Nipoppy: A framework for the organization and decentralized processing of neuroimaging-clinical data

Michelle Wang[1], Nikhil Bhagwat[1], Brent McPherson[1], Alyssa Dai[1], Rémi Gau[1], Qing Wang[2], Tristan Glatard[3], Jean-Baptiste Poline[1]

[1] McGill University, Montreal, Canada [2] Shanghai Mental Health Center, Shanghai, China
[3] Concordia University, Montreal, Canada

Email: michelle.wang6@mail.mcgill.ca, jean-baptiste.poline@mcgill.ca

Motivation

The **organization** and **processing** of neuroimaging and clinical data requires non-trivial effort and is prone to various errors.

Adoption of **best practices** and **clear documentation** can increase the **FAIRness** (findability, accessibility, interoperability, and reusability) (Wilkinson et al., 2016) of datasets and the reproducibility of studies.

We introduce **Nipoppy**, a collaborative, open and flexible framework that can help achieve **reproducible**, **decentralized** processing of ongoing neuroimaging-clinical studies.

Methods

The Nipoppy framework is made up of three components:

1) A **specification** for dataset organization (Fig. 1) that extends the Brain Imaging Data Structure (BIDS) standard (Gorgolewski et al., 2016) by providing additional guidelines for tabular (e.g., phenotypic) data and imaging derivatives. Two user-provided files are required: a **manifest** file listing the participant IDs and available visits in the dataset, and a **configuration** file with processing pipeline and parameter specifications (Fig. 1).

2) A **protocol** for data organization, curation and processing (Fig. 2), with steps that include the conversion of raw scanner output to BIDS, the processing of imaging data with existing or custom pipelines, the tracking of processing status, and the extraction of imaging-derived phenotypes (IDPs). Awareness of data availability enables the **automatic identification and processing of new data** with existing pipelines and configurations. We have also developed a **web dashboard** for interactive visualizations of imaging and phenotypic data availability (https://digest.neurobagel.org) (Fig. 3).

3) A **command-line interface** (Fig. 2) and **Python package** that provide user-friendly tools for applying the framework. The tools build upon existing technologies such as the **Apptainer** container platform (Kurtzer et al., 2017) and the **Boutiques** descriptor framework (Glatard et al., 2018) and makes them more accessible to the average neuroimaging researcher.

Results

Nipoppy has been used by several groups to organize data from studies such as the Parkinson's Progression Markers Initiative (PPMI) and the Quebec Parkinson Network (QPN). These datasets have been processed with pipelines such as fMRIPrep (Esteban et al., 2019), MRIQC (Esteban et al., 2017), and TractoFlow (Theaud et al., 2020) (Fig. 3).
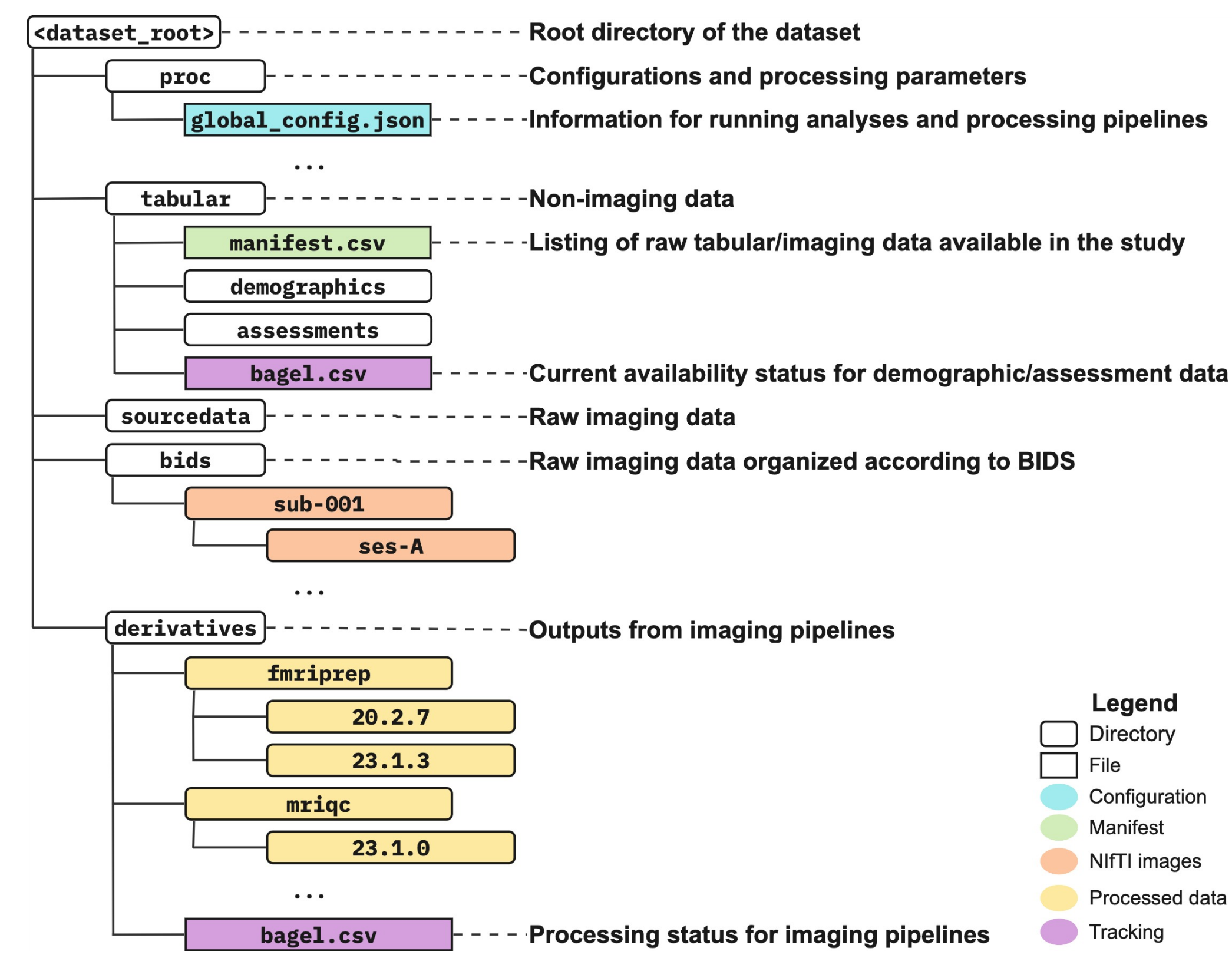
Conclusion

