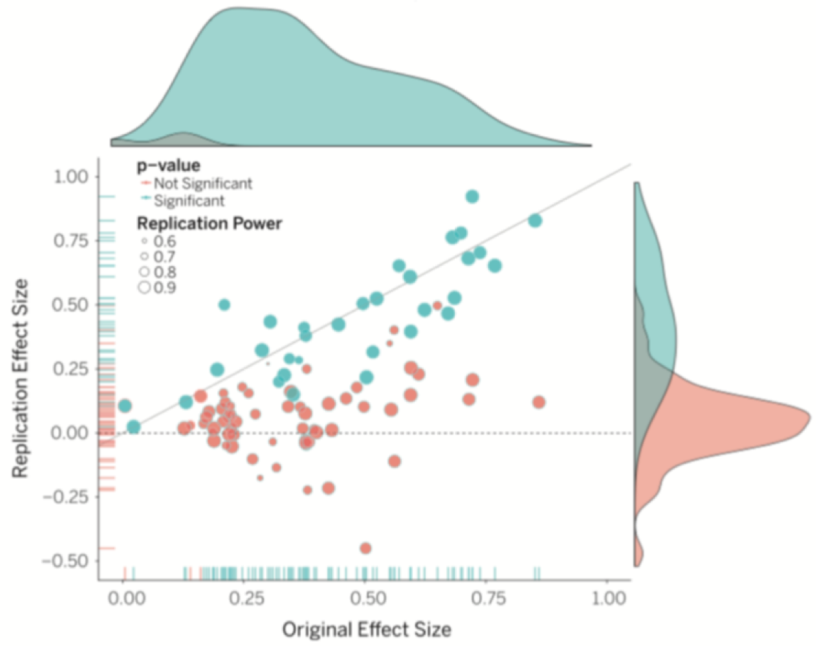
RESEARCH ARTICLE

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

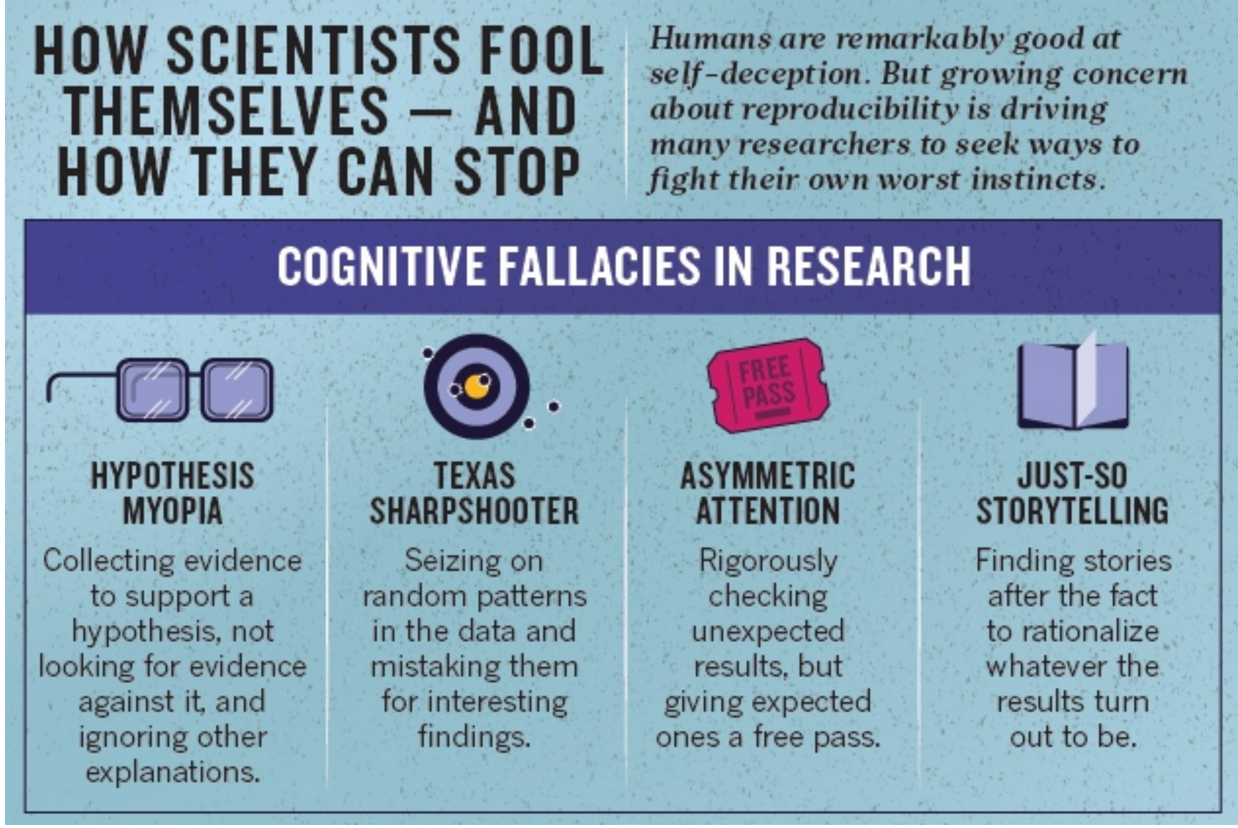
Incentives

Your career will benefit from

- Many publications
- High-impact publications
- Spectacular results

This *may* result in undesired behavior

- P-hacking
- Harking



Improving scientific procedures

Scaling up from pilot analysis to publication quality group analysis

Handling of data, scripts and results

Open Science

- BIDS for organizing your data
- Repositories for sharing your 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

Why *manage* research data?

Improve efficiency and quality of research

Researchers can use shared 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"

What is BIDS?

- BIDS is a way to organize raw data
  - To improve consistent and complete documentation
  - To facilitate (re)use by your future self and others

- BIDS is <u>not</u>
  - A new file format
  - A search engine
  - A sharing tool

File names and paths convey meaning

```
.
├── dataset_description.json
├── participants.tsv
├── README
├── sub-01
│   └── eeg
│       ├── sub-01_task-mytask_channels.tsv
│       ├── sub-01_task-mytask_eeg.edf
│       └── sub-01_task-mytask_events.tsv
├── task-mytask_eeg.json
└── task-mytask_events.json
```

Data organized in BIDS should be human and machine interpretable

```
.
├── dataset_description.json
├── participants.tsv
├── README
├── sub-01
│   └── eeg
│       ├── sub-01_task-mytask_channels.tsv
│       ├── sub-01_task-mytask_eeg.edf
│       └── sub-01_task-mytask_events.tsv
