The Reproducible Research Movement in Statistics

Victoria Stodden
Department of Statistics
Columbia University
vcs@stodden.net

Key Words: Reproducible Research, Open Data, Open Code, Open Science, Science Policy

It is now widely recognized that the traditional published article is insufficient to permit verification of computational results. The emergence of powerful computational hardware combined with vast data collection and storage capabilities presents many novel opportunities for researchers. Unfortunately the standards for communication of published computational findings have evolved in ways that make verification and validation next to impossible and impede the ability of others to build on past research [1-3]. A movement toward reproducible research – dissemination that includes sufficient experimental details such that results can be replicated by others in the field, i.e. the code and the data – has developed in many disciplines and research areas to address this shortcoming in research [9]. In this paper I will present a jointly developed set of standards to guide the dissemination of reproducible research, and discuss changes originating outside academia that affect computational and empirical research dissemination including recent journal publication and federal agency dissemination requirements.

Standards and Scientific Communication

In December of 2012, more than 70 computational scientists and other stakeholders such as journal editors and funding agency officials gathered at Brown University for the ICERM Workshop on "Reproducibility in Computational and Experimental Mathematics." This workshop provided an opportunity for a broad cross section of computational scientists from fields as diverse as mathematics to biology to discuss these issues and brainstorm ways to improve on current practices. The result is a series of recommendations intended to establish reproducible computational science as a standard [4]. The main recommendations are that:

1. It is important to promote a culture change that will integrate computational reproducibility into the research process.
2. Journals, funding agencies, and employers should support this culture change.
3. Reproducible research practices and the use of appropriate tools should be taught as standard operating procedure in relation to computational aspects of research.

Changing the Culture. Early in their career, bench scientists and experimental researchers are taught to maintain lab notebooks or computer logs for all experimental details, including protocol, procedures, equipment, data collection details and raw results, processing techniques, statistical methods, etc. At this stage, very few computational experiments are performed with the same care. In computational science, there is typically no record of workflow, computer hardware and software configuration, software used, or parameter settings. Source code is often lost, or is changed with no record of the revisions.

The research system must offer rewards for reproducible research at every level from departmental decisions to grant funding and journal publication, incorporating notions of code and data sharing into institutional

promotion and hiring and grant proposal review. The current academic research system places the primary emphasis on publication and little on reproducibility. This has the effect of penalizing those researchers that produce reproducible computational research. Software development has been characterized as support of science rather than doing *real science*. The result is that scientists are discouraged from spending time writing, testing, or releasing code. With the ever-increasing pervasiveness of computation and software across the research landscape, such attitudes and practices must change.

Funding Agencies, Journals, and Employers Must Support This Change. Software and data should be "open by default" and access only restricted if openness conflicts with other considerations such as confidentiality. Grant proposals involving computational work could be required to detail standards for: dataset and software documentation including reuse (some agencies already have such requirements [5]); persistence of resulting software and dataset preservation and archiving; sharing resulting software among reviewers and other researchers. Funding agencies could add "Reproducible Research" to the list of specific examples that proposals could include in their "Broader Impact" statements. Software and dataset curation should be explicitly included in grant proposals and recognized as a scientific contribution by funding agencies, and funds made available to support it. Although some exist already, further templates for data management plans could be made widely available that include making software accessible, perhaps by institutional archiving and library centers [6].

Editors and reviewers must insist on the full disclosure of computational details and rigorous verification and validity testing should be expected by authors [7]. Some experimental details might appear on a website with a persistent URL. Authors need to state intended exceptions for full disclosure, such as for proprietary, medical, or other confidentiality issues, upon submission, and reviewers and editors must agree such exceptions are reasonable. All software and data used in a paper should be citation, not merely mentioned in the text or in a footnote. Proper citation is essential both for improving reproducibility and to provide credit for making available software and data, which is a key component in encouraging the desired culture change [8].

The third source of influence on the research process stems from tenure and promotion committees and research managers at research labs. Software and dataset contributions should be rewarded as part of expected computational research practices. Data and code citation practices should also be recognized and expected in computational research.

