

Introducing the Associate Editor of Reproducibility

Ben Marwick 

Advances in Archaeological Practice is pleased to introduce Dr. Ben Marwick as the Associate Editor of Reproducibility, following the creation of this editorial role at *American Antiquity* (Martin 2024).

When a manuscript that uses the R, Python, or similar open-source programming language for data analysis and visualization is accepted with revisions, and when the authors opt in to a reproducibility review, the Associate Editor of Reproducibility (AER) will be asked to review the code and data. The AER can conduct the reproducibility review or invite peer reviewers to do the code review. Here are the typical steps that the AER will undertake for a code review, which should take less time than a traditional peer review:

1. Obtain the code and data files and record in the review where and when these files were obtained; for example, “The code and data files were downloaded from <https://zenodo.org/records/10623970> on 20 Feb 2024.” This step is important because sometimes authors have multiple copies (e.g., GitHub and Zenodo), and the files can change rapidly during the review process. Ideally the files will be obtained from a version-controlled repository via a DOI link. Authors should include the DOI that links to their code and data in their Data Availability Statement. Authors using data that cannot be shared for ethical or legal reasons can have their code reviewed using synthetic data.
2. Record the version number for R or Python that is being used for the review. Authors should state in the text of their manuscript what version they used, and the reviewer will either use that version or the most recent one.
3. Review and comment on the author’s README, if there is one. Authors should include a plain text document called README that briefly describes the purpose and contents of their files, gives version numbers of the key software packages used, and provides any instructions that typical users will need to get the code working on their computers. If running the code takes a long time, the approximate time to generate the results should be noted by the authors in the README.
4. Review and comment on the structure of the compendium (see Marwick et al. 2018) and file names. Authors should follow these general principles: it should be easy to navigate (names for files and folders should make it easy for others to understand what they contain), and the methods (i.e., code files) should be separate from the data. File formats should be appropriate for the type of data they contain (e.g., a table of numeric data is more accessible to others in a CSV file than a

PDF or Word document). Folder and file names should not have spaces and punctuation other than hyphens and underscores. If there are multiple files, their names should indicate the order that the files should be accessed (e.g., 01-prepare-data.r, 02-analyse-data.r).

5. Run the code and record any errors that appear. If the AER can solve the error with minimal effort, they will do so and record the solution in their review. If the AER cannot solve the problem with roughly 5–10 minutes of investigation, they should record the line number of the code that caused the error and the full text of the error to share with the authors so they can arrive at a solution. It is strongly recommended that authors reporting on analyses that use machine learning or Bayesian methods include a Dockerfile in their compendium that specifies the entire computational environment of their analysis. In our experience, doing so substantially reduces errors in the review process, and thus the time required for the reproducibility review.
6. If the code can be run successfully, this will be clearly stated in the review, and the specific figures and tables that could be reproduced will be noted. For very long running code, the AER can choose to run it for a short time (e.g., a few hundred of several hundred thousand iterations) to verify that the basic workflow is sound. The AER will not be expected to fully reproduce time-consuming, highly computationally intensive analyses. When the code is not fully run, this will be noted in the review.
7. The AER will note whether the code style is readable—that is, the authors made effective use of white space to organize the code into meaningful units—and code comments are adequate (e.g., the authors made it easy to see the correspondence between code outputs and figures and tables in the manuscript).

The AER will then send the review to the editors and the authors. It is confidential and for the benefit of the authors. If the AER found that the authors’ code and data could be used to reproduce some or all of the results in the article, the AER prepares a very short report (two to three sentences) describing their attempt; for example, “The Associate Editor for Reproducibility (Ben Marwick) downloaded all materials and reproduced results in all figures and tables.” If only certain figures or tables could be reproduced, they will be itemized. The AER report is included in the published Version of Record of the article after the Data Availability Statement for readers to see under the heading “Reproducibility Statement.” The AER will also recommend an Open Data Badge for the article (see <https://osf.io/tyyxz/wiki/1.%20View%20the%20Badges/>).

Advances in Archaeological Practice 12(2), 2024, pp. 61–62

Copyright © The Author(s), 2024. Published by Cambridge University Press on behalf of Society for American Archaeology
DOI:10.1017/aap.2024.15

When the AER finds that the code did not work, the authors have two options. They can do nothing, in which case a Reproducibility Statement will be attached to the end of their article that reports that their analyses was not reproducible. Or the authors can work with the AER to revise their code and then confirm with the AER that the code works. The article will then have a Reproducibility Statement confirming that the AER could reproduce the results presented by the authors.

In our experience, most fixes are straightforward and can be handled quickly, so we assume that most authors would be willing to make changes to their code to make it work and receive a Reproducibility Statement that validates their analysis. Given that papers were already accepted before the code and data review, rejection is only possible at this stage when very significant problems are identified.

To conclude, the purpose of these reviews is to support the reproducibility and reusability of research code and data for future researchers. The reproducibility reviewers are not “data thugs”

hunting for flaws or fraudulent research. On the contrary, these reviews aim to validate and publicly celebrate authors’ efforts to share high-quality, useful research to benefit the community long into the future. This initiative contributes to making archaeology more transparent and accessible and a source of trustworthy and reliable information about the human past.

REFERENCES CITED

- Martin, Debra L. 2024. Editor’s Corner. *American Antiquity* 89(2):163–164.
Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. Packaging Data Analytical Work Reproducibly Using R (and Friends). *American Statistician* 72(1):80–88. <https://doi.org/10.1080/00031305.2017.1375986>.

AUTHOR INFORMATION

Ben Marwick ■ Department of Anthropology, University of Washington, Seattle, WA, USA (bmarwick@uw.edu)