"Setting the Default to Reproducible" in Computational Science Research

June 3, 2013

Following a late-2012 workshop at the Institute for Computational and Experimental Research in Mathematics, a group of computational scientists have proposed a set of standards for the dissemination of reproducible research.

Victoria Stodden, Jonathan Borwein, and David H. Bailey

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 apply the basic tenets of the scientific method to computational procedures.

Courtesy of S. Harris, ScienceCartoonsPlus.com.

As a result, computational science is facing a credibility crisis [1,2,4,5]. 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 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 gave a broad cross section of computational scientists their first opportunity to discuss these issues and brainstorm ways to improve on current practices; the result was a series of recommendations for establishing really reproducible computational science as a standard [13]. Three main recommendations emerged from the workshop discussions:

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 careers, bench scientists and experimental researchers are taught to maintain notebooks or computer logs of every work detail---including design, procedures, equipment, raw results, processing techniques, statistical methods of analysis. Unfortunately, few 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 either lost or revised with no record of the revisions. These practices not only cripple the reproducibility of results; ultimately, they 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, with little attention to reproducibility. It penalizes those who devote the time needed to produce 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 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
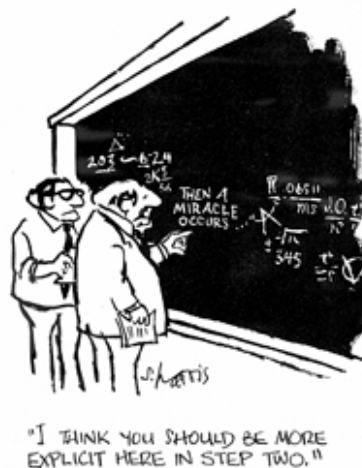
practices must change.

Support from Funding Agencies, Journals, and Employers

Software and data should be "open by default," in the absence of conflicts with other considerations, such as confidentiality. Grant proposals involving computational work should be required to provide such details as standards for: dataset and software documentation, including reuse (some agencies already have such requirements [11]); persistence of resulting software and dataset preservation and archiving; standards for sharing resulting software among reviewers and other researchers.

Funding agencies should add "reproducibility" to the specific examples, such as "Broader Impact" statements, that proposals could include. 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 that include making software open and available could be provided, 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 [6]. Some details might be relegated to a website, with assurances that this information will persist and remain accessible. Exceptions arise, as in the case of proprietary, medical, or other confidentiality issues, but authors need to state this upon submission, and reviewers and editors must agree that the exceptions are reasonable. Better standards are needed for including citations of software and data in the references of a paper, instead of inline or as footnotes. Proper citation is essential both for improving reproducibility and for ensuring credit for work done in developing software and producing data, which is a key component in encouraging the desired culture change [10].

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 above, 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 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 laboratory notebooks and follow the scientific method. Students should be encouraged and formally taught to adopt appropriate tools. Many tools are available or under development to help in replicating earlier results (of the researcher or others). Some tools ease literate programming and publishing of computer code, either as commented code or notebooks. Others capture the provenance of a computation or the complete software environment. Version control systems are not new, but current tools facilitate their use for collaboration and archiving complete project histories. For a description of current tools, see the workshop report [13] or wiki [8].

One of us teaches a graduate seminar that requires students to replicate results from a published paper [9]. This is one way to introduce tools and methods for replication into the curriculum, and it gives students first-hand appreciation for the importance of incorporating principles of reproducibility into the scientific research process.

Conclusions

Recent events in economics and psychology illustrate the current scale of error and fraud in scientific research [3]. Following the lead of the United Kingdom, Australia, and others, the United States recently mandated public release of publicly funded research, including data [12]. We hope that this will help bring about the needed cultural change in favour of consistently reproducible computational research. While different types and degrees of reproducible research were discussed at the ICERM workshop, an overwhelming majority argued that the community must move to "open research": research that uses accessible software tools to permit (a) auditing of computational procedures, (b) replication and independent verification of results, and (c) extension of results or application of methods to new problems.

References

[1] 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.

[2] D.H. Bailey and J.M. Borwein, *Exploratory experimentation and computation*, Notices Amer. Math. Soc., 58:10 (2011), 1410–1419.

[3] D.H. Bailey and J.M. Borwein, *Reliability, reproducibility and the Reinhart–Rogoff error*, April 19, 2013; http://experimentalmath.info/blog/2013/04/reliability-reproducibility-and-the-reinhart-rogoff-error/.

[4] J.M. Borwein and D.H. Bailey, *Mathematics by Experiment*, 2nd edition, AK Peters, Natick, MA, 2008.

[5] D. Donoho, A. Maleki, M. Shahram, V. Stodden, and I. Ur Rahman, *Reproducible research in computational harmonic analysis*, Comput. Sci. Engin., 11:1 (2009), 8–18.

[6] D. Fanelli, *Redefine misconduct as distorted reporting*, Nature, February 13, 2013; http://www.nature.com/news/redefine-misconduct-as-distorted-reporting-1.12411.

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

[9] D. Ince, Systems failure, Times Higher Education, May 5, 2011; http://www.timeshighereducation.co.uk/416000.article.

[10] R. LeVeque, I. Mitchell, and V. Stodden, *Reproducible research for scientific computing: Tools and strategies for changing the culture*, Comput. Sci. Engin., 14:4 (2012), 13–17.

[11] 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.

[12] OSTP, *Expanding public access to the results of federally funded research*, February 22, 2013; http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.

[13] 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*, February 2, 2013; http://www.davidhbailey.com/dhbpapers/icerm-report.pdf.

Victoria Stodden is an assistant professor of statistics at Columbia University. Jonathan Borwein is Laureate Professor of Mathematics at the University of Newcastle, Australia. David H. Bailey is a senior scientist in the Department of Computational Research at Lawrence Berkeley National Laboratory and a research fellow in the Department of Computer Science at the University of California, Davis.

- See more news from this issue