Open Science with OSF

UTD Brain Hack 2019

Kendra Seaman, PhD

Open Science Framework

- Preregistrations
- Sharing project information
- You’re turn!

Preregistration

Preregistration

- Specify plans *in* advance:
  - hypotheses
  - data collection plan
  - analysis plan

Preregistration

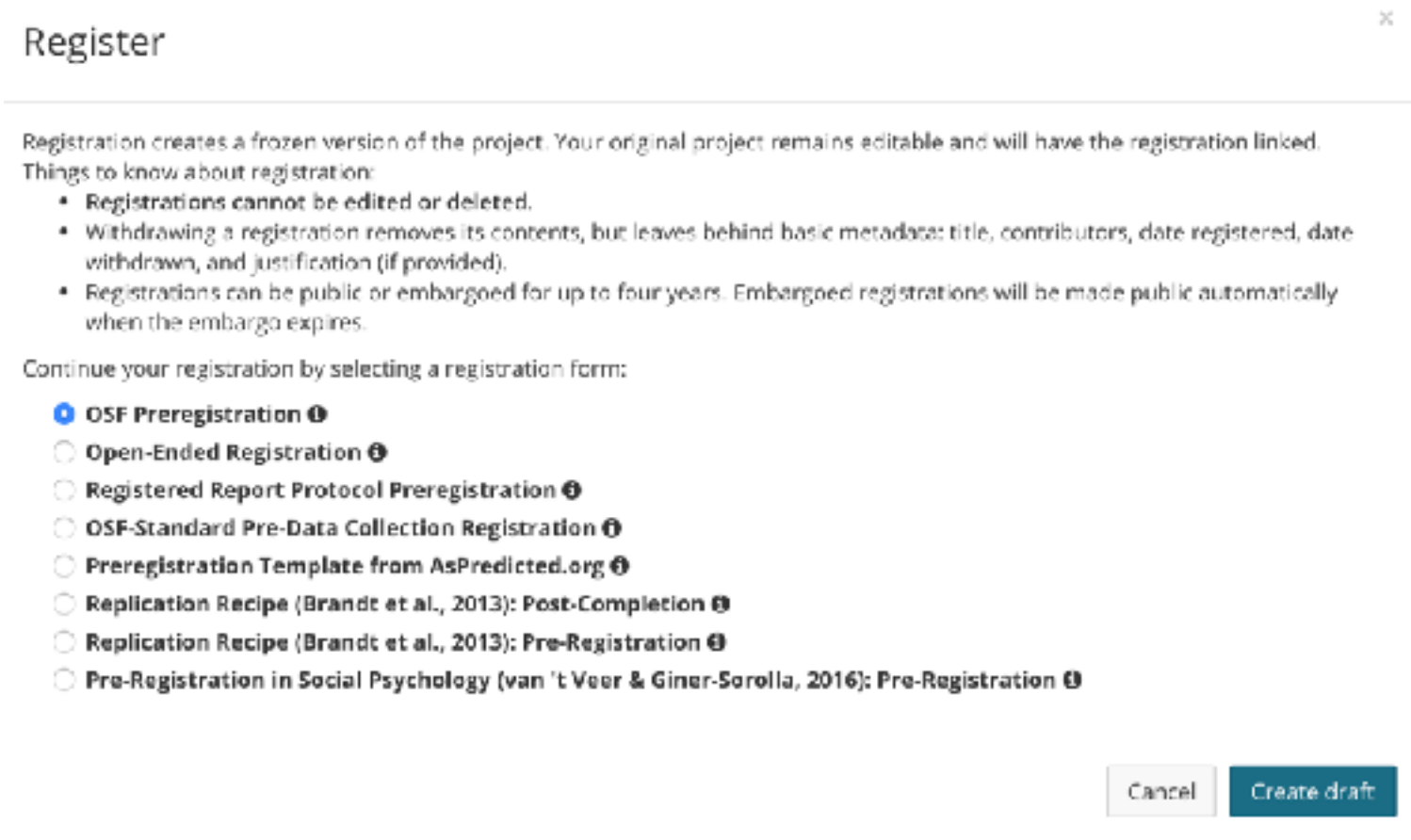
- Specify plans *in* advance:
  - hypotheses
  - data collection plan
  - analysis plan
- Allows you to separate **exploratory** (hypothesis-generating) from **confirmatory** (hypothesis-testing) research

Preregistration

- Specify plans *in* advance:
  - hypotheses
  - data collection plan
  - analysis plan
- Allows you to separate **exploratory** (hypothesis-generating) from **confirmatory** (hypothesis-testing) research
- Can preregister at various stages:
  - Before data collection begins
  - Before beginning analysis of an existing data set
  - After being asked to collect more data in peer review

Preregistration

CENTER FOR OPEN SCIENCE



[fMRI Preregistration - https://osf.io/6juft/](https://osf.io/6juft/)

Preregistration Components

- Study Info (title, authors, hypotheses)
- Design Plan (study type, design, randomization)
- Sampling Plan (existing data?, data collection procedures, sample size, stopping rule)
- Variables (manipulated and measured variables)
- Analysis Plan (statistical models, transformations, data exclusions, missing data, exploratory analyses)

