RESEARCH DATA MANAGEMENT

👩‍💻🧠👨‍💻

FOR NEUROIMAGERS

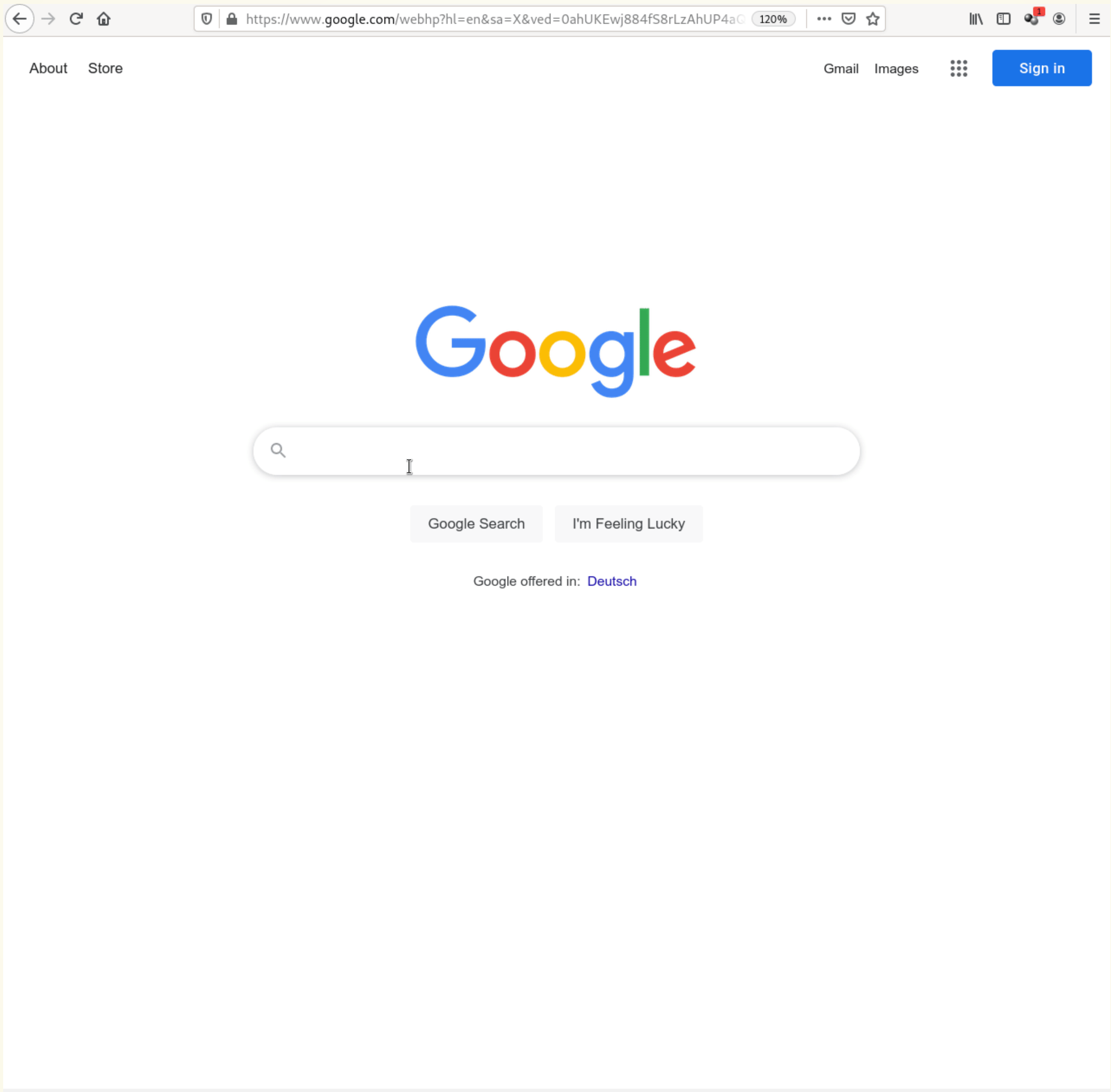
Adina Wagner
@adswa@mas.to 🐦 @AdinaKrik

Psychoinformatics lab,
Institute of Neuroscience and Medicine (INM-7)
Research Center Jülich
Institute of Experimental Psychology, HHU Düsseldorf

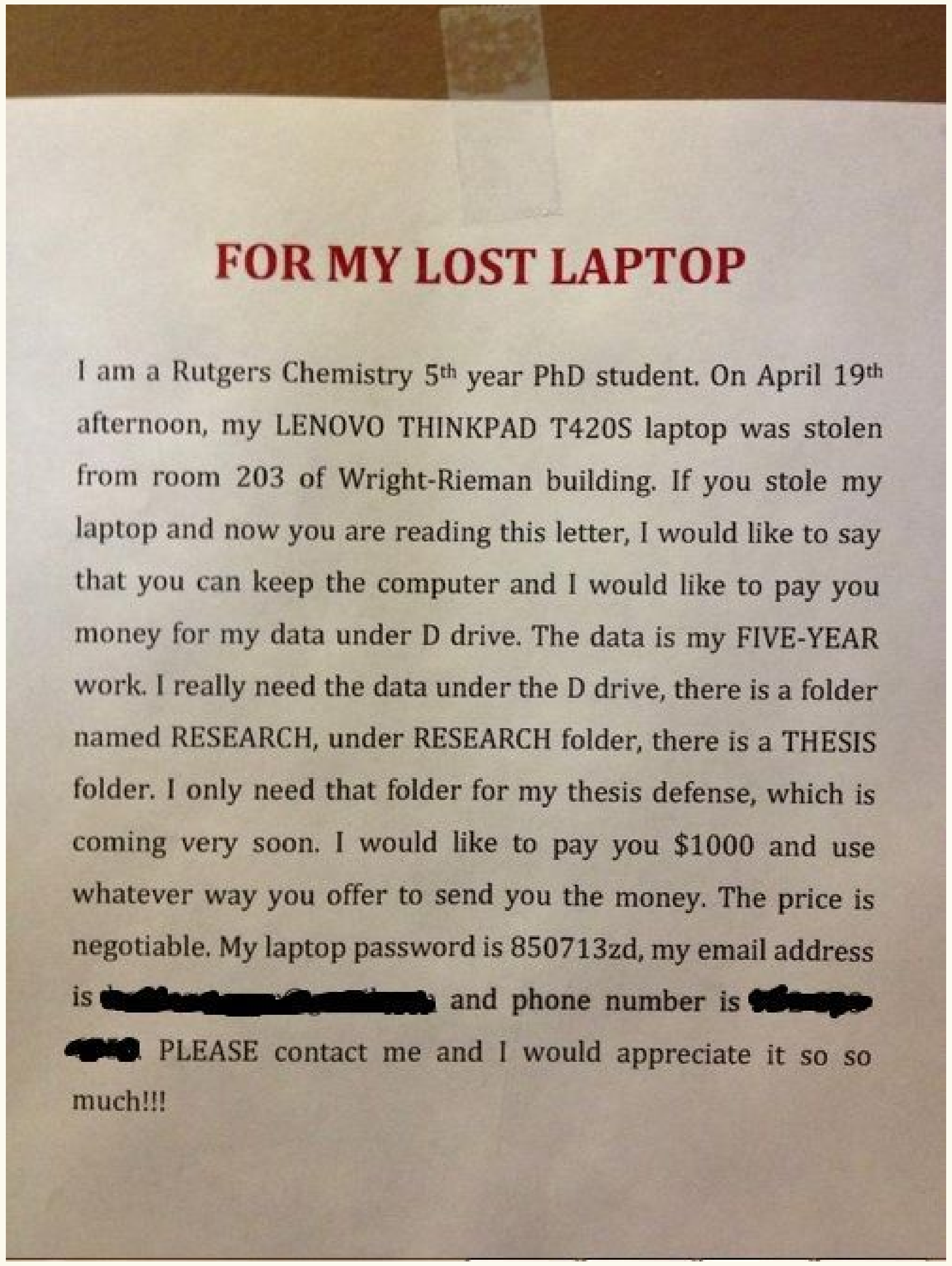
Slides: DOI 10.5281/zenodo.7419377 (Scan the QR code)

RESEARCH DATA MANAGEMENT?

RESEARCH DATA MANAGEMENT?


https://www.google.com/webhp?hl=en&sa=X&ved=0ahUKEwj884fS8rLzAhUP4a
120%
About
Store
Gmail
Images
Sign in
Google
Google Search
I'm Feeling Lucky
Google offered in: Deutsch

RESEARCH DATA MANAGEMENT?

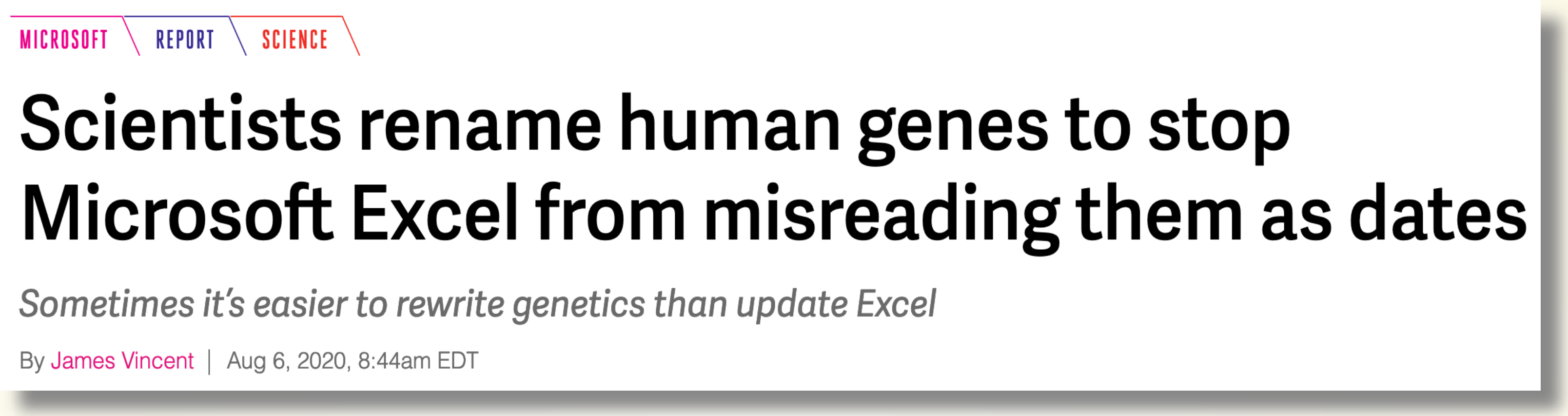


FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you $1000 and use whatever way you offer to send you the money. The price is negotiable. My laptop password is 850713zd, my email address is [redacted] and phone number is [redacted]. PLEASE contact me and I would appreciate it so so much!!!

Image credit: adapted from https://dribbble.com/shots/3090048-Front-end-vs-Back-end

RESEARCH DATA MANAGEMENT?

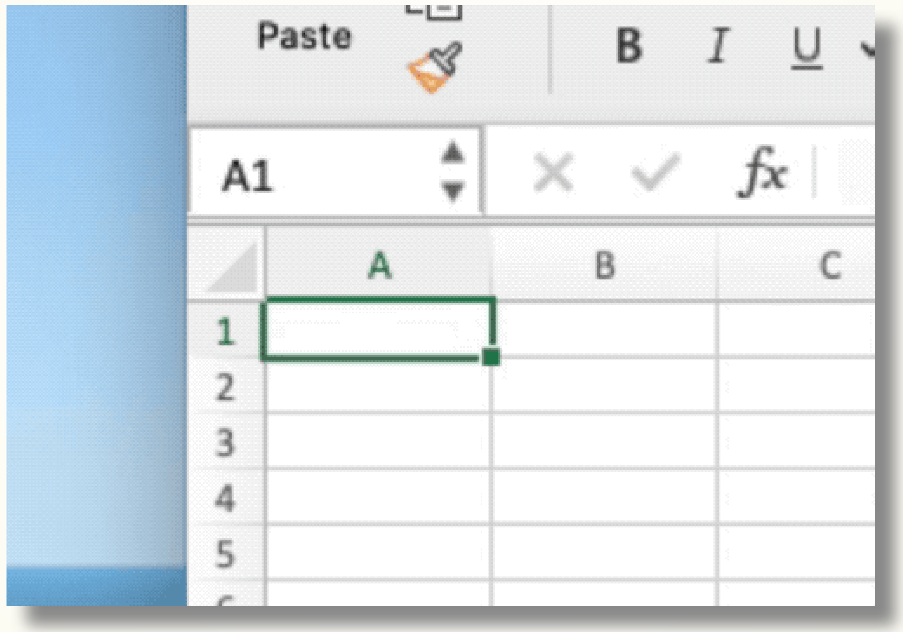


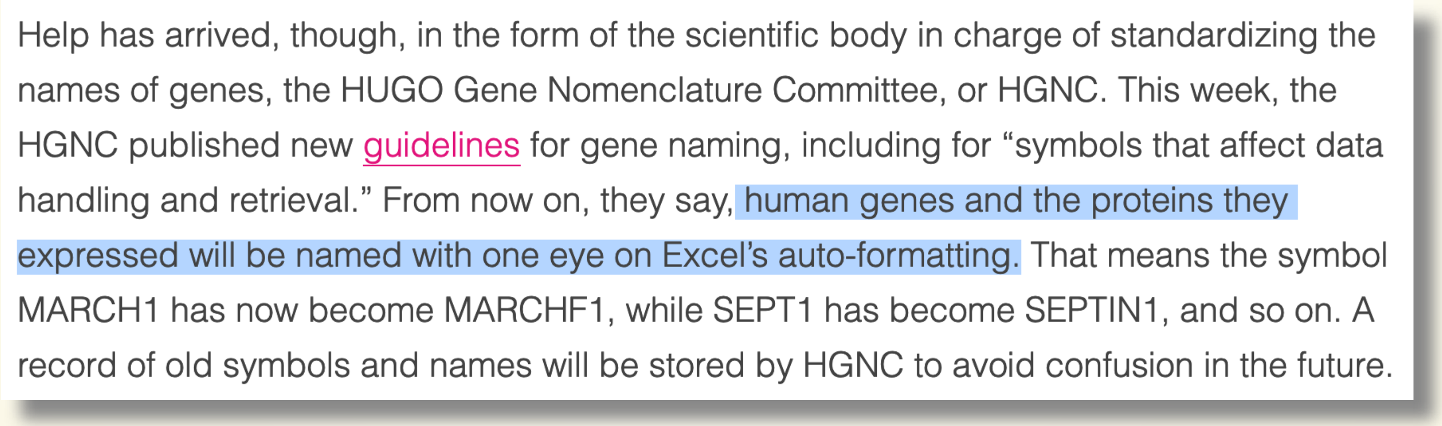
MICROSOFT \ REPORT \ SCIENCE

Scientists rename human genes to stop Microsoft Excel from misreading them as dates

