

Sharing's caring

Our data-sharing policies continually evolve. Here we provide an overview of our current approaches.

Research data are a cornerstone of the scientific ecosystem. Their value independent from papers and reports is increasingly recognised by the academic community: funders are building stronger mandates around the sharing of data arising from their grants; more initiatives aim to provide appropriate credit for datasets as unique resources; new tools are allowing communities to collate and centralize data to help make research more efficient; and publishers are implementing new policies to help underpin and further drive such efforts.

The Nature Portfolio is no different in this respect. Our journals have worked hard to introduce standards and practices around research data that support the communities we work with. These policies are continually assessed and iterated as the landscape changes, reflecting the goals of the different fields we serve.

At the heart of the Nature journals' [data policy](#) is the desire to increase openness and transparency in the availability of materials. We believe that researchers should be able to replicate and build on the studies we publish.

At a minimum, this means that our authors must make materials, data, code and protocols available to any readers who request them. This should be done promptly and without undue qualifications.

To expand this principle, in 2016 we introduced the data availability statement (DAS). This should describe how a reader can access the minimum dataset required to interpret, verify and extend the research presented in an article. It should include information about data generated during the study, third-party data referenced as part of the analysis and any restrictions to data, as applicable. Articles that are sent to peer review must carry a DAS.

For third-party data, authors must ensure that they have permission to use it in publications, especially for proprietary data. We recognise that such data cannot always be made widely available. In such cases, the DAS should outline these restrictions, the data source (where possible) and any costs that may be associated with the data.

However, where data are only available on request from the authors, we ask that the DAS explain why the data cannot be made publicly available. Again, this

description should appear at the outset of the submission process.

A similar approach applies to custom code. In such cases, a code availability statement should outline the availability of such materials.

In both cases, authors should ensure that all necessary materials — or associated restrictions — are available for the editor and reviewers.

We strongly encourage authors to make their data publicly available where possible. Such practice helps your peers to more readily extend your work and make new breakthroughs and discoveries. It also helps the community more easily identify errors and work towards improved reporting standards that ultimately benefit everyone.

There are many ways in which data can be made publicly available.

In the first instance, authors should check the requirements of their funding body or institutions. Increasingly, many such bodies host data repositories where materials can be stored, catalogued and linked in accordance with their best practices.

Beyond these requirements, Nature journals encourage the use of community or disciplinary repositories, or of general outlets like Figshare, Zenodo or Dryad. A list of approved and recommended repositories is maintained by our sister title *Scientific Data* at <https://www.nature.com/sdata/policies/repositories>.

A key factor in the choice of repository is persistence and version control. We encourage our authors to make use of DOI-minting repositories; these provide greater assurance that the resources will still be discoverable later on.

We encourage similar practice for custom code. Although we recognise the prevalence and value of platforms such as GitHub, they do not provide assurance that code will be permanently accessible in the way that a DOI-issuing service does.

Not all data needs to be stored in repositories, of course. In many cases, all the data necessary to interpret, verify and build on a study may be contained within the paper already, including its Supplementary Information.

To further support such simpler cases, our journals introduced the use of Source Data files. These optional files correspond

to figures in the paper and present their underlying data. This could be something as simple as the coordinates presented in a line graph or the values of a bar chart, where this is deemed valuable.

Where statistical aggregates are present in figures, however, Source Data should provide readers with the constituent data points that form the aggregation — not just the average values plotted. This allows the reader to gain insight into the variation in the measurements taken, which may be obscured by the aggregation, especially with small numbers of measurements.

When presenting statistical aggregates, it's important to consider whether there is validity in doing so. Our [editorial policy](#) requests that individual data points are plotted in cases where there are ten or fewer measurements.

Having a small number of measurements is not necessarily a problem — practical considerations can frequently mean that only a small number of devices or characterisations can be achieved. However, presenting an average with error bars on only a small number of samples (and certainly on fewer than five) risks misleading a reader over the distribution of your measurements. Such practice should be avoided, in favour of presenting the constituent data.

When averages and error bars are presented, it remains important to describe the sample size, the type of average (mean, median and so on) and the definition of the error bars (standard deviation, standard error on the mean, confidence intervals and so on).

Our policies around data sharing and transparency are intended to help the scientific community. This spirit should be kept in mind while navigating the manuscript drafting and submission process. As an author, it can be helpful to think about how you as a reader would want to make use of your paper and what materials you would want to access to help further your own work. In such ways, we can hopefully continue to improve data sharing practices for the benefit of all. □

Published online: 22 April 2021
<https://doi.org/10.1038/s41560-021-00825-6>