

# A deep dive into open neuroimaging workflows: towards a better understanding of the landscape of open codes in brain imaging

## Stage M2

### Supervision:

[Camille MAUMET](#), Research scientist at Inria, [Empenn team](#) ([camille.maumet@inria.fr](mailto:camille.maumet@inria.fr))

[Elise BANNIER](#), Research engineer at CHU Rennes, [Empenn team](#) ([elise.bannier@irisa.fr](mailto:elise.bannier@irisa.fr))

**Location:** Rennes - [INRIA](#) / [IRISA](#)

**Team:** Empenn

**Research themes:** Knowledge representation, open workflows, open data, brain imaging, Neuroinformatics, Reproducibility

**Keywords:** Analytical variability, FAIR data, data processing, open science, reproducibility, brain

**Duration:** 4-6 months

## Context

10 years ago, a series of publications pointed to the difficulty of reproducing scientific findings. This reproducibility crisis was a wake-up call for scientific communities to rethink how we practise research, and an important driver towards greater transparency and robust results. In particular, the FAIR principles, which stand for "Findable Accessible Interoperable and Reusable" (Wilkinson et al. 2016) have increasingly been used to describe and share open data. But for shared data to be truly reusable they must be accompanied by metadata that provide a complete description of the workflows that led to their creation.

Historically in neuroimaging, only a coarse description of those workflows was available in the final Methods section of the published manuscript. But in recent years, open science practices have flourished and an increasing number of workflows have been shared in the form of scripts and made available using software forges such as Github, Gitlab, etc (Poldrack et al. 2017). But so far those scripts are scattered which makes it extremely challenging to retrieve them and perform any type of query.

**The goal of this internship will be to use Software Heritage (Di Cosmo and Zacchiroli 2017) - the universal software archive supported by UNESCO - to search and retrieve neuroimaging workflows.** This will effectively make it possible to provide an overview of the open workflow landscape in neuroimaging and answer questions such as: "How many open neuroimaging workflows are available", "Are there datasets that are most often used by these workflows", "What type of imaging modalities are most often used by those workflows", etc.

This work will be co-supervised by Dr. Camille Maumet and Dr. Elise Bannier and is part of the PEPR Santé numérique ShareFAIR.

## Main activities

To reach this goal you will :

- Familiarise yourself with Software Heritage ;
- Learn about neuroimaging workflows in Magnetic Resonance Imaging ;
- Read scientific papers on data sharing
- Develop code to extract information from Software Heritage about neuroimaging workflows ;
- Perform a quantitative analyses to better understand the landscape of open neuroimaging workflows

## Références

- Di Cosmo, Roberto, and Stefano Zacchiroli. 2017. "Software Heritage: Why and How to Preserve Software Source Code." In *iPRES 2017 - 14th International Conference on Digital Preservation*, 1–10. Kyoto, Japan. <https://hal.science/hal-01590958>.
- Poldrack, Russell A., Chris I. Baker, Joke Durnez, Krzysztof J. Gorgolewski, Paul M. Matthews, Marcus R. Munafò, Thomas E. Nichols, Jean-Baptiste Poline, Edward Vul, and Tal Yarkoni. 2017. "Scanning the Horizon: Towards Transparent and Reproducible Neuroimaging Research." *Nature Reviews Neuroscience* 18 (2): 115–26. <https://doi.org/10.1038/nrn.2016.167>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.