Reproducibility guidelines for invited reproducible papers

Dr. Juan J. Lastra-Díaz
Reproducibility Editor of Information Systems
email: jlastra@invi.uned.es

Science is suffering from a reproducibility crisis in many fields of research [5], including computational sciences. As a consequence of the complexity and interdependency of software projects, it is difficult to create and deploy self-contained reproducible experiments portable over different systems and in time. Fortunately, there is a set of mature reproducibility tools, such as Reprozip[1] [2] and Docker[2] to bridge this aforementioned gap. In addition, our reproducibility effort incentivizes authors [1] to produce reproducible papers by offering a second publication to establish that a parent paper is reproducible, thus increasing its visibility.

A reproducible paper is a companion article of a previously published article that includes a description and link of software and data resources together with a very detailed repoducibility protocol to allow the replication of all results (including relative times but excluding exact times) and conclusions reported in the original article on the reviewers' hardware and operating system. One difference between a reproducible paper and any standard research article is that it will be co-authored by the original authors and the reproducibility reviewers. Figure 1 shows a summary of the submission and review process encouraged by our reproducibility initiative, whilst table 1 introduces some basic definitions setting our main reproducibility criteria and artifacts. The guidelines below provide an in-depth description of the entire process. If you have questions, please contact the reproducibility editor of Information Systems Juan J. Lastra-Díaz.

Why should any author adopt these reproducibility guidelines?

A goal of our reproducibility initiative is to encourage the authors to adopt a *reproducibility-centered research methodology* as a means of increasing the quality, rigor, impact and credibility of all their scientific communications, as well as the confidence of the authors on their own findings and submissions. By adopting the reproducibility practices introduced here, authors will contribute to making comparisons with their work easier, in addition to encouraging the adoption, citation and reuse of their research. The availability of reproducible experiments also contributes to speeding up the integration of newcomers, especially graduate students, who spend a considerable part of their time trying to set up experiments.

The rest of the document provides a detailed description of the submission and review processes for any invited reproducible paper.

[1] https://www.reprozip.org/
[2] https://www.docker.com/

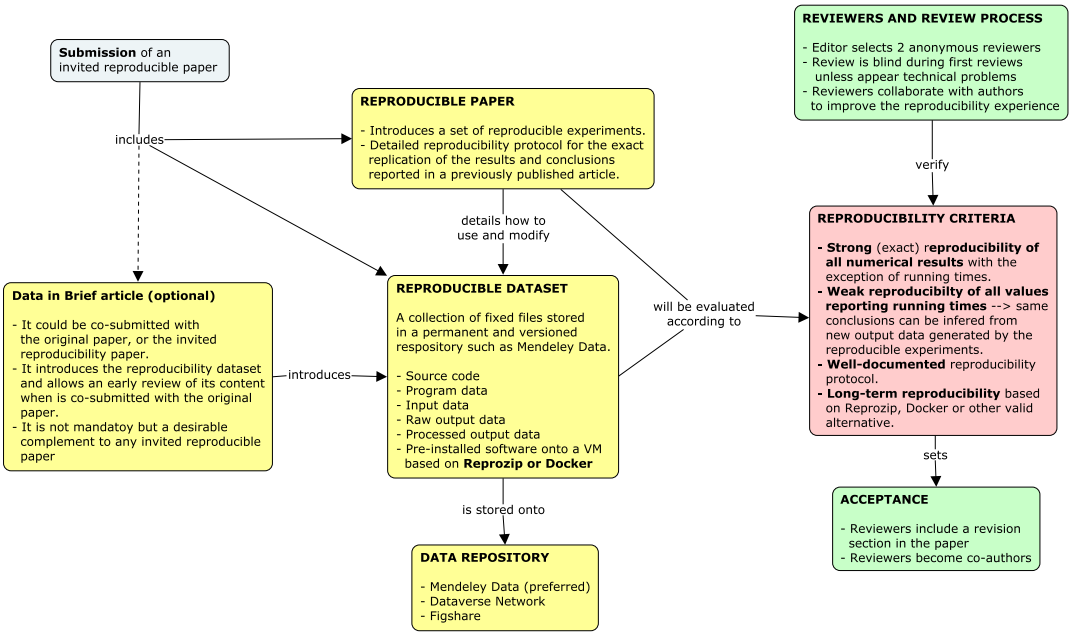

Figure 1: Concept map detailing the main features and artifacts involved in the submission and review process for invited reproducible papers.

