# nature research

# Data availability statements and data citations policy: guidance for authors

## Policy summary

All manuscripts reporting original research must include a data availability statement. Authors are also encouraged to include formal citations to datasets in article reference lists where deposited datasets are assigned Digital Object Identifiers (DOIs) by a data repository.

This policy builds upon our long-standing policy on data availability, which requires that authors make materials, data, code, and associated protocols promptly available to readers without undue qualifications. The preferred way to share large data sets is via public repositories. For certain types of data, data sharing is mandatory. This policy does not introduce new data sharing mandates but aims to make the circumstances for data availability more transparent to readers.

### ▶ Writing a data availability statement

Data availability statements should provide a statement about the availability of data supporting the results reported in the article. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the methods or findings reported in the article.

The statement should be placed at the end of the Methods section (titled, 'Data availability'), after the code availability statement if one is present. For papers that do not have a Methods section, data availability statements should be provided as a separate section before the References or Acknowledgements, whichever comes first.

Data availability statements should include, where applicable, accession codes, other unique identifiers and associated web links for publicly available datasets, and any conditions for access of non-publicly available datasets. Where figure source data are provided, statements confirming this should be included in data availability statements. Depending on the data described in the publication, data availability statements commonly take one of the following forms or may be a composite of the statements below:

- The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article (and its supplementary information files).
- The datasets generated during and/or analysed during the current study are not publicly available due to [REASON(S) WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
- No datasets were generated or analysed during the current study.
- The data that support the findings of this study are available from [THIRD PARTY NAME] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [THIRD PARTY NAME].

### ▶ Published examples

| Statement type | Example statement text | Published example(s) |
| --- | --- | --- |
| Data generated during the study available in a public repository (mandated deposition, DOIs not provided) | *[Data type e.g. "Sequence"] data that support the findings of this study have been deposited in [repository name e.g. "GenBank"] with the [primary] accession codes [list accession codes e.g. "KP253039" or "KP253039 (http://www.ncbi.nlm.nih.gov/nuccore/KP253039)"]* | *Nature Communications* <br> *Nature Communications* |
| Data available in a public (institutional, general or subject specific) repository that issues datasets with DOIs (non-mandated deposition) | *The [data type] data that support the findings of this study are available in [repository name e.g "figshare"] with the identifier(s) [data DOI(s) e.g. "doi:10.6084/m9.figshare.1499292_D8"][Reference number]* | *Nature Communications* <br> *Nature Physics* |
| Data available in a public (institutional, general or subject specific) repository that does not issue datasets with DOIs (non-mandated deposition) | *The [data type e.g. "Snow pack depth"] data that support the findings of this study are available in/from [repository/resource name e.g. "GlobSnow"], [hyperlink to dataset(s)/data source e.g. "http://www.globsnow.info/"].* | *Nature Geoscience* |
| Data available on request from the authors | *The data that support the findings of this study are available from the corresponding author upon reasonable request.* | *Nature Communications* <br> *Nature Neuroscience* |
| Authors can confirm that all relevant data are included in the paper and/or its supplementary information files | *The authors declare that [the/all other] data supporting the findings of this study are available within the paper [and its supplementary information files].* | *Nature Communications* |
| Data are available on request due to privacy or other restrictions | *The data that support the findings of this study are available on request from the corresponding author [author initials]. The data are not publicly available due to [state restrictions e.g. "them containing information that could compromise research participant privacy/consent"].* | *Nature Communications* |
| Standard statement where figure source data are provided (must be used in combination with another statement) | *Source data for figure(s) [number(s)] are provided with the paper.* | *Nature Cell Biology* |

| Statement type | Example statement text | Published example(s) |
|---|---|---|
| Data subject to third party restrictions | *The data that support the findings of this study are available from [third party name] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [third party name].* | *Nature Energy* |
| Data generated at a central, large scale facility – common practice in some physical sciences | *Raw data were generated at the [facility name] large-scale facility. Derived data supporting the findings of this study are available from the corresponding author upon request.* | *Nature Physics* |
| Data sharing not applicable – for example if an article describes entirely theoretical research | *Data sharing not applicable to this article as no datasets were generated or analysed during the current study.* | *Nature Communications* |

More examples are available on the Springer Nature Research Data Policy website.

# ▶ Writing a data citation

By "data citation" we mean the text and identifying information for a publicly available dataset or data resource that is formally cited in an article's reference list. These are very similar to citations to papers, book chapters and other scholarly works.

Citations of datasets should include the minimum information recommended by DataCite and follow Nature Research style i.e. authors, title, publisher (repository name), identifier. Dataset identifiers including DOIs should be expressed as full URLs. For example:

- Hao, Z., AghaKouchak, A., Nakhjiri, N. & Farahmand, A. Global Integrated Drought Monitoring and Prediction System (GIDMaPS) Data sets. *figshare* http://dx.doi.org/10.6084/m9.figshare.853801 (2014)
- See a published example in *Nature Communications*.

For more detailed guidance and Frequently Asked Questions see this document.