Publishing Standards for Computational Science: “Setting the Default to Reproducible”

BY VICTORIA STODDEN, JONATHAN BORWEIN, & DAVID H. BAILEY, 21-4-2013

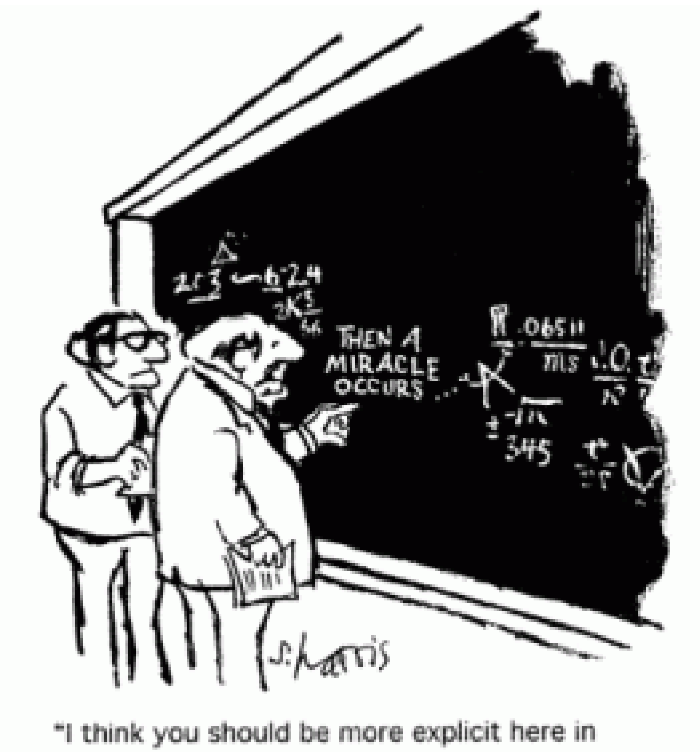
A group of computational scientists has developed a set of standards to guide the dissemination of reproducible research. Computation is now central to the scientific enterprise, and the emergence of powerful computational hardware combined with a vast array of computational software, presents novel opportunities for researchers. Unfortunately the scientific culture surrounding computational work has evolved in ways that make it difficult to verify findings, efficiently build on past research, or even to apply the basic tenets of the scientific method to computational procedures.

As a result computational science is facing a credibility crisis [1-4]. The enormous scale of state-of-the-art scientific computations, using tens or hundreds of thousands of processors, presents unprecedented challenges. Numerical reproducibility is a major issue, as is hardware reliability. For some applications, even rare interactions of circuitry with stray subatomic particles matter.

In December of 2012, more than 70 computational scientists and 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 the first opportunity for a broad cross section of computational scientists to discuss these issues and brainstorm ways to improve on current practices, resulting in a series of recommendations to establish really reproducible computational science as a standard [5]. The main recommendations that emerged from the workshop discussions 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 notebooks or computer logs of every work detail—design, procedures, equipment, raw results, processing techniques, statistical methods of analysis, etc. Unfortunately few



“I think you should be more explicit here in step two.”

computational experiments are documented so carefully. Typically, there is no record of workflow, computer hardware and software configuration, parameter settings, or function invocation sequences. Source code is often lost, or is revised with no record of the revisions. While crippling reproducibility of results, these practices ultimately impede the researchers' own productivity.

The research system must offer institutional rewards for producing reproducible research at every level from departmental decisions to grant funding and journal publication. The current academic and industrial research system places primary emphasis on publication and project results and little on reproducibility. It penalizes those devoting time to producing really reproducible research. It is regrettable that software development is often discounted. It has been compared to, say, constructing a telescope, rather than doing *real science*. Thus, scientists are discouraged from spending time writing or testing code. Sadly, NSF-funded projects on average remain accessible on the web only about a year after funding ends. Researchers are busy with new projects and lack the time or money to preserve the old. With the ever-increasing importance of computation and software, such attitudes and practices must change.

Funding Agencies, Journals, and Employers Should Support This Change. Software and data should be "open by default" unless it conflicts with other considerations, such as confidentiality. Grant proposals involving computational work should be required to provide details such as standards for: dataset and software documentation, including reuse (some agencies already have such requirements [6]); persistence of resulting software and dataset preservation and archiving; standards for sharing resulting software among reviewers and other researchers.

Funding agencies should add "reproducible research" to the list of specific examples that proposals could include in their requirements such as "Broader Impact" statements. Software and dataset curation should be explicitly included in grant proposals and recognized as a scientific contribution by funding agencies. Templates for data management plans could be made available that include making software open and available, perhaps by funding agencies, or by institutional archiving and library centers [7].

Editors and reviewers must insist on rigorous verification and validity testing, along with full disclosure of computational details [8]. Some details might be relegated to a website, with assurances this information will persist and remain accessible. Exceptions exist, such as where proprietary, medical, or other confidentiality issues arise, but authors need to state this upon submission, and reviewers and editors must agree such exceptions are reasonable. There is also a need for better standards on how to include citations for software and data in the references of a paper, instead of inline or as footnotes. Proper citation is essential both for improving reproducibility and in order to provide credit for work done developing software and producing data, which is a key component in encouraging the desired culture change [9].

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

Teaching and Tools for Reproducible Research. Proficiency in the skills required to carry out reproducible research in the computational sciences should be taught as part of the scientific methodology, along with teaching modern programming and software engineering techniques. This should be a standard part of any computational research curriculum, just as experimental or observational scientists are taught to keep a laboratory notebook and follow the scientific method. Adopting appropriate tools should be encouraged, and formally taught. Many tools exist and are under development to help in replicating past results (by the researcher or others). Some ease literate programming and publishing of computer code, 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 [5] or the workshop wiki [10].

One of us teaches a graduate seminar requiring students to replicate results from a published paper [11]. This is one 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.

Conclusions. Recent events in economics and psychology illustrate the current scale of error and fraud [12]. The United States has recently followed the lead of the United Kingdom, Australia and others in mandating public release of publicly funded research, including data [13]. We hope this helps bring about the needed cultural change in favour of consistently reproducible computational research. While different types and degrees of reproducible research were discussed, an overwhelming majority argued the community must move to "open research": research using accessible software tools to permit (a) auditing of computational procedures, (b) replication and independent verification of results, and (c) extending results or applying methods to new problems.

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. A. Abbott and Nature Magazine, "Disputed Results a Fresh Blow for Social Psychology," Scientific American, April 30, 2013. http://www.scientificamerican.com/article.cfm?id=disputed-results-a-fresh-blow-for-social-psychology
5. 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.
6. 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
7. For examples see: http://scholcomm.columbia.edu/data-management/data-management-plan-templates/ and http://www2.lib.virginia.edu/brown/data/NSFDMP.html

8. 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.
9. 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.
10. ICERM Reproducibility Workshop Wiki: http://wiki.stodden.net/ICERM_Reproducibility_in_Computational_and_Experimental_Mathematics:_Readings_and_References
11. D. Ince, System Failure, Times Higher Education, May 5, 2011. http://www.timeshighereducation.co.uk/416000.article
12. 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.
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.