1. Invited reproducible papers

1.1. Before you Begin

Submission to the Reproducibility Section of Information Systems is by invitation only. Our online submission system guides authors through the process of entering the reproducibility paper. The system converts article files to a single PDF file to be used in the peer-review process. All correspondence, including notification of the editor's decision and requests for revision, is sent by e-mail.

Submission of the reproducible experiments. Reproducible experiments should be submitted through any permanent, versioned and recognized public data repository with a Creative Commons Attribution 4.0 International License[3] for all files. We note that the version of the source code and software binaries used to build the reproducible experiments should be jointly submitted as a snapshot in the same reproducibility dataset provided as supplementary material. In addition, authors can keep it in a public code repository such as GitHub. However, it is essential to fix a full snapshot of the software and data with the aim of providing a permanent and self-contained reproducibility dataset which allows the reproduction of the experiments over time. Thus, we urge the authors to publish a single reproducibility dataset in Mendeley Data, or some other well-known repositories such as the Dataverse Network or Figshare whenever it is not possible to put the data in Mendeley. It is the responsibility of the authors to ensure that the submitted software and data do not violate any licensing

[3] http://creativecommons.org/licenses/by/4.0/

Concept	Definition
Reproducibility dataset	A reproducibility dataset is a single and self-contained collection of software and data resources made up of a set of fixed files which are made publicly available in a permanent and versioned data repository, and whose aim is to allow the independent reproduction of a set of software experiments.
Reproducible experiment	A reproducible experiment is a reproducibility dataset together with a setup protocol which provides a set of detailed step-by-step instructions for obtaining and deploying all required reproducibility resources onto a target computer, as well as setting up and running all experiments, and finally checking that the reproduced results match the original results and conclusions previously reported by the authors.
Weakly reproducible experiment	Given a set of previously reported experimental results and conclusions, a computational experiment is weakly reproducible if it allows to confirm all previously reported conclusions but it is unable to reproduce all results exactly.
Strongly reproducible experiment	Given a set of previously reported experimental results and conclusions, a computational experiment is strongly reproducible if it allows to confirm both previously reported results and conclusions exactly.
Adaptable reproducible experiment	A strongly reproducible experiment that allows configuration parameters to be varied so that an observer can look at the sensitivity of the results to particular parameter settings or obtaining results for unexplored datasets or methods.

Table 1: Some basic definitions of reproducibility concepts used herein.

agreement, and that their datasets respect ethical guidelines on data protection and privacy.

1.2. Preparation

Authors of invited papers should submit a single and self-contained reproducibility dataset together with a reproducibility paper providing the following content:

1. *Software.* The software that is necessary to reproduce the experiments including the system under test, data generation and processing scripts, experimentation scripts, plot generation scripts and whichever other material needed to reproduce the previously published results.
2. *Data.* All the original datasets used for training, parameter tuning, validation and testing, original input files, original raw output files, and original processed data files corresponding to the data tables shown in their original paper.

3. *Experimental setup.* A detailed description of the experimental platform and the configuration used in all experiments, as well as the overall running times of the experiments on the original platform.
4. *Reproducibility paper.* A reproducibility paper introducing the reproducible experiments by providing a detailed set of step-by-step instructions for setting-up and running the experiments, as well as for generating the plots and data tables contained in the paper starting from the raw output data generated by the former ones. Every data table or plot should be reproduced from raw output data by running some data processing script based on an accessible software tool, such as R statistical program, RStudio or Python. This paper should also introduce the reproducibility dataset and the information detailed above.
5. *Missing and complementary information.* The reproducible paper provides a good opportunity to provide further data tables or technical details to clarify the methods and experiments previously reported in the original paper. For instance, some papers introduce a large set of plots in which the numerical data cannot be easily interpreted and compared with future replications of the same experiments. Thus, the reproducibility paper could provide a set of complementary data tables including the raw and processed data introduced by the plots as supplementary material.

