# Data Descriptor Template

**Scope Guidelines**

**Data Descriptors** submitted to *Scientific Data* should provide detailed descriptions of valuable research datasets, including the methods used to collect the data and technical analyses supporting the quality of the measurements. Data Descriptors focus on helping others reuse data, rather than testing hypotheses, or presenting new interpretations, methods or in-depth analyses. Relevant datasets must be deposited in an appropriate public repository prior to Data Descriptor submission, and their completeness will be considered during editorial evaluation and peer review. The data must be made publicly available without restriction in the event that the Data Descriptor is accepted for publication (excepting reasonable controls related to human privacy issues or public safety).

### Title

*110 characters maximum, including spaces*

Titles should avoid the use of acronyms and abbreviations where possible. Colons and parentheses are not permitted.

### Authors

Firstname Lastname1, Firstname Lastname2

**Affiliations**

1. institution

2. institution

corresponding author(s): Firstname Lastname (email@address)

### Abstract

*170 words maximum*

The Abstract should succinctly describe the study, the assay(s) performed, the resulting data, and their reuse potential, but should not make any claims regarding new scientific findings. No references are allowed in this section.

### Background & Summary

*700 words maximum*

The Background & Summary should provide an overview of the study design, the assay(s) performed, and the data generated, including any background information needed to put this study in the context of previous work and the literature, and should reference literature as needed. The section should also briefly outline the broader goals that motivated collection of the data, as well as their potential reuse value. We also encourage authors to include a figure that provides a schematic overview of the study and assay(s) design.

### Methods

The Methods should include detailed text describing any steps or procedures used in producing the data, including full descriptions of the experimental design, data acquisition assays, and any computational processing (e.g. normalization, image feature extraction). Related methods should be grouped under corresponding subheadings where possible, and methods should be described in enough detail to allow other researchers to interpret and repeat, if required, the full study. Specific data outputs should be explicitly referenced via data citation (see Data Records and Data Citations, below).

Authors should cite previous descriptions of the methods under use, but ideally the method descriptions should be complete enough for others to understand and reproduce the methods and processing steps without referring to associated publications. There is no limit to the length of the Methods section.

**Code availability**

For all studies using custom code in the generation or processing of datasets, a statement must be included in the Methods section, under the subheading "Code availability", indicating whether and how the code can be accessed, including any restrictions to access. This section should also include information on the versions of any software used, if relevant, and any specific variables or parameters used to generate, test, or process the current dataset.

### Data Records

The Data Records section should be used to explain each data record associated with this work, including the repository where this information is stored, and to provide an overview of the data files and their formats. Each external data record should be cited using the data citation format presented at the end of this template (e.g. "Data resulting from Method X can be found in xxxxx.txt (Data Citation 1)"). A data citation should also be placed in the subsection of the Methods containing the data-collection or analytical procedure(s) used to derive the corresponding record.

Tables should be used to support the data records, and should clearly indicate the samples and subjects (study inputs), their provenance, and the experimental manipulations performed on each (please see Tables and Submitting Experimental Metadata, below). They should also specify the data output resulting from each data-collection or analytical step, should these form part of the archived record.

### Technical Validation

The Technical Validation section should present any experiments or analyses that are needed to support the technical quality of the dataset. This section may be supported by figures and tables, as needed. *This is a required section*; authors must provide information to justify the reliability of their data.

Possible content **may include:**

* experiments that support or validate the data-collection procedure(s) (e.g. negative controls, or an analysis of standards to confirm measurement linearity)
* statistical analyses of experimental error and variation
* phenotypic or genotypic assessments of biological samples (e.g. confirming disease status, cell line identity, or the success of perturbations)
* general discussions of any procedures used to ensure reliable and unbiased data production, such as blinding and randomization, sample tracking systems, etc.
* any other information needed for assessment of technical rigour by the referees

Generally, this **should not include:**

* follow-up experiments aimed at testing or supporting an interpretation of the data
* statistical hypothesis testing (e.g. tests of statistical significance, identifying differentially expressed genes, trend analysis, etc.)
* exploratory computational analyses like clustering and annotation enrichment (e.g. GO analysis).

### Usage Notes

*This section is optional*

The Usage Notes should contain brief instructions to assist other researchers with reuse of the data. This may include discussion of software packages that are suitable for analysing the assay data files, suggested downstream processing steps (e.g. normalization, etc.), or tips for integrating or comparing the data records with other datasets. Authors are encouraged to provide code, programs or data-processing workflows if they may help others understand or use the data. Please see our [code availability policy](http://www.nature.com/sdata/policies/editorial-and-publishing-policies#code-avail) for advice on supplying custom code alongside Data Descriptor manuscripts.

