

Julien Chiquet ¹ Université Paris-Saclay/AgroParisTech/INRAE

Chloé Azencott  MinesParisTech

François-David Collin  IMAG/CNRS

Ghislain Durif  LBMC/CNRS/ENS Lyon

Mathurin Massias  INRIA Lyon OCKHAM

Pierre Neuval  IMT/CNRS

Nelle Varoquaux 

Date published: 3/22/23 Last modified: 3/22/23

Abstract

Computo is a newborn journal published by the [French Statistical Society](#) which aims at promoting computational/algorithmic contributions in statistics and machine learning. In order to achieve this goal, Computo goes beyond classical static publications by leveraging technical advances in literate programming and scientific reporting. We present in this talk the solution that we put in place so far to support both editorial and scientific reproducibility, that relies on community-based tools like Quarto, Github actions and open review. See our webpage <https://computo.sfds.asso.fr/>, and github group <https://github.com/computorg>.

Keywords: Reproducible research – Statistics – Machine-Learning – Quarto – R – Python – Julia

1 About

1.1 Aims and scope

Computo has been created in the context of a reproducibility crisis in science [1–5], which calls for improvements of scientific research implementation [6, 7] and for higher standards in the publication of scientific results. *Computo* aims at promoting computational and algorithmic contributions in statistics and machine learning (ML) that provide insight into which models or methods are the most appropriate to address a specific scientific question. The journal welcomes the following types of contributions:

1. New methods with original stats/ML developments, or numerical studies that illustrate theoretical results in stats/ML;
2. Case studies or surveys on stats/ML methods to address a specific (type of) question in data analysis, neutral comparison studies that provide insight into when, how, and why the compared methods perform well or less well;
3. Software/tutorial papers to present implementations of stats/ML algorithms or to feature the use of a package/toolbox/library.

¹Corresponding author: julien.chiquet@inrae.fr

1.2 An open access journal with reproducible contributions

Computo is free for readers and authors. It is a diamond open access journal [8] which means that all contents are freely published without any fee for the authors, and are freely available without any charge to the reader or their institution. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author². This is in accordance with the [Budapest Open Access Initiative \(BOAI\)](#) definition of open access and with the [Plan S](#) initiative [9].

The reproducibility of numerical results is a necessary condition for publication in *Computo*. In particular, submissions must include all necessary data (e.g. via Zenodo repositories) and codes. For contributions featuring the implementation of methods/algorithms, the quality of the provided code is assessed during the review process.

The reviews are open, i.e. visible to any reader after acceptance of the contribution. Reviewers may choose to remain anonymous or not.

2 Submission

2.1 Overview

Submissions require both scientific content (equations, codes and figures, data) and a proof that this content is reproducible. This is achieved by means of a notebook system, a virtual environment fixing the dependencies and continuous integration (plus, if needed, an external data repository to store data files such as [Zenodo](#) [10] or [OSF](#) [11]). A *Computo* submission is thus a git repository containing

- the source file of the notebook with auxiliary files: a markdown file with yaml metadata and some statics files, e.g. figures or small .csv data tables;
- configuration files to set up the dependencies in a virtual environment;
- configuration files to set up the continuous integration rendering the final documents.

The compiled notebook (HTML and PDF) is directly generated in the git repository via continuous integration (e.g., Github action or Gitlab CI) and published, if successful, to a web page (e.g. gh-page). The PDF and the git repository address are then submitted to our submission platform.

On the use of [Quarto](#)

Our choice to rely on Quarto has been carefully considered. Leveraging its experience in literate programming [12] with [Rmarkdown](#) [13] and relying on top community tools like universal document converter [Pandoc](#) [14], RStudio/Posit developed [Quarto](#) [15], a standalone, language-agnostic publishing tool^a. This point was decisive for the editorial committee of *Computo*, which in no way wishes to privilege one language over another. We thus encourage authors to use Quarto, which hopefully offer numerous publishing oriented features, a centralized workflow which facilitates the final production, and a literate programming experience well adapted to VScode/Rstudio/Neovim users.

^aincluding support for R, Python, Julia and Observable JS.

²as long as original authors are duly cited.

Warning

Computo is not just about publishing a notebook and proving that it can be compiled with CI! This part of the process is what we call “*Editorial Reproducibility*”. “*Scientific*” or “*numerical*” reproducibility^{a b} of the analyses is also mandatory, on top of classical peer-review evaluation.

^aWe don’t ask people reproducing their data... yet!

^bWe also don’t ask for bit-wise reproducibility but at least statistical reproducibility.