The authors should preferably use license-free software to avoid licensing and access limitations.

1.2.1. Software

Submission of the source code. The source code for the software components must be submitted together with installation scripts. If the code lives in a repository hosting service, in addition to submitting the source code files in their reproducibility dataset, authors can also provide the URL together with an identifier for the version used in the paper (e.g.: version number, commit id, etc.). We suggest that authors host their code on GitHub, but even in that case it is mandatory to provide a snapshot of the software binaries and source code in the reproducibility dataset.

Portability and long-term reproduction. In addition to the source code, we urge the authors to submit a Virtual Machine (VM), in which all appropriate software components are pre-installed and can be reproduced on a wide variety of platforms, such as Linux-based systems, MacOs and Windows. For this reason, we encourage the authors to submit their experiments using either ReproZip[4] or Docker[5] with the aim of providing lightweight and easily portable reproducibility packages which can be reproduced over the long term.

Next, we provide three reproducible papers published in our section as examples of the aforementioned approach. For instance, Lastra-Díaz et al. [4] introduce two reproducibility methods, the first one based on the compilation and execution of the software components used to run the original experiments whose source code and binaries are published in Mendeley Data and GitHub,

[4] https://www.reprozip.org/
[5] https://www.docker.com/

and the second one based on a Reprozip package deployed into a Docker container. On the other hand, Fariña et al. [3] introduce a reproducibility protocol based on the building and execution of a Docker container in addition to provide their source code in a GitHub repository, and finally, Wolke et al. [6] introduce a set of reproducible experiments based on a Docker image published in Mendeley Data.

Raw input and output data, intermediate data and final processed data. All the input data and parameters used by the software must be included. In addition, we encourage authors to include in their reproducibility dataset both the raw output data generated by their experiments and the output data resulting from their data analysis pipeline with the aim of making the comparison of the outcomes to the reviewers easier. Likewise, for those scenarios in which the raw output data is the result of a complex processing pipeline, we encourage authors to consider the publication of any intermediate data which could be helpful for the understanding, debugging and independent replication of their methods and experiments.

1.2.2. Reproducibility Paper

The reproducibility paper must contain a detailed description of the submitted software and reproducibility dataset, including the following information:

1. *Experimental environment.* Detailed description of the computational environment originally used to run the experiments (i.e. operating system, memory resources, etc.) as well as the overall running times reported in the original computational environment. Reviewers and readers should know in advance the expected running times with the aim of allocating proper time and resources to reproduce the experiments.
2. *Adaptability.* Explanations regarding the different data and input parameters that can be used with the aim of allowing the reuse and modification of the experiment in unexplored experimental setups.
3. *Experiment deployment.* Detailed instructions for downloading, installing and compiling the software. This should be done even if either a virtual machine with all the pre-installed software is provided or the software is already provided in binary form. The software could be provided in their binary form to accelerate the reproduction tasks; however, it is mandatory to provide detailed downloading and compiling instructions to allow the reconstruction of the software from its original source code.
4. *Compilation.* Detailed description of the platform used to create the software, paying special attention to the versions of all software tools and components involved in the process, as well as all dependencies for its compilation and execution.
5. *Virtual-Machine deployment.* Detailed instructions for downloading, installing and executing the Virtual Machine-based components. Even when the execution of the experiments is based on a Docker image or Reprozip package deployed into Docker, it is mandatory to provide detailed setup and running instructions on at least one specific platform such as UBUNTU. Instructions for other Docker-complaint platforms as Windows and MacOS are welcome. Authors should provide detailed instructions without assuming any previous Docker or Reprozip knowledge on the part of the readers.