Sometimes it's easier to rewrite genetics than update Excel

By James Vincent | Aug 6, 2020, 8:44am EDT





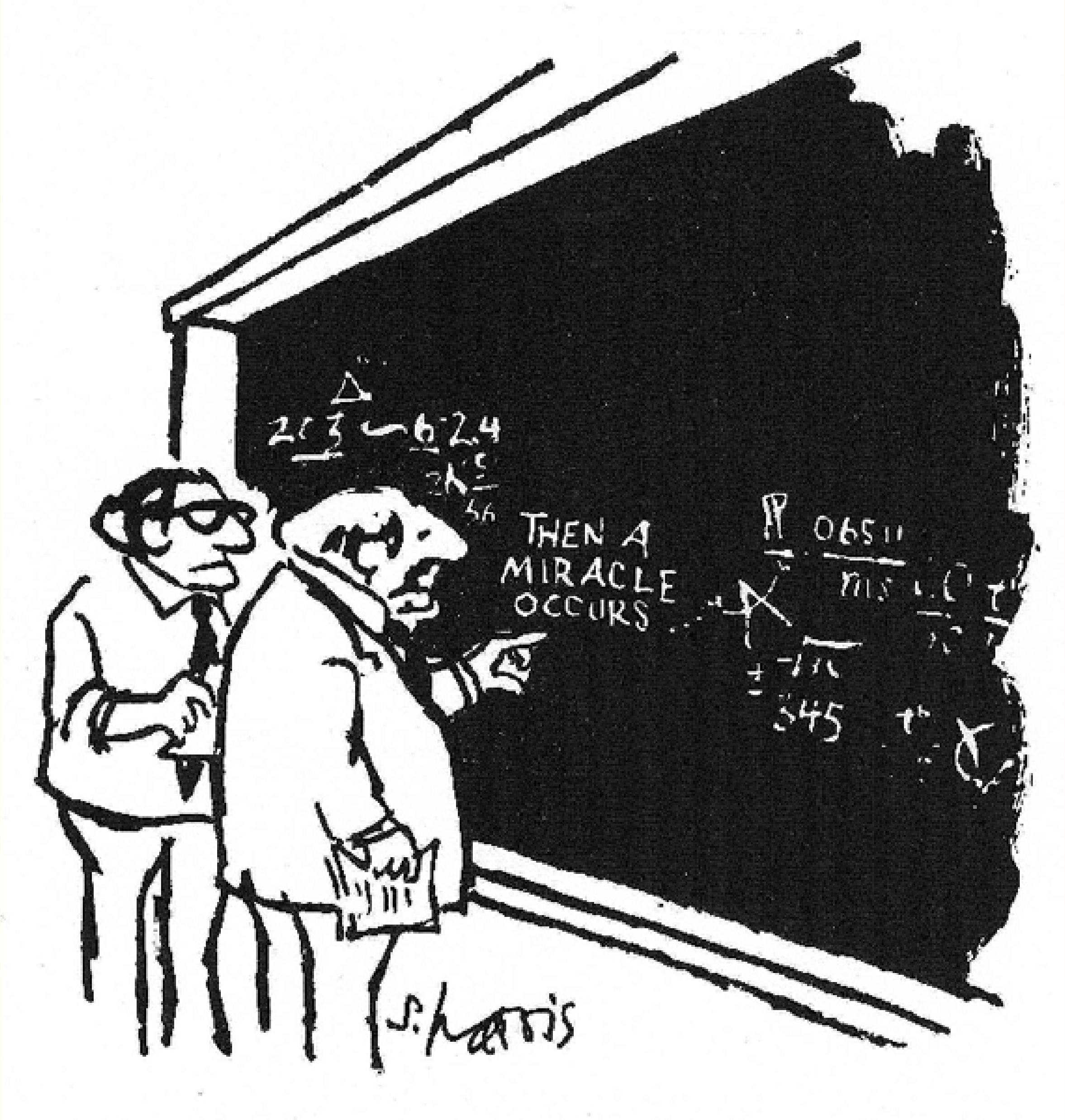
Help has arrived, though, in the form of the scientific body in charge of standardizing the names of genes, the HUGO Gene Nomenclature Committee, or HGNC. This week, the HGNC published new guidelines for gene naming, including for "symbols that affect data handling and retrieval." From now on, they say, human genes and the proteins they expressed will be named with one eye on Excel's auto-formatting. That means the symbol MARCH1 has now become MARCHF1, while SEPT1 has become SEPTIN1, and so on. A record of old symbols and names will be stored by HGNC to avoid confusion in the future.

www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misreading-dates

RDM - FOR WHOM?

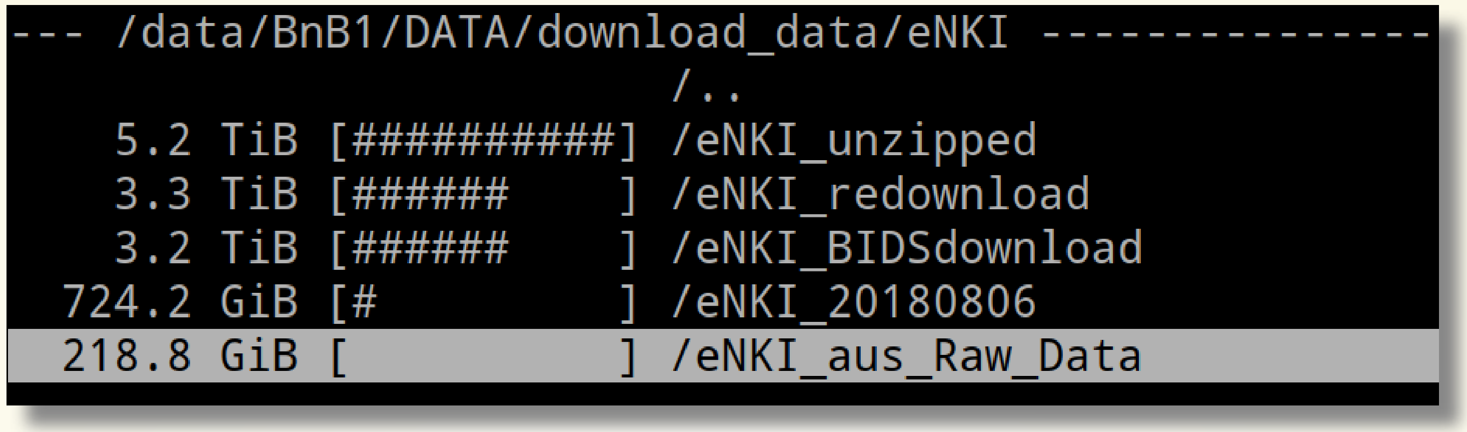
Funders & publishers require it

RDM - FOR WHOM?



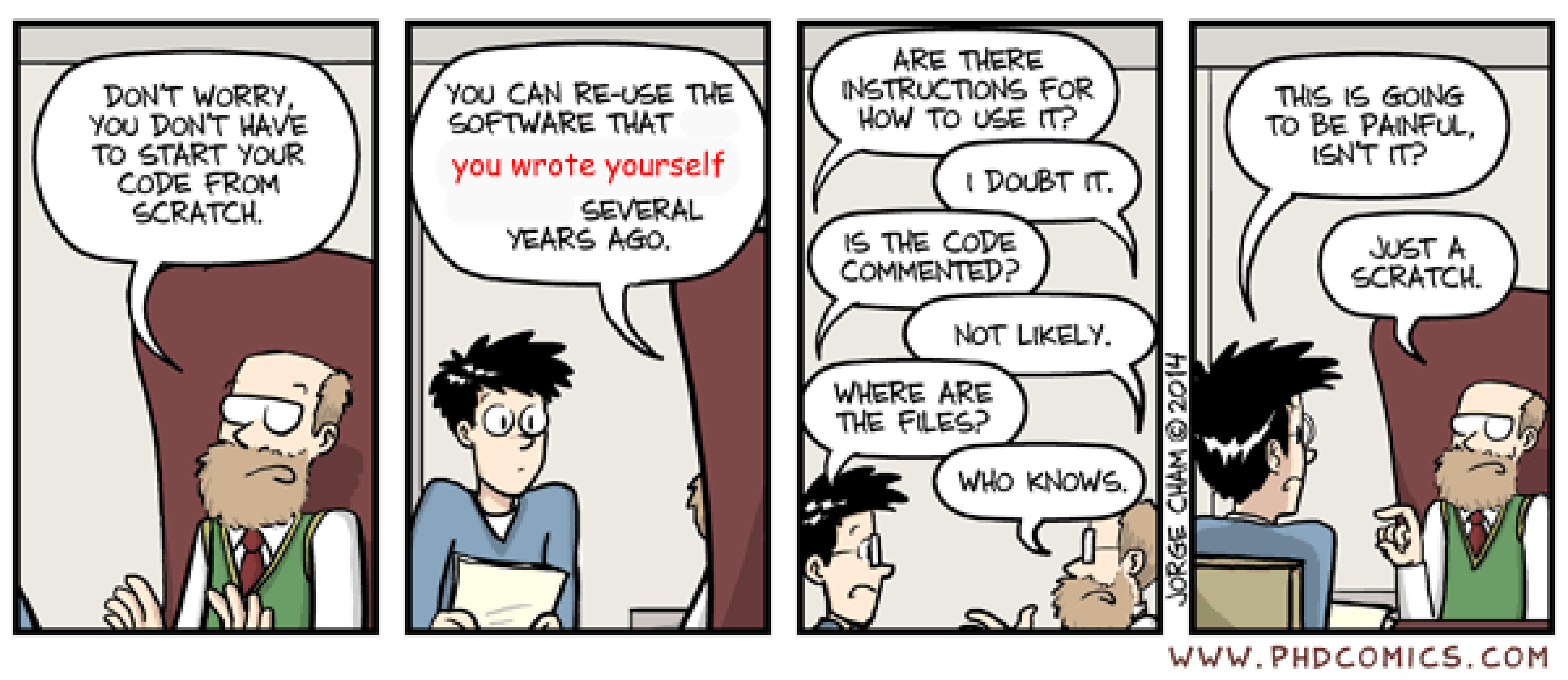
"I think you should be more explicit here in step two."

RDM - FOR WHOM?

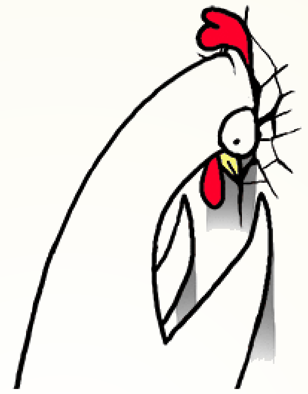


```
--- /data/BnB1/DATA/download_data/eNKI ---------------
                              /..
    5.2 TiB [##########] /eNKI_unzipped
    3.3 TiB [######    ] /eNKI_redownload
    3.2 TiB [######    ] /eNKI_BIDSdownload
  724.2 GiB [#         ] /eNKI_20180806
  218.8 GiB [          ] /eNKI_aus_Raw_Data
```

RDM - FOR WHOM?


DON'T WORRY, YOU DON'T HAVE TO START YOUR CODE FROM SCRATCH.
YOU CAN RE-USE THE SOFTWARE THAT you wrote yourself SEVERAL YEARS AGO.
ARE THERE INSTRUCTIONS FOR HOW TO USE IT?
I DOUBT IT.
IS THE CODE COMMENTED?
NOT LIKELY.
WHERE ARE THE FILES?
WHO KNOWS.
JORGE CHAM © 2014
THIS IS GOING TO BE PAINFUL, ISN'T IT?
JUST A SCRATCH.
WWW.PHDCOMICS.COM

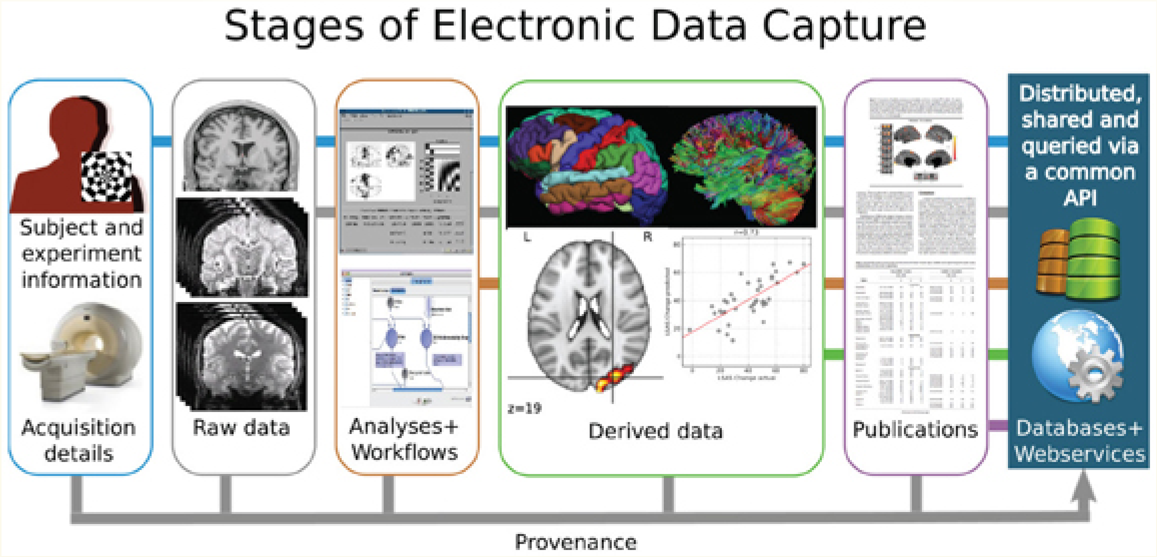
RDM - FOR WHOM?



