

All about data sharing

Making data available is essential for validating and furthering scientific discoveries. Helping authors navigate whether, how and in what form to share the data is also essential.

If we have learned something during the last year, it is that sharing data and scientific results is essential for addressing a public health emergency. The World Health Organization [noted this](#) early in the COVID-19 pandemic, but had already emphasized this need during the outbreak of Ebola virus disease in 2015 in a [statement](#) issued after the convening of stakeholders from academia, the industry, governments and publishers.

Beyond tackling public health emergencies, sharing scientific findings and the data that underpin them is inherent to the research endeavor itself. Data sharing is necessary for validating results, for following up on and extending discoveries, and for establishing and disseminating scientific knowledge. It can also lead to more collaborations, as others engage more deeply with the data underlying a publication. Additional benefits are the credit to the original investigators and testament to the impact of their work—for example, sharing data in a paper by providing links to a repository has been associated with a 25% increase in citations¹.

Over the past several years, many major funding agencies, including the National Institutes of Health and National Science Foundation in the USA, the European Research Council, and the Wellcome Trust, Cancer Research UK and the Research Councils in the UK, have adopted more stringent data-sharing requirements for their grantees. Many publishers also have relevant mandates, including the Nature Portfolio journals. However, the days when a dataset could be provided within the few pages of a scientific publication are long gone. Especially in multidisciplinary fields such as cancer research, the results reported in one paper are, more often than not, supported by several distinct and complex datasets, from classic cell and molecular biology experiments and animal work, to large-scale ‘-omics’ data and analyses of clinical samples obtained from human participants. Navigating whether, how and in what form to share the data associated with a paper can be complicated.

Nature Cancer authors can find all information related to our [data availability policy](#) on the dedicated webpage of the Nature Portfolio and can receive help for their research data-related queries through

the [Springer Nature Research Data Help Desk](#). Moreover, they can obtain specific advice for the datasets in their submitted manuscripts from the *Nature Cancer* editors.

Making data available to readers without undue qualifications is a condition for publication in *Nature Cancer*. Sharing the data should be the norm unless there are justifiable restrictions on availability, which we ask authors to disclose at the time of submission. If these restrictions are found to be unduly prohibitive, we may decline further consideration of the manuscript. As a general rule, we advise authors to ensure that all relevant data are available either within the manuscript files or through deposition in appropriate public repositories. For certain data types, we mandate deposition and ask authors to provide access to editors and referees before we send a manuscript to peer review. These include gene expression, DNA and RNA sequencing and proteomic data; nucleic acid and protein sequences; and macromolecular structures, genetic polymorphism and linked genotype and phenotype data. We also strongly encourage public provision of other data types for which community-endorsed repositories exist, such as metabolomics and imaging data. In cases for which discipline-specific repositories do not exist, we recommend the use of unstructured, generalist repositories such as Figshare, Dryad and Zenodo. In general, we may require deposition in a public repository of any dataset and data type that is deemed to be central to the main message of the study or essential for reproducibility of the reported findings. To help authors identify the appropriate public repository for their data, our sister journal *Scientific Data* maintains a [dedicated list](#) of approved and recommended data repositories. Apart from the deposition of datasets before peer review, we also advise inclusion with the submitted manuscript files of unprocessed images of gels and blots, and of all raw numerical data behind graphs and statistical analyses. We ask authors to include these source data files when we invite a revision and ultimately require them for publication of the study.

To inform readers of the conditions of availability for the data that underlie the reported findings, the Nature-branded journals have been mandating the inclusion of data availability statements in all published

primary research papers since 2016 (ref. ²). These statements are stand-alone sections of the paper that explain how the minimum dataset that supports the reported results can be accessed by others. Therein authors list information (including accession numbers, distinct object identifiers, references and links) for accessing datasets generated for their particular study; public or previously published datasets re-analyzed in the study; and source data that may be included with the manuscript. For clinical trial data in particular, we ask authors to provide detailed information on data sharing, following the relevant [recommendations](#) of the International Committee of Medical Journal Editors. Authors must also state any restrictions and specific conditions for data access, whether these relate to controlled access or lack of access—for example, due to ethical, legal or privacy concerns for human data or because of data provenance from third parties.

This detailed information on data deposition and availability, including the full data availability statement, is provided by authors before peer review in the [Reporting Summary](#), a document they are required to complete to aid evaluation of the paper by editors and referees. The Reporting Summary is updated in revision and is ultimately published with the manuscript. Throughout the submission and peer-review process, *Nature Cancer* editors are available to advise on data sharing and also evaluate any aspect that might impede further consideration of the study. When offering publication of the study, we also give detailed guidance for optimal reporting of data availability information in the manuscript.

Modern scientific discovery requires that data be available, discoverable and re-usable in the longer term. The policies and initiatives outlined here aim to help authors achieve these objectives to enhance the reproducibility, reach and impact of their work. □

Published online: 25 May 2021
<https://doi.org/10.1038/s43018-021-00217-5>

References

1. Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K. & McGillivray, B. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0230416> (2020).
2. *Nature* <https://www.nature.com/news/announcement-where-are-the-data-1.20541> (2016).