

Curating legacy resting-state EEG datasets using RS-BIDSify

Dan Brady

Research Software Engineer



Background

- Experiences of the #EEGManyLabs and EEG101 teams have shown that there are many researchers with “dormant” resting-state datasets
 - Old datasets, stored in idiosyncratic structures, unlikely to be analysed
- Would be very beneficial to have these out of the “file-drawer” and into the public domain

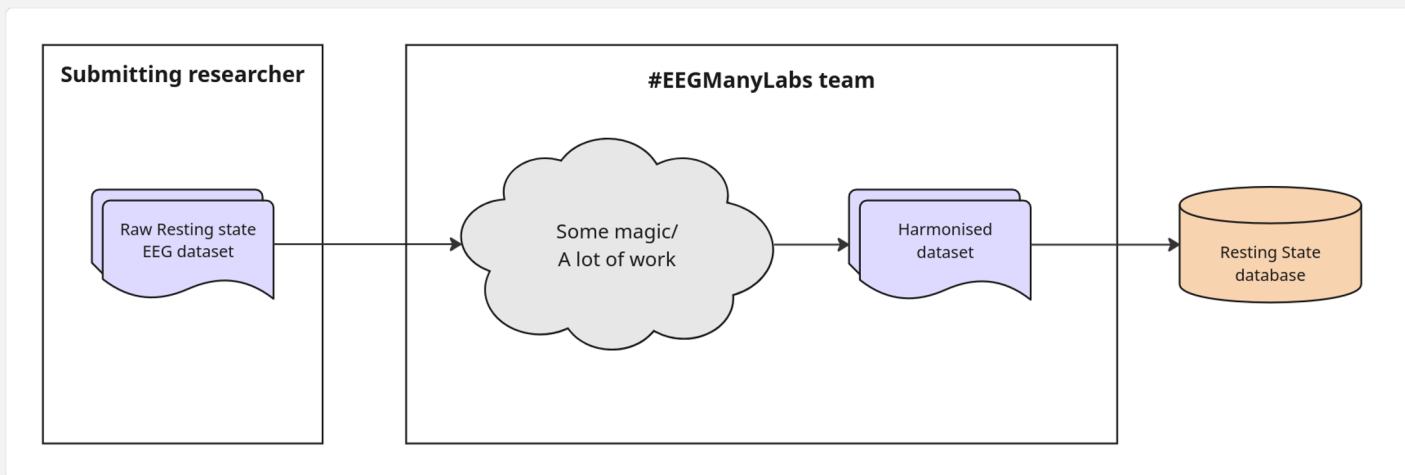
Overarching project aim:

**Curate a harmonised database of
donated resting-state EEG
datasets**



Background

- The idea of harmonising neuroimaging datasets into a database for broader use has been around for at least three decades (Gibbons, 1992; Chicurel, 2000)
- In practice this requires lots of work to successfully undertake



Pipeline design constraints

- The pipeline is specifically focussed on curating existing resting-state datasets
- Donation process should require minimal effort from the donator
 - Single form to fill in to gather metadata
 - Minimal requirements around data structure
- Final datasets should be Brain Imaging Data Structure (BIDS) compliant
- Use of non-proprietary tools for pipeline

