Achieving Transparency and Replicability: A Data Curation, Verification, and Publication Workflow

Thu-Mai Christian
Odum Institute for Research in Social Science
University of North Carolina at Chapel Hill
thumai@email.unc.edu

Sophia Lafferty-Hess
Odum Institute for Research in Social Science
University of North Carolina at Chapel Hill
slaffer@email.unc.edu

ABSTRACT
In this poster, we illustrate the workflow developed by the Odum Institute for Research in Social Science Data Archive to support the curation and verification of replication data files for the *American Journal of Political Science*.

General Terms
Preservation strategies and workflows

Keywords
Data curation, Data quality, Replication, Verification

1. POSTER SUMMARY
In a move to promote open scientific inquiry, several major journals have issued policies requiring authors to make the data underlying results presented in their published articles openly available to others, which enables verification and replication of findings. In doing so, the journals protect the integrity of the scientific record [4] while also enhancing the visibility and impact of research [5]. Unfortunately, submission of dataset files to a repository has not been adequate to ensure the long-term preservation and reuse of these data for these purposes.

In the frequently cited article, "Replication, Replication," Gary King asserts that the "replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author" [3]. In an effort to uphold King's replication standard, the editorial staff of the *American Journal of Political Science (AJPS)* recently issued a revision to its data availability policy, which had already required authors to upload replication files to a designated open access repository prior to submission of the final manuscript for publication [1]. Despite this initiative, the *AJPS* editorial staff has recognized the varying quality of replication files currently housed in the *AJPS* data repository [2]. In response to this, the new *AJPS* replication policy revision stipulates that article publication is contingent not only on the submission of supporting files—including the data, programming code, codebooks, and other explanatory text—but also the successful replication of tables and figures in the final manuscript using the submitted files [1].

The Odum Institute for Research in Social Science Data Archive has been tasked to perform the verification of replication datasets and ensure the comprehensiveness of submissions. Even with guidance provided to authors on how to prepare replication files, the quality of data submissions has varied, with only a fraction able to reproduce tables and figures in final manuscript drafts on the first attempt. Missing codebooks, incomplete or non-commented programming code, rounding errors, mismatched figures, and an array of other issues have added complexity to both the publication and data curation and verification workflows. Because of this, it has been necessary to develop a standard, integrated workflow that relies significantly on cooperation of and coordination between the author, editor, and data archive in order to ensure that submitted files meet quality standards for both replication and preservation and reuse.

This poster will outline the human-driven workflow to archive, verify, and link replication data to their associated journal publications, as well as its integration into the scholarly publication workflow. The poster will also describe critical issues, key lessons, and potential opportunities for archives working to preserve scholarly assets to help sustain the research enterprise.

2. ACKNOWLEDGMENTS
The authors would like to thank the editorial staff of *AJPS* and the editor of *AJPS*, Dr. William Jacoby, for their assistance in the development and implementation of the data curation, verification and publication workflow. We would also like to thank the statistical consultants at the Odum Institute who perform the meticulous task of verifying that the submitted replication files can successfully reproduce the findings in final manuscripts. They play a key role in achieving the goal of research transparency and replicability.

3. REFERENCES
[1] Jacoby, W. (2015, May 26). The AJPS replication policy: Innovations and revisions. Retrieved from http://ajps.org/2015/03/26/the-ajps-replication-policy-innovations-and-revisions/

[2] Janz, N. (2015, May 4). Leading journal verifies articles before publication--So far, all replications failed [Blog]. Retrieved from https://politicalsciencereplication.wordpress.com/2015/05/04/leading-journal-verifies-articles-before-publication-so-far-all-replications-failed/

[3] King, G. (1995). Replication, replication. PS: Political Science & Politics, 28(3), 444. http://doi.org/10.2307/420301

[4] National Research Council & Committee on National Statistics. (1985). Sharing research data. (S. E. Fienberg, M. E. Martin, & M. L. Straf, Eds.). Washington, D.C.: National Academy Press. Retrieved from http://www.nap.edu/openbook.php?record_id=2033

[5] Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. http://doi.org/10.7717/peerj.175