6. *Running and postprocessing.* Instructions for running the experiments and producing the plots and tables contained in the paper. All numerical and graphical results presented in the primary paper should be automatically reproduced from the raw output data. We urge the authors to use data processing scripts based on R statistical package, Python or other well-established data processing tools for this purpose.
7. *Verified and failed running environments.* Description of the software and hardware settings in which the original results do or do not hold, such as system parameters, workload characteristics and minimal verified configurations.
8. Limitations of the software execution, if any.
9. Any other useful information regarding the software.

Ease of Reproducibility. We urge the authors to adopt the point of view of any person who neither has knowledge of their software nor even is familiar with their line of research. That is, the ultimate goal of the authors should be that any reader should be able to set up and reuse their experiments in a couple of hours or less. Thus, we emphasize that the reproducibility protocol should be as automatic and well-documented as possible. In this spirit, we invite the authors to include diagrams, workflows or any complementary information which help to achieve the aforementioned goal.

The article may have up to 12 pages and follow the formatting guidelines for regular Information Systems articles, including a title, abstract and highlights, and the following sections: (1) an introduction which puts in context the experiments, as well as their motivation, contributions and novelty, highlighting their value for the research community; (2) a section introducing the software; (3) main section introducing the reproducibility protocol, dataset and experiments; (4) a section detailing how to extend or modify the experiments to explore new experimental setups; (5) optionally a discussion of the experiments; and (6) conclusions. Further supplementary raw and processed data tables or information could be moved to an appendix provided as supplementary material in order to avoid exceeding the aforementioned page limit if needed. The original paper and the tools used in the reproducibility process should be added to the references section. Authors are encouraged to add discussion sections about the process of making their experiment reproducible.

1.3. Submission process

Authors should follow the recommendations for the submission detailed below.

1.3.1. Submission of the Reproducibility Dataset

All the software and data making up the reproducibility dataset must be first published through Mendeley Data[6], whenever possible, as a single and self-contained dataset. Authors should publish their datasets in other well-known repositories such as the Dataverse Network or Figshare if Mendeley cannot handle it for any reason (e.g. size limits). After publishing the dataset, authors

[6] https://data.mendeley.com/

must include the corresponding link, citation and dataset version in the reproducibility report. Note that Mendeley Data has a limit of 10 Gb for file, but multiple files can be submitted in a single dataset. Should there be any issues with packaging and size limit, please contact the reproducibility editor.

Co-submission of a Data in Brief article. We strongly recommend the co-submission of a Data in Brief article introducing the companion reproducibility dataset, either with the invited reproducible paper or with the original paper being reproduced.

1.3.2. Submission of the Reproducibility Paper

Authors should select the issue "Invited Reproducibility Paper" when submitting the paper at Editorial Manager system of Elsevier.

1.4. Review Process

Editor will select two reviewers matching the field of expertise as done with any research article. Review process will follow the standard practices of any blind peer-review process until the acceptance of the paper, unless any technical difficulty occurs during the setting up and running of the experiments. If this happens, there will be an open and non-blind collaboration and discussion between authors and reviewers to solve any issues and improve the reproducibility experience.

The main goals of the review process are as follows:

1. *Strong and weak reproducibility.* To verify that all results and claims presented by the original paper can be exactly reproduced. This means to verify that all numerical results, data tables and plots can be exactly reproduced by following the detailed reproducibility protocol provided by the authors. Thus, we consider herein a strongly reproducible experiment by all measures except running times. However, we expect that the ratios between reproduced time-measurements allow to confirm the same qualitative conclusions reported by any comparative analysis carried-out in the original paper.
2. *Ease of reproducibility.* To verify that the reproducibility protocol and reproducibility dataset provided by the authors are precise, complete, self-contained and allow the strict reproduction of all results reported by the authors in their original paper. Ideally, the authors should provide, as much as possible, an automatic reproducibility protocol which makes the setup and running of all experiment easy.
3. *Adaptable Reproducibility.* To establish the versatility of the software and reproducibility dataset by allowing the evaluation of new input datasets or previously unconsidered methods, with the aim of promoting the reuse of software assets. The hope is that any reproducible experiment not only allows the reproduction of the author's results, but also becomes both a standard and extensible experimentation platform for any researcher working in the same line of research. Thus, it is important that the submitted paper provides a detailed explanation of how anyone could modify the experimental settings to generate new experimental setups.