Design decisions

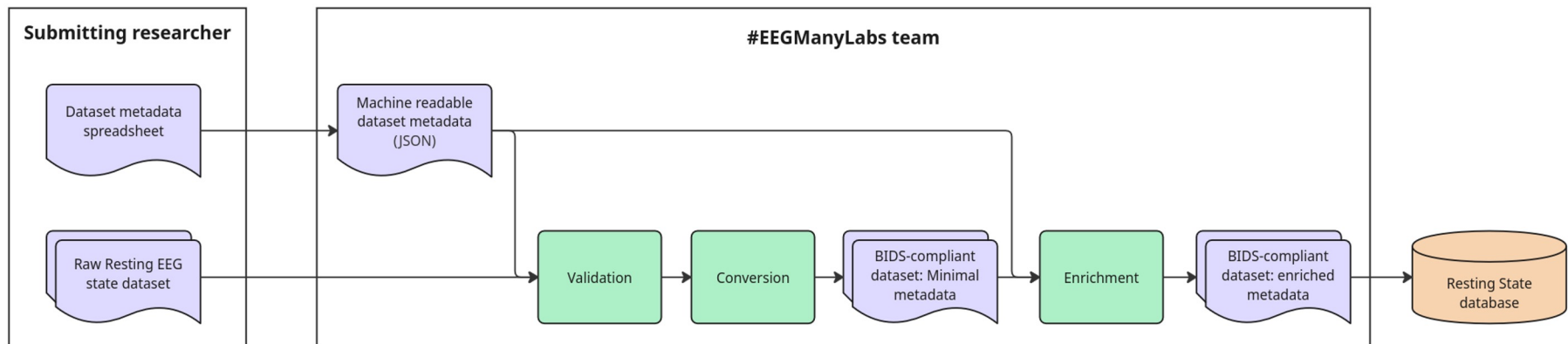
Donated data requirements:

- Spreadsheet will be used to collect metadata about datasets, including:
 - Acquisition information
 - Resting state conditions
 - Participant information
 - Phenotype information
- Basic directory structure:
 - Subject/Task/File
 - Task directory is optional

Design decisions

Processing pipeline:

- Built using Python, MNE, MNE-BIDS
- Metadata collected in spreadsheet will be converted into a machine readable JSON file
- The metadata collected will be used to enrich the BIDS data as much as possible
- Data validation and internal checks will be used to inform user of inconsistencies



Processing steps

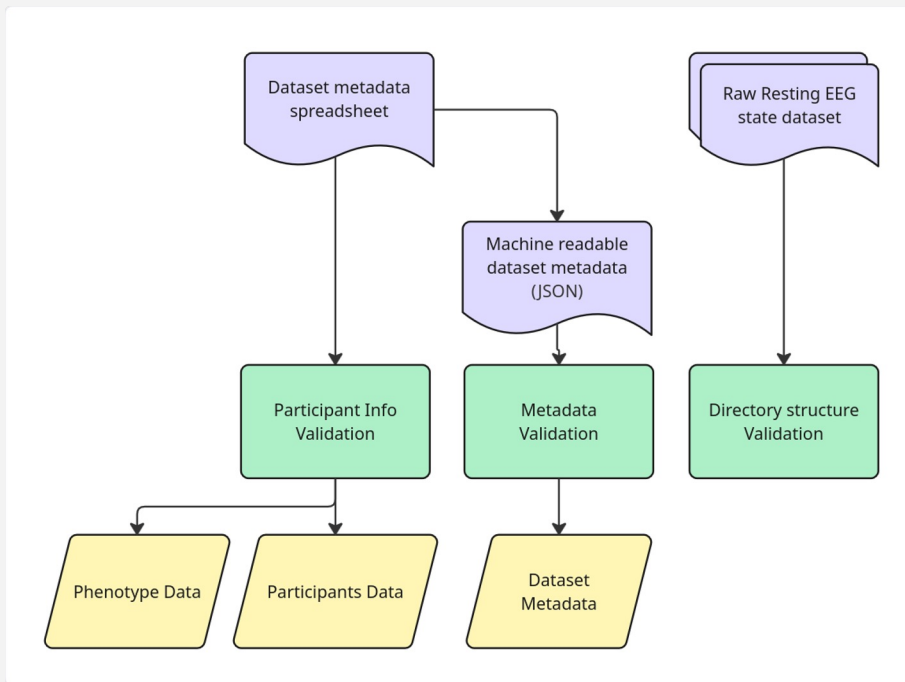


Validation

Reads and validates dataset information:

