

UC Berkeley

UC Berkeley Previously Published Works

Title

Changing computational research. The challenges ahead.

Permalink

<https://escholarship.org/uc/item/7r13j4sn>

Journal

Source Code for Biology and Medicine, 7(1)

ISSN

1751-0473

Authors

Neylon, Cameron

Aerts, Jan

Brown, C

et al.

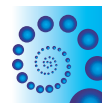
Publication Date

2012-05-28

DOI

<http://dx.doi.org/10.1186/1751-0473-7-2>

Peer reviewed



EDITORIAL

Open Access

Changing computational research. The challenges ahead

Cameron Neylon^{1,2*}, Jan Aerts³, C Titus Brown⁴, Simon J Coles⁵, Les Hatton⁶, Daniel Lemire⁷, K Jarrod Millman⁸, Peter Murray-Rust⁹, Fernando Perez⁸, Neil Saunders¹⁰, Nigam Shah¹¹, Arfon Smith¹², Gaël Varoquaux¹³ and Egon Willighagen¹⁴

The past year has been an interesting one for those interested in reproducible research. There have been great examples of replicability [1,2] in research communication, and examples of horrifying failure of reproducibility (as described in [3]) with serious questions being raised on the ability of our current system of research communication to guarantee, or even encourage, that published research be reproducible or replicable.

When we launched the call for papers for Open Research Computation in late 2010 we saw a clear need for higher standards. Computational research should stand out as an exemplar of just how reproducible research can be, yet it falls short more often than not. With modern computational tools it is entirely possible to provide packages which allow direct replication of results. It is possible to provide data and code in the form of a functional virtual machine image along with automated tests to ensure everything is working as expected. But alongside this we can support the reader's ability to modify and re-purpose tools, to run them against new data, indeed to support efforts to deliberately break the system to identify its limitations. In short, to do what we are supposed to do as scientists – replicate, reproduce, and test the limits of our models and understanding.

We deliberately set the bar high, because we felt it should be high, and because we felt that current standards were, in general, not high enough. Over the past year commentaries [4-6] have supported these principles, recognizing that there are serious problems – but few have actually backed up those words with actions. As with data, so with code, journal statements requiring that it be available often lack substance – how is it to be made available – and policies generally lack teeth.

As we looked at papers for ORC we set higher standards. We demanded that testing of the code be described. We required that we be able to fork the relevant code before formal acceptance of the paper. We looked hard at the documentation. Are the papers and the code being published today perfect? No. Are they an improvement over the average? Absolutely. Indeed a big part of that improvement is in ensuring that those imperfections can be identified, and worked on for the future.

But arguably we may have set the bar too high. Creating code and documentation to the level that we wanted to see is hard work. And the bottom line is that relatively few groups or projects are willing to put in that work, particularly for publication in an, as yet unproven, journal. So we struggled to get submissions at the level we wanted. And here we ran into the second problem.

In approaching our goals in the form of a journal we made an implicit compromise. We wanted to do something radical, but we did it in a form that was familiar and conventional. This was a deliberate tactical move. But problems arise from the straightjacket that a conventional journal form creates. The first is the issue of indexing. To get indexed requires a journal to publish in each month of a calendar year. If you're trying something new this isn't so straightforward.

In the end we have decided to fold ORC into another BMC journal, Source Code in Biology and Medicine, as a thematic series. This has a lot of advantages. It means there is less pressure to immediately get the submission numbers up, making it possible to take a longer term approach, and adapt over time as interest, demand and standards change. We can pull papers in from across the BMC portfolio and offer the "badge" of ORC certification as an extra bonus. In this way we get most of the advantages of a journal but avoid many of the pitfalls. But part of our strategy was the thinking that the journal could have a high Journal Impact Factor (JIF) if a reasonable

* Correspondence: cameron.neylon@stfc.ac.uk

¹Science and Technology Facilities Council, Didcot, Harwell Oxford, UK

²Public Library of Science, San Francisco, USA

Full list of author information is available at the end of the article

proportion of papers, being highly used tools, got significant numbers of citations. But you only get JIF's for individual journals – not for collections of papers. And republishing articles that have “already been published” (whatever *that* means in a web based world) is a definite no-no.

Many of us have felt for a long time that the construct of the journal places artificial constrictions on what we can do in research communication. The restrictions placed on what counts as a journal, what is allowed in a journal, by indexing services are a real drag on innovation. This raises the question of what is required to place innovation at the heart of the system we use to communicate research. How can we build the systems and infrastructure that we use in a way that actively encourages innovation?

In part, the answer lies with the papers that we are publishing today, and that we will continue to review for acceptance into this thematic series. Show our working, use open tools, enable others to replicate and to fork our work and our tools and our systems. Embed within those systems the measures of reputation and use and re-use that support the most successful open source projects and arguably also the research projects that are most successful at using the web as a resource.

Raising the standards of computational research is an important task, and one that we will continue to pursue by identifying and celebrating papers describing code that reaches those high standards that we have set. But equally we hope to keep learning from the process about what we can apply both to research more generally beyond pure computation and to the process of communicating that research.

We need more than just reproducible computational research, we desperately need a step change in our expectations and in the incentives for communicating research in a reproducible form more generally. We need educators and the materials to support them in raising awareness and experience. And we need the development of policy and standards that help us move towards a world where reproducibility and replicability are minimum standards not aspirations. ORC will not be doing this as a separate journal, but through the thematic series we will continue to promote the principles and the lessons we can learn. The problems are real and we need to tackle them.

Author details

¹Science and Technology Facilities Council, Didcot, Harwell Oxford, UK. ²Public Library of Science, San Francisco, USA. ³Leuven University, Heverlee, Belgium. ⁴Michigan State University, East Lansing, MI, USA. ⁵University of Southampton, Southampton, UK. ⁶Oakwood Computing Associates Ltd, Surrey, UK. ⁷LICEF Research Centre, Montreal, QC, Canada. ⁸University of California, Berkeley, USA. ⁹University of Cambridge, Cambridge, UK. ¹⁰CSIRO, Sydney, Australia. ¹¹National Center for Biomedical Ontology, Stanford, USA. ¹²Adler Planetarium, Chicago, USA. ¹³INRIA, Saclay, France. ¹⁴Maastricht University, Maastricht, Netherlands.

Received: 2 May 2012 Accepted: 28 May 2012
Published: 28 May 2012

References

1. Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH: A single pass approach to reducing sampling variation, removing errors, and scaling de novo assembly of shotgun sequences. [<http://ged.msu.edu/papers/2012-diginorm/>].
2. Brown CT: Our Approach to replication in computer science. [<http://ivory.idyll.org/blog/apr-12/replication-i.html>].
3. Begley CG, Ellis LM: Drug development: raise standards for preclinical cancer research. *Nature* 2012, **483**:531–533.
4. Ince DC, Hatton L, Graham-Cumming J: The case for open computer programs. *Nature* 2012, **482**:485–488.
5. Peng RD: Reproducible research in computational science. *Science* 2011, **334**:1226–1227.
6. Stodden V: Trust your science? Open your data and code. *Amstat News* 2011, July, 21–22 [<http://magazine.amstat.org/blog/2011/07/01/trust-your-science/>].

doi:10.1186/1751-0473-7-2

Cite this article as: Neylon *et al.*: Changing computational research. The challenges ahead. *Source Code for Biology and Medicine* 2012 **7**:2.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