├── task-mytask_eeg.json
└── task-mytask_events.json
```

- Reuse established file formats (nifti, brainvision, tsv, json, ...)
- Metadata should be easy to read by humans in a text editor
- Metadata (for large collections) should be queryable by computers

TSV files for additional data

- Tab Separated Values

```
.
├── dataset_description.json
├── participants.tsv
├── README
├── sub-01
│   └── eeg
│       ├── sub-01_task-mytask_channels.tsv
│       ├── sub-01_task-mytask_eeg.edf
│       └── sub-01_task-mytask_events.tsv
├── task-mytask_eeg.json
└── task-mytask_events.json
```

onset	duration	trial_type	response_time	stim_file
18	0	right	116.666666667	left_hand.png
22	0	right	100	right_hand.png
27	0	left	33.3333333333	right_hand.png
35	0	right	16.6666666667	right_hand.png

```
.
├── dataset_description.json
├── participants.tsv
├── README
├── sub-01
│   └── eeg
│       ├── sub-01_task-mytask_channels.tsv
│       ├── sub-01_task-mytask_eeg.edf
│       └── sub-01_task-mytask_events.tsv
├── task-mytask_eeg.json
└── task-mytask_events.json
```

JSON files for metadata

- JavaScript Object Notation
- JSON files as “sidecar”
  - Usually placed next to the file they are supposed to describe

```
{
    "onset": {
        "Description": "Onset of the event",
        "Units": "seconds"
    },
    "duration": {
        "Description": "Duration of the event",
        "Units": "seconds"
    },
    "trial_type": {
        "Description": "Which hand was lifted by the participant.",
        "Levels": {
            "left": "Left",
            "right": "Right"
        }
    }
}
```