Preregistration

- Examples - My Preregistrations
  - Behavior (boundSkew) - https://osf.io/vwhxb
  - fMRI (socialAL) - https://osf.io/b3au5
  - Meta-Analysis (timeprefsmeta) - https://osf.io/e4anc

Sharing on OSF

Sharing on OSF

- What to share?
  - Stimuli
  - Data – make sure it's deidentified!
  - Code – can link to GitHub repo
  - Literature – can link to Mendeley or Zotero library
  - Data products (posters, preprints) – links to PsyArXiv, but not BioRxiv
  - Notes and documentation

Sharing on OSF

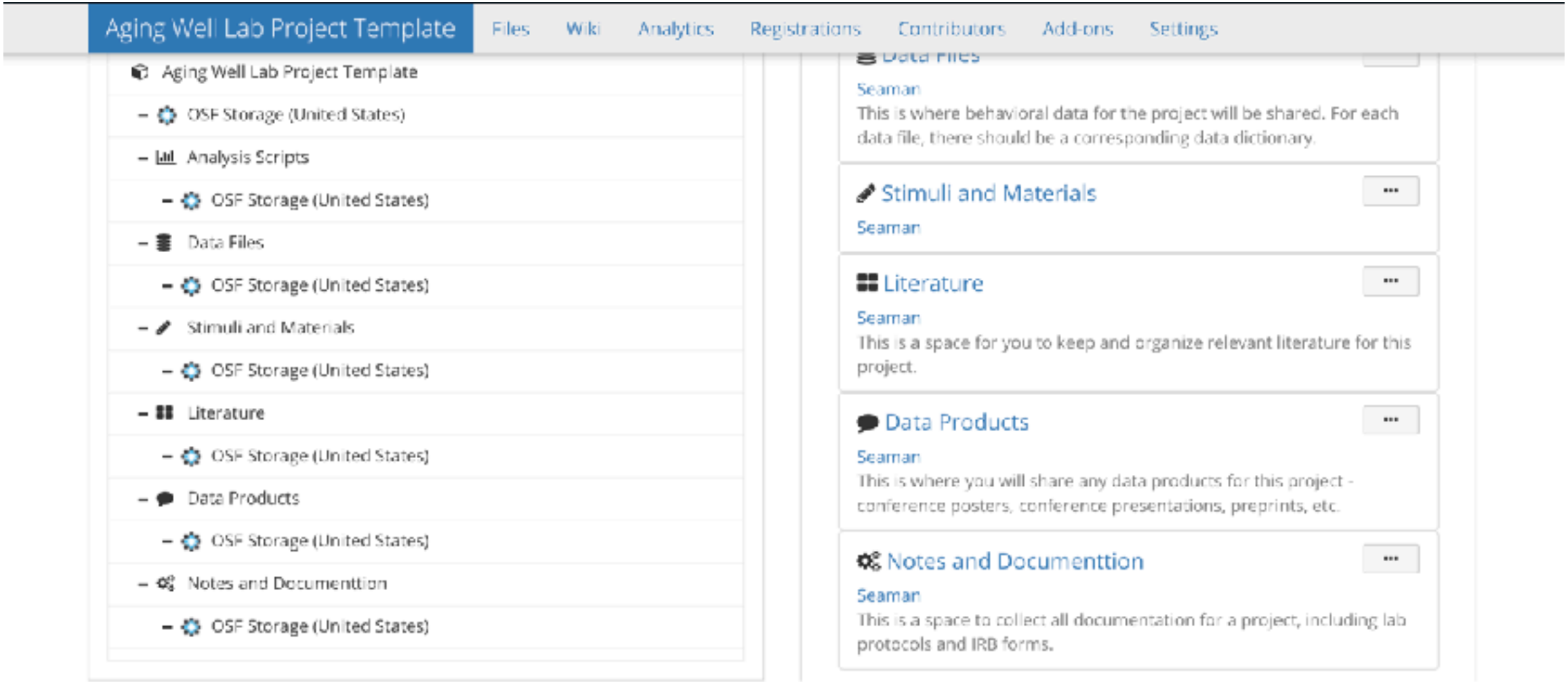
- What to share?
  - Stimuli
  - Data – make sure it's deidentified!
  - Code – can link to GitHub repo
  - Literature – can link to Mendeley or Zotero library
  - Data products (posters, preprints) – links to PsyArXiv, but not BioRxiv
  - Notes and documentation
- What not to share?
  - Neuroimaging data (unless you want to…)

Sharing on OSF

- What to share?
  - Stimuli
  - Data – make sure it's deidentified!
  - Code – can link to GitHub repo
  - Literature – can link to Mendeley or Zotero library
  - Data products (posters, preprints) – links to PsyArXiv, but not BioRxiv
  - Notes and documentation
- What not to share?
  - Neuroimaging data (unless you want to…)