RDM IN NEUROIMAGING

Some peculiarities of our field...

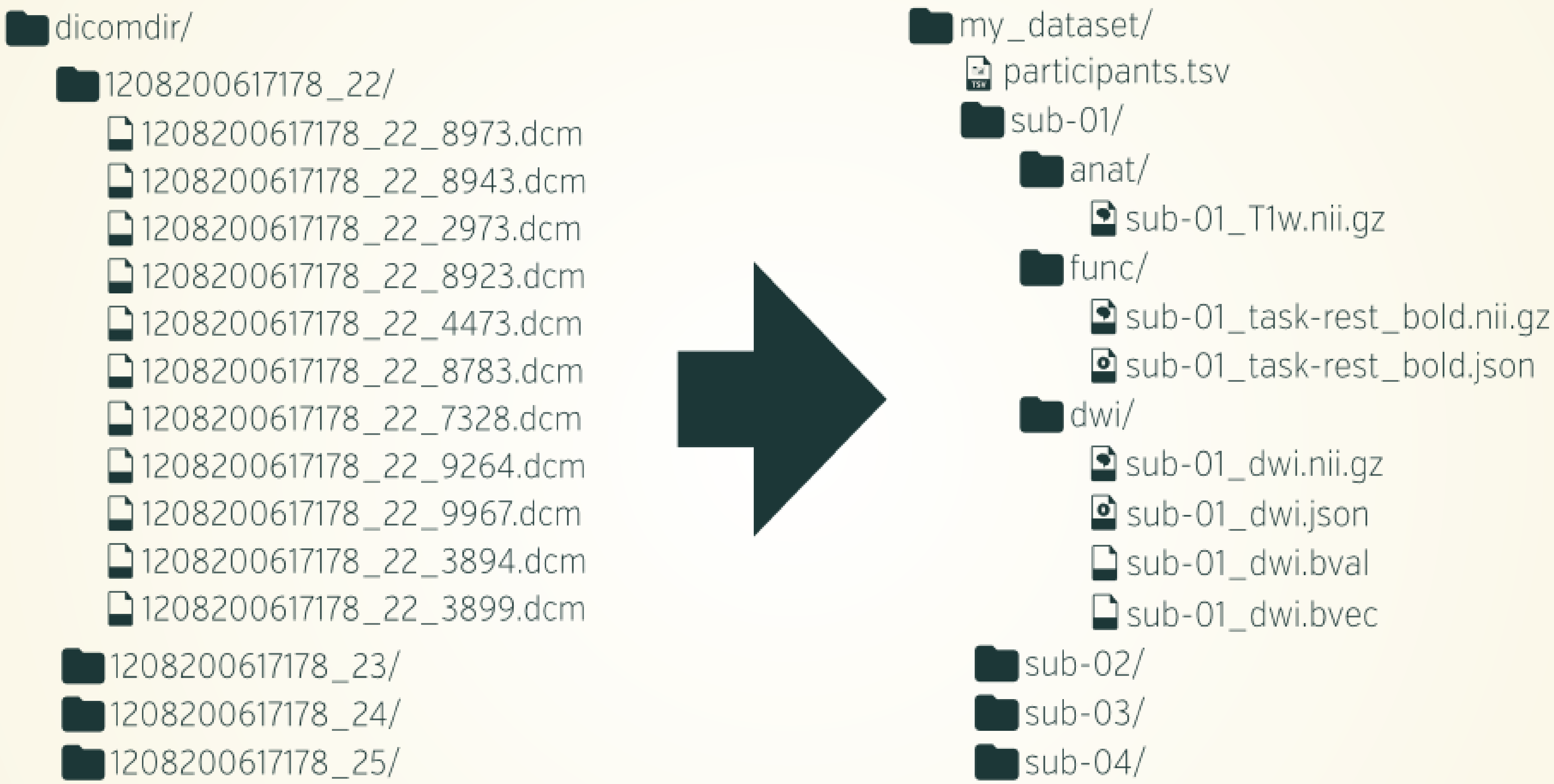
- Depending on acquisition hardware and analysis software, some data are in proprietary formats (e.g., Neuromag, brain voyager, brain vision)
- Depending on field, data can be sizeable (e.g., (f)MRI, CT, EEG, PET, MEG)
- Heterogenous data from complex acquisitions with multiple data channels and modalities
- Datasets are getting bigger and bigger (Bzdok & Yeo, 2017), e.g. multi-modal imaging, behavioral + genetics data in HCP (humanconnectome.org) or UKBiobank (ukbiobank.ac.uk/)
- Some data fall under General Data Protection Regulation (GDPR)
- Complex, multi-stepped analyses


Stages of Electronic Data Capture
Subject and experiment information
Acquisition details
Raw data
Analyses+ Workflows
Derived data
Publications
Distributed, shared and queried via a common API
Databases+ Webservices
Provenance

... make RDM more difficult, but also more relevant

Image credit: BIDS, CC-BY

BIDS
BRAIN IMAGING DATA STRUCTURE



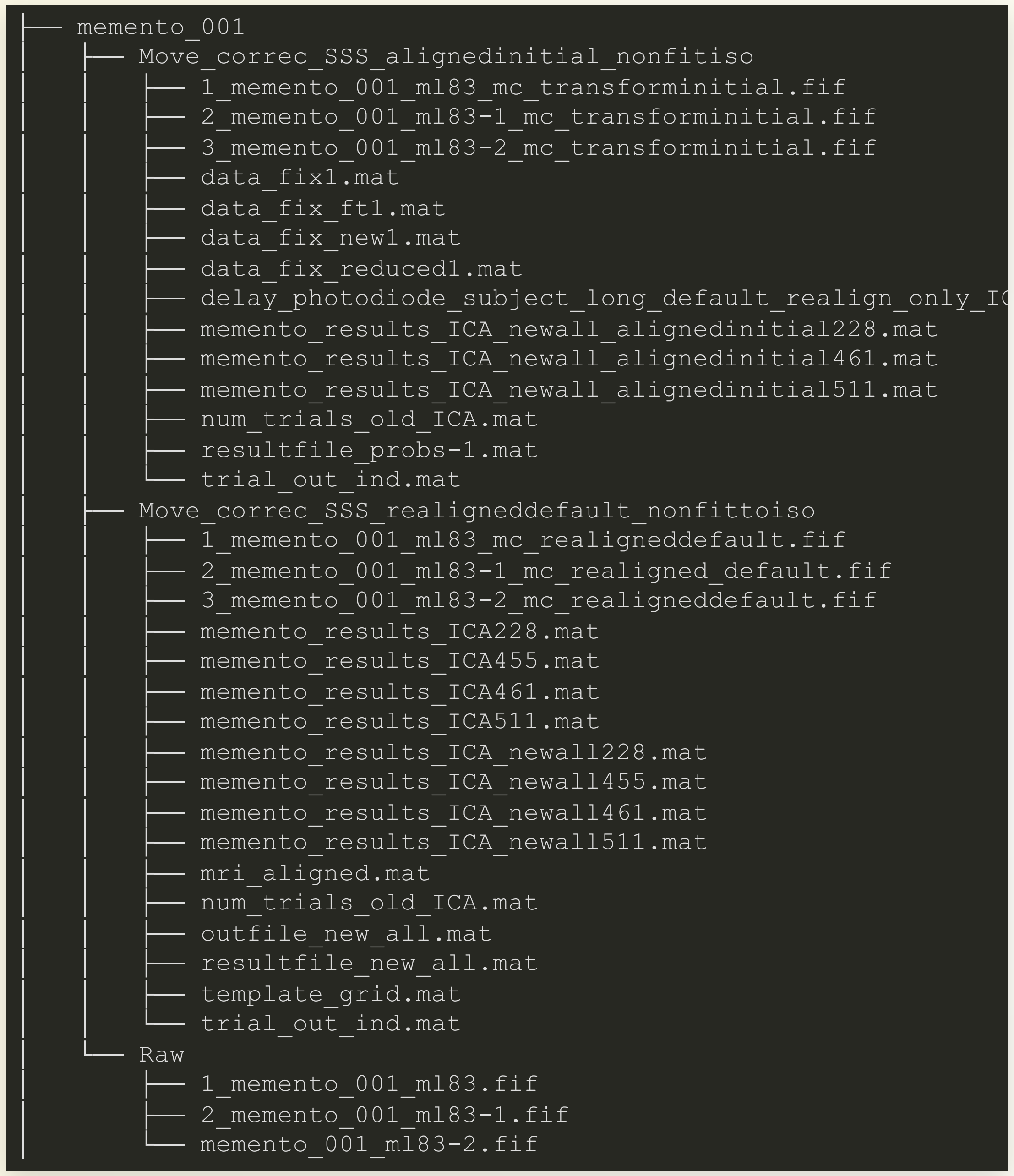
```
dicomdir/
    1208200617178_22/
        1208200617178_22_8973.dcm
        1208200617178_22_8943.dcm
        1208200617178_22_2973.dcm
        1208200617178_22_8923.dcm
        1208200617178_22_4473.dcm
        1208200617178_22_8783.dcm
        1208200617178_22_7328.dcm
        1208200617178_22_9264.dcm
        1208200617178_22_9967.dcm
        1208200617178_22_3894.dcm
        1208200617178_22_3899.dcm
    1208200617178_23/
    1208200617178_24/
    1208200617178_25/
```

```
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/
            sub-01_dwi.nii.gz
            sub-01_dwi.json
            sub-01_dwi.bval
            sub-01_dwi.bvec
    sub-02/
    sub-03/
    sub-04/
```

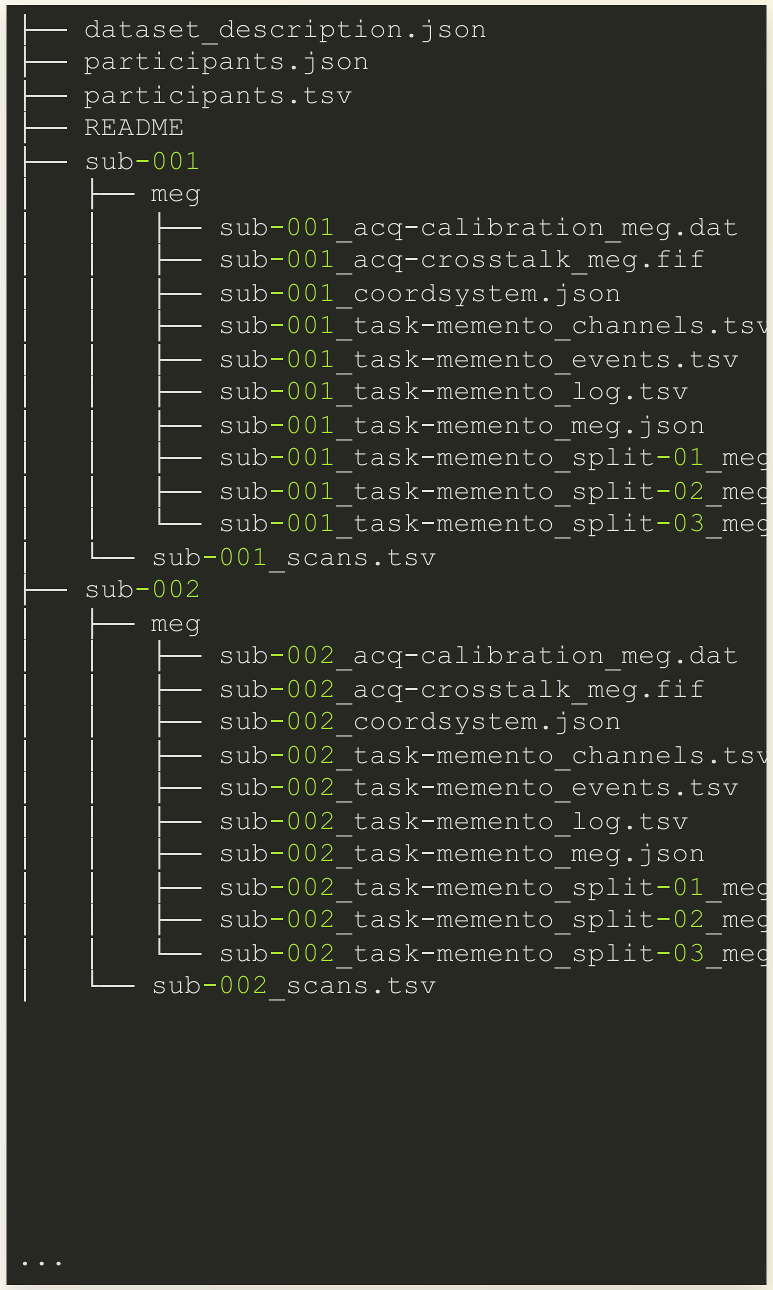
BIDS is an established and evolving community standard for a multitude of neuroimaging data (MRI, (i)EEG, MEG, ...). It defines a data organization, naming schemes for files, and meta data descriptors.