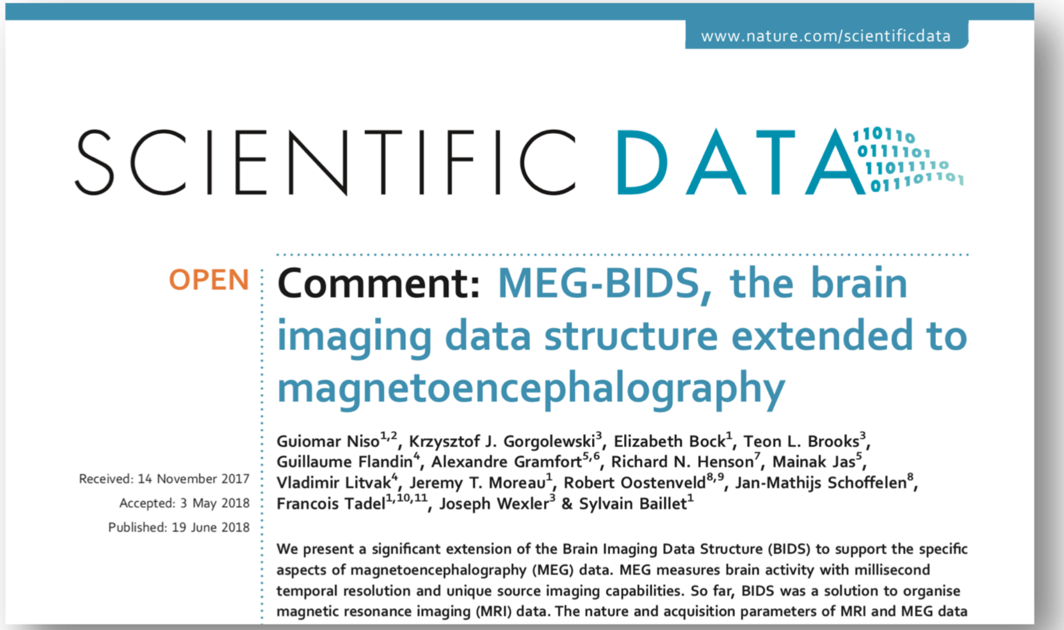
Inheritance Principle

```
.
├── dataset_description.json
├── participants.tsv
├── README
├── sub-01
│   └── eeg
│       ├── sub-01_task-mytask_channels.tsv
│       ├── sub-01_task-mytask_eeg.edf
│       └── sub-01_task-mytask_events.tsv
├── task-mytask_eeg.json
└── task-mytask_events.json
```

- *Any metadata file may be defined at any directory level. The values from the top level are inherited by all lower levels unless they are overridden by a file at the lower level.*

BIDS Extension Proposals (BEPs)

- “Original” BIDS was only for (f)MRI data
- Open standard that is extensible - but backward compatible!
- Examples
  - BEP002: Magnetoencephalography
  - **BEP006: Electroencephalography**
  - **BEP010: intracranial EEG**
  - BEP020: Eye tracking
  - BEP024: CT Scan
  - … currently 26 BEPs
- Process:
  - Small team with informal leaders
  - Discuss and get consensus
  - Submit, peer-review and publish
  - Merge BEP with main specification


www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Received: 14 November 2017
Accepted: 3 May 2018
Published: 19 June 2018

Comment: MEG-BIDS, the brain imaging data structure extended to magnetoencephalography

Guiomar Niso[1,2], Krzysztof J. Gorgolewski[3], Elizabeth Bock[1], Teon L. Brooks[3], Guillaume Flandin[4], Alexandre Gramfort[5,6], Richard N. Henson[7], Mainak Jas[5], Vladimir Litvak[4], Jeremy T. Moreau[1], Robert Oostenveld[8,9], Jan-Mathijs Schoffelen[8], Francois Tadel[1,10,11], Joseph Wexler[3] & Sylvain Baillet[1]

We present a significant extension of the Brain Imaging Data Structure (BIDS) to support the specific aspects of magnetoencephalography (MEG) data. MEG measures brain activity with millisecond temporal resolution and unique source imaging capabilities. So far, BIDS was a solution to organise magnetic resonance imaging (MRI) data. The nature and acquisition parameters of MRI and MEG data