- Project template: https://osf.io/ce8p4/

Sharing on OSF



Sharing on OSF - My Journey

- Post-hoc
  - ageHDisc - https://osf.io/94225/
- With publication -
  - ageSkew - https://osf.io/vbw3v/ - NOT RECOMMENDED
  - Subval - https://osf.io/26mqt/
- Planned in advance
  - Lab Template - https://osf.io/ce8p4/

Sharing on OSF - Share when you're ready

- You can embargo your preregistration for up to 4 years.
  - How to (Step 4. Register): https://help.osf.io/hc/en-us/articles/360019738834-Create-a-Preregistration

- You can make a project (or any component of a project) private:
  - Example: https://osf.io/d56b2/
  - How to: https://help.osf.io/hc/en-us/articles/360018981414-Control-Your-Privacy-Settings

Sharing Best Practices: README

- Text file at the top of each directory
  - General info about the contents of that directory.
  - Contents with a *brief* description of each file in the directory.

Sharing Best Practices: Data Dictionary

Sharing Best Practices: Data Dictionary

- Ideally, there is one of these for each dataset shared.

Sharing Best Practices: Data Dictionary

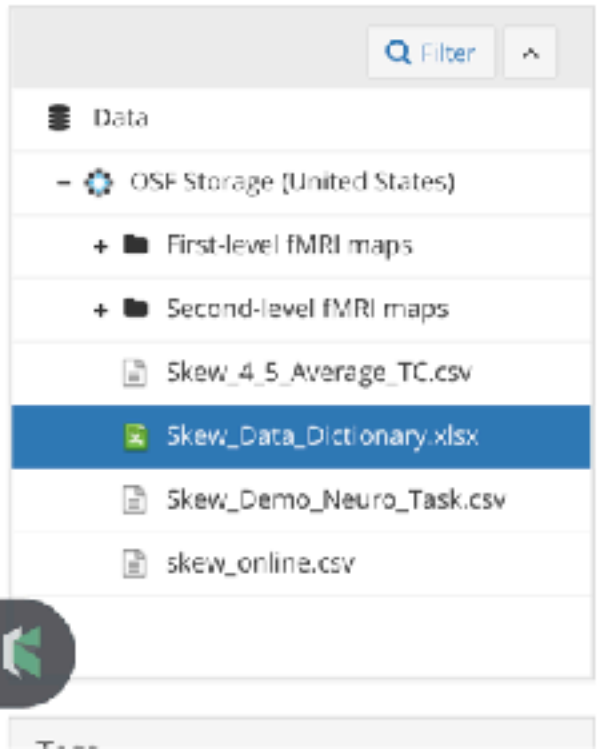
- Ideally, there is one of these for each dataset shared.
- Provides metadata for your dataset
  - Variable names
  - Readable variable names
  - Measurement units
  - Allowed values
  - Description

Sharing Best Practices: Data Dictionary

CENTER FOR OPEN SCIENCE

Skew_Data_Dictionary.xlsx (Version: 3)

Check out | Delete | Download | Share | View | Revisions



Demo-Neuro-Task | 4_5_Average_TC | Skew_online

Show rows with cells including:

Variable Name	Measurement Units	Allowed Values	Description of Variable
ID	string	ssnnnnnn	Subject ID based on initials (two strings) and date...
age	numeric	18-85	Age (in years)
gender	string	m = male, f = female	Gender
numeracy	numeric	0-9	Numeracy Inventory - 9 items
trailDIFF	numeric		Trail Making Test (B-A)
dsSUM	numeric		WAIS III Digit Span Test (Forward + Backward)
LNseq	numeric		Letter-Number Sequencing
MMSE	numeric	0-30	Mini-Mental Status Exam
shipley_vocab	numeric	0-40	Shipley Vocabulary Test
defraud_detect	numeric	1-7	SaFE Question 39: How able are you to detect a f...
defraud_likely	numeric	1-7	SaFE Question 40: How likely are you to make a f...
defraud_resist	numeric	1-7	SaFE Question 41: How able are you to resist hig...