The Nipoppy framework can be used to establish a **decentralized data processing network**. Nipoppy also creates files compatible with the **Neurobagel** ecosystem for distributed dataset harmonization and search. The ultimate goal is for Nipoppy to become a **community standard**, and we welcome feedback and contributions from the broader neuroimaging community.
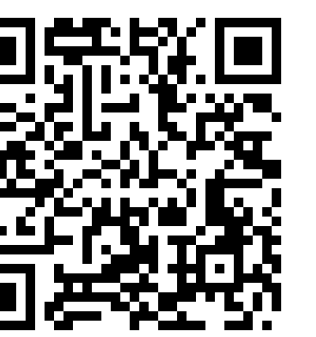
References

Esteban, O. et al. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, 12(9).

Esteban, O. et al. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1).

Glatard, T. et al. (2018). Boutiques: A flexible framework for automated application integration in computing platforms. *GigaScience*, 7(5).

Gorgolewski, K. J. et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1).

Kurtzer, G. M. et al. (2017). Singularity: Scientific containers for mobility of compute. *PLOS ONE*, 12(5).

Theaud, G. et al. (2020). TractoFlow: A robust, efficient and reproducible diffusion MRI pipeline leveraging Nextflow & Singularity. *NeuroImage*, 218, 116889.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018.

The Nipoppy framework

Data organization specification

Check out Nipoppy on GitHub!

Figure 1: Nipoppy specification (left) and example configuration files (right).

Protocol and software tools
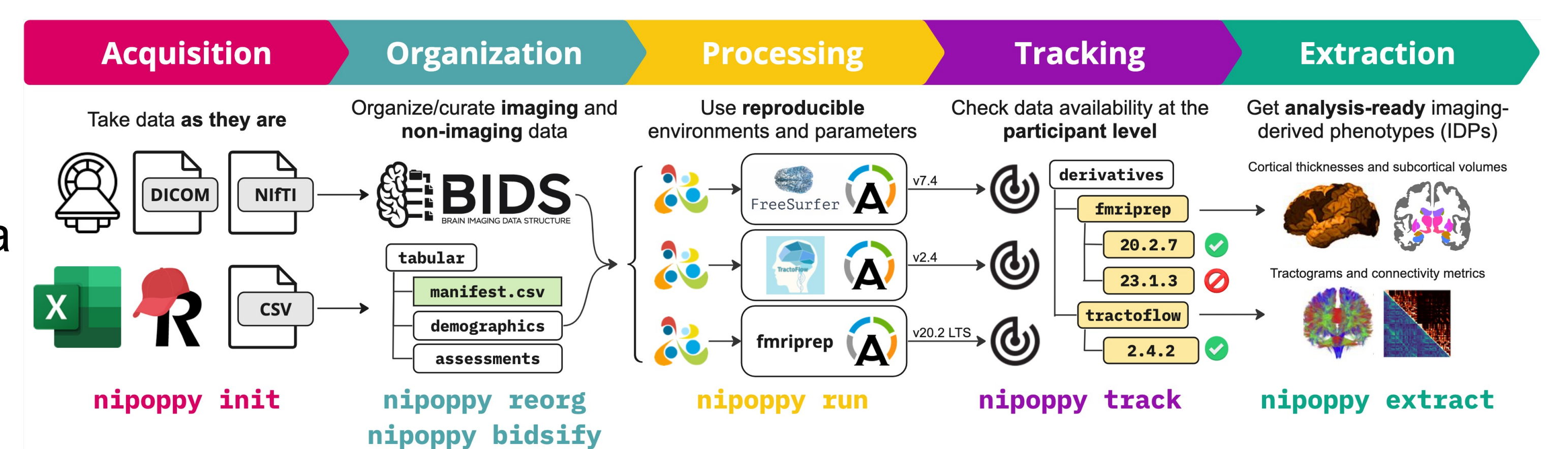
Figure 2: Main steps in the Nipoppy protocol and associated commands.

Figure 3: Nipoppy dashboard (https://digest.neurobagel.org) showing the availability of select imaging (top) and non-imaging (bottom) data for the PPMI dataset, stratified by session.

Acknowledgements