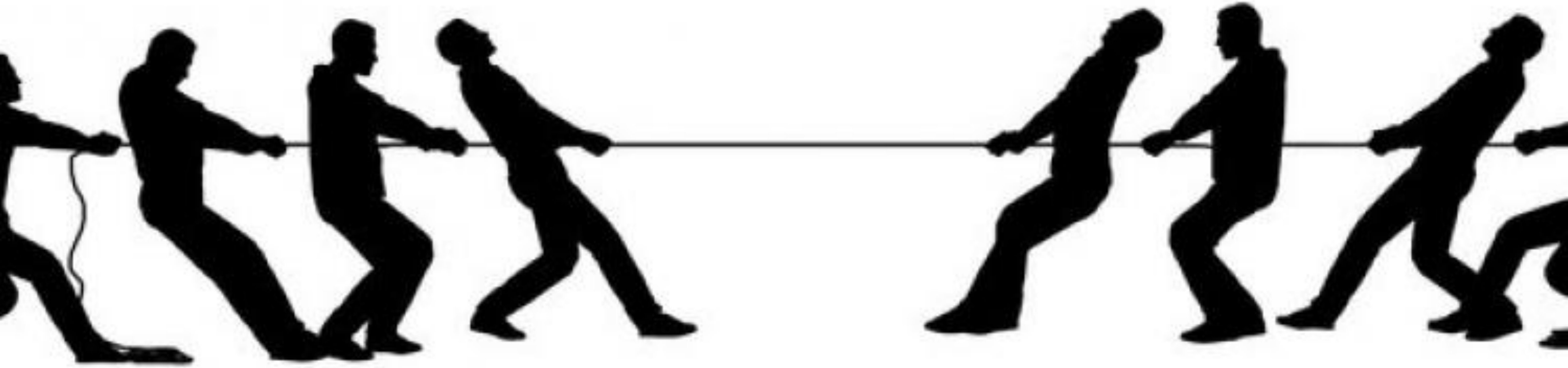
BEP006: Electroencephalography

- Work started in 2017
- Dec 2018: Preprint
- Apr 2019: Merged with main specification
- Publication under review
- In close collaboration with BEP010 for intracranial EEG

Preprint already published at https://psyarxiv.com/63a4y/

http://bids.neuroimaging.io

Open data versus privacy



Personal data

name

address

date of birth

phone number

license plate

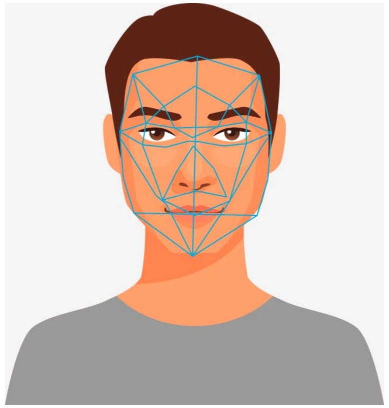
IP address

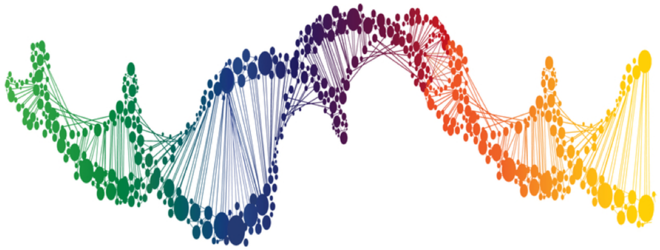
…

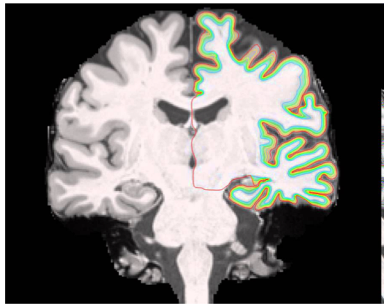
Crime Scene Investigation
http://www.abc.net.au/news/2017-09-19/csi/8960590

