A Commitment to Open Source in Neuroscience

Padraig Gleeson,[1] Andrew P. Davison,[2] R. Angus Silver,[1] and Giorgio A. Ascoli[3,*]
[1]Department of Neuroscience, Physiology and Pharmacology, University College London, Gower Street, London WC1E 6BT, UK
[2]Unité de Neuroscience, Information et Complexité, Centre National de la Recherche Scientifique, FRE 3693, 91198 Gif-sur-Yvette, France
[3]Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA
*Correspondence: ascoli@gmu.edu
https://doi.org/10.1016/j.neuron.2017.10.013

Modern neuroscience increasingly relies on custom-developed software, but much of this is not being made available to the wider community. A group of researchers are pledging to make code they produce for data analysis and modeling open source, and are actively encouraging their colleagues to follow suit.

Developing custom software for extracting, translating, analyzing, and visualizing experimental data, as well as code for modeling and simulating the mechanisms underlying the examined phenomena, is a crucial element of the work behind many publications in neuroscience today. This code can often be complex and involve many processing steps that cannot be fully described in the accompanying publications. Releasing these scripts is generally not a prerequisite for publication. Nevertheless, making them publicly available would increase the reproducibility and scientific rigor of the results described, and potentially accelerate the pace of research by making it easier to build on previous work.

Reproducibility Crisis

The inability to reproduce the findings of many published studies in neuroscience has been highlighted recently (Baker, 2016; Open Science Collaboration, 2015), and there is general agreement that this is a problem that needs to be tackled by the whole field. While there are practical reasons why not all of the experimental steps to acquire the data can be easily reproduced (specialist hardware, access to transgenic animals or reagents, etc.), this does not necessarily apply to the computational analyses carried out on the data. Such analyses are becoming more and more sophisticated, and there is a widening gap between the raw experimental data and the figures in the publication. Where the code is made available, these steps are much more transparent; not having access to underlying code can make replication of the study impossible.

Benefits of Openly Releasing Code

Making the code associated with a scientific publication openly available provides the opportunity for anyone in the community to reuse, build on, and improve the software the authors have developed (and will continue to use) in their labs; the publication becomes an advertisement for the usefulness of the software (Claerbout; paraphrased in Buckheit and Donoho, 1995). The willingness to share one's code and receive constructive feedback contributes to the reliability and scientific value of results obtained. As more groups share, test, and contribute to one another's software, openly releasing code will lead to a distributed and freely available network of tools, databases, and related resources for data analysis and model development, making neuroscience research more efficient and reliable.

Our Commitment

While there is growing consensus that such open sharing of code should happen, some members of the neuroscience community have decided to make an active, public commitment to this effect. As an outcome of the September 2016 conference at Janelia Research Campus on Collaborative Development of Data-Driven Models of Neural Systems, we wrote an open letter pledging to release promptly, completely, and freely all computer code, model scripts, and parameters necessary to reproduce the analyses and simulations from our future publications (http://opensourceforneuroscience.org/). The signatories of the letter commit to making all software applications (tools, libraries, etc.) they develop for experimental data analysis or model construction open source at time of publication, whether or not the software is the main subject of the paper. Importantly, if and when asked to serve as peer reviewers, we will henceforth ask authors about the availability of any code they have developed for data analysis and modeling that is essential to reproducing the results of their paper and require that this be shared publicly upon acceptance (see also Morey et al., 2016).

Code Sharing Doesn't Need to Be Onerous

Many researchers feel reluctant to share code, believing it may not be sufficiently well written or documented to be useful to others (Barnes, 2010). This shouldn't be a barrier to releasing it. There should be no obligation to support code after making it available; all that the authors are claiming is that it can be used to reproduce the results of that specific publication. Making the source files available to read allows others to find the parameters, algorithms, and/or assumptions used in the analysis or model that may be missing from the paper.

Releasing the core scripts and/or libraries as an open source package before the publication (without any data specific to the publication) allows early feedback from the community and can increase the usability and quality of the code. It also means junior researchers receive more instant recognition for their contributions, without having to wait until publication of their work.