bids.neuroimaging.io

BIDS
BRAIN IMAGING DATA STRUCTURE



```
├── memento_001
│   ├── Move_correc_SSS_alignedinitial_nonfitiso
│   │   ├── 1_memento_001_ml83_mc_transforminitial.fif
│   │   ├── 2_memento_001_ml83-1_mc_transforminitial.fif
│   │   ├── 3_memento_001_ml83-2_mc_transforminitial.fif
│   │   ├── data_fix1.mat
│   │   ├── data_fix_ft1.mat
│   │   ├── data_fix_new1.mat
│   │   ├── data_fix_reduced1.mat
│   │   ├── delay_photodiode_subject_long_default_realign_only_IC
│   │   ├── memento_results_ICA_newall_alignedinitial228.mat
│   │   ├── memento_results_ICA_newall_alignedinitial461.mat
│   │   ├── memento_results_ICA_newall_alignedinitial511.mat
│   │   ├── num_trials_old_ICA.mat
│   │   ├── resultfile_probs-1.mat
│   │   └── trial_out_ind.mat
│   ├── Move_correc_SSS_realigneddefault_nonfittoiso
│   │   ├── 1_memento_001_ml83_mc_realigneddefault.fif
│   │   ├── 2_memento_001_ml83-1_mc_realigned_default.fif
│   │   ├── 3_memento_001_ml83-2_mc_realigneddefault.fif
│   │   ├── memento_results_ICA228.mat
│   │   ├── memento_results_ICA455.mat
│   │   ├── memento_results_ICA461.mat
│   │   ├── memento_results_ICA511.mat
│   │   ├── memento_results_ICA_newall228.mat
│   │   ├── memento_results_ICA_newall455.mat
│   │   ├── memento_results_ICA_newall461.mat
│   │   ├── memento_results_ICA_newall511.mat
│   │   ├── mri_aligned.mat
│   │   ├── num_trials_old_ICA.mat
│   │   ├── outfile_new_all.mat
│   │   ├── resultfile_new_all.mat
│   │   ├── template_grid.mat
│   │   └── trial_out_ind.mat
│   └── Raw
│       ├── 1_memento_001_ml83.fif
│       ├── 2_memento_001_ml83-1.fif
│       └── memento_001_ml83-2.fif
│
```

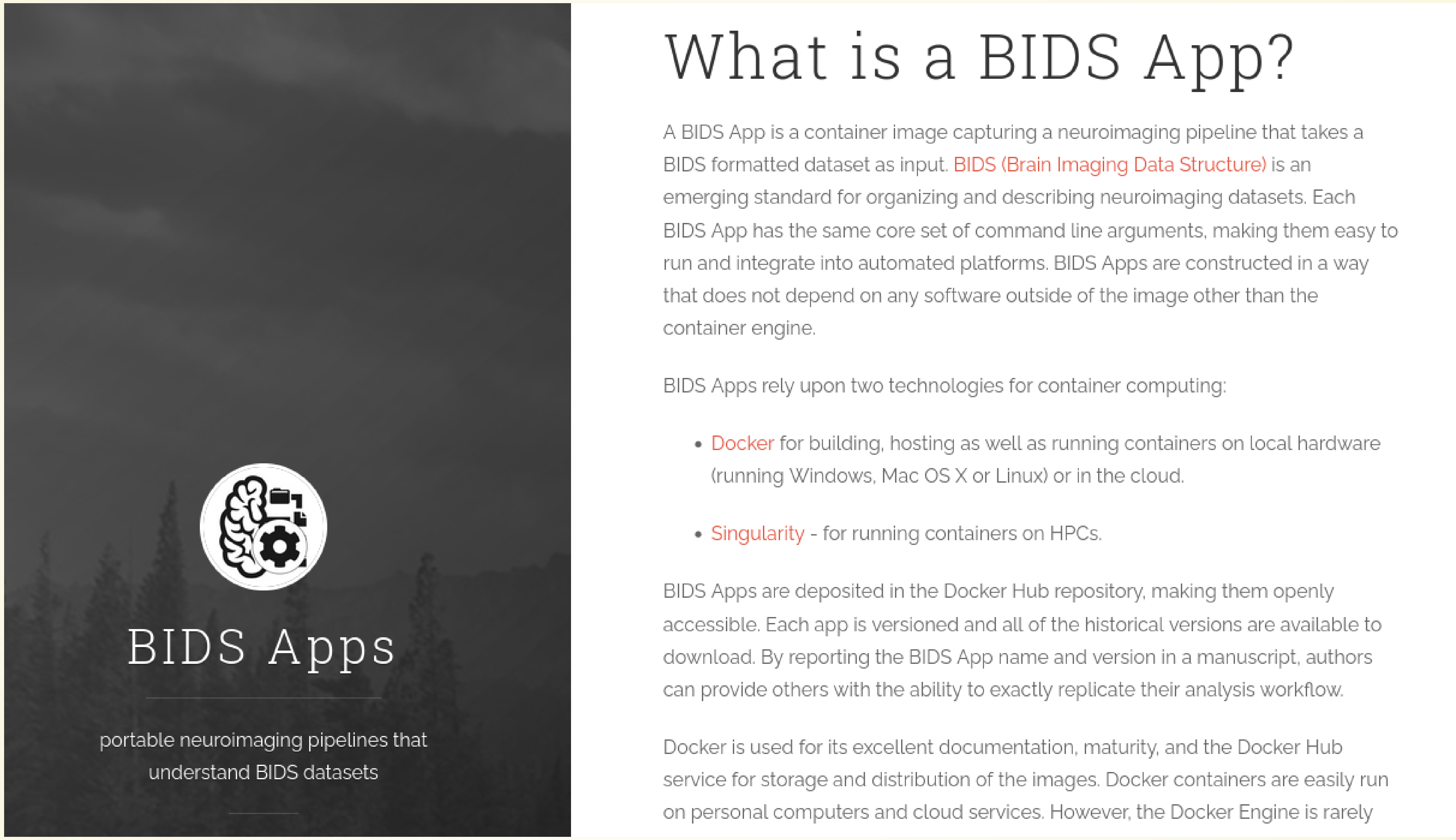


```
├── dataset_description.json
├── participants.json
├── participants.tsv
├── README
├── sub-001
│   ├── meg
│   │   ├── sub-001_acq-calibration_meg.dat
│   │   ├── sub-001_acq-crosstalk_meg.fif
│   │   ├── sub-001_coordsystem.json
│   │   ├── sub-001_task-memento_channels.tsv
│   │   ├── sub-001_task-memento_events.tsv
│   │   ├── sub-001_task-memento_log.tsv
│   │   ├── sub-001_task-memento_meg.json
│   │   ├── sub-001_task-memento_split-01_meg
│   │   ├── sub-001_task-memento_split-02_meg
│   │   └── sub-001_task-memento_split-03_meg
│   └── sub-001_scans.tsv
├── sub-002
│   ├── meg
│   │   ├── sub-002_acq-calibration_meg.dat
│   │   ├── sub-002_acq-crosstalk_meg.fif
│   │   ├── sub-002_coordsystem.json
│   │   ├── sub-002_task-memento_channels.tsv
│   │   ├── sub-002_task-memento_events.tsv
│   │   ├── sub-002_task-memento_log.tsv
│   │   ├── sub-002_task-memento_meg.json
│   │   ├── sub-002_task-memento_split-01_meg
│   │   ├── sub-002_task-memento_split-02_meg
│   │   └── sub-002_task-memento_split-03_meg
│   └── sub-002_scans.tsv



...
```

BIDS is an established and evolving community standard for a multitude of neuroimaging data (MRI, (i)EEG, MEG, ...). It defines a data organization, naming schemes for files, and meta data descriptors.

bids.neuroimaging.io

BIDS
BRAIN IMAGING DATA STRUCTURE



BIDS Apps

portable neuroimaging pipelines that understand BIDS datasets

What is a BIDS App?

A BIDS App is a container image capturing a neuroimaging pipeline that takes a BIDS formatted dataset as input. BIDS (Brain Imaging Data Structure) is an emerging standard for organizing and describing neuroimaging datasets. Each BIDS App has the same core set of command line arguments, making them easy to run and integrate into automated platforms. BIDS Apps are constructed in a way that does not depend on any software outside of the image other than the container engine.

BIDS Apps rely upon two technologies for container computing:

- Docker for building, hosting as well as running containers on local hardware (running Windows, Mac OS X or Linux) or in the cloud.
- Singularity - for running containers on HPCs.

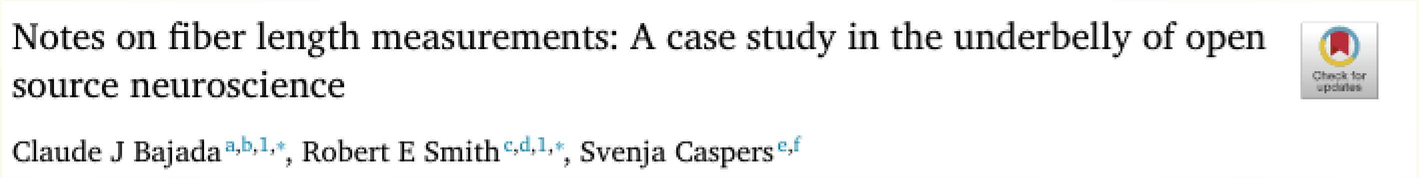
BIDS Apps are deposited in the Docker Hub repository, making them openly accessible. Each app is versioned and all of the historical versions are available to download. By reporting the BIDS App name and version in a manuscript, authors can provide others with the ability to exactly replicate their analysis workflow.

Docker is used for its excellent documentation, maturity, and the Docker Hub service for storage and distribution of the images. Docker containers are easily run on personal computers and cloud services. However, the Docker Engine is rarely

BIDS is an established and evolving community standard for a multitude of neuroimaging data (MRI, (i)EEG, MEG, ...). It defines a data organization, naming schemes for files, and meta data descriptors.
bids.neuroimaging.io

OPEN {SOFTWARE,STANDARDS}

Ok
YOU, 6 MONTHS LATER
ERR
ERR



Notes on fiber length measurements: A case study in the underbelly of open source neuroscience

Claude J Bajada[a,b,1,*], Robert E Smith[c,d,1,*], Svenja Caspers[e,f]

doi.org/10.1016/j.neuroimage.2022.119738

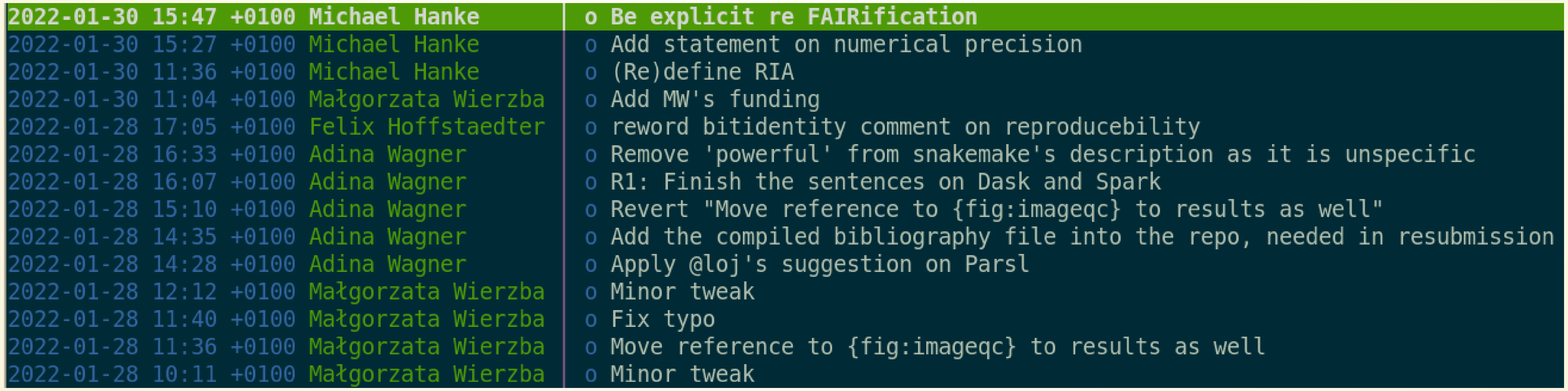
- remove accessibility barriers
- allow transparent digital provenance

VERSION CONTROL

Image credit: CC-BY Scriberia & The Turing Way

Version control

- keep things organized
- keep track of changes
- revert changes or go back to previous states
- collect and share digital provenance
- industry standard: Git

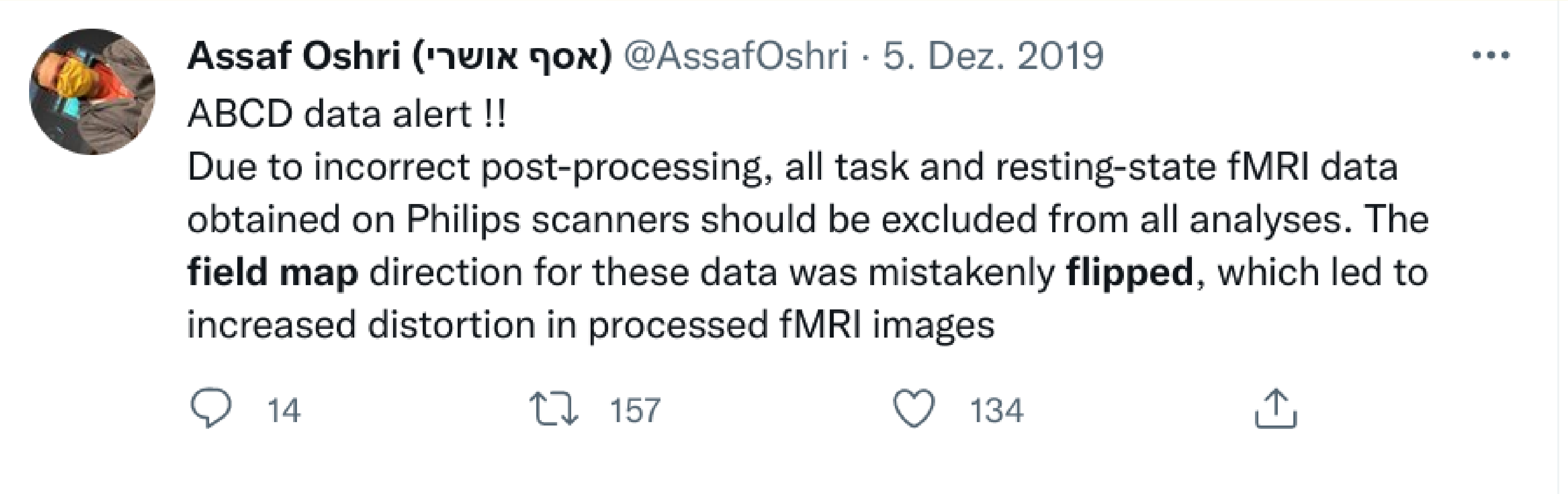


```
2022-01-30 15:47 +0100 Michael Hanke        o Be explicit re FAIRification
2022-01-30 15:27 +0100 Michael Hanke        o Add statement on numerical precision
2022-01-30 11:36 +0100 Michael Hanke        o (Re)define RIA
2022-01-30 11:04 +0100 Małgorzata Wierzba   o Add MW's funding
2022-01-28 17:05 +0100 Felix Hoffstaedter   o reword bitidentity comment on reproducebility
2022-01-28 16:33 +0100 Adina Wagner         o Remove 'powerful' from snakemake's description as it is unspecific
2022-01-28 16:07 +0100 Adina Wagner         o R1: Finish the sentences on Dask and Spark
2022-01-28 15:10 +0100 Adina Wagner         o Revert "Move reference to {fig:imageqc} to results as well"
2022-01-28 14:35 +0100 Adina Wagner         o Add the compiled bibliography file into the repo, needed in resubmission
2022-01-28 14:28 +0100 Adina Wagner         o Apply @loj's suggestion on Parsl
2022-01-28 12:12 +0100 Małgorzata Wierzba   o Minor tweak
2022-01-28 11:40 +0100 Małgorzata Wierzba   o Fix typo
2022-01-28 11:36 +0100 Małgorzata Wierzba   o Move reference to {fig:imageqc} to results as well
2022-01-28 10:11 +0100 Małgorzata Wierzba   o Minor tweak
```