Sharing Best Practices: Data Dictionary

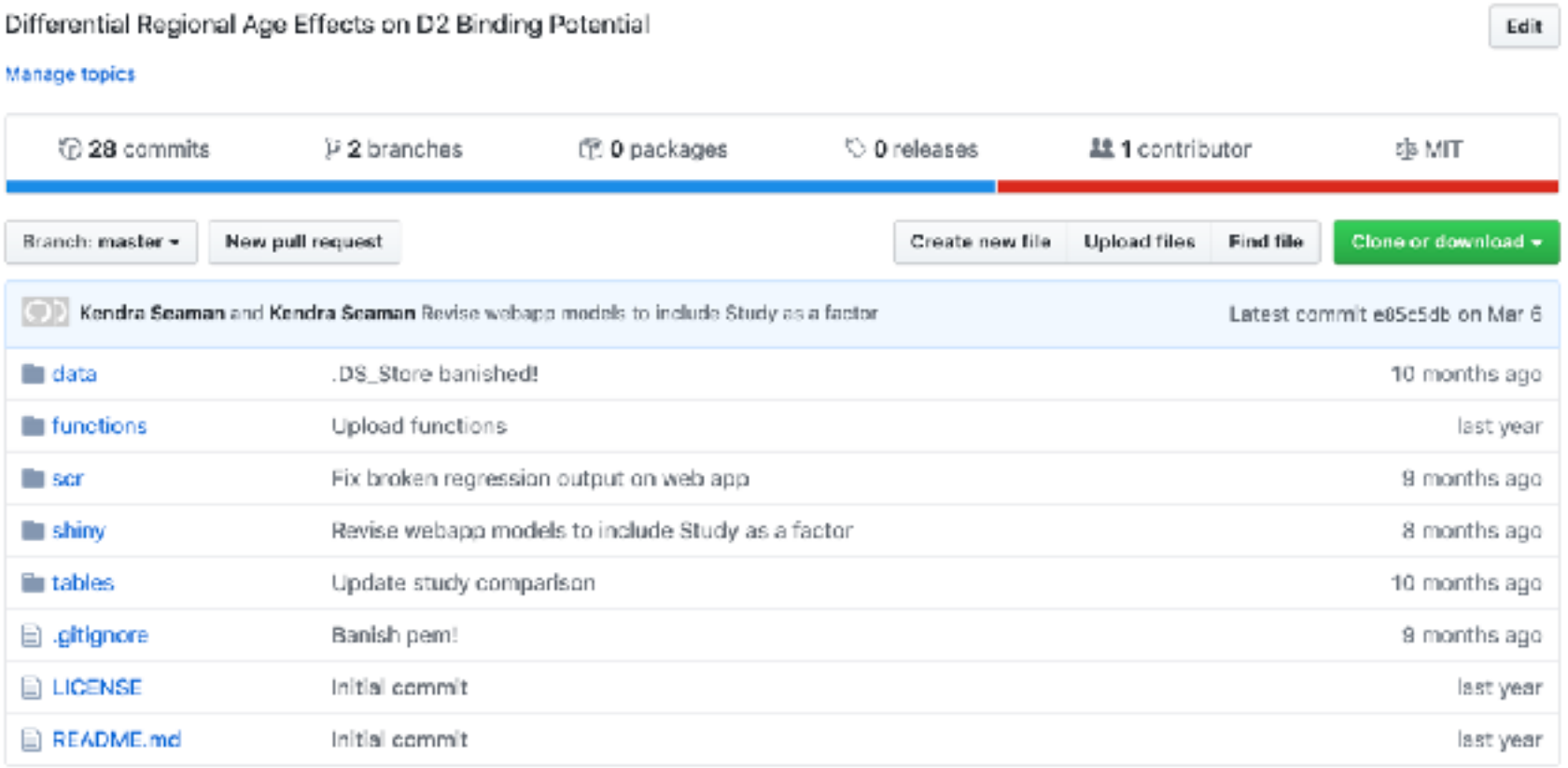
- Ideally, there is one of these for each dataset shared.
- Provides metadata for your dataset
  - Variable names
  - Readable variable names
  - Measurement units
  - Allowed values
  - Description

Sharing Best Practices: Data Dictionary

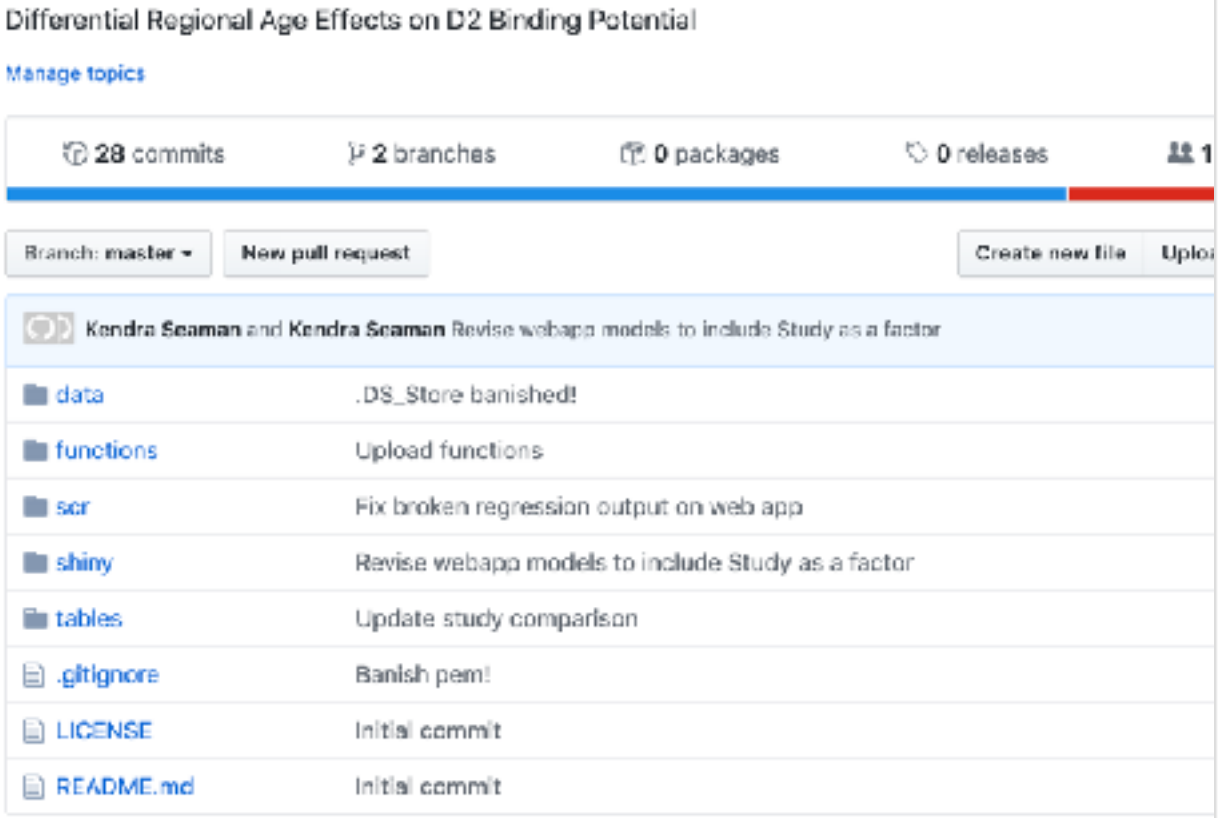
- Ideally, there is one of these for each dataset shared.
- Provides metadata for your dataset
  - Variable names
  - Readable variable names
  - Measurement units
  - Allowed values
  - Description
- Tutorial: https://help.osf.io/hc/en-us/articles/360019739054-How-to-Make-a-Data-Dictionary

