ReScience C: A Journal for Reproducible Replications in Computational Science

Nicolas P. Rougier[1,2,3(✉)] and Konrad Hinsen[4,5(✉)]

[1] INRIA Bordeaux Sud-Ouest, Talence, France
Nicolas.Rougier@inria.fr
[2] Institut des Maladies Neurodégénératives, Université de Bordeaux,
CNRS UMR 5293, Bordeaux, France
[3] LaBRI, Université de Bordeaux, Bordeaux INP, CNRS UMR 5800, Talence, France
[4] Centre de Biophysique Moléculaire, CNRS UPR 4301, Orléans, France
Konrad.Hinsen@cnrs.fr
[5] Synchrotron SOLEIL, Division Expériences, Gif sur Yvette, France

Abstract. Independent replication is one of the most powerful methods to verify published scientific studies. In computational science, it requires the reimplementation of the methods described in the original article by a different team of researchers. Replication is often performed by scientists who wish to gain a better understanding of a published method, but its results are rarely made public. ReScience C is a peer-reviewed journal dedicated to the publication of high-quality computational replications that provide added value to the scientific community. To this end, ReScience C requires replications to be reproducible and implemented using Open Source languages and libraries. In this article, we provide an overview of ReScience C's goals and quality standards, outline the submission and reviewing processes, and summarize the experience of its first three years of operation, concluding with an outlook towards evolutions envisaged for the near future.

Keywords: Open Science · Computational science · Reproducibility

1 Introduction

The question of how to attain reliable outcomes from unreliable components pervades many aspects of life. Scientific research is no exception. Individual research contributions are prone to mistakes, and sometimes fraud, and therefore error detection and correction mechanisms are required to reach a higher level of reliability at the collective level. The two main methods for error detection are critical inspection, starting with peer review of article submissions but continuing well after publication, and independent replication of published work. But replication is more than a verification technique. For the researchers performing the replication, it yields a level of understanding and insight that is impossible to achieve by other means. This is in fact the main motivation for much replication work, verification being merely a side effect.

The power but also the limitations of replication as an approach to verification are best illustrated by the recent discussion of replication crises in various scientific domains [3,5,6,9], which are all based on the observation of frequent failures to replicate published scientific findings. However, a replication failure does not necessarily mean that the original study is flawed. First of all, it could well be the replication work that is at fault. But it is also possible that both the original and the replication work are of excellent quality and yet yield different conclusions, if some important factor has escaped everyone's attention and accidentally differs between the two studies (see [10,12] for a recent example that led to a seven-year search for the cause of the disagreement). In this situation, independent replication can become the starting point of completely new lines of research.

Replication is thus an important contribution to science, and its findings should be shared with the scientific community. Unfortunately, most journals do not accept replication studies for publication because originality is one of their selection criteria. For this reason, we launched ReScience in 2015 (now called ReScience C for reasons explained later) as a journal dedicated to replications of computational research. In this article, we outline its mode of operation and summarize our experience from the first few years. A more complete account, also containing more background references, has been published recently [11].

2   Terminology: Reproducible Replications

The replication crisis has given rise to an active debate in various domains of science, in which some terms, in particular "reproducible" and "replicable", are used with very different meanings. We therefore explain the definitions that we are using in this article and more generally in ReScience C. Our definitions are formulated in the specific context of computational science, and are not easily transferable to experimental science [4].

A computation is *reproducible* if the code and input data is available together with sufficient instructions for someone else to re-do or *reproduce* the computation. The only point in reproducing the computation is to verify its reproducibility, which in turn is evidence that the archived code and data is (1) complete and (2) indeed the code and data that was used in the original published study. A failed reproduction means that the description of the original code and data is incomplete or inaccurate. A frequent form of incompleteness is the lack of a detailed description of the computational environment, i.e. the infrastructure software (operating system, compiler, ...) or code dependencies (libraries, ...) that were used in the original work. Reproducible computations are the most detailed and accurate possible description of a computational method within the current state of the art of computational science.

A *replication* of computational work involves writing and then running new software, using only the description of a method published in a journal article, i.e. without using or consulting the software used by the original authors, which may or may not be available. Successful replication confirms that the method

description is complete and accurate, and significantly reduces the probability of an error in either implementation. A replication failure can be caused by such errors or by an inexact or incomplete method description. It requires further investigation which, as explained above, can even lead to new directions of scientific inquiry.

A *reproducible replication* is a replication whose code and data has been archived and documented for reproducibility. It is especially useful in the still dominant situation that the target of the replication was not published reproducibly. In that case, the replication provides not only verification, but also the missing code and data.

