

# Code deposition is unskippable



**As computational analysis becomes ever-more ubiquitous for researchers, the deposition of the underlying code is now an expected part of publication. Shortcomings in code sharing can lead to delays in peer review and publication, as well as reproducibility issues that are easily avoided with author preparation.**

It is well-appreciated that computational analysis of data is important for the genetics and genomics fields, and thus to many of our submissions. As large-scale genome-wide association studies (GWAS) and genomics profiling studies have generated progressively larger datasets that are then subjected to increasingly complex analytic workflows, these computational aspects now constitute a vital part of many studies. As the centrality of computational work has become apparent, expectations from the community are that – as well as the underlying data – the analysis code should be shared upon publication to ensure reproducibility.

In line with this, *Nature Genetics* has a [policy](#) regarding the availability of computer code. Before publication, we require ‘Code availability’ statements, which explain how code will be shared, and the deposition of all code required to reproduce the major results of an analysis at a DOI-minting repository such as [Zenodo](#) before publication. Because code repositories are dynamic, future changes to repositories may affect the reproducibility of a manuscript in the long run – a concern that

applies to both the authors’ own websites and those for commonly reused pipelines. Thus, requesting a DOI such as this will provide a permanent and exact snapshot of your repository at the time of publication.

In our observations of peer review, there is an expectation that well-documented code also be made available to the referees. This is an especially noticeable trend for manuscripts presenting a new software method or tool – one factor in our editorial assessment of these studies is whether a tool will be widely used, and we can and do request that reviewers try out the software. If referees find that the authors’ software cannot be installed or generates unresolvable errors when run, this is likely to present a notable barrier for a positive decision. In this respect, providing a program as a pre-compiled binary or R/Python package is often explicitly acknowledged by reviewers and is therefore a point of interest to us editorially.

Analysis scripts are of course distinct to a complete piece of software, but are also increasingly requested, especially for studies where computational results form a major part of the findings. A recent survey<sup>1</sup> showed that examining code is a common way for computational biologists to understand an article better, and those who have tried to write a methods section for a computational analysis will no doubt appreciate that crafting informative, yet concise text can be a difficult prospect.

We acknowledge that providing analysis code as executable scripts, or sharing one’s workflow as a Docker container or similar, can require substantial amounts of time that may be viewed as ‘non-productive’.

In the context of very large datasets that require substantial computational resources such as high-performance computing clusters for analysis, the idiosyncrasies of each researcher’s computing environment, and other such factors, we do not mandate that deposited code be executable. But the value of even basic, adequately documented and commented scripts (including version numbers of tools and packages used) to a reader attempting to reproduce an analysis cannot be underestimated. Using tools that are designed to support reproducible research, such as [R Markdown](#), package management systems including [Conda](#), or executable notebooks such as [Jupyter](#), when conducting computational studies can substantially ease the amount of further work required to make code shareable.

Our sister journals have also adopted [checklists](#) and [guidelines](#) that we suggest authors consider when preparing their code for submission and publication. In the same way that standards regarding data deposition have evolved over time, such that broad sharing is now generally assumed the norm, these principles on code deposition are becoming more and more prevalent. As a journal, *Nature Genetics* is fully supportive of open science and we encourage our authors to bear this in mind when conducting their research, for the betterment of the field as a whole.

Published online: 12 May 2023

## References

1. Cadwallader, L. & Hrynaszkiewicz, I. *PeerJ* **10**, e13933 (2022).