For studies involving privacy or safety controls on public access to the data, this section should describe in detail these controls, including how authors can apply to access the data, what criteria will be used to determine who may access the data, and any limitations on data use.

### Acknowledgements

The Acknowledgements should contain text acknowledging non-author contributors. Acknowledgements should be brief, and should not include thanks to anonymous referees and editors or effusive comments. Grant or contribution numbers may be acknowledged.

### Author contributions

Each author’s contribution to the work should be described briefly, on a separate line, in the Author Contributions section.

### Competing interests

A competing financial interests statement is required for all papers accepted by and published in *Scientific Data*. If there is no conflict of interest, a statement declaring this must still be included in the manuscript.

### Figures

Figure images should be provided as separate files and should be referred to using a consistent numbering scheme through the entire Data Descriptor. In most cases, a Data Descriptor should not contain more than three figures, but more may be allowed when needed. We discourage the inclusion of figures in the Supplementary Information – all key figures should be included here in the main Figure section.

For initial submissions, authors may choose to supply a single PDF with embedded figures.

Authors are encouraged to consider creating a figure that outlines the experimental workflow(s) used to generate and analyse the data output(s).

### Figure Legends

### Figure legends begin with a brief title sentence summarizing the purpose of the figure as a whole, and continue with a short description of what is shown in each panel and an explanation of any symbols used. Legends must total no more than 350 words, and may contain literature references. The first sentence of the legend will be used as the title for the figure. It should contain no references of any kind, including to specific figure panels, data citations, bibliographic citations or references to other figures or panels.

### Tables

Authors are encouraged to provide one or more tables that provide basic information on the main ‘inputs’ to the study (e.g. samples, participants, or information sources) and the main data outputs of the study; see the additional information on providing metadata on page 6. Tables in the manuscript should generally not be used to present primary data (i.e. measurements). Tables containing primary data should be submitted to an appropriate data repository.

Authors may provide tables within the Word document or as separate files (tab-delimited text or Excel files). Legends, where needed, should be included in the Word document. Generally, a Data Descriptor should have fewer than ten tables, but more may be allowed when needed. Tables may be of any size, but only tables that fit onto a single printed page will be included in the PDF version of the article (up to a maximum of three).

### References

Bibliographic information for any works cited in the above sections, using the standard Nature referencing style.

### Data Citations

Data citations provide bibliographic information for any data records described or used in the manuscript. See further details below.

### Additional Formatting Information

**Referencing Figures, Tables, and other content**

**The Word document may reference Figures (e.g. Fig. 1), Tables (e.g. Table 1), and Supplementary Information (e.g. Supplementary Table 1, or Supplementary File 2, etc.). When information from metadata documents must be referred to, it should also be included in the main manuscript as Tables, and formatted in a way that suits human readability.**

**Citation format**

References should be numbered sequentially, first throughout the text, then in tables, followed by figures and, finally, boxes; that is, references that only appear in tables, figures or boxes should be last in the reference list. Only one publication is given for each number. Only papers that have been published or accepted by a named publication or recognized preprint server should be in the numbered list; preprints of accepted papers in the reference list should be submitted with the manuscript. Published conference abstracts, numbered patents, and archived code with an assigned DOI may be included in the reference list. Grant details and acknowledgments are not permitted as numbered references. Footnotes are not used.

Scientific Data uses standard Nature referencing style. All authors should be included in reference lists unless there are six or more, in which case only the first author should be given, followed by ‘et al.’. Authors should be listed last name first, followed by a comma and initials (followed by full stops, '.') of given names. Article titles should be in Roman text; only the first word of the title should have an initial capital and the title should be written exactly as it appears in the work cited, ending with a full stop. Book titles should be given in italics and all words in the title should have initial capitals. Journal names are italicized and abbreviated (with full stops) according to common usage. Volume numbers and the subsequent comma appear in bold. The full page range should be given where appropriate. See the examples below:

**Journal Article**:

1. Schott, D. H., Collins, R. N. & Bretscher, A. Secretory vesicle transport velocity in living cells depends on the myosin V lever arm length. *J. Cell Biol*. **156**, 35‐39 (2002).

**Book** ‐ Book titles should be given in italics and all words in the title should have initial capitals:

1. Hogan, B. *Manipulating The Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, 1994)

**Publicly available preprint:**

1. Babichev, S. A., Ries, J. & Lvovsky, A. I. Quantum scissors: teleportation of single-mode optical states by means of nonlocal single photon. Preprint at http://arXiv.org/quant-ph/0208066 (2002).