(Biometric) data

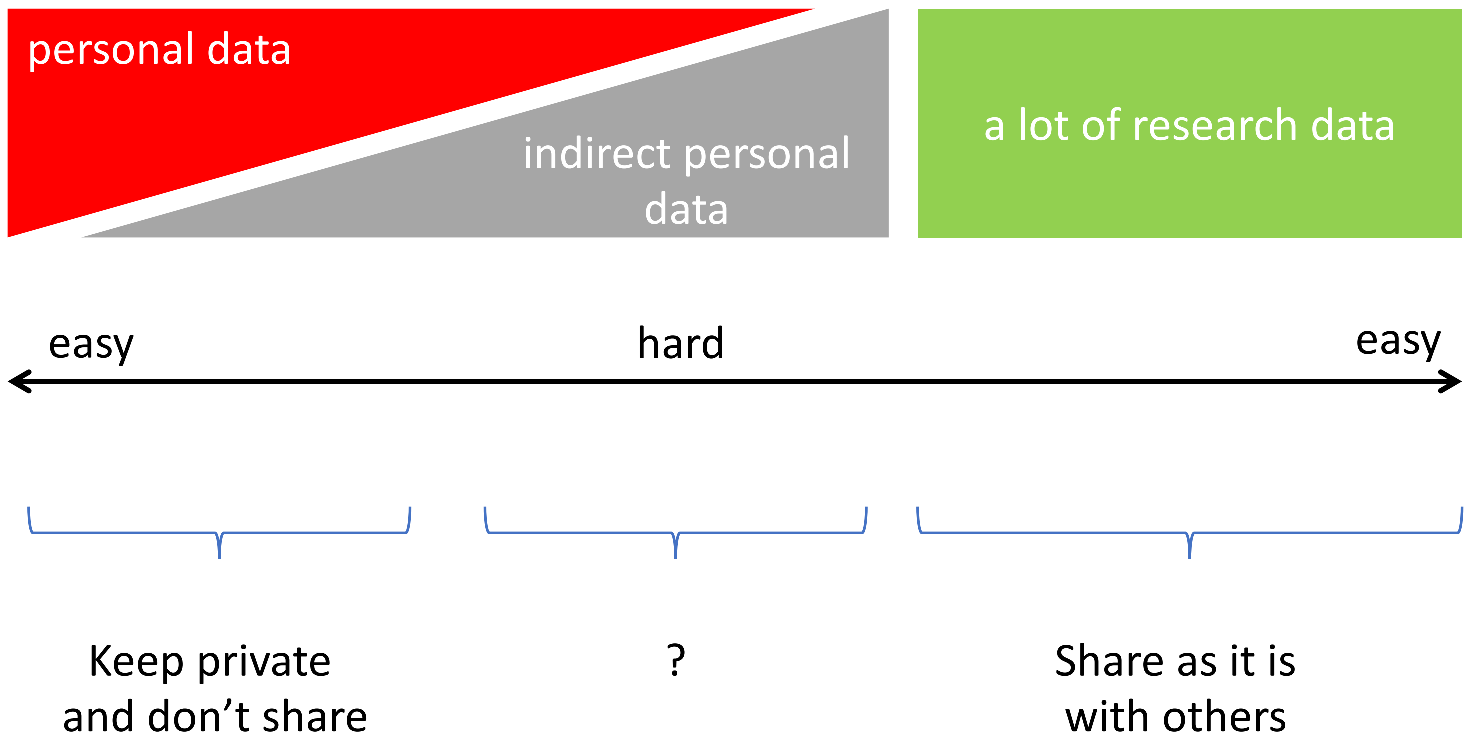
facial details
dental record
fingerprint
genetics
cortical folding pattern
clinical data
gait/movement pattern
responses on questionnaires







