



Reproducibility in the analysis of brain data

What:

The WG Neuroimaging aims to create a community around the special case of reproducibility in the analysis of brain data.

The stark realisation that scientific results do not always readily replicate has led some to investigate the root causes of the so-called “reproducibility crisis”. Such self-critical appraisal has been so far more prevalent in Psychology and Neuroscience than in other disciplines, and typically highlight statistical issues, like inadequate statistical designs, as well as poor computational training; problems that are only likely to worsen as data grow larger, become more widely shared, and advanced techniques are imported from fields of engineering, like machine learning.

Specifically, neuroimaging data, in both clinical and fundamental research, have the particularity that they involve a large number of processing steps on a very heterogeneous set of equipment and infrastructures, from the

moment they are gathered in proprietary devices (magnetic resonance imaging scanners, electro-encephalography systems, etc) through preprocessing, analysis to annotation, curation and finally deposited into open repositories for others to use in upstream research. A lot of this pipeline remains an error-prone, manual process that relies on the researcher's voluntary (and unpaid) efforts to acquire an understanding of the infrastructure available and their technical knowledge to use it, to ensure the traceability and provenance of the data, the reproducibility and replicability of the work, and the production of FAIR open datasets.

The successful integration of such data into routine neuroimaging practice thus requires neuroscientists to develop skills that fall outside of ordinary training curricula, which should also include data curation, data handling, high performance and on-demand computation (in the "cloud"), semantic web annotation, as well as statistics suitable for large scale inference. The researchers who have been the most receptive to exploring and developing such techniques are typically early career researchers, motivated by the desire to learn, apply and share robust practices. The WG seeks to alleviate some of the biggest challenges they face: They are not formally trained and teach themselves these new data practices following online resources, in isolation and on a voluntary basis. We want to fill this gap of support, by pooling interests, experiences and expertise into a platform available globally.

Continuous activities of the RDA ambassadors on the EOSC Future website:

Over the 18-month of the WG:

- A call for participation in the community will be announced, and support will be arranged where it is need, in the form of financial support to plan and budget training, scaffolding and agreement with host institutions to carve training time for participants in the community, etc.
- Monthly meetings will be planned with the objective of scoping needs of the community and form working groups. Such projects may include, for instance:
 - the creation of training curricula, including teaching material, exercises and pedagogical documentation;
 - the review of typical analytical pipelines, and identification of best practices;
 - or the evaluation of current data sharing practices, and scoping of what might a sustainable practice look like as neuroimaging incorporates tools and methods in database management, machine learning and advanced statistical methods.

Who:

Etienne Roesch is an applied statistician and researcher at the University of Reading

Scientific Domain:

Software engineering

Your Promotion and Networking:

Promotion of domain specific Open Science practices/EOSC/RDA at meetings and conferences, publications and presentations. For example: ORCID/LinkedIn profiles, project homepages, document repositories etc.

<http://repro.school>

Your Outputs:

<https://orcid.org/0000-0002-8913-4173>

<https://etienneroes.ch>

Country:

United Kingdom

EOSC FUTURE FUNDING PLATFORM