3   ReScience C

The definition of a replication given above should be sufficient to show that performing replications is a useful activity for a researcher. Moreover, whether successful or not, a replication yields additional insight into the problem that are worth sharing with the scientific community. For example, minor omissions or inaccuracies are inevitable in the narratives that make up for most of a journal article, meaning that replication authors have to do some detective work whose results are of use to others.

Unfortunately, the vast majority of scientific journals would not consider such work for publication, with the possible exception of a failed replication of particularly important findings, because novelty is for them an important selection criterion. Moreover, the reviewing process of traditional scientific journals, designed in the 20th century for experimental and theoretical but not for computational work, cannot handle the technical challenges posed by a verification of reproducibility and successful replication. For these reasons we created the ReScience C journal (at the time called simply ReScience) in September 2015 as a state-of-the-art venue for the publication of reproducible replication studies in computational science.

The criteria that a submission must fulfill for acceptance by ReScience C are the following:

- It must aim at reproducing all or a significant part of the figures and tables in an already published scientific study.
- The text of the article must discuss which results were successfully replicated and which, if any, could not be replicated. It should also provide a description of problems that were encountered, e.g. additional assumptions that had to be made.
- The complete source code of the software used for the replication must be provided, and should have only Open Source software as dependencies in order to allow full inspection of the complete software stack.
- In order to ensure the independence of the replication, its authors should not include any authors of the original study, nor their close collaborators.

A newly submitted replication is assigned to a member of the editorial board, which at this time is composed of 12 scientists from different research domains. The handling editor recruits two reviewers from a pool of currently nearly 100 volunteers. The reviewing process consists of a dialog between the reviewers, the authors, and the handling editor whose goal is to improve the submission to the point that it can be accepted. In particular, the reviewers verify that they can reproduce the results from the supplied code and data, and judge if the replication claims made by the authors are valid subject to the criteria of their scientific domain. The entire reviewing process is openly conducted on the GitHub platform, meaning that contributions are open to read for anyone, and anyone with a GitHub account can participate by leaving a comment. Once the submission is deemed acceptable, it is added to the table of contents and to the ReScience archive, with links to the submission repository, the review, and a PDF version which permits the article to handled like a standard scientific paper in personal and institutional databases and bibliography management software. An additional copy is deposited on Zenodo [2], which, being an archiving platform, makes stronger promises about long-term preservation than GitHub, whose primary goal is to support dynamic development processes. An additional advantage is that Zenodo issues a DOI that serves as a persistent reference.

The outstanding feature of this reviewing process, even compared to other journals practicing open peer review, is the rapid interaction between reviewers and authors that does not require the constant intervention of the handling editor. This rapid exchange has turned out to be essential in the quick resolution of the technical issues that inevitably arise when dealing with software and data.

Another outstanding feature of ReScience C is its reliance on no other infrastructure than two digital platforms, GitHub and Zenodo, which are both free to use. Considering that editors and reviewers as well as authors are unpaid volunteers, this means that ReScience C has so far been able to operate without any budget at all, and thereby avoid being subjected to any political pressure. We note however that this may not always be true for the individual volunteers contributing to ReScience C because the open reviewing process provides no anonymity. It is therefore imaginable that authors or reviewers of a ReScience article pointing out a mistake in prior work by an influential scientist could be exposed to sanctions by that scientist in grant or tenure reviews.

4   Learning from the Past to Prepare the Future

After three years of operation, our original ideas for ReScience C have turned into concrete practical experience which has mostly confirmed our expectations. It has also shown a few weaknesses, most of which concern technical details, which we are currently addressing in an overhaul of the ReScience C publishing workflow. In the following we summarize this experience and the conclusions we have drawn from it, referring to the full account [11] for the details.

ReScience C has so far published 27 articles. Most submissions are from computational neuroscience, the other represented domains are neuroimaging,

computational ecology, and computer graphics. No submission was ever rejected. All submitted replications were successful, but this is probably due to a selection bias: publishing a failed replication is equivalent to publicly accusing the authors of the target work of having made a mistake, which is a potential source of conflict. One idea we have put forward to alleviate this obstacle is pre-publication replication. In that scenario, researchers submit their original work to a new type of journal, for which we use the name CoScience to indicate that we imagine it as the successor of ReScience. The journal then invites other scientists to do a replication, and publishes the original work and the replication together as a single joint work by the original authors and the replication team.

Achieving reproducibility has been much more challenging than expected. It is the reviewers' task to verify reproducibility, but our experience has shown that this is not sufficient to ensure that someone else can reproduce the work as well. Reviewers typically work in the same field as the authors and are likely to have similar software installation on their computers, meaning that unlisted dependencies can easily go unnoticed. There are a few approaches that would improve reproducibility, but each has its downsides as well. IPOL [8] provides online execution via its Web site, which is extremely convenient for both reviewers and readers. However, it is feasible only because IPOL's narrow domain scope (signal processing) makes the restriction to a small number of computational environments (C/C++, Python, Matlab) acceptable. We could also impose higher technical reproducibility guarantees on authors, e.g. the submission of an archived environment in the form of a virtual machine or a container image, which would also open the door to online execution via services such as Binder [1]. Such a requirement might, however, also become an additional barrier discouraging researchers from publishing their replication work.

