

Open Science

Our lab is committed to supporting and contributing to open science.

Why open science?

The [Center for Open Science](#) writes that:

An open exchange of ideas accelerates scientific progress towards solving humanity's most persistent problems. The challenges of disease, poverty, education, social justice, and the environment are too urgent to waste time on studies lacking rigor, outcomes that are never shared, and findings that are not reproducible.

Unfortunately, climate change adaptation is a field with especially severe problems of reproducibility and replicability ([pollack_transparency:2023?](#)). Making our work open and reproducible helps us to ensure that:

1. our assumptions and modeling choices are transparent and documented, enabling scientific critique and discussion
2. our research findings are accessible to the public, including people who aren't able to afford expensive academic journal subscriptions
3. we can build on our own work in the future.

Lots of agencies, including [NASA](#), [the White House](#), the [NSF](#), and many others are placing increasing emphasis on open science. Training group members in open science is a priority for our lab!

Definitions

The ideas of replicability, reproducibility, and open science are closely related.

- **Reproducibility** means obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis.
- **Replicability** means obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. If an analysis is not reproducible, it's hard for others to replicate it.
- **Open science** is the idea that scientific research should be transparent and accessible to everyone.

See [this National Academies News Release](#) for more details.

Scientific reproducibility

Unless there are specific and compelling reasons (which should be documented and reviewed by peers), the necessary components of a scholarly work to provide reproducibility should be provided in a publicly accessible location and potentially as part of the scientific record. In the case of a simulation model, this would include things like:

1. a specification of the computing environment
2. the source code that ran the simulation
3. the parameter file or definitions file, and if applicable initial conditions, that ran the simulation
4. analysis code that generated plots for the paper

Timing

We will make our data and code publicly available no later than with publication of our work, and often earlier.

Closed-source tools

Although our preference always is to use fully open-source tools, we occasionally use tools that we have a license to use but to disseminate. In such cases, some additional work is needed.

- If we are using proprietary software, the exact version of the external software should be documented and instructions for how to install it identically should be made clear. Wherever possible, we avoid using proprietary software!
- If we are using proprietary data, we should make clear the terms of use and how others can obtain the data. We should also make a reproducible example that uses open data to demonstrate the methods. See

Toolkit

We use lots of tools to facilitate reproducible science. Although it's now a few years old Wilson et al. (2017) provides an excellent summary of the tools and practices that can be used to facilitate reproducible science.

The specific tools used need to be chosen based on the specific needs of the project. For computing, we often use tools like Python and Julia for programming, and Git or GitHub for version control. We use tools like Snakemake for workflow management. Final datasets and repositories can be published on sites like Zenodo. Preprints can be published on places like Arxiv or EarthArxiv. We try not to be ideological about the specific tools we use, but rather

to ensure that the tools we use are appropriate for the task at hand and that we are using them in a way that is consistent with open science principles.

Software Licenses

Software and other research materials that we make available to the public must declare a license for re-use.

All code generated in the lab should be open-source with a permissive (non-copyleft) license (MIT, BSD, or Apache). In some cases, copyleft licenses like the GPL may be appropriate (the key distinction is that these require that derivative works also be open-source).

If contributions to upstream copyleft projects are made, those contributions can be licensed in accordance with the upstream project license.

Importantly, in the lab we should cultivate an atmosphere of respect for licensing terms and ensuring that we are at all times in compliance with those terms. For help choosing a license, see <https://choosealicense.com/licenses/> or read Stodden (2009).

Openness to the public

We seek to make our research accessible to all levels of an inquiring society, amateur or professional. This necessarily involves communicating methods and results to peers using scientific tools, and to the public through the channels.

Scientific communication

We communicate findings to our peers in academia and the broader research community through formal and informal writing, presentations, and code. In keeping with our open science principles, we:

1. publish in open-access journals where appropriate, and make post-prints freely available when not
2. make preprints of work available on open-access servers at the time of submission (and update with later versions of the paper)
3. summarize findings in technical blog posts
4. publish conference slides and posters on permanent repositories like figshare and Zenodo
5. make GitHub repositories for papers publicly available at the time of submission to a journal
6. make software available through permissive licenses
7. write and share clear and well-organized scientific code

Public communication

Some of our work is of general public interest or has the potential to inform policy. It would be unethical to allow only the best-educated members of the public to have access to this research – particularly since the effects of floods, unsafe water, and inadequate infrastructure disproportionately affect members of vulnerable and marginalized communities. To attempt to ensure that all members of the public have access to our research, we:

1. summarize findings in non-technical blog posts
2. translate papers and communications as appropriate
3. use social media to share key findings
4. write 2-page policy briefs or create video summaries of our work

as appropriate. It can be very challenging to write about complex technical topics in a manner that is readily understandable to the public.

Stodden, V. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science Engineering*, 11(1), 35–40. <https://doi.org/10.1109/mcse.2009.19>

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, 13(6), e1005510. <https://doi.org/10.1371/journal.pcbi.1005510>