

Data availability statements and data citations policy: Frequently Asked Questions (FAQs)

The policy and guidance for authors on compliance is available [here](#).

Q: Why have data availability statements?

A: Sharing of research data is a condition of publication in Nature journals and the preferred way to share data is via public repositories, although this is not mandated for all data types. Requiring data availability statements signals our commitment to making the availability of data associated with Nature journal articles transparent to readers. The mandatory data availability statement consolidates information on availability of data associated with the manuscript, whether data are in repositories, available on request or included with supplementary information or figure source data files. This policy is consistent with our drive to support reproducibility and robustness in research. This policy also aims to improve the consistency of information on data availability in published articles.

Data availability statements also support researchers' compliance with the requirements of funding agencies. [Many funding agencies](#), including the National Science Foundation (NSF), National Institutes of Health (NIH), Wellcome Trust, Research Councils UK (RCUK) and Bill and Melinda Gates Foundation require data sharing as a condition of grants. Some of these funding agency policies, such as the UK's Engineering and Physical Sciences Research Council (EPSRC)'s policy, also require statements about data accessibility and links to data to be included in articles.

Where public data archiving is a mandatory requirement of journals, there is [some evidence](#) that including data availability statements with persistent links to data in published articles is the most effective approach to ensuring public data availability and policy compliance.

Requiring data availability statements is also consistent with the [Research Data Policies](#) set out by our publisher, Springer Nature.

Q: What are the benefits to authors of data availability statements?

A: Data availability statements help authors demonstrate compliance with funder policies on data archiving, where applicable, and also shows compliance with Nature Research editorial policies.

More consistent links between public datasets and journal articles increases the visibility of research.

Q: What are the benefits to readers of data availability statements?

A: Data availability statements provide more transparent and consistent information about where and how data supporting published articles are available. This supports the reuse, where possible, of data for further research and validation or reanalysis of findings by other researchers.

Q: Does this policy alter data sharing policies at Nature Research?

A: No, Nature Research's [policies on the availability of data](#), and other materials to support reproducible research, are unchanged by the introduction of data availability statements. While we are not introducing new data sharing mandates with this policy, the introduction of data availability statements supports our long-standing policies on data availability as a condition of publication in Nature journals and it aims to make the conditions for data availability more transparent to our readership. Many articles published in the Nature family of journals already include data availability statements in some form, and this new policy standardises our approach.

Q: What data should be described in the data availability statement?

A: Data availability statements should provide a statement about where data supporting the results reported in the article can be found. Nature Research journals strongly encourage provision of a "minimal data set" underlying the figures provided in the paper that is necessary to support the central findings of the study, and to interpret, analyse or reproduce the methods and findings. This minimal dataset may be provided through deposition in public repositories or associated with the paper in supplementary information files and figure source data files. This would include datasets generated or analysed during the study, and source data that are necessary to interpret, replicate and build upon the methods and findings reported in the article.

Q: How should data about identifiable human research participants or other sensitive data be described?

A: In the absence of consent for publication or complete anonymisation, public sharing of data about potentially identifiable human participants in research is usually not possible due to the need to protect research participant privacy. Where human or other data are only available on request, at minimum, a named group or individual to whom enquiries about data access can be made must be provided. This may often be the corresponding author but responsibility for data access should ideally not lie solely with an individual, whose contact information could change.

A limited number of data repositories that archive sensitive data – such as dbGAP, the European Genome-phenome Archive (EGA), the Cancer Imaging Archive, and UK Data Archive – provide permanent online summary records (metadata records) about sensitive datasets. In these cases, the summary record or accession code should be linked or cited in the data availability statement. Such repositories may, also, have application procedures for obtaining access to data, which may only be granted to researchers whose request is approved. Such restrictions should be documented in the data availability statement. See further [guidance on managing research data](#), [data anonymisation](#), and [linking sensitive data to publications](#).

The locations of endangered species or fossils or other sensitive material can be a reason for data being available on request. In such cases it

may be appropriate not to name an individual responsible for providing data access but the research institution or an appropriate body at the institution.

Q: Is public sharing of all research data mandatory?

A: A condition of publication in a Nature Research journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript. The preferred way to share large data sets is via public repositories. For certain types of data, data sharing is mandatory. See our [data availability policy](#) for more information.

Our requirement, as of September 2016, for all papers to include a data availability statement does not introduce new data sharing mandates however it makes the conditions for data availability more transparent to readers.

Q: Why cite data in the references?

A: Citing and referencing data in publications supports reproducible research, by increasing the transparency and provenance tracking of data generated or analysed during research. Citing data formally in reference lists also helps facilitate the tracking of data reuse and may help assign credit for individuals' contributions to research. It also helps recognise datasets as a legitimate published scholarly work, which are admissible in research assessment procedures.

Data citations in reference lists are not mandated by the Nature journals but there is growing support in some research communities, and in the publishing industry, for the citation of data in published articles. The publisher of the Nature Research journals along with numerous other organisations is a signatory of the Joint Declaration on Data Citation Principles, which stress the importance of data resources in scientific communication and the need for citation to facilitate credit and attribution to those who contribute to data generation.

Q: What should readers do if they cannot access data when the data availability statement says data are available?

A: As per our existing [policy on availability of data and materials](#), readers should contact the Chief Editor of the journal if they encounter problems in obtaining access to data supporting publications.

Q: What is the best repository for research data?

A: Please refer to our [data and materials availability policy](#) and [list of recommended repositories](#) for information.

Q: Can data be deposited in institutional repositories?

A: We support the use of institutional repositories in preparing data availability statements and data citations, where authors are required to use them as part of funder or institution policy and where institutional repositories are able to provide DataCite DOIs for hosted data. See our [list of recommended repositories](#) for more information.

Q: Are the data citations fully machine readable?

A: Currently, we are focusing on improving the accessibility and human readability of data citations and links in our articles. We recognise the importance of systems that enable full machine readability of data citations while ensuring robust links between articles and supporting data. We are investigating the implementation of machine readable data citations in our content at a broad scale.

Q: When is this policy being introduced?

A: Five Nature journals (*Nature Cell Biology*, *Nature Communications*, *Nature Geoscience*, *Nature Neuroscience*, *Nature Physics*), piloted the policy beginning in March 2016. In September 2016 *Nature*, and several other journals introduced the policy, signalling our commitment to roll out the policy to all Nature Research journals by early 2017.

Further information about the Research Data Policies of Springer Nature journals is available [here](#).