The building blocks of a scientific result are rarely static

Data changes

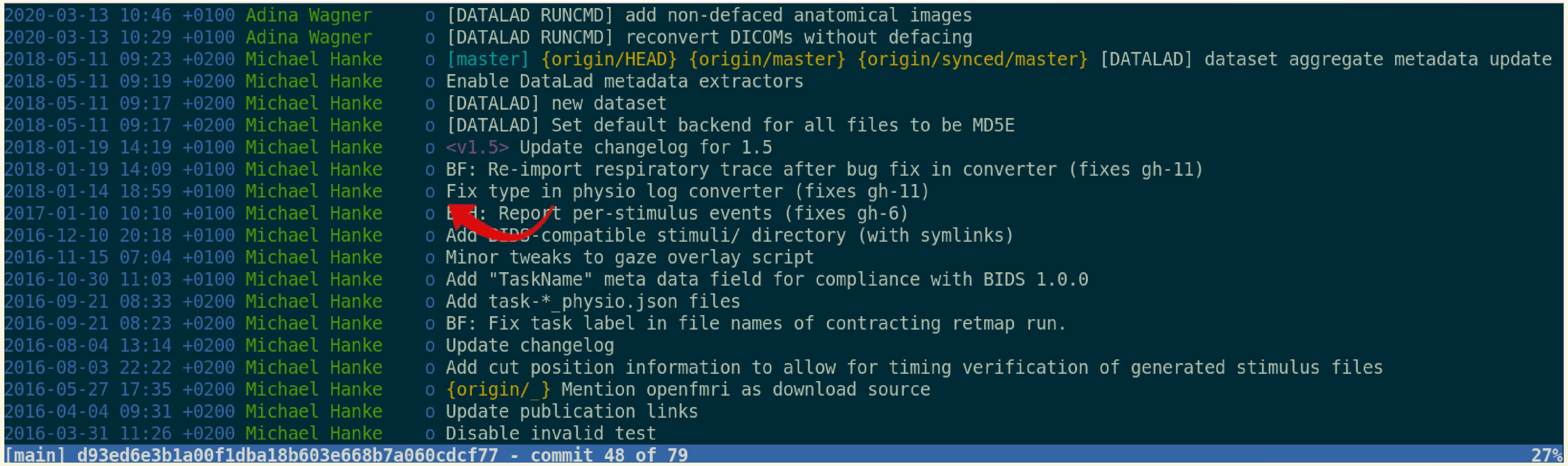
(errors are fixed, data is extended, naming standards change, an analysis
requires only a subset of your data...)



Assaf Oshri (אסף אושרי) @AssafOshri · 5. Dez. 2019

ABCD data alert !!
Due to incorrect post-processing, all task and resting-state fMRI data obtained on Philips scanners should be excluded from all analyses. The **field map** direction for these data was mistakenly **flipped**, which led to increased distortion in processed fMRI images

git-annex and DataLad version control large data

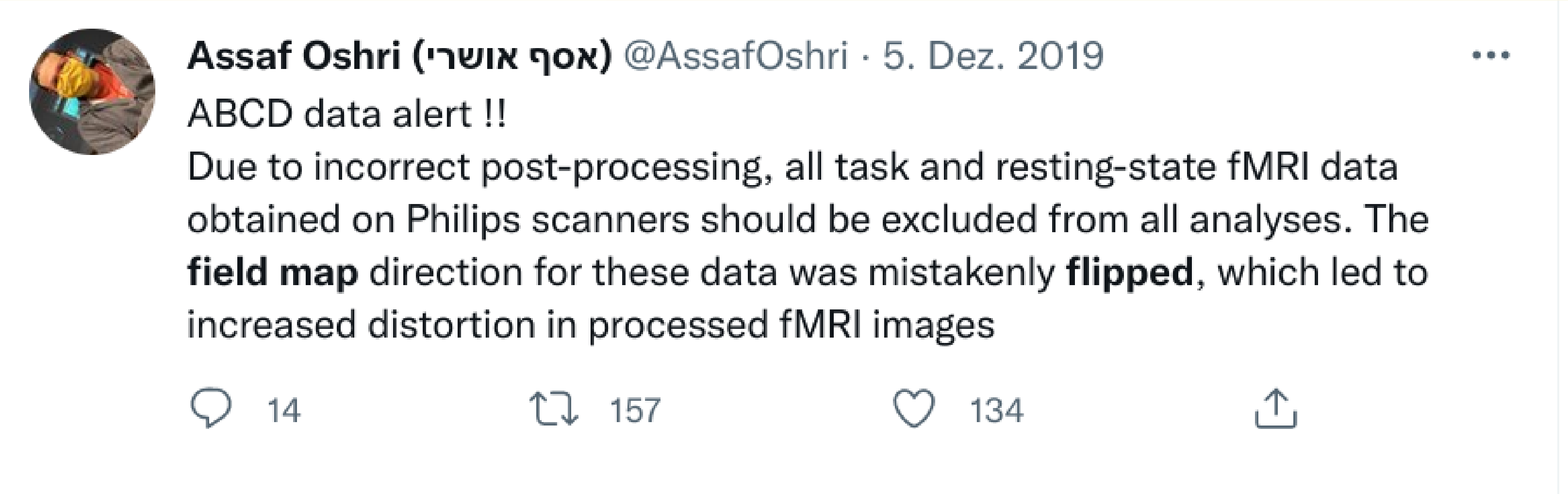


```
2020-03-13 10:46 +0100 Adina Wagner      o [DATALAD RUNCMD] add non-defaced anatomical images
2020-03-13 10:29 +0100 Adina Wagner      o [DATALAD RUNCMD] reconvert DICOMs without defacing
2018-05-11 09:23 +0200 Michael Hanke     o [master] {origin/HEAD} {origin/master} {origin/synced/master} [DATALAD] dataset aggregate metadata update
2018-05-11 09:19 +0200 Michael Hanke     o Enable DataLad metadata extractors
2018-05-11 09:17 +0200 Michael Hanke     o [DATALAD] new dataset
2018-05-11 09:17 +0200 Michael Hanke     o [DATALAD] Set default backend for all files to be MD5E
2018-01-19 14:19 +0100 Michael Hanke     o <v1.5> Update changelog for 1.5
2018-01-19 14:09 +0100 Michael Hanke     o BF: Re-import respiratory trace after bug fix in converter (fixes gh-11)
2018-01-14 18:59 +0100 Michael Hanke     o Fix type in physio log converter (fixes gh-11)
2017-01-10 10:10 +0100 Michael Hanke     o ENH: Report per-stimulus events (fixes gh-6)
2016-12-10 20:18 +0100 Michael Hanke     o Add BIDS-compatible stimuli/ directory (with symlinks)
2016-11-15 07:04 +0100 Michael Hanke     o Minor tweaks to gaze overlay script
2016-10-30 11:03 +0100 Michael Hanke     o Add "TaskName" meta data field for compliance with BIDS 1.0.0
2016-09-21 08:33 +0200 Michael Hanke     o Add task-*_physio.json files
2016-09-21 08:23 +0200 Michael Hanke     o BF: Fix task label in file names of contracting retmap run.
2016-08-04 13:14 +0200 Michael Hanke     o Update changelog
2016-08-03 22:22 +0200 Michael Hanke     o Add cut position information to allow for timing verification of generated stimulus files
2016-05-27 17:35 +0200 Michael Hanke     o {origin/_} Mention openfmri as download source
2016-04-04 09:31 +0200 Michael Hanke     o Update publication links
2016-03-31 11:26 +0200 Michael Hanke     o Disable invalid test
[main] d93ed6e3b1a00f1dba18b603e668b7a060cdcf77 - commit 48 of 79                                                        27%
```

The building blocks of a scientific result are rarely static

Data changes

(errors are fixed, data is extended, naming standards change, an analysis
requires only a subset of your data...)

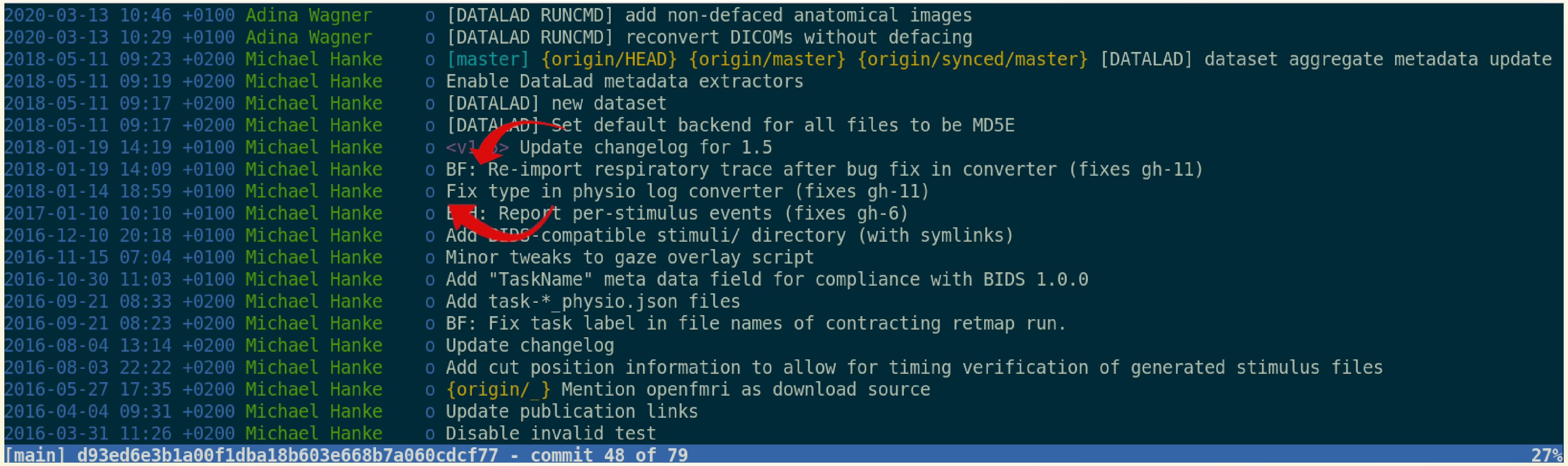


Assaf Oshri (אסף אושרי) @AssafOshri · 5. Dez. 2019

ABCD data alert !!
Due to incorrect post-processing, all task and resting-state fMRI data obtained on Philips scanners should be excluded from all analyses. The **field map** direction for these data was mistakenly **flipped**, which led to increased distortion in processed fMRI images

14 157 134

git-annex and DataLad version control large data

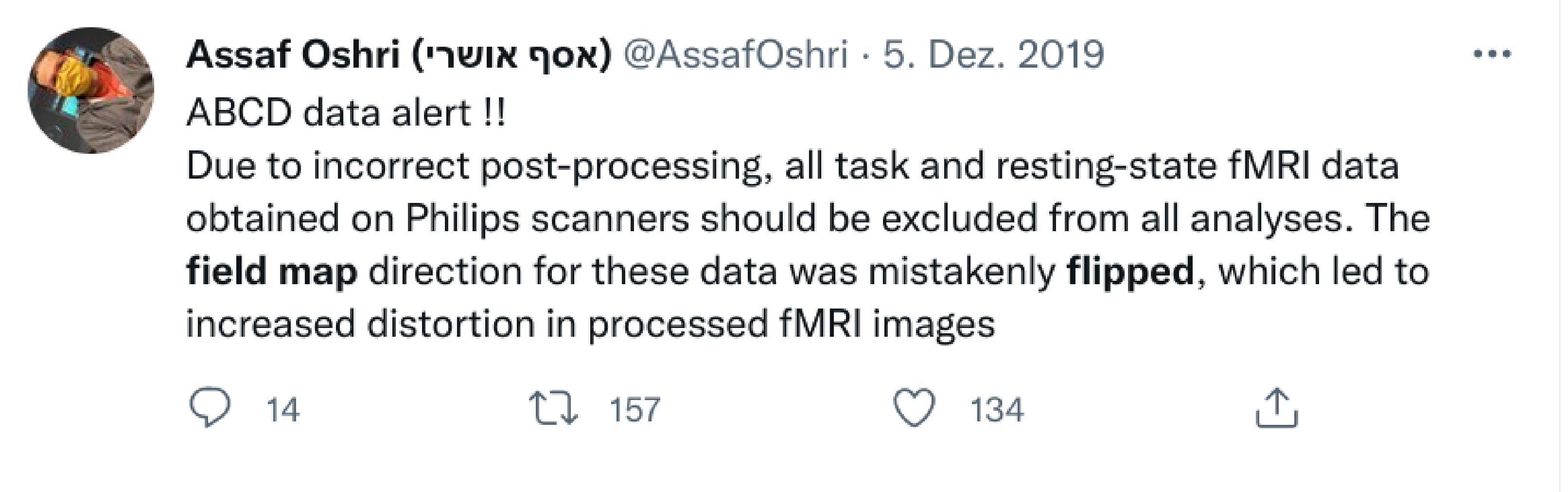


```
2020-03-13 10:46 +0100 Adina Wagner     o [DATALAD RUNCMD] add non-defaced anatomical images
2020-03-13 10:29 +0100 Adina Wagner     o [DATALAD RUNCMD] reconvert DICOMs without defacing
2018-05-11 09:23 +0200 Michael Hanke    o [master] {origin/HEAD} {origin/master} {origin/synced/master} [DATALAD] dataset aggregate metadata update
2018-05-11 09:19 +0200 Michael Hanke    o Enable DataLad metadata extractors
2018-05-11 09:17 +0200 Michael Hanke    o [DATALAD] new dataset
2018-05-11 09:17 +0200 Michael Hanke    o [DATALAD] Set default backend for all files to be MD5E
2018-01-19 14:19 +0100 Michael Hanke    o <v1.5> Update changelog for 1.5
2018-01-19 14:09 +0100 Michael Hanke    o BF: Re-import respiratory trace after bug fix in converter (fixes gh-11)
2018-01-14 18:59 +0100 Michael Hanke    o Fix type in physio log converter (fixes gh-11)
2017-01-10 10:10 +0100 Michael Hanke    o ENH: Report per-stimulus events (fixes gh-6)
2016-12-10 20:18 +0100 Michael Hanke    o Add BIDS-compatible stimuli/ directory (with symlinks)
2016-11-15 07:04 +0100 Michael Hanke    o Minor tweaks to gaze overlay script
2016-10-30 11:03 +0100 Michael Hanke    o Add "TaskName" meta data field for compliance with BIDS 1.0.0
2016-09-21 08:33 +0200 Michael Hanke    o Add task-*_physio.json files
2016-09-21 08:23 +0200 Michael Hanke    o BF: Fix task label in file names of contracting retmap run.
2016-08-04 13:14 +0200 Michael Hanke    o Update changelog
2016-08-03 22:22 +0200 Michael Hanke    o Add cut position information to allow for timing verification of generated stimulus files
2016-05-27 17:35 +0200 Michael Hanke    o {origin/_} Mention openfmri as download source
2016-04-04 09:31 +0200 Michael Hanke    o Update publication links
2016-03-31 11:26 +0200 Michael Hanke    o Disable invalid test
[main] d93ed6e3b1a00f1dba18b603e668b7a060cdcf77 - commit 48 of 79                                                              27%
```

