Hail, software!

Software is essential to computational science research, and yet it hasn't achieved first-class status when it comes to citations. It's time for all of us in the research community to change this behavior.

From designing complex mathematical models to using sophisticated computing infrastructure, computational science tasks are tested and validated with computer code, or more broadly speaking, software. One cannot think of computational science research without thinking of its associated software, be it a set of scripts that assess hypotheses discussed in a paper, or a full-fledged system containing several features that help researchers perform scientific activities. It is not an overstatement to say that computational science depends on software development.

For this reason, when it comes to the writing of a scientific manuscript, its associated software deserves praise. The software component doesn't just complement the manuscript: the software is as important as the manuscript. While the paper provides details about the research conducted and the science behind it, including background, motivation, algorithms, and a discussion on the relevant results and limitations, the software allows for the reported results to be properly validated, reproduced and reused in the future. There are many software implementation idiosyncrasies that are hard to capture in a manuscript, but those make all the difference when trying to reuse a research compendium. For instance, the programming language and the optimization techniques used by the researchers might make the manipulation of a particular dataset possible or the execution of the proposed algorithm substantially efficient, which adds to the original contributions of the manuscript by providing a usable resource to the community. Furthermore, in computational science, the software is often the main research outcome — for instance, when authors propose a tool to perform a previously intricate task or to make it more computationally efficient — and thus, in those cases, considering a manuscript without the software is meaningless.

If software is so essential, why hasn't it attained the same level of importance that manuscripts do when it comes to citations?

Properly citing software, similar to how we cite papers, theses and books, has many benefits. First and foremost, it gives proper attribution and recognition to those involved in the development of the software. Note that the software development might include researchers and developers who are not necessarily co-authors of the associated manuscript, meaning that they might not be involved in the research that uses that software. Therefore, citing the manuscript only — instead of citing both the manuscript and the software — doesn't support proper attribution and credit. Second, just as manuscript citation helps to track how the associated research is being used, software citation helps to track the usage of the software. Perhaps most importantly, it helps to track how the code is being repurposed by other researchers, which encourages reproducibility: building on the work of others is very common, and software citation can help identify these cases, thus fostering further collaborations. Finally, software citation gives the well-deserved praise to software products, adequately elevating their importance in computational science as first-class research objects.

It is worth mentioning that data, which correspond to the digital records used and produced by a research study, are also essential research objects, as they provide evidence for the reported results. Similar to software citation, data citation gives attribution to those researchers who collect and process these records, and helps to track where these records are being used. However, while data citation has been recommended, and sometimes mandated, over the past several years by major science policy bodies[1], software citation still lacks proper endorsement from scholarly organizations, funders and publishers.

The good news is that software citation is gaining traction in the community. Just recently, the FORCE11 Software Citation Implementation Working Group published guidelines and best practices to help authors on how to properly cite software[2].

For instance, as part of the guidelines, the group argues that the use of persistent identifiers (PIDs) is essential, not only because, as the name suggests, these are long-lasting references, but also because these references accompany relevant and descriptive metadata about the resource. In addition, PIDs allow for the resources to be properly indexed and tracked, which is one of the main goals with software citation.

Nature Portfolio journals are taking steps to ensure that authors follow these guidelines and use repositories that can mint digital object identifiers (DOIs), a type of PID. By utilizing these repositories, authors should publish the software associated with a manuscript and properly cite it in the reference list of the manuscript. The importance of using DOIs cannot be overstated: they provide a long-lasting, persistent reference for readers, ensuring the longevity of the reproducibility of the results. In addition, software, especially in open source communities, is often characterized by an ongoing development that goes beyond the manuscript publication, and therefore, it is important to capture the exact version that was used to generate the results published and discussed in the manuscript. Authors can — and should! — include links to the corresponding code repositories (such as GitHub and GitLab), in order to foster collaboration and community-building, but these links do not replace the relevance of the DOIs.

At *Nature Computational Science*, we will do our best to make sure that the code associated with a manuscript has a DOI and is properly cited. We encourage you, our reader and future author and contributor, to also do your part: think about how important software is to your research and daily tasks, and don't let it down! ❐

References
1. Cousijn, H. et al. *Sci. Data* **5**, 180259 (2018).
2. Katz, D. S. et al. *F1000Research* **9**, 1257 (2021).