Transforming the Computational Sciences to Achieve Reproducible Research

An Opinion Piece for Data and Code Sharing in Computational Science November 21st 2009 Roundtable

Ian M. Mitchell & Michael P. Friedlander
Department of Computer Science
University of British Columbia
{mitchell,mpf}@cs.ubc.ca

The reproducibility of research results is a key feature of the scientific method, a feature which is in large part responsible for the ability of scientists to build upon the results of one another. Unfortunately, such reproducibility appears to be lacking in much of the computational science that is done today. It is unclear to us how, why, and when the social norm of reproducibility disappeared in computational science—it isn't even clear that it ever existed. Since computational science has taken a place alongside theory and experiment as a pillar of scientific discovery, it seems important that it be subject to the same expectations as theory and experiment. Introducing reproducibility into the computational sciences is vital to the health of the field, and will require a major transformation in the norms and processes of the individuals and organizations that constitute the field.

Retired Harvard Business School professor John P. Kotter has published widely on the transformation process in organizations (see, for example, Kotter, Harvard Business Review, pp. 96–103, January 2007). He argues that there are eight key steps through which the process passes, and skipping steps too often leads to failure. In this piece we consider those steps in the context of transforming the computational sciences into a field where the reproducibility of computational results is a highly valued component of research, and hence worthy of additional expenditure of effort.

Before examining the steps, we identify four constituencies which seem important to this effort: the producers of computational research, the editors and publishers of those results, the consumers of those results (typically other researchers), and the funding agencies. We omit at least two other identifiable groups from this list—industry and university administration—because we are unsure to what extent they might directly impact this process.

For consumers of research, reproducibility is clearly beneficial and appears to be free of costs. Funding agencies wield the biggest stick in the business; however, we can't comment on this constructively because we are not personally familiar enough with the process by which funding-agency mandates—such as the NIH dictum on open access—are made. Consequently, we restrict our comments below primarily to the producers and editors/publishers constituencies.

1. **Establish a sense of urgency.** The members of this roundtable feel the urgency, but it does not seem to be prevalent in the community. The Bermuda Declaration occurred with the threat that a patent land-rush through the human genome would lock it away from public research. There is no such obvious threat in computational science. Perhaps the consumers and/or funding agencies could provide this urgency, but it is unclear to us how that process can best get started.

2. **Form a powerful guiding coalition**. In this case, the coalition must include influential members of the producer and editor/publisher communities. The former would be prominent researchers willing to publicly adopt a policy of reproducibility in their research; we already have some examples among the roundtable participants and supporters. The latter would be editors-in-chief willing to confirm and identify reproducible results, and publishers willing to adopt intellectual property provisions compatible with this goal.

3. **Create a vision.** The Toronto and Bermuda statements provide high level guidance, although the devil is in the details. To get the most opportunities to spread the vision, the core of it must be brief and compelling.

4. **Communicate the vision.** Talk about it every chance we get: presentations, publications, grant applications, press releases, department social hour, graduate student meetings, undergraduate classes—never miss an opportunity to step up on the soapbox. Also, practice what we preach by visibly demonstrating reproducibility in our own research. It is an excellent question whether we, as advocates, need to demonstrate reproducibility in all aspects of our own work. Do occasional failures significantly weaken our message? Is it appropriate for us to impose the extra work on graduate students when it is not yet a broadly held norm in the field?

5. **Empower others to act on the vision.** Create processes that encourage reproducibility. For producers these structures will include software maintenance and documentation tools, as well as clear delineation of what constitutes reproducibility (written in a format that can be given to new grad students). For editors/publishers, these structures will include that same delineation, as well as methods of deciding whether a submission meets those standards and methods of identifying published articles which do. Because we cannot expect everybody to achieve (or even endorse) full reproducibility immediately, it seems to us that it will be harder to remove current structures which permit results which are not reproducible.

6. **Plan for and create short-term wins.** On the producer side, we need popular software packages that trumpet the reproducibility vision. On the editor/publisher side, we need journals and conference proceedings that are willing to review for reproducibility and prominently identify those articles that achieve it.

7. **Consolidate improvements and produce more change.** Each short-term win (or identifiable loss) must be critically examined to understand what elements of the vision, processes and/or structures worked, and which did not. Appropriate adjustments to strategies will hopefully lead to more wins. We can all start this critical review process in our own research. When we fail to meet reproducibility standards in our own research, why? What tools and/or techniques allow us to achieve reproducibility with the minimum of effort?

8. **Institutionalize the new approaches.** New projects should always include the explicit goal of reproducibility. Editorial succession at journals and conferences should maintain or even improve the level of support for reproducible research. Introduce reproducibility into graduate curricula. This stage is still many years and iterations away.

Kotter believes that these phases are sequential, but it seems clear to us that through efforts such as this roundtable, several of the phases are progressing in parallel. Because Kotter's emphasis is on companies and organizations with much more hierarchical management structures than the computational sciences community, perhaps a parallel approach is more appropriate to the setting in which we find ourselves.

Item 3 (create a vision) is the stated objective of the roundtable, but we are also concerned about elements of items 1 (creating urgency), 2 (a powerful guiding coalition) and 5 (empowering others). As an example of a potential approach to item 5 for the editors/publishers, consider the following proposal. A journal could have a small group of trained reproducibility editors (REs) working in parallel with the associate editors (AEs); the RE would judge submissions solely on their adherence to reproducibility standards, while the AE (with the aid of reviewers) would judge the submissions' other academic qualities. Because the RE need not judge the originality, contribution and correctness of the contribution, he or she does not need to be as expert in the subject as a regular reviewer might be. Furthermore, the RE can communicate directly with the author to educate about reproducibility options, identify the appropriate choice, and achieve that level of reproducibility, all in parallel with a traditional AE-driven review process. Submissions that meet a certain reproducibility standard would then be published with a prominent logo on their first page identifying that property.

We do not pretend to be experts on institutional change, and merely present this structure in the hopes of identifying some strengths and weaknesses of what we can see is the important, but still marginal, process of reforming the computational sciences to improve reproducibility.