4. *Long-term reproducibility.* To verify that the experiments are reproducible in the long-term by providing a pre-installed virtual machine based on Reprozip, Docker or any other valid alternative which includes all the software dependencies of the experiments.

Reviewers may ask for a revision if the submitted software and paper does not meet the adaptive reproducible standard or is not portable. This will allow authors to correct any error and provide any missing information about their software. For instance, reviewers will check if all the necessary instructions are available, if all the parameters and different settings are explained and well documented, and if there is any problem in installing, compiling, and running the experiment and provided scripts.

1.5. Acceptance criteria

The acceptance criteria for any invited reproducible paper are to achieve adaptive or strong reproducibility, portability across platforms, and sufficient and clear documentation.

1.5.1. What happen if your work does not fit any acceptance criteria above?

In the unexpected case that your work does not fit the reproducibility criteria detailed above, it will be rejected. However, this information will not be made public in any way nor impact your original paper. You should not be afraid of any negative consequence derived from an "unexpected" rejection. Thus, there is no downside to submitting your work.

1.5.2. Who decides on acceptance or rejection?

Like any other research article, the final acceptance decision is taken by the reproducibility editor based on the final reports submitted by all reviewers and his critical analysis of all evidence and claims reported by the former ones and the authors respectively.

1.6. After Acceptance

Once the reproducible paper is accepted for publication, the publication process continues as follows:

1. Authors follow reviewers' suggestions to improve their software specification and paper, although it is likely that some suggestions have been already made during the first reviews.
2. Reviewers will add a section detailing their experience in reproducing the artifacts provided by the authors which will be called "Revision comments" and located before the conclusions.

Revision comments. The editor will incorporate a summary of the review discussion into the reproducibility paper, and the reviewers will become co-authors of the reproducibility paper. Reviewers will add a section detailing their experience on reproducing the artifacts provided by the authors.

The paper will be published in the Reproducibility Section of Information Systems following the same procedure as regular papers. The software and datasets will be maintained in Mendeley Data, or any other accepted repository as the Dataverse Network or Figshare, and a reference to the dataset will be created for the reproducibility paper.

1.7. Recognition and advertising

Once a reproducible paper is accepted and published in Information Systems, the journal will carry out the following actions:

1. The reproducible paper as well as the reproducibility dataset and original paper, will be linked in the Elsevier databases. Optionally and in addition, the co-submitted Data in Brief article would also be linked to the aforementioned resources once published.
2. The reproducible paper will be included in a permanent Virtual Special Issue (VSI) devoted to reproducible research.
3. Elsevier will showcase and promote your reproducible paper.

References

[1] Chirigati, F., Capone, R., Rampin, R., Freire, J., Shasha, D., 2016a.
A collaborative approach to computational reproducibility.
Information Systems 59, 95–97.

[2] Chirigati, F., Rampin, R., Shasha, D., Freire, J., 2016b.
ReproZip: computational reproducibility with ease, in: Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD), bigdata.poly.edu. pp. 2085–2088.

[3] Fariña, A., Martínez-Prieto, M.A., Claude, F., Navarro, G., Lastra-Díaz, J.J., Prezza, N., Seco, D., 2019.
On the reproducibility of experiments of indexing repetitive document collections.
Information systems 83, 181–194.

[4] Lastra-Díaz, J.J., García-Serrano, A., Batet, M., Fernández, M., Chirigati, F., 2017.
HESML: a scalable ontology-based semantic similarity measures library with a set of reproducible experiments and a replication dataset.
Information Systems 66, 97–118.

[5] Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., du Sert, N.P., Simonsohn, U., Wagenmakers, E.J., Ware, J.J., Ioannidis, J.P.A., 2017.
A manifesto for reproducible science.
Nature Human Behaviour 1, 0021.

[6] Wolke, A., Bichler, M., Chirigati, F., Steeves, V., 2016.
Reproducible experiments on dynamic resource allocation in cloud data centers.
Information Systems 59, 98–101.