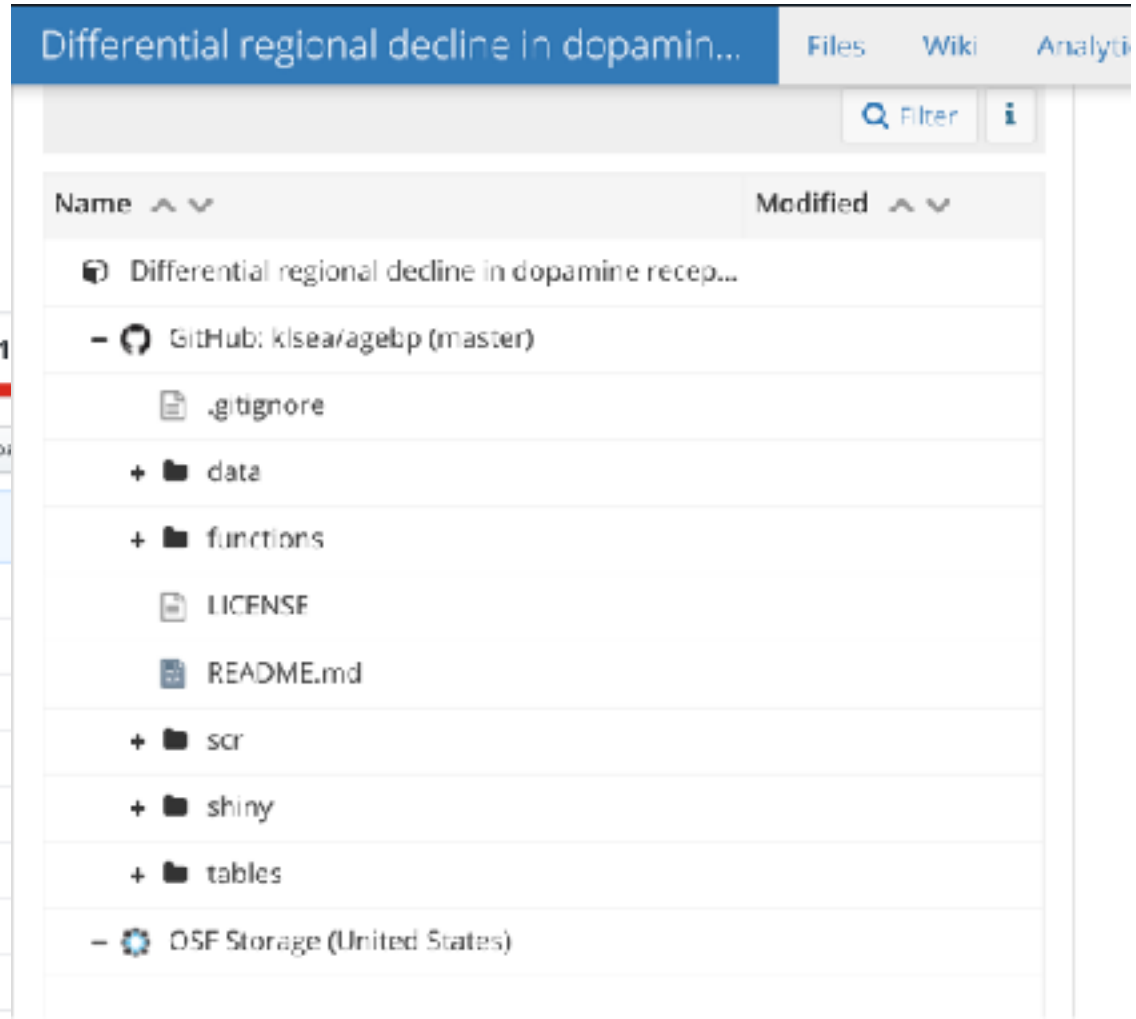
Sharing Best Practices - Link with GitHub

