Editor's review: "CoRR — The Cloud of Reproducible Records"

Lorena A. Barba[1]

[1]George Washington University

April 28, 2020

Summary

This is the editor's review of manuscript CiSESI-2018-02-0016: *"CoRR — The Cloud of Reproducible Records,"* submitted to *Computing in Science and Engineering* Reproducible Research (RR) Track. We requested via email to the corresponding author that a preprint be made available on arXiv or an equivalent service, but did not receive a reply. It is policy of the RR Track to require a preprint; see our launch editorial (Barba & Thiruvathukal, 2017).

We received two reviews for this manuscript. The first, by Andrew P. Davison (Davison, 2018), recommends a minor revision. Davison notes that the work fills a gap in the landscape of reproducible computational research, and could help expand the usage of runtime provenance capture of computations. The second review, by Fernando Chirigati (Chirigati, 2018), recommends a major revision. Chirigati says that proposing a common platform for managing records from computational-provenance tools (like Sumatra and ReproZip) is an interesting idea, but omissions in the manuscript prevent a full assessment of the contribution.

Editorial decision

From my point of view, the most concerning comment in Chirigati's review is that he attempted to create an account on the web platform, failed to receive a confirmation email, and was unable to log in. At this time (8 April 2019), the site at https://corr-root.org still seems broken: it has an expired certificate and won't immediately load unless a visitor overrides the browser's security measures. I also tried to register received no verification email, and cannot log in. The platform also shows just 6 users and 21 records. So, even though the concepts presented in the paper are interesting, and reviewers allow for a revised version to be published, the platform is still immature and thus seems not ready for a formal publication. We encourage the authors to work on fixing the site and building community. We would be willing to consider the paper again once the project appears more stable. If and when the authors are ready for the work to be considered again for CiSE RR Track, the following comments should be addressed.

Comments

The main modification needed in the paper is regarding terminology: I agree with both reviewers that the attempt to coin a new term is unfortunate and best avoided. Moreover, "computational version control" does not communicate well the process that the paper is taking about, which is runtime capture of computational provenance. The usage of "provenance" for scientific workflows is established over many years. For example, see a summary of the 2006 NSF workshop titled "Challenges of Scientific Workflows" in (Gil et al., 2007).

1

The second major aspect of revision is the length of the paper, the wordiness and use of subjective modifiers, and overall copy editing. The section presenting the case study should be edited down rigorously. I urge the authors to aim for a much shortened version, 10 pages or less.

The reviewers pose various other aspects to improve. The authors should consider these comments carefully, in addition those provided below.

Note: The site has an expired certificate. I tried registering, but received no verification email, and the system does not let me log in, with an error message: *"Your account is pending verification from the admin."*

Acronym usage:

Instead of using the acronym "SVC" for software version control, spell out "version control" (you can often omit software, as this is the default usage). Also avoid the acronym "R&R", which makes some sentences especially awkward when used as a verb.

Extra comments:

Page 1:

(Line numbers in parenthesis.)

(50) "The uncovering of recent frauds" >> Don't conflate fraud with reproducibility issues. Fraud is a crime. It's unhelpful to lump it into the issues addressed in this paper.

(50) "irreproducibility facts and concerns. . . " >> awkward; should it be "irreproducible" (the adjective form)? Should it be "results" rather than "facts"?

(51) "have enforced" >> who has enforced? Unclear who is the subject of this sentence ("the uncovering"?). Maybe say "have led to new requirements."

(51) "major Computational Science journals" >> delete "major" and remove capitalization

(53) "SVC is witnessing a large growth in adoption" >> do you have any data to support such a statement? If not, avoid.

(57) "a form of version control" > the parallel with version control is tenuous; the tools do provenance capture, and scientific workflow management.

Page 2:

(12) "improved open access" >> open access is a different problem, not addressed by version control (address elsewhere in the text)

(39) "large amount of funding that is currently being channeled into. . . " >> do you have data to support this? If not, re-phrase accordingly.

(42) "the current standard" >> there is no standard; rephrase.

Page 3:

(27) "these tools are experiencing rapid adoption" >> do you have data to back this up? If not, re-phrase.

(50 "the front end is designed . . . commonly used web browsers" >> Seems irrelevant for the paper, and is just good practice. Delete.

References:

Ref. 12 is a 404 webpage. Only cite public sources.

Copy edits:

(Line numbers in parenthesis.)

Page 1:

(24) "greatly increasing the pace" >> delete "greatly" (subjective, hyperbolical)

(28) "significantly enhance" >> delete "significantly" (idem)

(34) "Acutely, SVC is not designed" >> delete "Acutely" (subjective)

(36) "Currently, there are many efforts focused on developing computational version control (CVC) tools to help scientists... " >> Several current efforts are underway to develop tools for computational provenance, aimed at helping scientists ... (remove syntactic expletive "there are" and avoid coining a new term).

(38) "adequately address" >> delete "adequately" (unnecessary adverb)

(40 "the broader tracking issues" >> delete "broader" (unnecessary modifier)

With this sample set of copy edits, the authors should work through the rest of the paper looking for similar subjective wording, and unnecessary modifiers.

References

Reproducible Research for Computing in Science & Engineering. (2017). *Computing in Science & Engineering*, *19*(6), 85–87. https://doi.org/10.1109/mcse.2017.3971172

Review: CoRR — The Cloud of Reproducible Records. (2018). Authorea Inc. https://doi.org/10.22541/au.154126764.42551559

Review: CoRR — The Cloud of Reproducible Records. (2018). Authorea Inc. https://doi.org/10.22541/au.154267039.94077105

Examining the Challenges of Scientific Workflows. (2007). *Computer*, *40*(12), 24–32. https://doi.org/10.1109/mc.2007.421