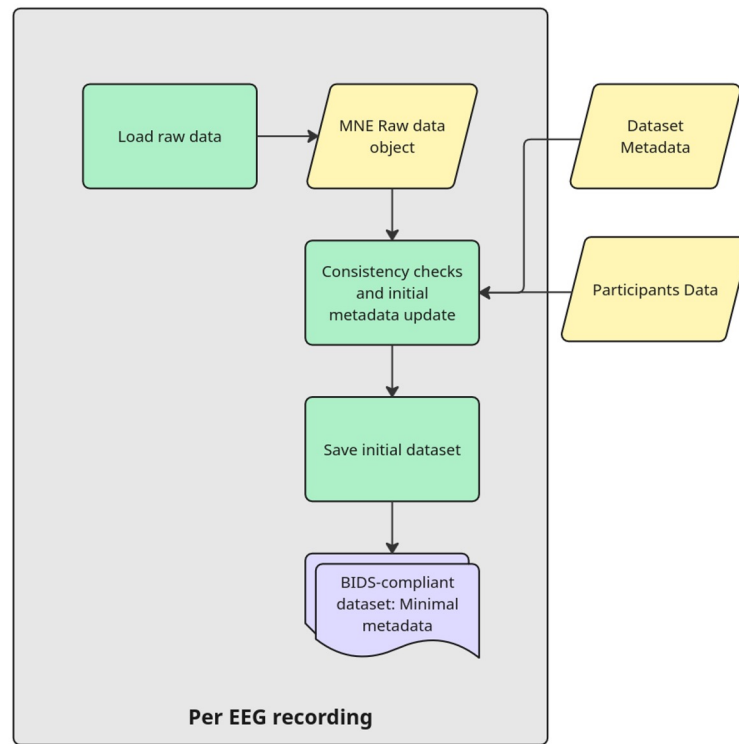
- Metadata
- Dataset structure
- Participant information

This step occurs before any EEG file is loaded for processing



Checking and Conversion

- Loads a Raw EEG recording
- Run consistency checks
- Initial metadata update
- Write minimally BIDS-compliant dataset



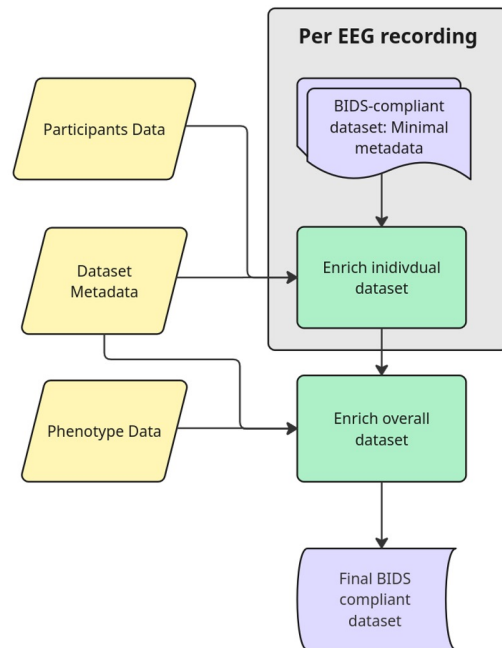
Enrichment

For each participant:

- Add additionally collected metadata to correct file

Finally:

- Update dataset description file
- Add phenotype dataset and codebook to dataset.



Walkthrough



Thanks

#EEGManylabs Core Team:

- Faisal Mushtaq
- Yuri Pavlov
- Dominik Welke
- Jasper Van Den Bosch
- Fangzhou Hu

N8-CIR Interns

- Ariana Williams
- Isaac Tugwell

ElectroTools Co-Is:

- Mahnaz Arvaneh
- Amy Bland
- Xun He

Special thanks (BEP-006):

- Cyril Pernet
- Stefan Appelhoff
- Robert Oostenveld



University of Sheffield

Get in touch:

Email: daniel.brady@sheffield.ac.uk

Github: [ubdbra001](https://github.com/ubdbra001)