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

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

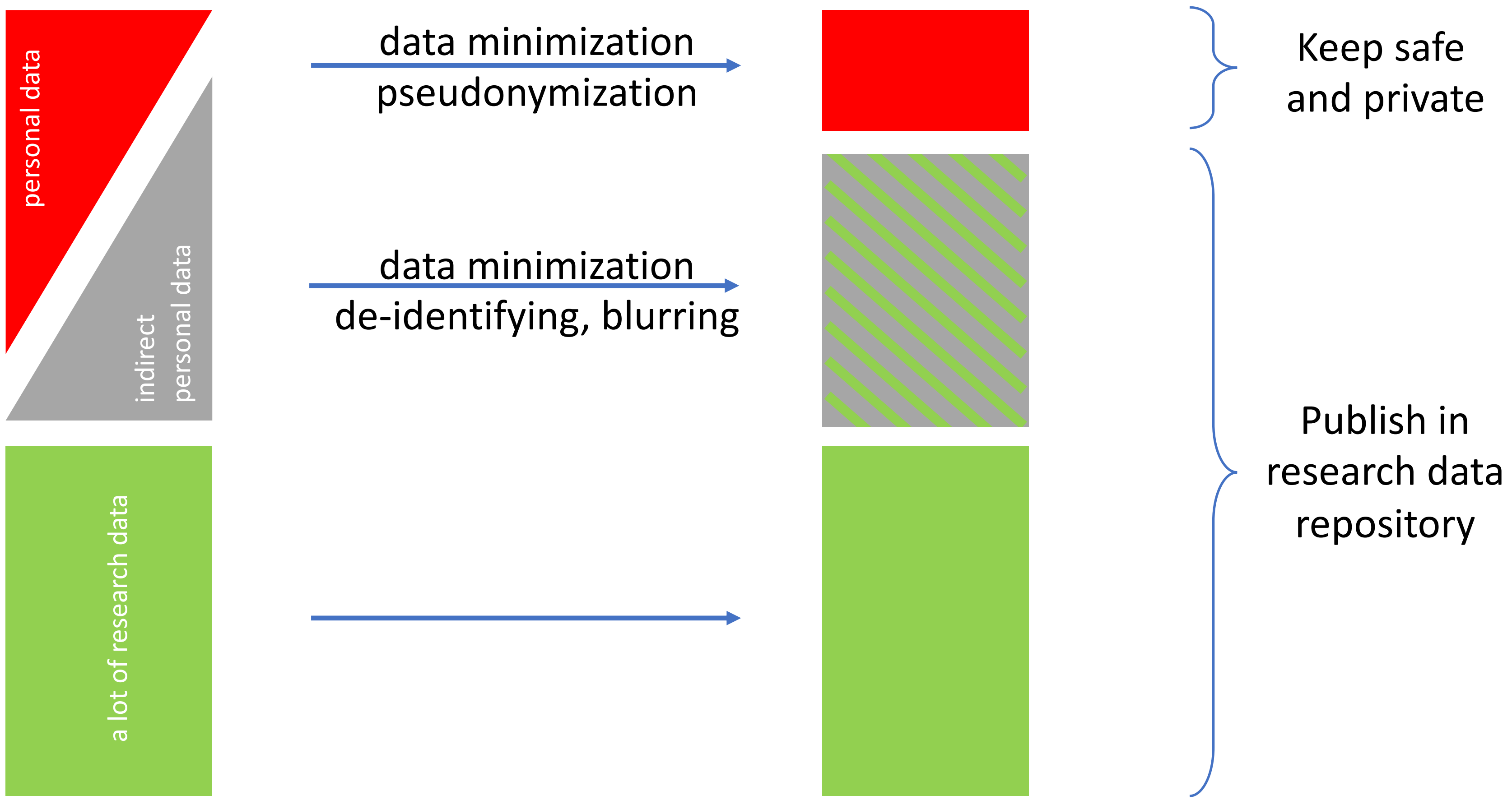
… for example blurring the age of the subject

