Statistical Journal of the IAOS 30 (2014) 91–93
DOI 10.3233/SJI-140818
IOS Press

The reproducible research movement in statistics

Victoria Stodden
Department of Statistics, Columbia University, NewYork, NY, USA
E-mail: victoria@stodden.net

Keywords: Reproducible research, open data, open code, open science, science policy

1. Introduction

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 as well as impeding the ability of others to build on past research [1–6]. 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 to address this shortcoming in research communication [11]. In this paper I 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 changes in journal publication and federal agency dissemination requirements.

2. Standards and scientific communication

In December of 2012, Brown University hosted more than 70 computational scientists and other stakeholders such as journal editors and funding agency officials for a workshop on "Reproducibility in Computational and Experimental Mathematics." This provided a unique opportunity for computational scientists from fields as diverse as mathematics and biology to discuss these issues and brainstorm ways to improve on current practices along with non-scientist stakeholders. The result is a series of recommendations intended to establish reproducible computational science as a standard [4–6]:

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

2.1. Changing the culture of computational research

Workshop discussants noted that bench scientists and experimental researchers are taught to maintain lab notebooks to record experimental details early in their career, including protocols, procedures, equipment, data collection details and raw results, processing techniques, and statistical methods. In computational science today, very few experiments are performed with the same attention to documentation. There is typically no record of workflow, computer hardware and soft-

[1]The results presented in this paper arose jointly from a collaborative workshop in which participants devised and shaped recommendations, and recap the workshop report [4].

ware configuration, software used, or parameter settings. Source code is typically lost or abandoned after experiments are completed, 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 recognition 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 or no emphasis on reproducibility. This has the effect of penalizing those researchers that take time to produce reproducible computational research. Software development has been frequently 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 programming across the research landscape, these attitudes and practices must change to include data and code production as bona fide scientific contributions worthy of institutional recognition and reward.

2.2. *Funding agencies, journals, and employers must support this change*

It should be expected that software and data be "open by default" and access only restricted in the case of confidentiality or other well-articulated concerns. Even in those cases openness and access should be maximized subject to those constraints. Grant proposals involving computational work could be required to detail standards for: dataset and software documentation including reuse (some agencies already have such requirements [7]); persistence of resulting software and dataset preservation and archiving; sharing software among reviewers and other researchers. Funding agencies such as the National Science Foundation could add "Reproducible Research" to the list of examples that proposals could include in their "Broader Impact" statements. Software and dataset curation should be expected and explicitly included in grant proposals and recognized as a scientific contribution by funding agencies, and funds made available to support it. Templates for data management plans could be made more widely available that include making software accessible, perhaps by institutional archiving and library centers [8]. Tenure and promotion committees and research managers at research labs should reward software and dataset contributions as part of expected computational research practices. Data and code citation practices should also be recognized and expected in computational research.

Editors and reviewers should expect a full disclosure of computational details and authors should be expected to carry out rigorous verification and validity testing [9]. Some experimental details might appear on a website with a persistent URL. Authors need to state any exceptions to full disclosure upon submission, such as for proprietary, medical, or other confidentiality issues, and reviewers and editors must agree to such exceptions prior to review. All software and data used in a paper should be cited, including version and access information, and not merely mentioned in the text or in a footnote as is done by some today. Proper citation is essential both for improving reproducibility and to provide credit for making available software and data, which is a key component in bringing about the culture change [10].

2.3. *Teaching the practice of reproducible research*

Today, the skills required to carry out and disseminate reproducible research in the computational sciences are not taught in a systematic organized way, and students frequently receive no training at all. These skills should be taught as part of scientific methodology, along with modern programming and software engineering techniques. Just as traditional experimental scientists are taught to keep a laboratory notebook and follow the scientific method, these skills should be a standard part of any computational science curriculum. Many software tools exist and are being actively developed to help in replicating past scientific findings, both by researchers and others. Some enable literate programming and the publishing of usable computer software, either as documented code or notebooks. Others capture provenance of a dataset or computation or the complete software environment. Interfaces to version control systems are making version control and sharing code easier, and allow for collaboration and the archiving of complete project histories. For a description of current tools see the ICERM workshop report [4] or the workshop wiki [11].

I have taught several graduate seminars requiring students to replicate results from a published paper [12]. 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 and communication process.

2.4. Recent policy steps

There are other stakeholders in the scientific community that are adapting the traditional methods for research dissemination to the new technology-driven changes, including policy makers in Washington, D.C. On February 22, 2013 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 [13]. On March 5, the Research subcommittee of the House Committee on Science, Space, and Technology convened a hearing on Scientific Integrity and Transparency. Recent events in economics and psychology illustrate the current scale of error and fraud [14,15]. 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. Changes that bring about reproducible computational science should be initiated by the computational science community rather than federal governments, but a failure to make such changes creates an opportunity for regulators and lawmakers to make new rules to enforce the dissemination of data and methods. We are starting to see this process happen.

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* **58** (Nov. 2011), 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://stodden.net/icerm_report.pdf.

[5] V. Stodden, D. Bailey and J. Borwein, "'Setting the Default to Reproducible' in Computational Science Research," SIAM News, June 2013.

[6] D. Bailey, J. Borwein and V. Stodden, "Set the Default to 'Open'," Notices of the AMS, June 2013.

[7] 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

[8] For examples see: http://scholcomm.columbia.edu/data-management/data-management-plan- templates/ and http://www2.lib.virginia.edu/brown/data/NSFDMP.html.

[9] 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.

[10] R. LeVeque, I. Mitchell and V. Stodden, Reproducible research for scientific computing: Tools and strategies for changing the culture. Computing in Science and Engineering, July 2012, pp. 13–17.

[11] ICERM Reproducibility Workshop Wiki: http://wiki.stodden.net/ICERM_Reproducibility_in_Computational_and_Experimental_Mathematics:_Readings_and_References

[12] D. Ince, System Failure, Times Higher Education, May 5, 2011. http://www.timeshighereducation.co.uk/416000.article.

[13] 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.

[14] 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.

[15] V. Stodden, What the Reinhart & Rogoff Debacle Really Shows: Verifying Empirical Results Needs to be Routine, http://themonkeycage.org/2013/04/19/what-the-reinhart-rogoff-debacle-really-shows-verifying-empirical-results-needs-to-be-routine/ April 19, 2013.