Sharing Best Practices - Link with GitHub



Sharing Best Practices - Link with GitHub





Sharing Best Practices - Link with GitHub

- My example - ageBP -
  - GitHub: https://github.com/klsea/agebp
  - OSF: https://osf.io/h67k4/

- More info: https://help.osf.io/hc/en-us/articles/360019929813-Connect-GitHub-to-a-Project

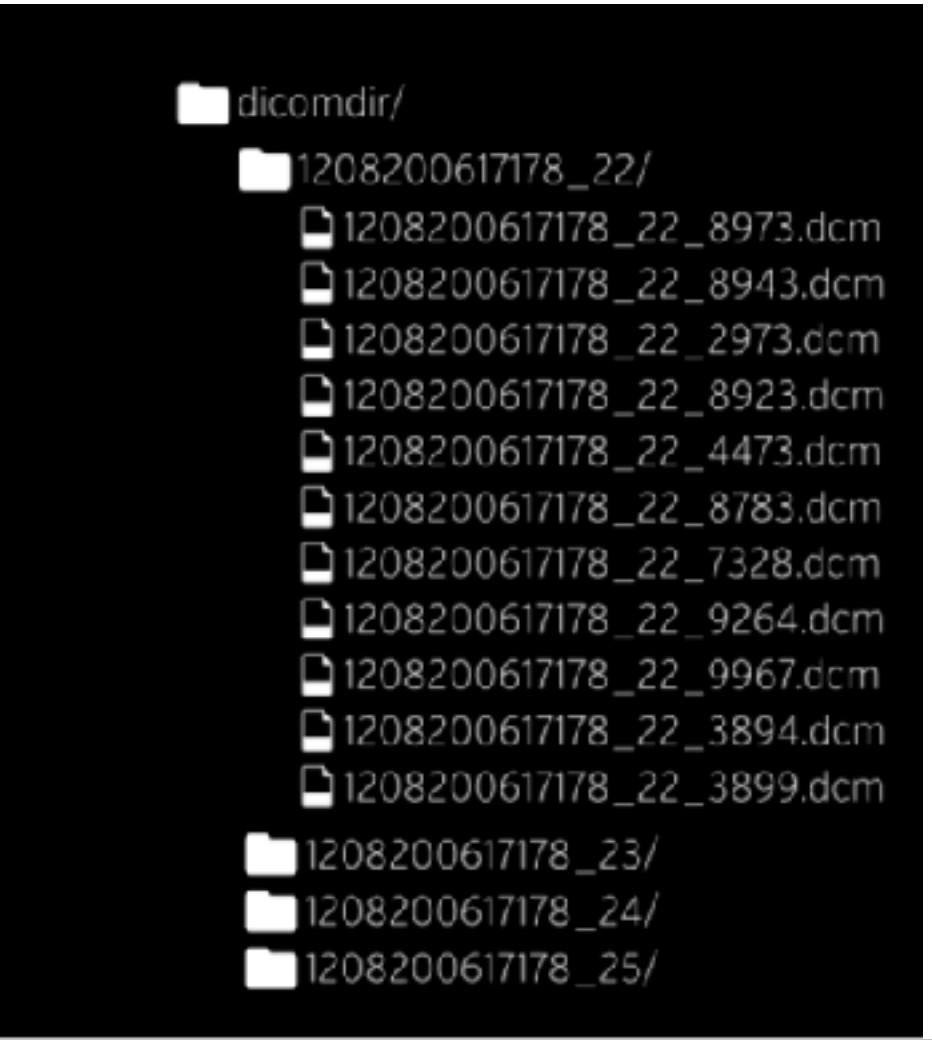
Sharing Best Practices - Other Resources

- Check out OSF website - https://help.osf.io/hc/en-us/categories/360001530634-Best-Practices