1 month bins

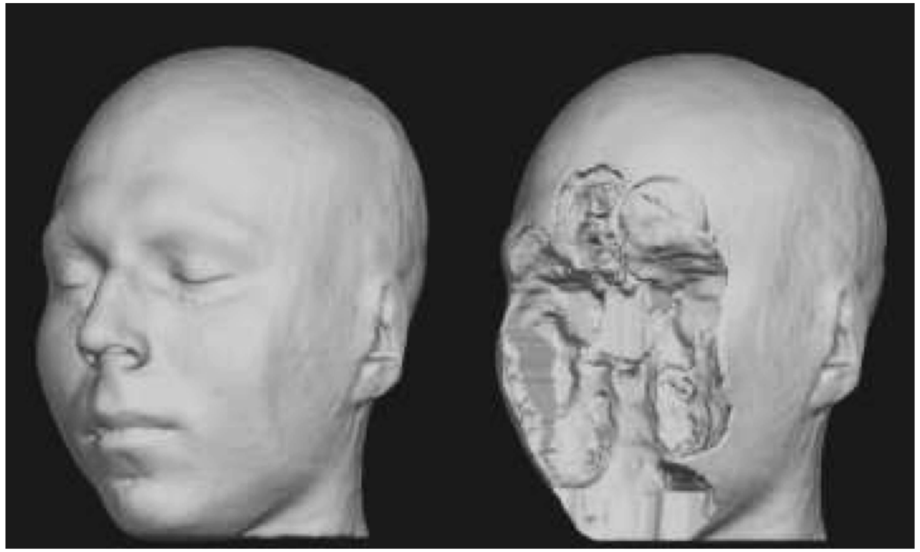
5 or 10 year bins

Personal and research data

Personal and research data



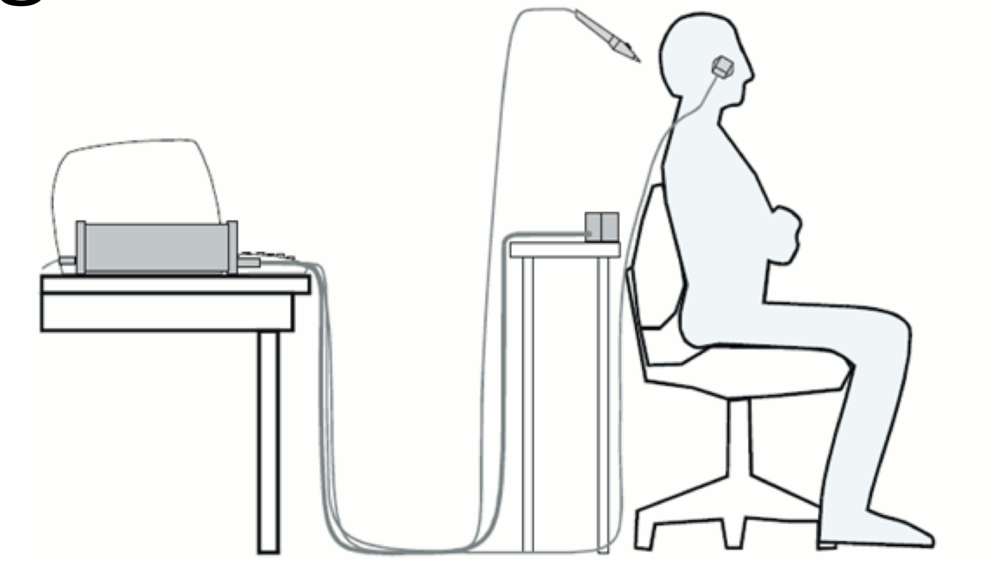
Sharing deidentified imaging data



facial details have been removed,
e.g. using ft_defacevolume or
ft_defacemesh.

nasion is missing, the outline of the
nose is missing, sometimes also the ears
are missing.

Coregistration between MEG/EEG and anatomy

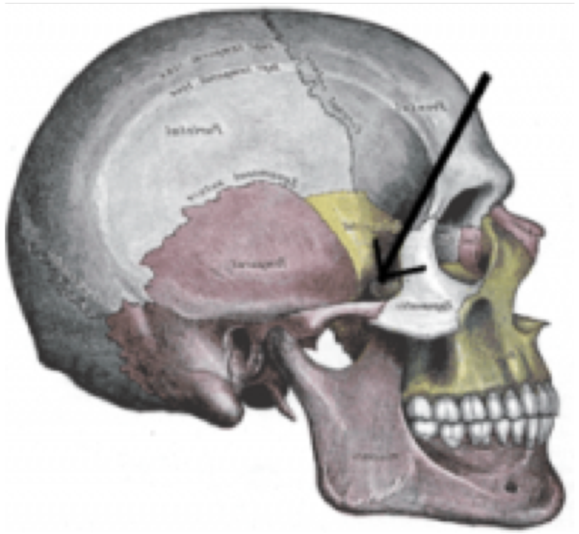


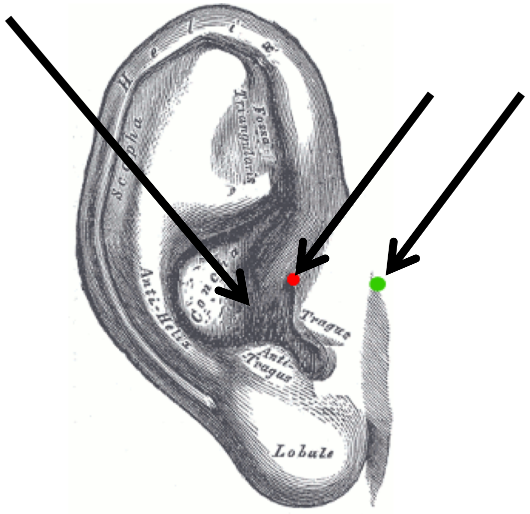
1) anatomical landmarks (lpa, rpa, nas)
2) HPI/HCL coil locations
3) scalp surface points