The open reviewing process has overall worked very well. The exchanges between reviewers and authors have been constructive and courteous without exception. The handling editor intervenes mainly at the beginning, by inviting reviewers, and at the end, by judging if the reviewers' feedback is satisfactory for publication. Occasionally, reviewers or authors ask the handling editor for help with specific, mostly technical, issues. Another common task for the handling editor is to gently nudge authors or reviewers towards completing their tasks within reasonable delays. It is rare for third parties to intervene, but in one case a reviewer suggested asking the author of the target study for the permission to re-use some data, which he did by commenting directly on the GitHub platform.

An unexpected and so far unresolved consequence of the open reviewing process is the impossibility to handle replications that process confidential data. In some fields of science, confidentiality is inevitable, be it for ethical reasons (e.g. in medical research) or for commercial ones (e.g. data on stock market transactions not freely available). This is an issue of wider concern for the Open Science community, and we hope that satisfactory solutions will emerge in the near future.

The use of the GitHub platform has turned out to be a good choice overall. Since a ReScience C submission combines a narrative and source code, with the

code taking center stage during the reviewing process, a platform designed for collaborative software development and code reviewing is a better match than the traditional manuscript management platforms used by scientific journals, which have no provision at all for reviewing code. We are, however, currently revising several technical details. Submissions currently take the form of a pull request to the ReScience repository, which is counter intuitive for an article submission. More importantly, the final steps of publishing in our current workflow are laborious and not automated, causing too much hassle mainly for the handling editor. In the future workflow, articles are submitted as individual repositories of which ReScience retains a fork upon acceptance.

Finally, an evolution that has motivated the name change from ReScience to ReScience C is the imminent launch of ReScience X, a new journal dedicated to replications of experimental research, under the auspices of Etienne Roesch and Nicolas Rougier. We hope that it will be able to profit from the experience gained with ReScience C, although the challenges it will face are of a quite different nature. ReScience C will continue to focus on improving computational research, joining forces with the wider Reproducible Research community wherever possible. For example, we envisage proposing the publication of dedicated issues to reproducibility-related workshops such as the Reproducible Research on Pattern Recognition workshop [7] (part of the International Conference on Pattern Recognition) or the Enabling Reproducibility in Machine Learning workshop (part of the International Conference on Machine Learning).

References

1. Binder. https://mybinder.org/
2. Zenodo. http://www.zenodo.org/
3. Baker, M.: 1,500 scientists lift the lid on reproducibility. Nature **533**(7604), 452–454 (2016). https://doi.org/10.1038/533452a
4. Hinsen, K.: Scientific software is different from lab equipment, May 2018. http://blog.khinsen.net/posts/2018/05/07/scientific-software-is-different-from-lab-equipment/
5. Ioannidis, J.P.A.: Why most published research findings are false. Plos Med **2**(8), e124 (2005). https://doi.org/10.1371/journal.pmed.0020124
6. Iqbal, S.A., Wallach, J.D., Khoury, M.J., Schully, S.D., Ioannidis, J.P.A.: Reproducible research practices and transparency across the biomedical literature. PLOS Biol. **14**(1), e1002333 (2016). https://doi.org/10.1371/journal.pbio.1002333
7. Kerautret, B., Colom, M., Monasse, P. (eds.): RRPR 2016. LNCS, vol. 10214. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56414-2
8. Limare, N., Oudre, L., Getreuer, P.: IPOL: reviewed publication and public testing of research software. In: 2012 IEEE 8th International Conference on E-Science, pp. 1–8. IEEE, Chicago, October 2012. https://doi.org/10.1109/eScience.2012.6404449
9. Munafò, M.R., et al.: A manifesto for reproducible science. Nat. Hum. Behav. **1**(1), 0021 (2017). https://doi.org/10.1038/s41562-016-0021
10. Palmer, J.C., et al.: Comment on "The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water" [I and II: J. Chem. Phys. **135**, 134503 (2011); J. Chem. Phys. **138**, 214504 (2013)]. J. Chem. Phys. **148**(13), 137101 (2018). https://doi.org/10.1063/1.5029463

11. Rougier, N.P., et al.: Sustainable computational science: the ReScience initiative. PeerJ Comput. Sci. **3**, e142 (2017). https://doi.org/10.7717/peerj-cs.142
12. Smart, A.G.: The war over supercooled water. Phys. Today (2018). https://doi.org/10.1063/PT.6.1.20180822a