Open and Reproducible Research

According to UNESCO, open science

> is the movement to make scientific research and data accessible to all. It includes practices such as publishing open scientific research, campaigning for open access and generally making it easier to publish and communicate scientific knowledge. Additionally, it includes other ways to make science more transparent and accessible during the research process. This includes open notebook science, citizen science, and aspects of open source software and crowdfunded research projects.
>
> The many advantages of this movement include:
>
> - Greater availability and accessibility of publicly funded scientific research outputs;
> - Possibility for rigorous peer-review processes;
> - Greater reproducibility and transparency of scientific works;
> - Greater impact of scientific research.

As The Center for Open Science says,

> Openness and reproducibility are core values of scholarship. Scholarly claims become credible via transparent communication of the supporting evidence and the process of acquiring that evidence. This way, independent observers can evaluate the quality of evidence for supporting the claim. By making the evidence and the process of acquiring it transparent, others can confirm that the provided evidence is consistent with the claim, or reproduce independent evidence using the same process. Ultimately, scholarly research is a public good. Anyone should be able to examine the content of the evidence (open data), the process by which it was obtained (open workflows), and the outcomes that were observed (open access).

To this end, our research group:

- uses open-source and reproducible tools;
- communicates our findings and methods through open-access channels;
- communicates findings of public relevance through appropriate channels

This page describes, at a high level, our procedures and philosophies concerning open and reproducible science. Our standard set of tools is documented elsewhere, and familiarization with them will be part of the onboarding process.

Alternative tools can be used if they have a particular feature set or if a collaborator would prefer. Always be inclusive rather than dogmatic!

Transparency

Group members are required to share any data and code used in the development of a paper. This does not have to happen immediately (and there may be good reasons to keep a code repository private for a time), but public release should generally occur by the time we are ready to submit a manuscript to a peer-reviewed journal. A paper is not ready for peer review if the data and code are not publicly available and documented.

Large ensembles of simulation outputs should also be made available, even if only a subset was used for a particular analysis. This facilitates transparency about how the choices involved in the subsetting process and their impact.

Reproducible Science

Reproducibility is an essential part of science. Your entire analysis, from model calibration to figure generation, should be made public and organized in a straightforward fashion. As an example, for a project which involves model-based uncertainty quantification and simulation, you should document the workflows needed to:

- install any required tools or packages;
- load and clean data;
- run and calibrate the model;
- generate simulations based on the calibration results;
- analyze simulation results;
- plot any figures included in the paper.

You should also include partial products of the analysis, such as calibration and simulation results, so that subsequent parts of the analysis can be reproduced without needing to re-run everything from scratch.

Ideally, your code has also been written and commented to make key choices clear.

When multiple versions of a model or other piece of software exist, you should document the specific versions that were used, or, even better, develop and provide a container that will set up the correct configuration.

Open vs. Closed-Source

In an ideal world, we would use nothing but open source software. Occasionally, we may need to use software which does not allow its source code to be distributed. In that case, in addition to documenting the version(s) of the relevant software, you should also provide documentation for how to install and configure them (a container is once again the preferred approach).

We try to avoid proprietary software, which requires payment, as that reduces the number of people who can potentially try to reproduce or build on the analysis. It may be that we encounter a situation where this is required; if so, you should discuss this with Vivek and your other collaborators to make sure that this is the only option.

Toolset for Reproducible Science

We try not to be dogmatic about particular choices of tools for reproducibility. Our workflow is heavily inspired by

> Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS Comput Biol 13(6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510

which you should read before starting to use any particular set of tools to understand why we use them and how they fit into the workflow. While we have some default choices (such as GitHub for code repositories), these are not exclusive. If you prefer another tool, feel free to use it. You should, however, discuss those choices ahead of time with collaborators as part of the paper-writing process to make sure everyone is comfortable with them.

Communication and Outreach

A key part of open science is increasing the accessibility of results, both to other scientists and to the public.

Scientific Communication

To increase access of results to the scientific community, we will:

- post preprints on an open-access preprint server at the time of submission, to be updated through the revision process;
- publish in open-access journals (or with open-access licenses) where possible and appropriate, and make post-prints readily available when not;

- summarize findings in blog posts;

- clearly document and organize code;

- make code repositories publically available at time of submission;

- make software, figures, and papers available through permissive licenses.

Public Communication

Our research is mission-oriented and stakeholder-driven. Moreover, it is often funded by public institutions. Withholding access to key results from the public is therefore not only an abandonment of our mission to do research which supports decision-making on key societal issues, but could also be seen as unethical. To help communicate our results to the public, we will:

- summarize findings in non-technical posts (on our blog or other outlets);

- use social media to share key findings;

- develop communications plans with stakeholders as part of our research design;

- issue press releases as appropriate.

It can be challenging to write about scientific and technical topics in a manner which is readily understandable to non-experts (Vivek certainly cannot claim to be an expert on this!). However, honing your ability to write in this style will benefit interest in your work and the quality and accessibility of your peer-reviewed papers. Seek out examples of popular science writing that you admire and try to learn what makes them so effective. There are likely other resources, both online and local, that you can utilize.

Last update: June 26, 2023