It’s Your Turn

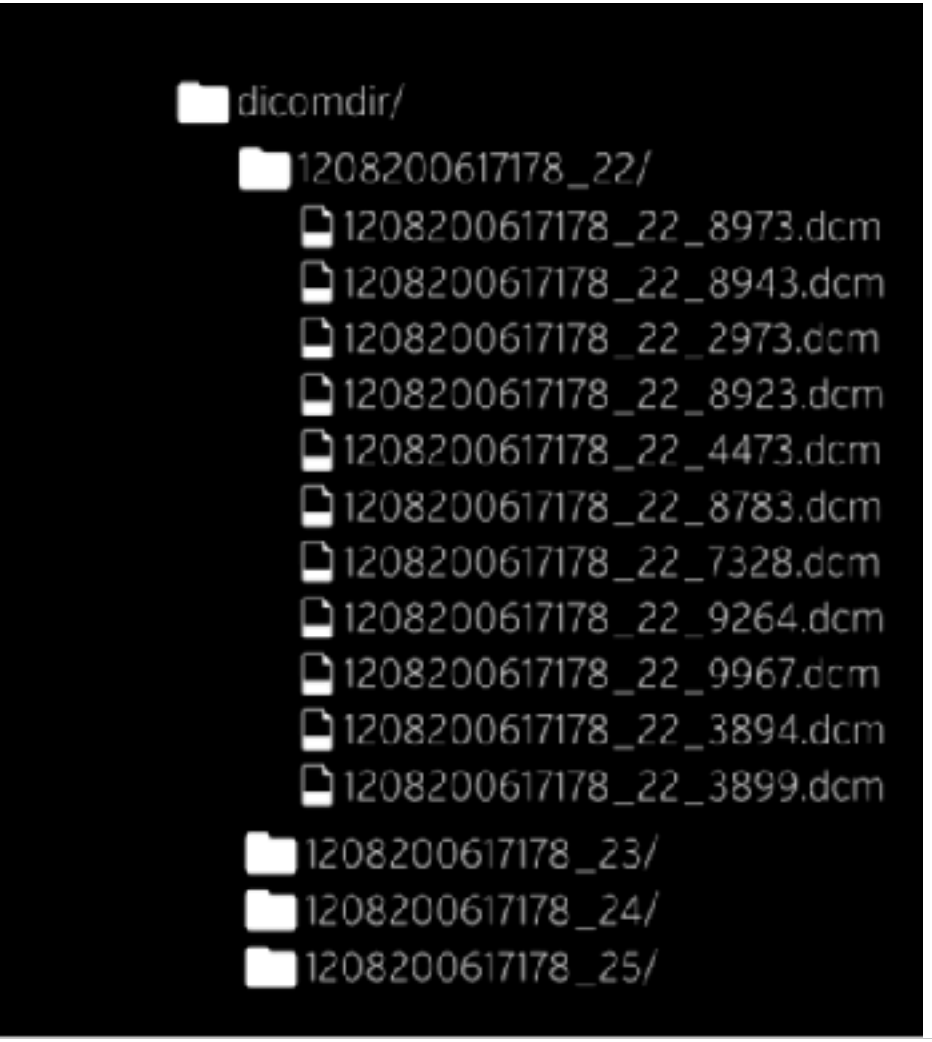
- Create an OSF account - https://help.osf.io/hc/en-us/articles/360019929673-Create-an-OSF-Account
  - UTD is an affiliated institution
- Create something!
  - OSF preregistration - https://help.osf.io/hc/en-us/articles/360021390833-Preregistration
  - OSF project page - https://journals.sagepub.com/doi/full/10.1177/2515245918757689
  - My lab project template (you’re welcome to use): https://osf.io/ce8p4/