Many Good Code-Sharing Solutions Exist

While simply making a zipped file available for download on the lab website is a good start, sharing code in an open, sustainable way while encouraging feedback is becoming easier and easier. GitHub is the choice of many open source projects (Perkel, 2016), and code can be shared there, updates made, and user issues answered, all through standard web browsers with no command line knowledge required. Other online version control repositories such as BitBucket and GitLab offer similar functionality. Zenodo can be used to give a permanent DOI to specific releases of one's code. Figshare is useful, too, for releasing (citable) datasets associated with the code. Jupyter notebooks can show how to interact with the code or data and can be the starting point for new users. More details of best practices for code sharing can be found in Eglen et al. (2017) and Prlić and Procter (2012).

Planning sooner rather than later what will be shared from your scripts, and in what form, will make it much easier to release them upon publication. Sharing the repository privately inside a lab, getting lab members used to committing code to a version control repository, merging changes, and opening and closing issues, before finally flipping the switch to make it visible to the rest of the world (and then continuing your development workflow exactly as before), is a good option to consider.

Funding Agencies and Journal Obligations

While altruistic motives and positive reasons for sharing code are clear, there are also more and more obligations on researchers to change their attitudes toward releasing the code they produce. Funding agencies such as Wellcome Trust in the United Kingdom and the NIH in the United States already have policies on code sharing and more will likely follow. Journals too are tightening their requirements on the code behind their publications (Bernard, 2017; Nature Methods Editorial, 2014). It will be more and more difficult to justify not being prepared to share code at the point of paper publication.

The signatories to this pledge will be asking about the status of sharing of code relevant to manuscripts during reviews. The eventual outcome of these requests will require case-by-case judgement calls, and will depend on personal perspectives on how to maximize scientific impact of the whole body of work. Nevertheless, this initiative shows that there are named scientists out there who will be demanding this of other researchers when they review their work.

Some Restrictions and Caveats

Of course, not all software can or should be shared with others. Software for controlling custom hardware, for managing data inside a lab environment, etc. will be of little use in another context. Software developed for commercial purposes is a special case. There may be valid reasons why this cannot be released, not least if there has been industry funding that restricts the rights of the code. However, if access to this code is required for verifying the claims of a paper, it will still have to be made available in a usable form at least to the reviewers.

There can be many interpretations of what "should" be shared, but there needs to be a transition from the old attitude that there is no obligation or motivation to release any code associated with a publication, to a mindset that the code release can improve the scientific worth of a publication as well as provide benefits for the lab and the community in general.

A Call to the Community

We invite all like-minded scientists, developers, users, and peers to join us in this pledge and encourage the wider community to sign the open letter at http://www.opensourceforneuroscience.org/.

REFERENCES

Baker, M. (2016). Nature *533*, 452–454.

Barnes, N. (2010). Nature *467*, 753.

Bernard, C. (2017). eNeuro *4*, https://doi.org/10.1523/ENEURO.0233-17.2017.

Buckheit, J.B., and Donoho, D.L. (1995). WaveLab and reproducible research. In Wavelets and Statistics (New York, NY: Springer), pp. 55–81.

Nature Methods Editorial (2014). Nat. Methods *11*, 211.

Eglen, S.J., Marwick, B., Halchenko, Y.O., Hanke, M., Sufi, S., Gleeson, P., Silver, R.A., Davison, A.P., Lanyon, L., Abrams, M., et al. (2017). Nat. Neurosci. *20*, 770–773.

Morey, R.D., Chambers, C.D., Etchells, P.J., Harris, C.R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C.C., Newman, D.P., Schönbrodt, F.D., et al. (2016). R. Soc. Open Sci. *3*, 150547.

Open Science Collaboration (2015). Science *349*, aac4716.

Perkel, J. (2016). Nature *538*, 127–128.

Prlić, A., and Procter, J.B. (2012). PLoS Comput. Biol. *8*, e1002802.