The building blocks of a scientific result are rarely static

Data changes

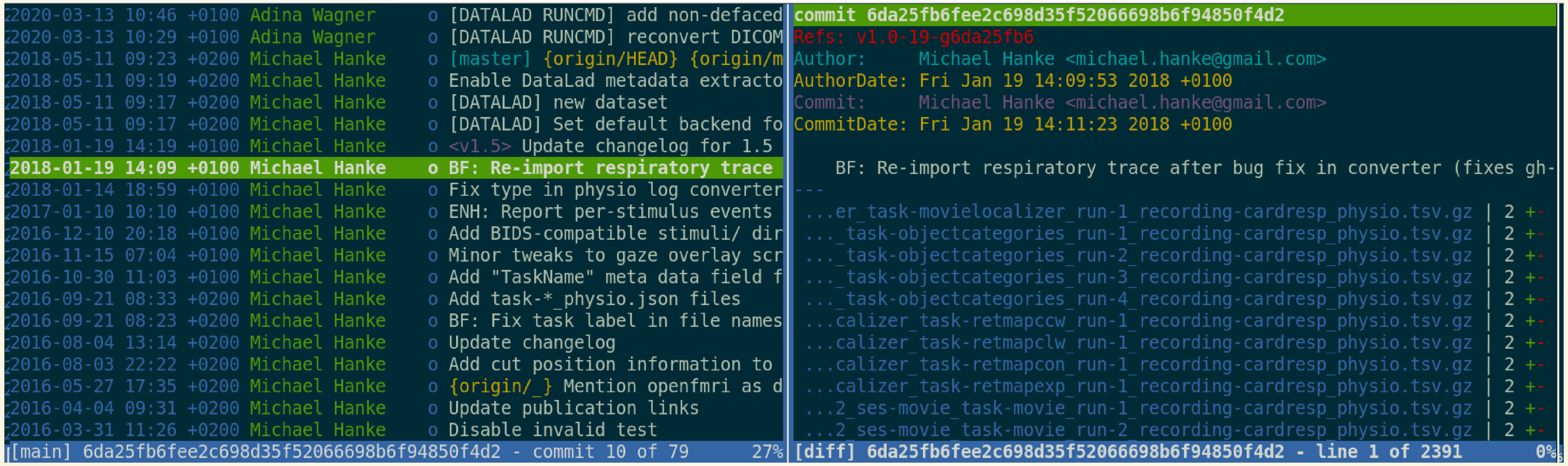
(errors are fixed, data is extended, naming standards change, an analysis
requires only a subset of your data...)



Assaf Oshri (אסף אושרי) @AssafOshri · 5. Dez. 2019

ABCD data alert !!
Due to incorrect post-processing, all task and resting-state fMRI data obtained on Philips scanners should be excluded from all analyses. The **field map** direction for these data was mistakenly **flipped**, which led to increased distortion in processed fMRI images

git-annex and DataLad version control large data



```
2020-03-13 10:46 +0100 Adina Wagner      o [DATALAD RUNCMD] add non-defaced
2020-03-13 10:29 +0100 Adina Wagner      o [DATALAD RUNCMD] reconvert DICOM
2018-05-11 09:23 +0200 Michael Hanke     o [master] {origin/HEAD} {origin/m
2018-05-11 09:19 +0200 Michael Hanke     o Enable DataLad metadata extracto
2018-05-11 09:17 +0200 Michael Hanke     o [DATALAD] new dataset
2018-05-11 09:17 +0200 Michael Hanke     o [DATALAD] Set default backend fo
2018-01-19 14:19 +0100 Michael Hanke     o <v1.5> Update changelog for 1.5
2018-01-19 14:09 +0100 Michael Hanke     o BF: Re-import respiratory trace
2018-01-14 18:59 +0100 Michael Hanke     o Fix type in physio log converter
2017-01-10 10:10 +0100 Michael Hanke     o ENH: Report per-stimulus events
2016-12-10 20:18 +0100 Michael Hanke     o Add BIDS-compatible stimuli/ dir
2016-11-15 07:04 +0100 Michael Hanke     o Minor tweaks to gaze overlay scr
2016-10-30 11:03 +0100 Michael Hanke     o Add "TaskName" meta data field f
2016-09-21 08:33 +0200 Michael Hanke     o Add task-*_physio.json files
2016-09-21 08:23 +0200 Michael Hanke     o BF: Fix task label in file names
2016-08-04 13:14 +0200 Michael Hanke     o Update changelog
2016-08-03 22:22 +0200 Michael Hanke     o Add cut position information to
2016-05-27 17:35 +0200 Michael Hanke     o {origin/_} Mention openfmri as d
2016-04-04 09:31 +0200 Michael Hanke     o Update publication links
2016-03-31 11:26 +0200 Michael Hanke     o Disable invalid test
[main] 6da25fb6fee2c698d35f52066698b6f94850f4d2 - commit 10 of 79      27%
```

```
commit 6da25fb6fee2c698d35f52066698b6f94850f4d2
Refs: v1.0-19-g6da25fb6
Author:     Michael Hanke <michael.hanke@gmail.com>
AuthorDate: Fri Jan 19 14:09:53 2018 +0100
Commit:     Michael Hanke <michael.hanke@gmail.com>
CommitDate: Fri Jan 19 14:11:23 2018 +0100

    BF: Re-import respiratory trace after bug fix in converter (fixes gh-
---
 ...er_task-movielocalizer_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ..._task-objectcategories_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ..._task-objectcategories_run-2_recording-cardresp_physio.tsv.gz | 2 +-
 ..._task-objectcategories_run-3_recording-cardresp_physio.tsv.gz | 2 +-
 ..._task-objectcategories_run-4_recording-cardresp_physio.tsv.gz | 2 +-
 ...calizer_task-retmapccw_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ...calizer_task-retmapclw_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ...calizer_task-retmapcon_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ...calizer_task-retmapexp_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ...2_ses-movie_task-movie_run-1_recording-cardresp_physio.tsv.gz | 2 +-
 ...2_ses-movie_task-movie_run-2_recording-cardresp_physio.tsv.gz | 2 +-
[diff] 6da25fb6fee2c698d35f52066698b6f94850f4d2 - line 1 of 2391       0%
```

LEAVING A TRACE

"Shit, which version of which script produced these outputs from which version of what data?"

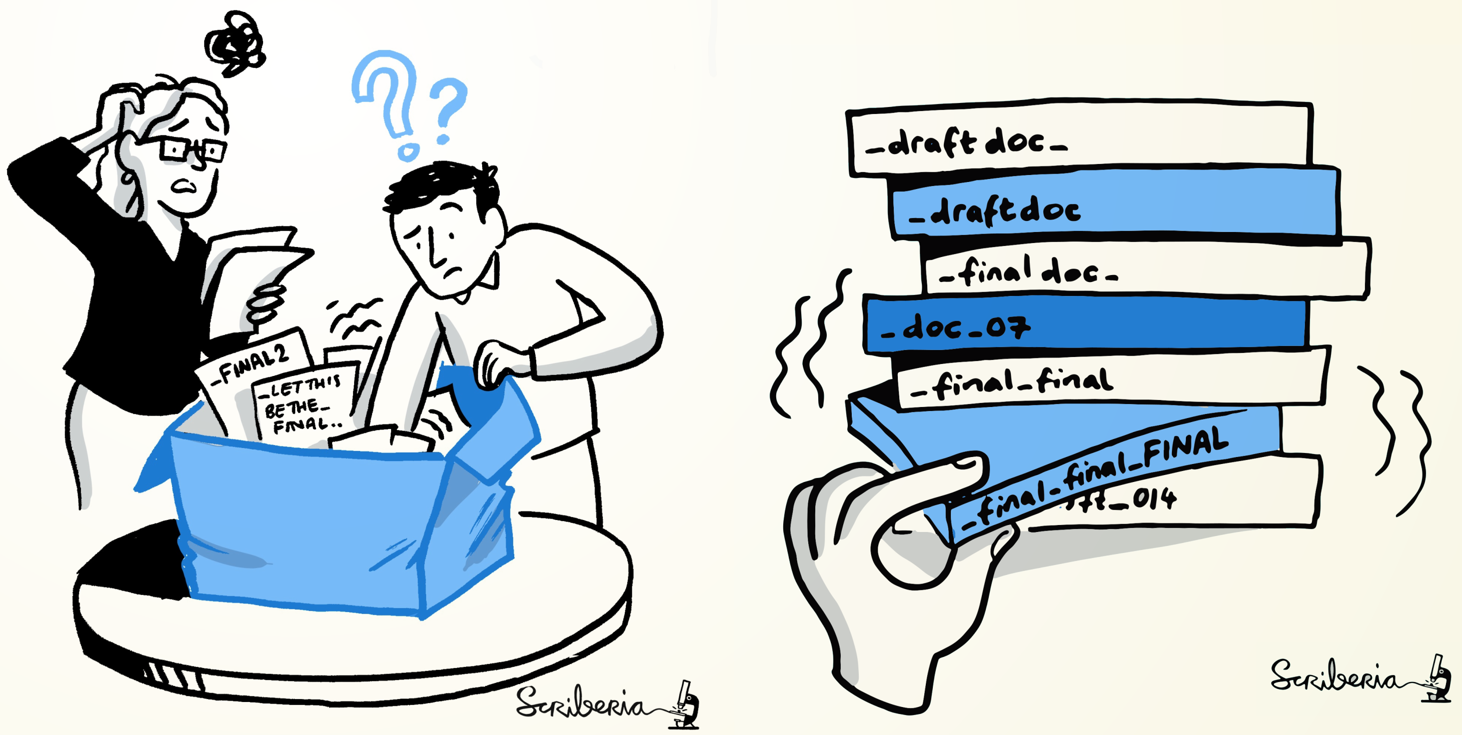


Image credit: CC-BY Scriberia and The Turing Way

LEAVING A TRACE

"Shit, why buttons did I click and in which order did I use all those tools?"

Image credit: CC-BY Scriberia and The Turing Way

LEAVING A TRACE

"Shit, why buttons did I click and in which order did I use all those tools?"

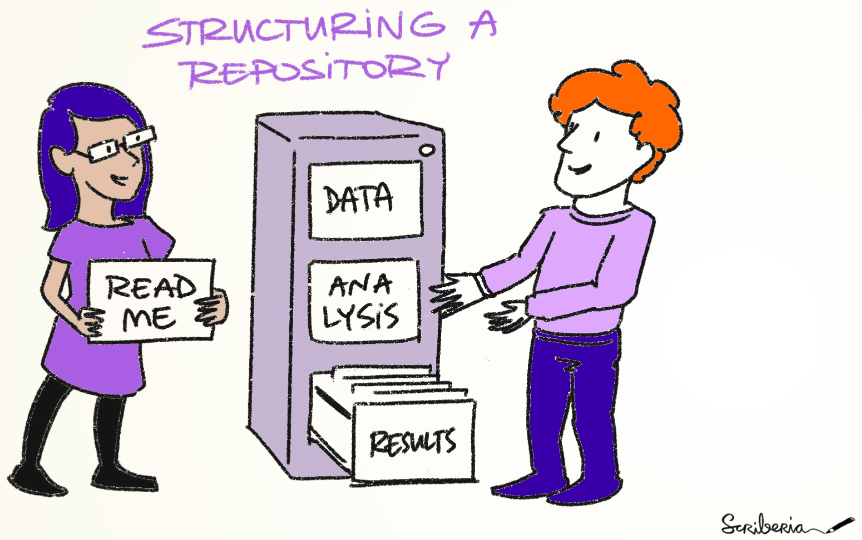
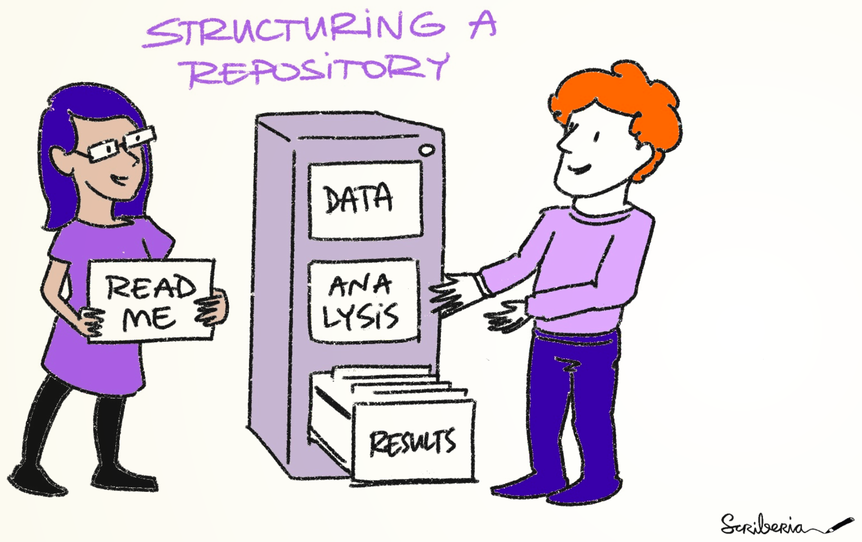


Image credit: CC-BY Scriberia and The Turing Way

1) Create an intuitive structure, and

LEAVING A TRACE

"Shit, why buttons did I click and in which order did I use all those tools?"