2.2 Available templates

Quarto-based templates are available from our [github group](#), which serves as material for starting to write a submission, and as a documentation for doing so. We developed a [Quarto extension](#) and dedicated templates for setting everything up either for R, Python and Julia users:

- [Computo Template for R users](#)³
- [Computo Template for Python users](#)
- [Computo Template for Julia users](#)

3 Reviewing and publication

Submitted papers are reviewed by external reviewers selected by the Associate Editor in charge of the paper. *Computo* strives for fast reviewing cycles, but cannot provide strict guarantees on the matter; the current time between submission and publication is under six months. In order to ensure an efficient reviewing process, authors are requested upon submission to suggest the names of four potential referees. To avoid conflicts of interests, recent co-authors or collaborators should not be suggested. At the moment, *Computo* relies on [Scholastica](#) for the review process. We plan to migrate very soon to the [openreview platform](#).

Once a manuscript is accepted, its reviews are made available on the finally rendered article as an issues associated to the git repository. Reviewers can choose to remain anonymous or not. *Computo*’s accepted papers are published electronically immediately upon receipt under [CC BY 4.0 license](#) [16]. Authors retain the copyright and full publishing rights without restrictions.

At the moment, we have [3 published contributions](#) and are processing 6 additional submissions. We strongly encourage you to submit your work to our journal.

4 Perspectives

Boundaries between “editorial” and “scientific/numerical” reproducibilities are not always clear and may be blurred or moving. For example, some contributions which computational workload completely fits the CI rendering process and provides also near-perfect reproducibility of their materials, if approved, achieve both reproducibilities in a single step. But other contributions, for example involving simulations, large amount of data or complex deep-learning models may require a more “hands-on” reproducibility process. One of the perspectives of *Computo* is to broaden the automated reproducibility scope to include more and more “scientific” reproducibility, and, ultimately, to provide authors a full array of computational resources to allow them to include their own scientific/numerical reproducibility developing process in the editorial one.

³Quarto R-based notebooks are similar to Rmarkdown notebooks which are also natively supported by Quarto.

5 References

1. **Ioannidis, J P A** 2005 Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8): e124. DOI: <https://doi.org/10.1371/journal.pmed.0020124>
2. **Steen, R G** 2011 Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*, 37(4): 249–253. DOI: <https://doi.org/10.1136/jme.2010.040923>
3. **Allison, D B, Brown, A W, George, B J, and Kaiser, K A** 2016 Reproducibility: A tragedy of errors. *Nature*, 530(7588): 27–29. DOI: <https://doi.org/10.1038/530027a>
4. **Bastian, H** 2016 Reproducibility Crisis Timeline: Milestones in Tackling Research Reliability. URL <https://absolutelymaybe.plos.org/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/>. [Online; accessed 22-March-2023]
5. **Whitfield, J** 2021 Replication Crisis. *London Review of Books*, 43(19). URL <https://www.lrb.co.uk/the-paper/v43/n19/john-whitfield/replication-crisis>. [Online; accessed 22-March-2023]
6. **Desquilbet, L L, Granger, S, Hejblum, B, Legrand, A, Pernot, P, Rougier, N P, Castro Guerra, E de, Courbin-Coulaud, M, Duvaux, L, Gravier, P, Le Campion, G, Roux, S, and Santos, F** 2019 *Vers une recherche reproductible*. Unité régionale de formation à l’information scientifique et technique de Bordeaux. URL <https://hal.science/hal-02144142>
7. **The Turing Way Community** 2022 The Turing Way: A handbook for reproducible, ethical and collaborative research. DOI: <https://doi.org/10.5281/zenodo.7625728>
8. **Ancion, Z, Borrell-Damián, L, Mounier, P, Rooryck, J, and Saenen, B** 2022 Action Plan for Diamond Open Access. DOI: <https://doi.org/10.5281/zenodo.6282403>
9. **Publications, C, Scientifique Ouverte, G d’Expertise Édition, and Construire La Biodiversité, G P** 2019 Recommendations for Plan S implementation by the ANR. DOI: <https://doi.org/10.52949/18>
10. **European Organization For Nuclear Research and OpenAIRE** 2013 Zenodo. DOI: <https://doi.org/10.25495/7GXX-RD71>
11. **Foster, E D and Deardorff, A** 2017 Open Science Framework (OSF). *Journal of the Medical Library Association : JMLA*, 105(2): 203–206. DOI: <https://doi.org/10.5195/jmla.2017.88>
12. **Knuth, D E** 1984 Literate programming. *The Computer Journal*, 27(2): 97–111.
13. **Allaire, J, Xie, Y, McPherson, J, Luraschi, J, Ushey, K, Atkins, A, Wickham, H, Cheng, J, Chang, W, and Iannone, R** 2023 *rmarkdown: Dynamic documents for R*. URL <https://pkgs.rstudio.com/rmarkdown/>
14. **MacFarlane, J, Krewinkel, A, and Rosenthal, J** *Pandoc*. URL <https://github.com/jgm/pandoc>
15. **Allaire, J J, Teague, C, Scheidegger, C, Xie, Y, and Dervieux, C** 2022 *Quarto*. DOI: <https://doi.org/10.5281/zenodo.5960048>
16. Creative Commons — Attribution 4.0 International — CC BY 4.0. URL <https://creativecommons.org/licenses/by/4.0/legalcode>. [Online; accessed 22-March-2023]