

## Credit for code

**Moving toward fully transparent research publications, we suggest several approaches to share research that is instantiated in software written for computers and other laboratory machines. Review, replication, reuse and recognition are all incentives to provide code.**

Software for biomedical research ranges from single scripts used to format data to complex suites of analytical tools. The biggest problem often encountered by editors, referees and readers is knowing where in the work any code has been employed. Thus, when submitting a manuscript to a journal for peer review, it is good practice to deposit annotated open source code in a recognized community repository that tests software under a set of standard conditions and to provide a unique resource locator. Because software rapidly versions and operates under diverse settings and with customization, it is also useful to offer the code actually used non-exclusively to the journal in a supplementary text or archived file. This latter solution is simple enough to deter malware while providing some guidance for determined readers.

Community repositories that carry out testing are ideal for commonly used programs (for example, those used in statistical analysis), and a fair proportion of the genetics community is fortunately familiar with the Comprehensive R Archive Network (<http://cran.r-project.org/>) and the principles of stewardship of modular software embodied in the Bioconductor suite (<http://www.bioconductor.org/>). The journal has sufficient experience with these resources to endorse their use by authors. We do not yet provide any endorsement for the suitability or usefulness of other solutions but will work with our authors and readers, as well as with other journals, to arrive at a set of principles and recommendations.

Two needs stand out. First, code should have permanent identifiers such as the Digital Object Identifiers used by publishers, and authors of code should receive attribution for their programs as well as for their publications. Second, data sets and the code to handle data should be stored together, as metadata can cover both and the repositories then become attractors for communities, sometimes even evolving into environments where stored code can be run by third parties on stored data (for example, *Nat. Genet.* 45, 1121–1126, 2013, doi:10.1038/ng.2761). We are pleased to see that the DataCite organization is indeed listing code storage as a feature in its index of data repositories (<http://www.datacite.org/repolist>). We think that this organization is a good one to engage with in coordinating efforts to attain the citation of data sets and code.

If these best practices are not possible, there are ways not to make the current situation worse. For example, it is a good idea to provide a unique resource locator for the version of the code at a corporate or institutional site. Much open source software is currently archived at the commercial site GitHub (<https://github.com/>). If none of these solutions are feasible, please do declare when there is code involved in the work, even if it is proprietary or unavailable, and provide equations or algorithms that enable a reader to understand and replicate analytical decisions made with the research software. ■