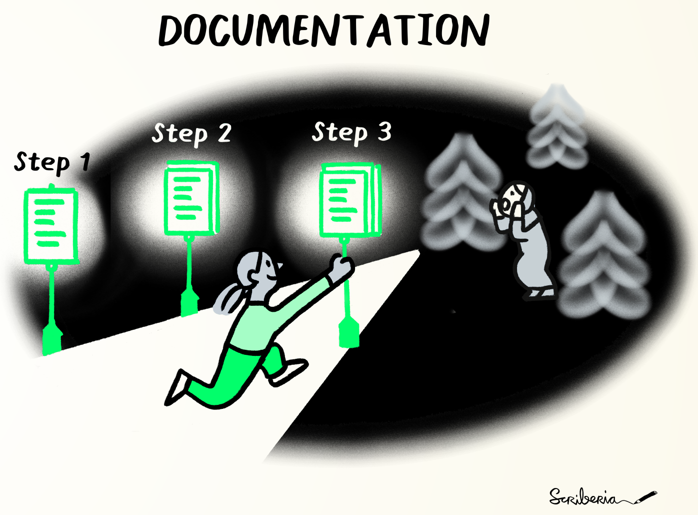


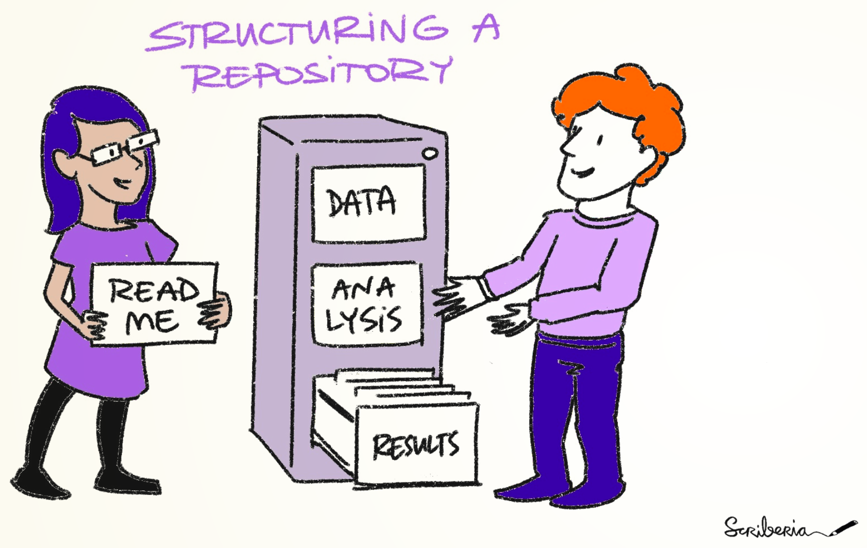
Image credit: CC-BY Scriberia and The Turing Way

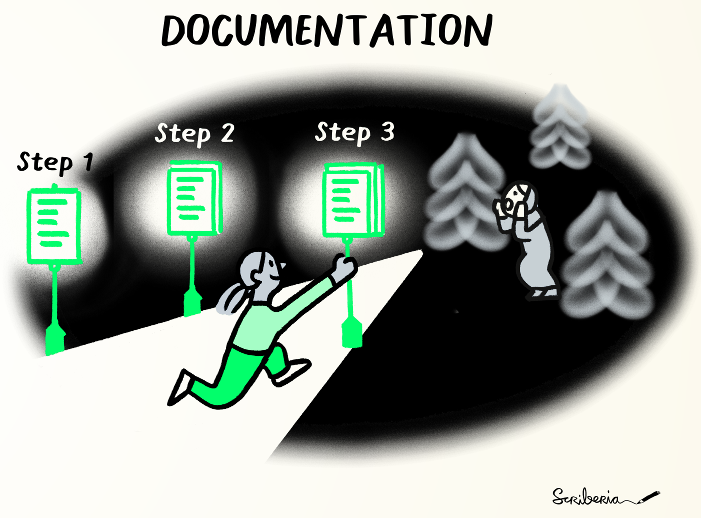
1) Create an intuitive structure, and

2) write (plenty! of) documentation as you go, and

LEAVING A TRACE

"Shit, why buttons did I click and in which order did I use all those tools?"





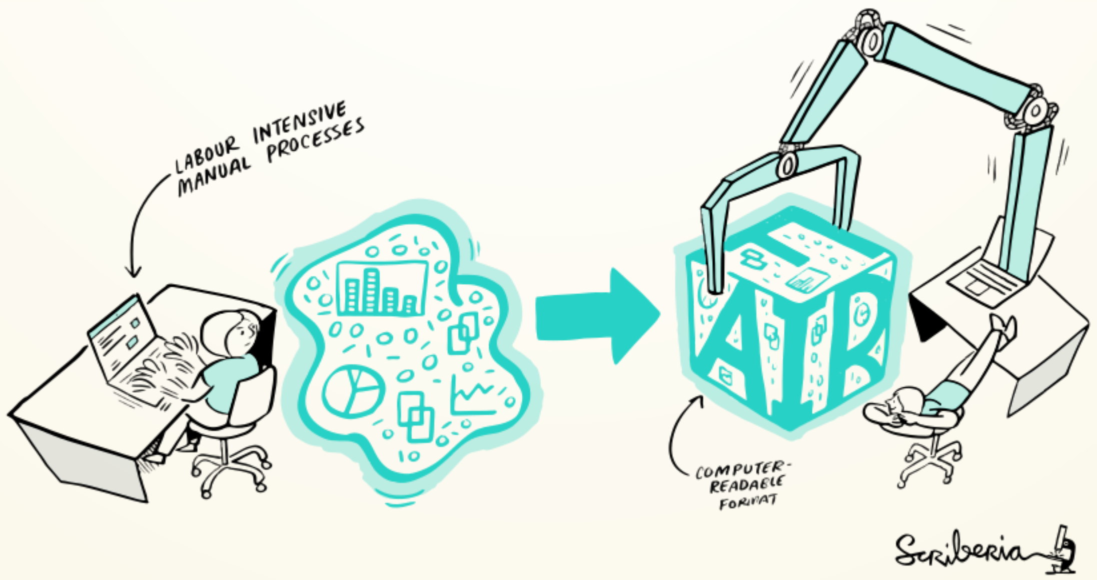


Image credit: CC-BY Scriberia and The Turing Way

1) Create an intuitive structure, and

2) write (plenty! of) documentation as you go, and

RESEARCH DATA MANAGEMENT IS TIED TO REPRODUCIBILITY

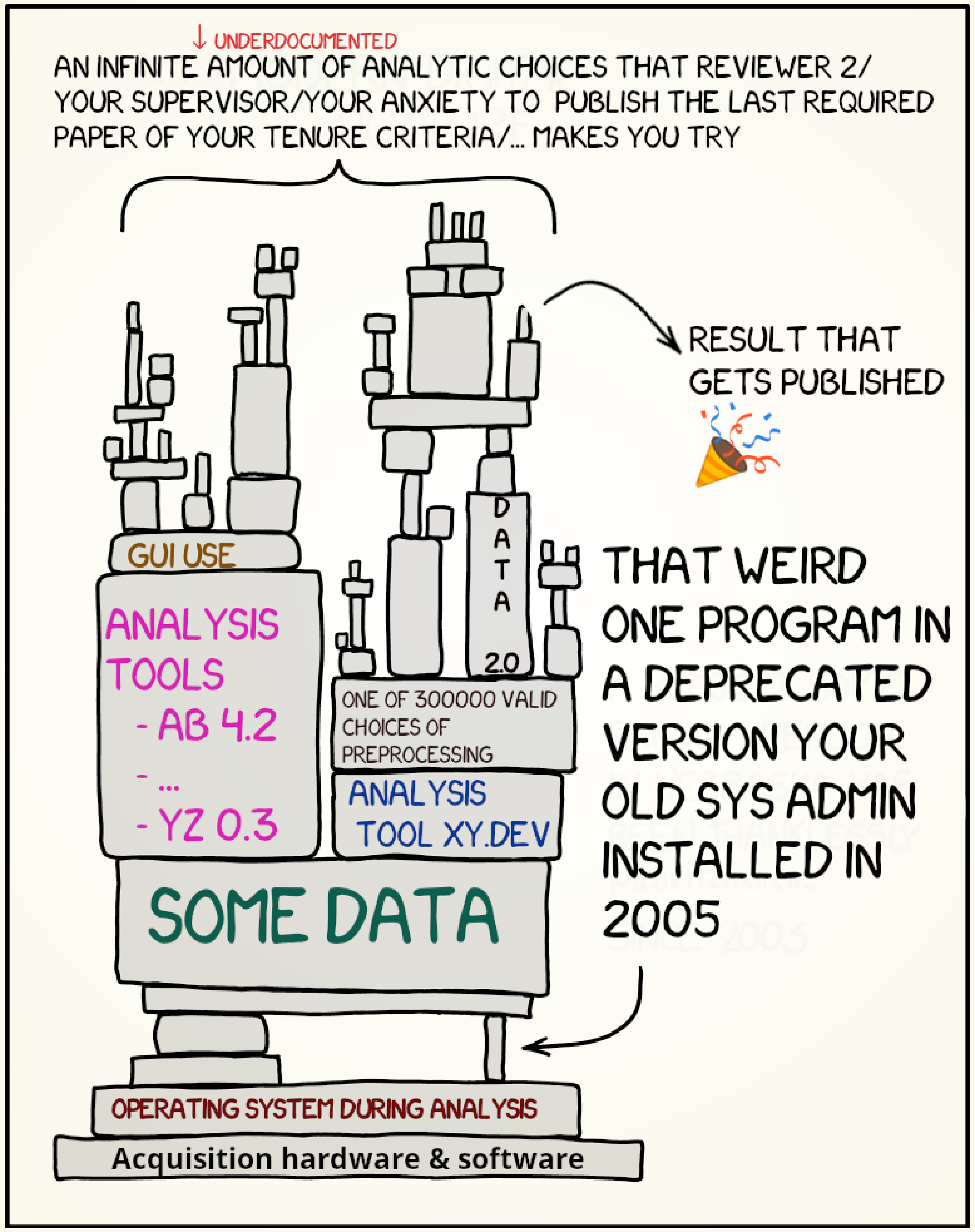

↓ UNDERDOCUMENTED
AN INFINITE AMOUNT OF ANALYTIC CHOICES THAT REVIEWER 2/ YOUR SUPERVISOR/YOUR ANXIETY TO PUBLISH THE LAST REQUIRED PAPER OF YOUR TENURE CRITERIA/... MAKES YOU TRY
RESULT THAT GETS PUBLISHED 🎉
GUI USE
ANALYSIS TOOLS
- AB 4.2
- ...
- YZ 0.3
DATA 2.0
ONE OF 300000 VALID CHOICES OF PREPROCESSING
ANALYSIS TOOL XY.DEV
THAT WEIRD ONE PROGRAM IN A DEPRECATED VERSION YOUR OLD SYS ADMIN INSTALLED IN 2005
SOME DATA
OPERATING SYSTEM DURING ANALYSIS
Acquisition hardware & software

Image credit: Based on xkcd.com/2347/ (CC-BY)

Reproducibility Management in Neuroscience - Specific Issues and Solutions (DOI 10.5281/zenodo.4285927)

BACK-UPS AND ARCHIVAL

Ensure that your data are regularily backed-up, and eventually deposited in an appropriate archive or repository

Back-ups

- Keep back-ups on different infrastructure, ideally even different physical locations
- Synchronize regularly
- My personal workflow: Distributed version control

Software

- E.g., Software Heritage , Zenodo (both have automatic GitHub integrations)

Data

- E.g., Zenodo, Gin.g-node.org Neurovault, DataVerse, Data DRYAD, FigShare

Further reading: the-turing-way.netlify.app/reproducible-research/rdm/rdm-sharing.html

DIGITAL OBJECT IDENTIFIERS

- Example: 10.5281/zenodo.7419377 (uniquely & persistently identifies this talk)
- Provide a persistent, trusted reference. Resolve any DOI at doi.org