Lab specific conventions for landmarks and markers

Landmarks:
anatomically recognizable points on a head

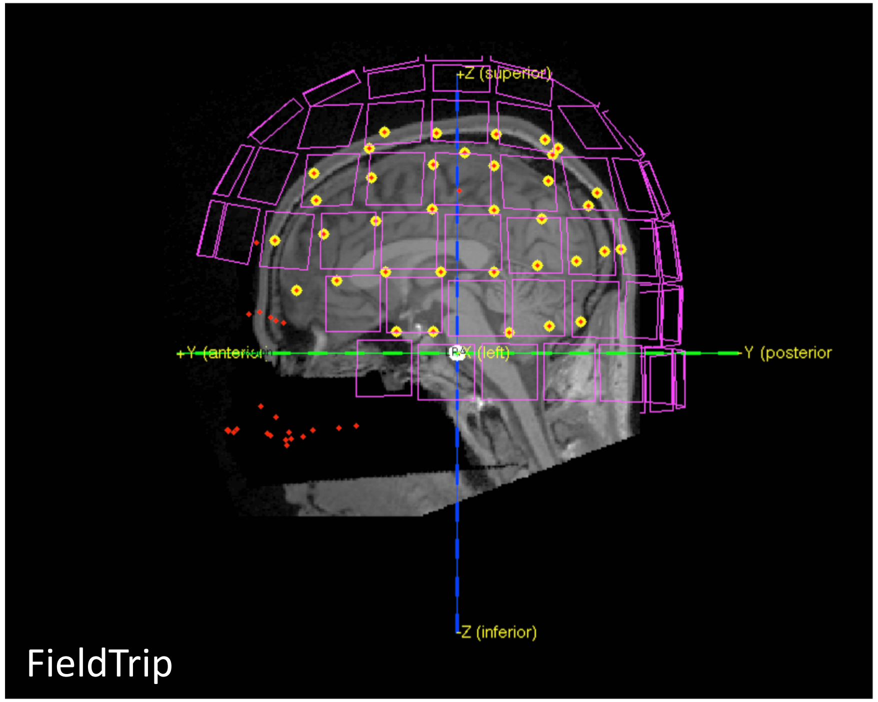
Markers (or fiducials):
points that are visible in multiple modalities, e.g. HPI coils or vitamin E capsules



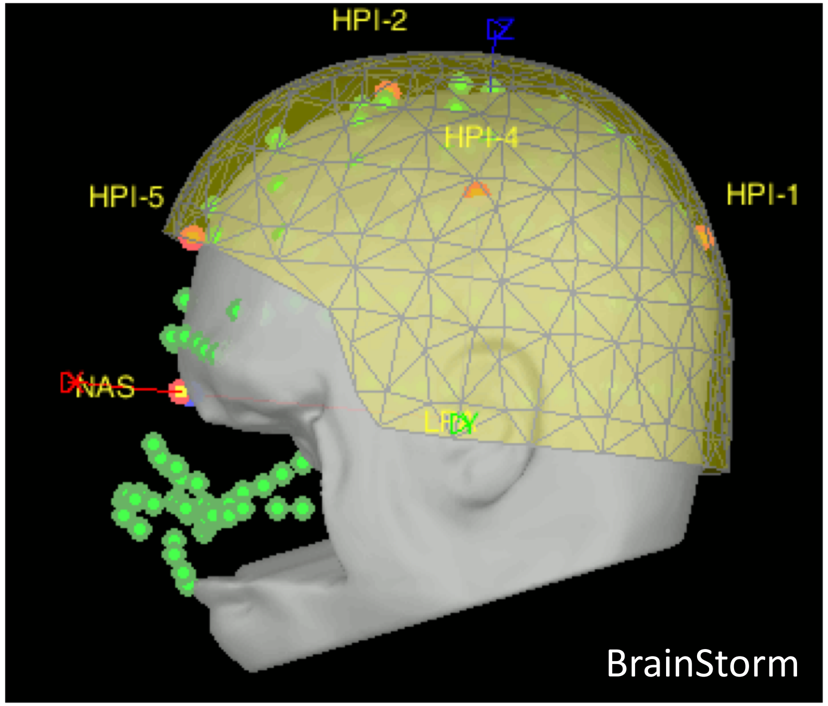


Coregistration

Redo the coregistration using limited data, or trust the coregistration that was provided.



FieldTrip



BrainStorm

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

Sharing of analysis details (code)

- Upload to “data” repositories
  - Advantage is that your code gets a DOI
- Manage on a code development website
  - Github, Gitlab, Bitbucket

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
```

Version Control with Git

Atlassian

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

software carpentry

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