

The ReproPub: A hybrid research object for supporting publication-level re-execution and generalization of neuroimaging research findings

David N. Kennedy
Eunice Kennedy Shriver Center, Department of Psychiatry
University of Massachusetts Medical School
Worcester, MA, United States
David.Kennedy@umassmed.edu (0000-0002-9377-0797)

Submission type: Abstract for oral communication

Keywords: reproducible research, re-execution, data publication, software publication, containers

ABSTRACT

In this report we introduce the ‘ReproPub’, a publication that includes the complete provenance of its experimental data, workflow, execution environment, and results. The ReproPub concept supports re-executability of the original finding, which, it is argued, supports a more systematic exploration of the generalizability of the finding and hence enhances the evaluation of its reproducibility.

BODY

ReproNim, a Center for Reproducible Neuroimaging Computation (repronim.org), is a NIH-funded Biomedical Technology Research Center (BTRC) that seeks to facilitate the “last mile” implementations of core re-executability tools in order to reduce the accessibility barrier and increase adoption of standards and best practices at the neuroimaging research laboratory level. ReproNim’s highest-level goal is to promote adoption of a more reproducible neuroimaging research process in order to promote ‘publication-level’ reproducibility and consequently ‘claims-level’ generalizability. Our premise is, however, that to truly approach these levels of generalizability and replication, we need to facilitate the evolution of the research publication from a pdf document that announces some observations and claims, to a document that completely describes the basis (data and process) upon which the experimental observations and claims are founded. The presentation of claims in a re-executable fashion, which we refer to as a **ReproPub**, facilitates an explicit and principled exploration of the generalizability of the claim; a claim that generalizes is one that by definition is reproducible (Figure 1).

A completely re-executable research publication requires the complete description (i.e. provenance) of the: experimental data, workflow, execution environment, and results that are used to establish the claim. In fact, each of these elements (experimental data, workflow, execution environment, and results) are themselves ‘research objects’¹ making the ReproPub an overarching mechanism to aggregate these subsidiary research objects together in support of a specific set of claims. Furthermore, they come together in a fashion such that each of these objects has its own history, evolution, creators, credit, and reusability. Provenance (the answering of “where did I come from”) for each of the element objects (and then the future reuse and citation of these elements) together creates a more explicit ‘graph’ of the research process, and enhances the community’s ability to refine, generalize, reason over, and aggregate support (or refutation) of specific claims.

Ghosh, et al, (2017) [1] published a ‘simple re-executable’ publication as a proof of concept. The key points were to document that even within the constraints of current publication practices, it is possible to aggregate the connection of a research publication to the data (via a DOI of the imaging data utilized), the processing workflow (via the DOI of a tagged GitHub release of the processing workflow), the analysis environment (via a DockerHub release of the Docker container of the analysis environment and workflow) and the

¹ <http://www.researchobject.org/>

archival of the complete results (as part of the tagged GitHub release). In this fashion, the publication both supports the ‘exact’ re-execution of the publication, by reuse of the exact data and processing environment; and enables the exploration of the sensitivity of the results by supporting comparison of results, if the analysis were to be performed using other execution environments or workflow design.

A ReproPub embraces many of the recent advances and evolutions in publication: treatment of data as a first-class object [2]; the principles of software citation [3]; the FAIR (findable, accessible, interoperable and reusable) [4] principles applied to the scientific process itself. By supporting the explicit use (aggregation) of research objects in support of building the ‘graph’ of scientific reasoning and the exploration of the stability and generalizability of findings and claims that emerge in a principled manner, we believe that this scientific literature can be rendered in a more reproducibility-supportive fashion. Culturally, this evolution of publication practice should be perceived as a plus for the scientific community. Specifically, what used to be one publication that referred to data, processing, and a set of claims, can now conceivably become numerous publications of distinct and independently creditable scientific output: a publication for the data, a publication for the processing approach, a publication for the complete results, in addition to the publication for the conclusions and claims.

This work was supported by: NIH-NIBIB P41 EB019936 (ReproNim), and NIH-NIMH R01 MH083320 (CANDIShare).

REFERENCES

- [1] S. S. Ghosh et al., “A very simple, re-executable neuroimaging publication,” *F1000Res*, vol. 6, p. 124, Jun. 2017.
- [2] L. B. Honor, C. Haselgrove, J. A. Frazier, and D. N. Kennedy, “Data Citation in Neuroimaging: Proposed Best Practices for Data Identification and Attribution,” *Front. Neuroinform.*, vol. 10, p. 34, 2016.
- [3] D. S. Katz et al., “Software Citation Implementation Challenges,” arXiv:1905.08674 [cs], May 2019.

- [4] M. D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data*, 15-Mar-2016. [Online]. Available: <https://www.nature.com/articles/sdata201618>. [Accessed: 20-Aug-2018].

Figure 1. The ReproPub as a traditional publication that includes explicit indication of the exact data, workflow, operating system and results in a fashion that can be re-executed to verify and explore the generalizability of the results.