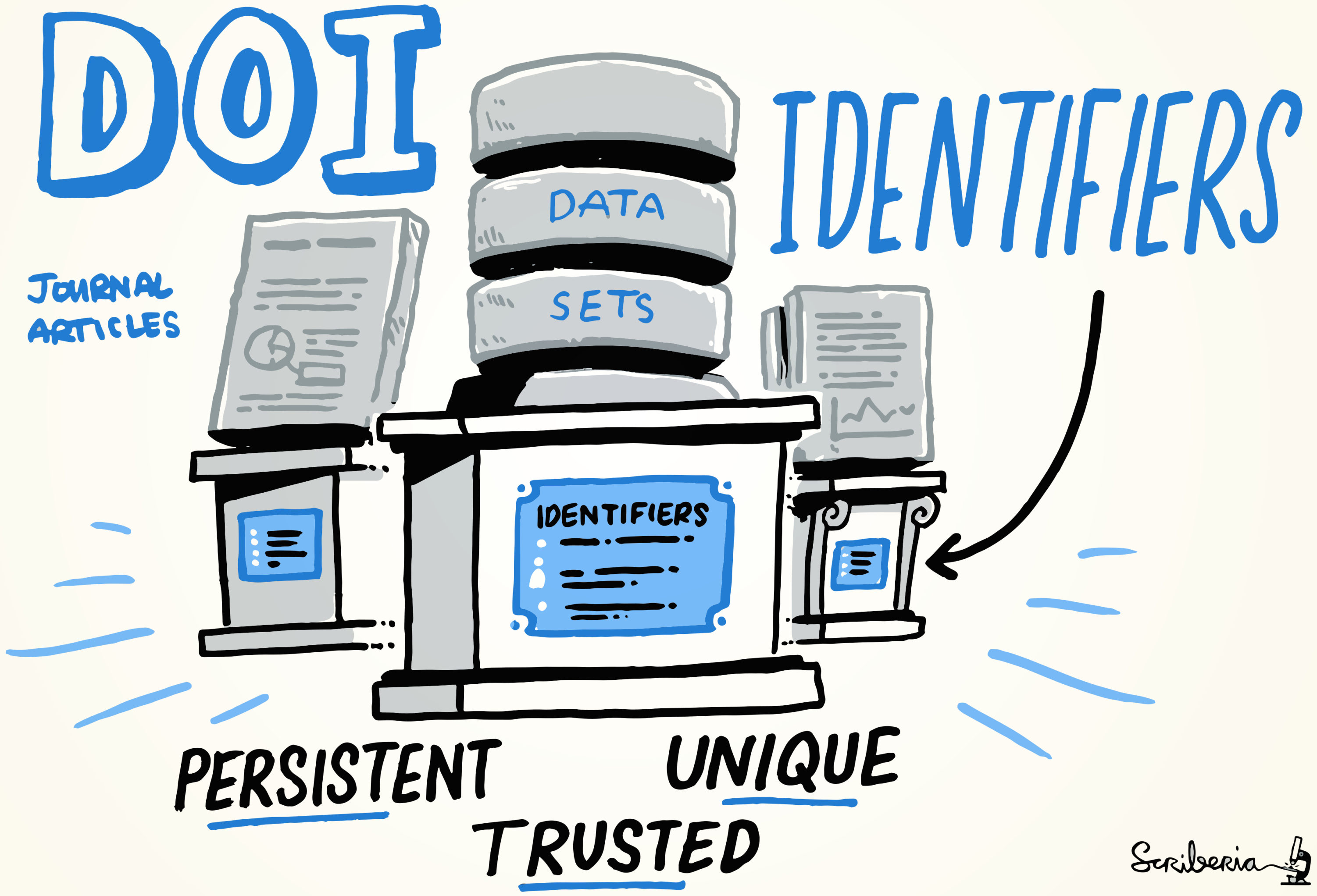


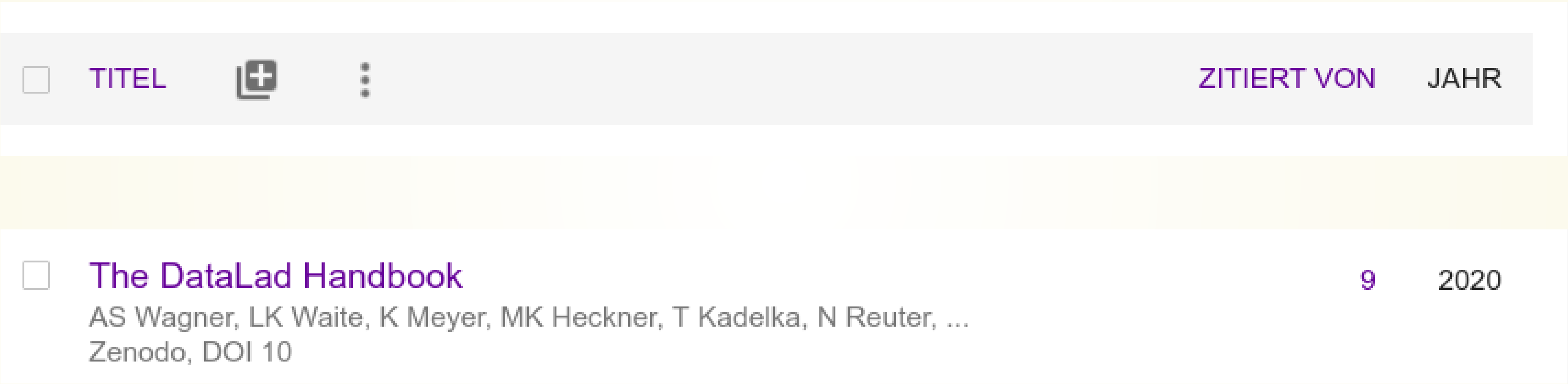
Image credit: The Turing Way

Get a DOI from

- free academic services and archives that you already use, such as Zenodo, FigShare, or OSF
- your own institutions (e.g., library, DataVerse, ...)
- Preprint servers, publishers
- My tip for large datasets: Gin.g-node.org

DIGITAL OBJECT IDENTIFIERS

- Example: 10.5281/zenodo.7419377 (uniquely & persistently identifies this talk)
- Provide a persistent, trusted reference. Resolve any DOI at doi.org
- Make your work citeable



□	TITEL	ZITIERT VON	JAHR
□	The DataLad Handbook AS Wagner, LK Waite, K Meyer, MK Heckner, T Kadelka, N Reuter, ... Zenodo, DOI 10	9	2020

Get a DOI from

- free academic services and archives that you already use, such as Zenodo, FigShare, or OSF
- your own institutions (e.g., library, DataVerse, ...)
- Preprint servers, publishers
- My tip for large datasets: Gin.g-node.org

LICENSES

- Everything you (co-)create has a copyright, and you're a/the copyright holder
- **Without a license your work is unusable by others**
- Use an established license rather creating one yourself
- Different licenses are suitable for different types of work

Software	Data (e.g., comics, text)
License picker	Creative Commons

Image credit: The Turing Way

Image credit: www.homoempresarius.com

Further reading: the-turing-way.netlify.app/reproducible-research/licensing

FAIR

Wilkinson et al., 2016: "The FAIR Guiding Principles for scientific data management and stewardship", doi.org/10.1038/sdata.2016.18

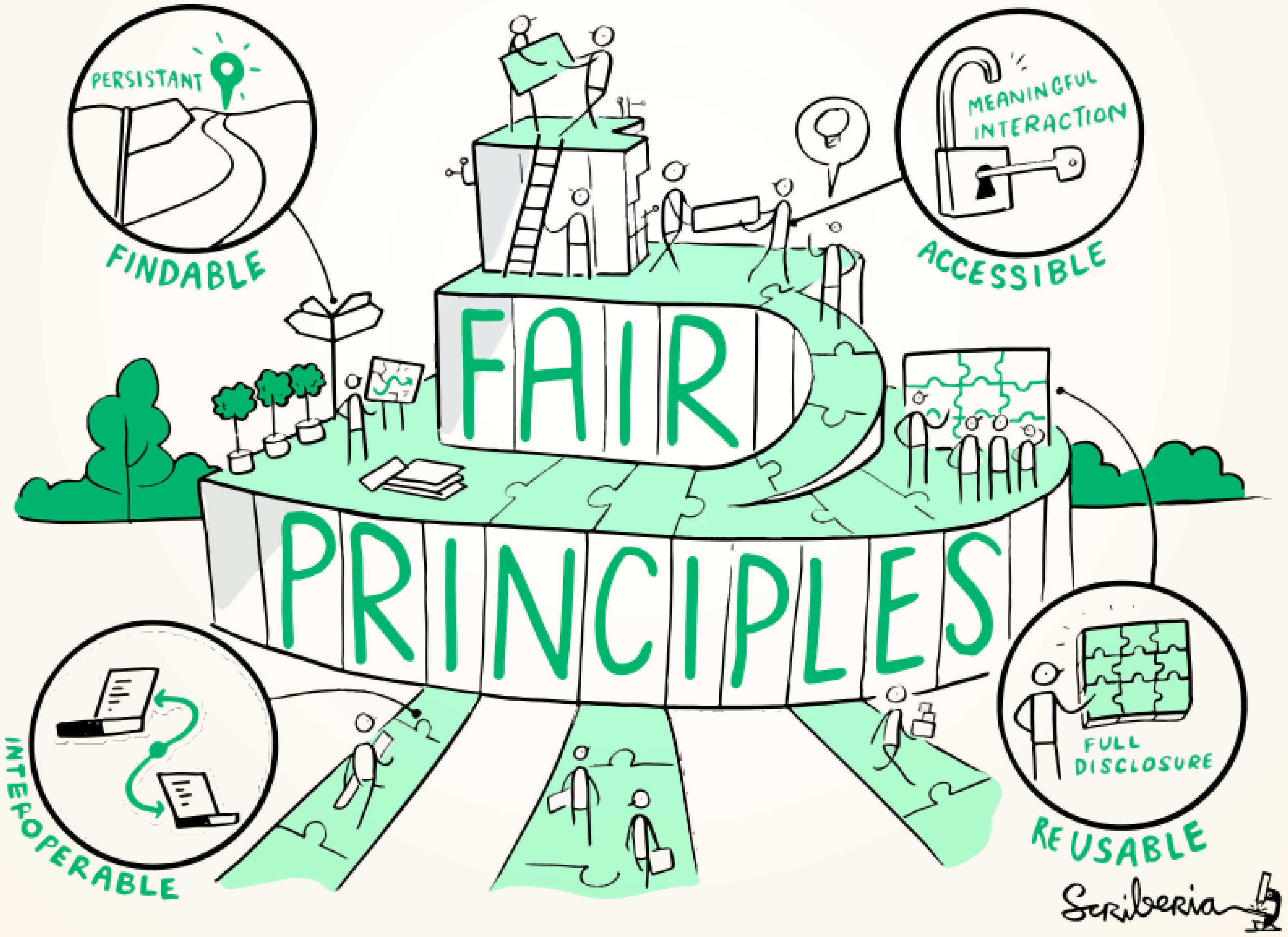


Image credit: Scriberia and The Turing Way (CC-BY)

DATA MANAGEMENT PLANS (DMP)

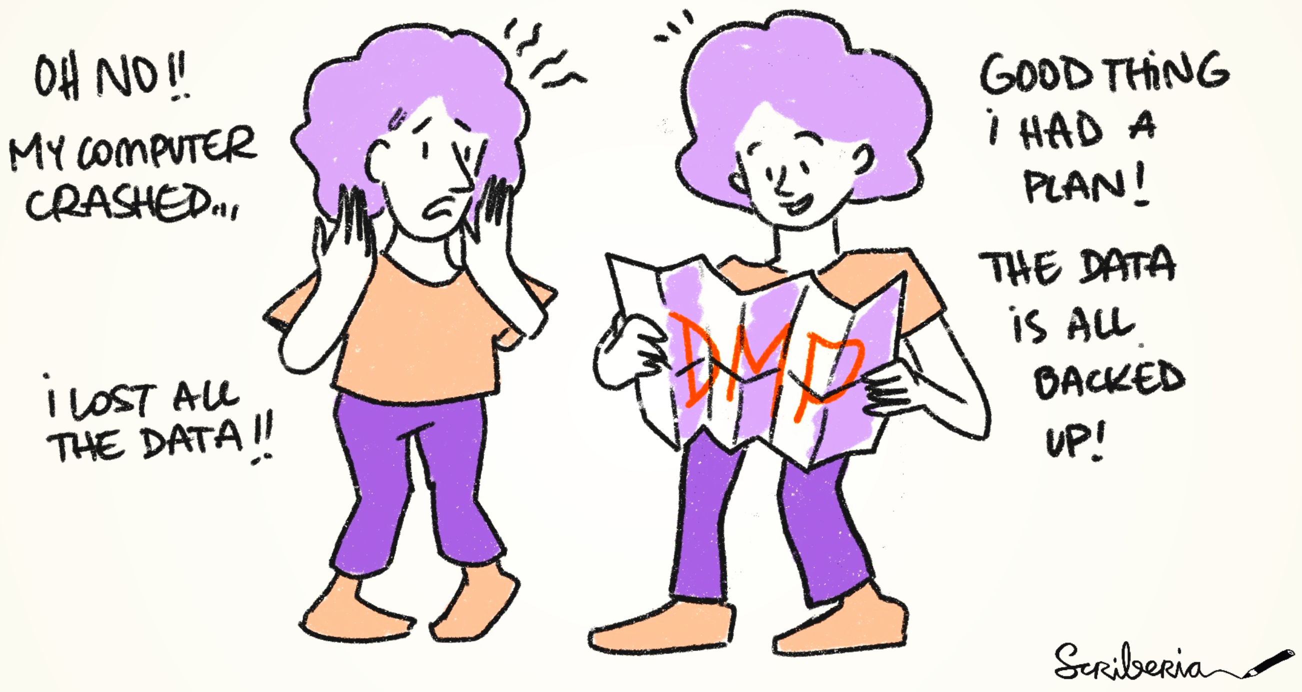


Image credit: Scriberia and The Turing Way (CC-BY)

A Data Management Plan (DMP) is a brief plan to define:

- how the data will be created or used
- how it will be documented
- who will be able to access it
- where it will be stored
- who will back it up
- whether (and how) it will be shared and preserved.

Further reading: https://ukdataservice.ac.uk/learning-hub/research-data-management

TAKE HOME MESSAGES

- RDM is a continuous process
- There is an ecosystem of resources, infrastructure, and experts to assist you at every step. Befriend your local librarian!
- The biggest beneficiary of RDM? Yourself
- The second biggest beneficiary of RDM? Yourself in 6 months
- The consequence of good RDM? Better science

Resources

- The Turing Way Handbook for Reproducible Science
- OpenAIRE Research Data Management Handbook
- RDM resources TU Delft
- Guide to FAIR
- Checklist: How FAIR are your data?
- Checklist for Writing a Data Management Plan