

Data availability is the golden rule in research



As the saying goes, sharing is caring – but when it comes to the data underlying primary research papers published at *Nature Microbiology*, sharing is also mandatory.

At the Nature Portfolio we share an [Editorial Values Statement](#) that encapsulates the guiding principles of the professional editors at each of our journals. In brief, we believe that science drives positive change and that journals support this through the curation, enhancement and dissemination of impactful research. At *Nature Microbiology*, our aim is to publish the most influential microbiological findings. In the long term, the impact of a paper can be judged by the research it inspires. On the most practical level, this is only possible if the data supporting the conclusions of a study are provided at publication. Viewed through this lens, publicly available and reproducible data are the fuel that drives the positive changes our Editorial Values Statement highlights. This is an ongoing commitment at the journal (*Nat. Microbiol.* **2**, 1573; 2017), crucial for bolstering public confidence in science, and one that we continue to refine as we strive to support our authors, readers and reviewers. In this Editorial, we present an overview of best practices, policies and preparation when it comes to data sharing in primary research.

One phrase that we frequently encounter in early drafts of data availability statements is some variation of “Data are available upon reasonable request.” We recognize that the provision of data and code can be complex and time consuming, and we acknowledge that science is competitive and scooping is a concern. However, the principle that research should be reproducible is a requirement at all journals published by [Springer Nature](#), and by many funders, and reflects community-driven initiatives for open science. According to the Nature Portfolio [editorial policies](#), as a condition of acceptance of research articles for publication, authors are required to make materials, data, code and associated protocols available to readers. Each paper is published along

with a data availability statement and code availability statement that can be found after the Methods section. The inclusion of these statements in published papers – piloted at select journals in 2016 and extended across the portfolio the following year – points readers to the “minimal data set” required to interpret or replicate the results of the study (*Nature* **537**, 138; 2016).

Given the complexity and multidisciplinary nature of research projects, keeping track of data and code is not straightforward. In a perfect world, planning for compliance with data and code availability policies would start at the inception of a research study, with adequate funding to ensure this is done appropriately. At a minimum, the design of experiments, meta-data collection and record keeping should be done with an eye towards which data require mandatory deposition as a condition of publication, as well as the reporting requirements for approved repositories.

Several data products require mandatory deposition as a condition of publication across the Nature Portfolio. These include DNA, RNA and protein sequences (including omics sequences), gene expression data, genetic polymorphisms, linked genotype and phenotype data, and macromolecular structures and crystallographic data. These data must be [deposited](#) to an approved repository for that data type. If there is no approved repository for a given data type, it may be deposited to a general data repository such as Figshare or Zenodo. All other data necessary to understand and replicate the conclusions of the paper – the data underlying all graphs and other display items, uncropped gels and blots, or additional representative micrographs – should be submitted as source data that will accompany the publication. Some data, for example those generated during clinical research or linked to patients, might require controlled access. These cases require data availability statements with a clear delineation of who can access the data and how, and more information can be found in our [reporting standards](#). In all cases, any potential restrictions on data availability must be brought to the attention of the editors at the time of initial submission.

Some data and materials do not have mandated deposition under our policies; however, we emphasize that making these resources available is good practice. For example, deposition of metabolomics data is not currently mandated – although such policies are regularly reviewed by the editorial community and are subject to change – and we strongly encourage making these data publicly available given the availability of repositories such as [Metabolomics Workbench](#) and [MetaboLights](#). In addition, any proprietary materials, such as plasmids or strains, generated and used in the study are also required to be made available to others without any undue qualifications. We also point readers to our sister journal, *Scientific Data*, a peer-reviewed and open-access journal for Data Descriptors, papers that support the reuse of data and provide official credit to those who share.

When it comes to the code used in a study, availability is key, just like any method or experiment used in the work – transparency enables evaluation of the data generated with code and ensures reproducibility. *Nature Microbiology* considers deposition of code to a DOI-minting repository and citation in the paper’s reference list to be the best practice. Zenodo fits these requirements, and it can be used to issue a persistent identifier for code stored on [GitHub](#). Adhering to these practices ensures that the code will remain publicly available.

Advanced planning in the data and code deposition department can help when it comes to peer review too, as referees should assess this material during review of the study. It will also help expedite the final steps towards publication. As in all matters related to the publishing process, the editors at the journal are your point of contact for any issues or concerns. We are happy to help, and we understand that data sharing can be complex. Overall, good data availability practices will help others read and cite your paper, ultimately ensuring that the research enterprise keeps running. When in doubt, our editorial advice is to err on the side of making data available – this benefits the authors and the community, and enables the advancement of science.

Published online: 4 April 2024