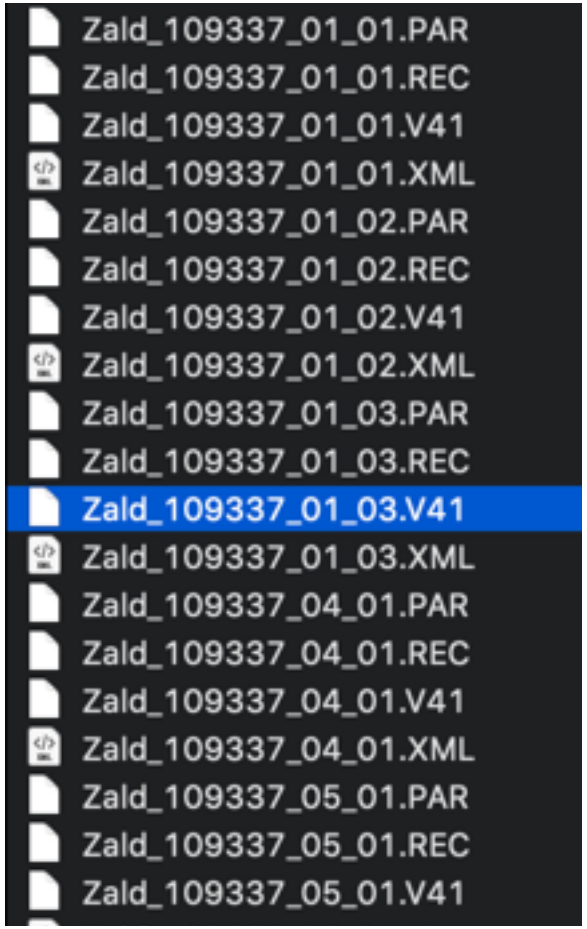
Standard format



https://bids.neuroimaging.io/

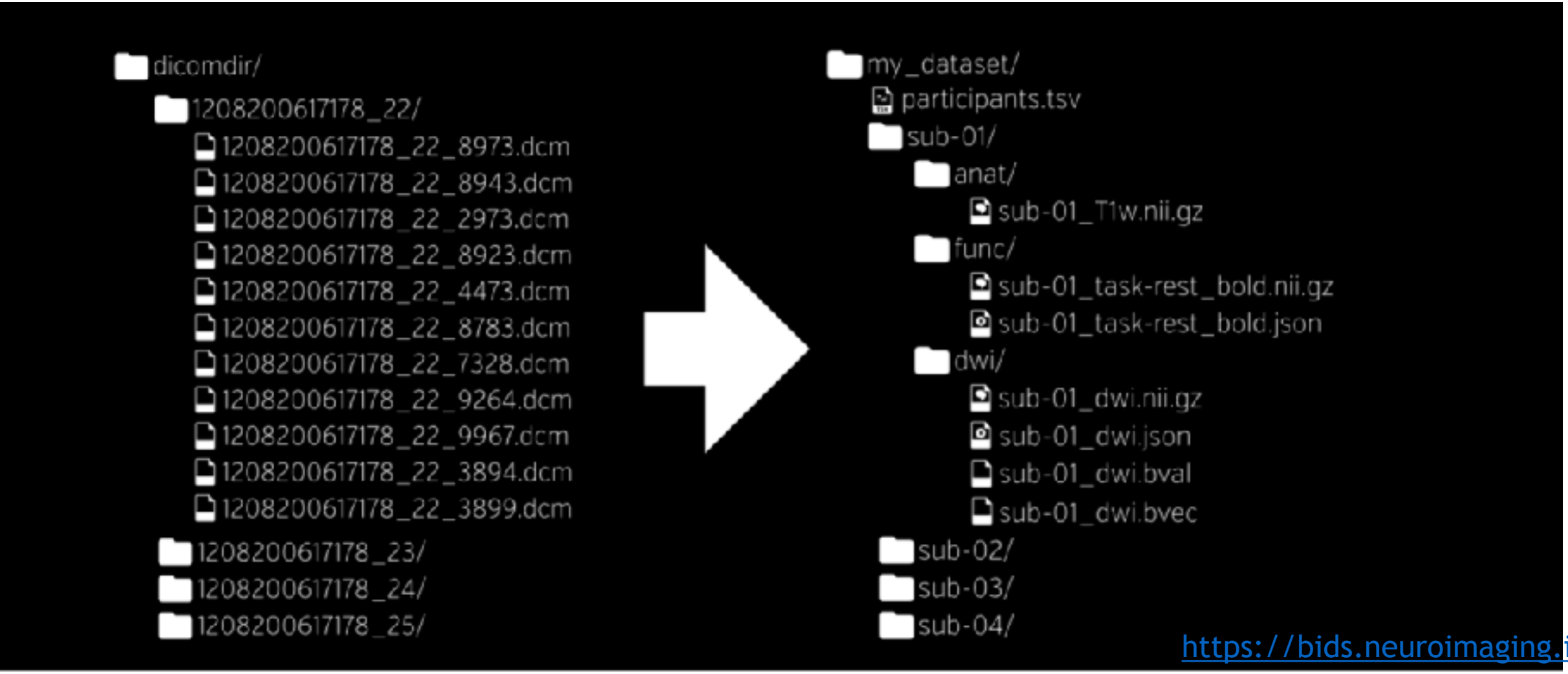
Standard format





https://bids.neuroimaging.io/

Standard format



https://bids.neuroimaging.io/

Standard format: BIDS

OpenNEURO

https://bids.neuroimaging.io/

Standard format: BIDS

OpenNEURO

- What is it?
  - All neuroimaging data is converted into as a .nii file.
  - Each .nii file has a corresponding .json with metadata for that file.
  - Critical metadata (e.g. subject ID) info included in filenames

https://bids.neuroimaging.io/

Standard format: BIDS

OpenNEURO

- What is it?
  - All neuroimaging data is converted into as a .nii file.
  - Each .nii file has a corresponding .json with metadata for that file.
  - Critical metadata (e.g. subject ID) info included in filenames
- Why does it help?
  - Metadata is both human AND machine readable
  - “Name Redundancy” reduces errors
  - Hierarchical structure and inheritance reduce *unnecessary* redundancy

https://bids.neuroimaging.io/

Standard format benefits

https://bids.neuroimaging.io/

Standard format benefits

- Data reusability
  - It’s easier for another researcher (or yourself in 6 months) to work with the data because the data and filenames are human-readable.

https://bids.neuroimaging.io/

Standard format benefits

- Data reusability
  - It’s easier for another researcher (or yourself in 6 months) to work with the data because the data and filenames are human-readable.
- Data sharing
  - Standardized with documentation so other researchers will understand the data organization and can more easily use the data

https://bids.neuroimaging.io/

Standard format benefits

- Data reusability
  - It's easier for another researcher (or yourself in 6 months) to work with the data because the data and filenames are human-readable.
- Data sharing
  - Standardized with documentation so other researchers will understand the data organization and can more easily use the data
- Study reproducibility
  - Standard organization will streamline the process to replicate the results from studies and write analysis pipelines that can be used across different datasets

https://bids.neuroimaging.io/

BIDS Resources

- Converter: https://github.com/nipy/heudiconv
- Validator: https://bids-standard.github.io/bids-validator/
- MRIQC: https://mriqc.readthedocs.io/en/stable/
- fMRIPrep: https://fmriprep.readthedocs.io/en/stable/