Teaching and Tools for Reproducible Research. The skills required to carry out and disseminate reproducible research in the computational sciences should be taught as part of scientific methodology, along with teaching modern programming and software engineering techniques. This should be a standard part of any computational science curriculum, just as experimental or observational scientists are taught to keep a laboratory notebook and follow the scientific method. Many software tools exist and are being actively developed to help in replicating past scientific findings, both by the researcher and by others. Some enable literate programming and the publishing of computer software, either as commented code or notebooks. Others capture provenance of a computation or the complete software environment. Version control systems are not new, but current tools facilitate use for collaboration and archiving complete project histories. For a description of current tools see the workshop report [4] or the workshop wiki [9].

One of us teaches a graduate seminar requiring students to replicate results from a published paper [10]. This is a simple way to introduce tools and methods for replication into the curriculum and students experience first hand how important it is to incorporate principles of reproducibility into the scientific research process.

Recent Policy Steps. Other stakeholders in the scientific community believe that traditional methods for research dissemination must adapt to the new technology-driven realities, including policy makers in Washington, D.C.. On February 22 of this year federal funding agencies were instructed by the Obama Administration to develop plans for enabling public access to both journal articles and digital datasets that arise from federal grants [12]. On March 5, the Research subcommittee of the House Committee on Science, Space, and Technology convened a hearing on Scientific Integrity & Transparency.[1] Recent events in economics and psychology illustrate the current scale of error and fraud [11]. I believe that the computational science community is best suited to decide how to make research code and data available and we hope the standards discussed here and in the workshop report [4] become an accepted and routine part of scientific research practice. The lead toward change should be taken by the computational science community, rather than federal governments.

A modified version of this article is forthcoming in Notices of the American Mathematical Society (June/July 2013).

References

1. D. Donoho, A. Maleki, M. Shahram, V. Stodden, and I. Ur Rahman. Reproducible research in computational harmonic analysis. Computing in Science and Engineering, 11, January 2009.
2. D. H. Bailey and J. M. Borwein, "Exploratory Experimentation and Computation," *Notices of the AMS*, vol. 58 (Nov. 2011), p. 1411-1419.
3. J. M. Borwein and D. H. Bailey, *Mathematics by Experiment*, Edition 2, AK Peters, 2008.
4. V. Stodden, D. H. Bailey, J. Borwein, R. J. LeVeque, W. Rider and W. Stein, "Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics," 2 Feb 2013, http://www.davidhbailey.com/dhbpapers/icerm-report.pdf.
5. B. Matthews, B. McIlwrath, D. Giaretta, and E. Conway. The significant properties of software: A study. 2008. http://www.jisc.ac.uk/media/documents/programmes/preservation/spsoftware_report_redacted.pdf
6. For examples see: http://scholcomm.columbia.edu/data-management/data-management-plan- templates/ and http://www2.lib.virginia.edu/brown/data/NSFDMP.html
7. D. Fanelli, "Redefine misconduct as distorted reporting," *Nature*, 13 Feb 2013, available at http://www.nature.com/news/redefine-misconduct-as-distorted-reporting-1.12411.
8. R. LeVeque, I. Mitchell, and V. Stodden. Reproducible research for scientific computing: Tools and strategies for changing the culture. Computing in Science and Engineering, p. 13–17, July 2012.
9. ICERM Reproducibility Workshop Wiki: http://wiki.stodden.net/ICERM_Reproducibility_in_Computational_and_Experimental_Mathematics:_Readings_and_References
10. D. Ince, System Failure, Times Higher Education, May 5, 2011. http://www.timeshighereducation.co.uk/416000.article
11. D. H. Bailey and J. M. Borwein, Reliability, Reproducibility and the Reinhart-Rogoff error, https://theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646, April 19, 2013.

[1] I testified as witnesses for this hearing. See http://science.house.gov/hearing/subcommittee-research-scientific-integrity-transparency for video and the written testimony.

12. OSTP, “Expanding Public Access to the Results of Federally Funded Research,” 22 Feb 2013, available at http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.