**Code:**

1. Gallotti, R. & Barthélemy, M. Source code for: The multilayer temporal network of public transport in Great Britain. *Figshare* https://dx.doi.org/10.6084/m9.figshare.1249862.v1 (2014).

**Online material** ‐ Stable documents hosted on the web may be cited in the main reference list, using the format below. Websites or dynamic web resources should be cited by embedding the URL in the main article text:

1. Manaster, J. Sloth squeak. *Scientific American Blog Network* http://blogs.scientificamerican.com/psi-vid/2014/04/09/sloth-squeak (2014).

**Technical or government report:**

1. Akutsu, T. *Total Heart Replacement Device.* Report No. NIH-NHLI-69 2185-4 (National Institutes of Health, 1974).

**Data Citations**

In-text data citations should be of the form (Data Citation 1), referring by number to a data record listed at the end of the document. Data citations may occur anywhere in the manuscript (except the Abstract), but in general each dataset central to the Data Descriptor should be cited at least once in the Data Records, and also indicated in the associated portion of the Methods section.

Data citations should be listed in the order they are cited in the manuscript. All authors or dataset submitters should be listed for each record in the “Data Citations” section. Please do not use “*et al.*” to shorten author lists.

Please use full repository names and check for consistency with the names in our list of [recommended data repositories](http://www.nature.com/sdata/policies/repositories).

Please write DOIs using full URL notation.

For data with a digital object identifier (DOI) this should be in the format:

Lastname1, Initial1A. Initial1B., Lastname2, Initial2A. Initial2B., … & LastnameN, InitialNA. Initial NB. *Repository\_name* DOI (YYYY).

Example citation for data with a DOI:

1. Perkins, A. D., Lee, M., & Tanentzapf, G. *Figshare* http://dx.doi.org/10.6084/m9.figshare.806269 (2014).

For data identified by accession ID, repositories may not provide the data creator (author). In these cases the data citation format should be *Repository\_name* Accession\_ID (YYYY)

Example citation for data with an accession identifier:

1. *GenBank* PRJNA244495 (2014).

**Metadata Records**

All manuscripts published in *Scientific Data* are accompanied by detailed machine-readable metadata files in ISA-Tab format depicting the workflow used to generate the accompanying datasets. In order to facilitate the creation of these records, authors are asked to submit one or more tables presenting, at a minimum, the samples and subjects employed in the study, the experimental, observational and analytical manipulations performed on each, and the data outputs resulting from these manipulations with their manuscript. Please also see Data Records and Tables, above; please note that, should you wish to create more complex metadata records, template Study and Assay files are available either in a ZIP archive with this document or for separate download from the “[Submission guidelines](http://www.nature.com/sdata/publish/submission-guidelines)” page of the *Scientific Data* website.

Here, we provide four generic ‘Table 1’ examples, including two experimental study examples, one observational study example, and an example for an aggregated dataset of the type that may result from a meta-analysis.

**Experimental study Table 1 example**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subjects** | **Protocol 1** | **Protocol 2** | **Protocol 3** | **Protocol 4** | **Data** |
| Mouse1 | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |
| Mouse2 | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |
| Mousen | Drug treatment | Liver dissection | RNA extraction | RNA-Seq | GEOXXXXX |

**Experimental study with replicates Table 1 example**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **Protocol 1** | **Protocol 2** | **Samples** | **Protocol 3** | **Data** |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep1a | Microarray hybridization | GEOXXXXX |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep2a | Microarray hybridization | GEOXXXXX |
| CellCulture1 | Drug treatment | RNA extraction | TechnicalRep3a | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep1b | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep2b | Microarray hybridization | GEOXXXXX |
| CellCulture2 | Drug treatment | RNA extraction | TechnicalRep3b | Microarray hybridization | GEOXXXXX |

**Observational study Table 1 example**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample** | **Geographical location** | **Geoposition** | **Protocol**  | **Data** |
| Body of water 1 | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile1 |
| Body of water 2 | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile2 |
| Body of water n | location name | latitude, longitude, altitude | Measurement of surface temperature | dataFile3 |

**Data aggregation study Table 1 example**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source** | **Sample** | **Sample number** | **Temporal range** | **Protocol 1** | **Protocol 2** | **Data** |
| Database URL 1 | Dataset 1 | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile1 |
| Database URL 1 | Dataset 2 | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile1 |
| Database URL 2 | Dataset n | Number of samples in the dataset | Range of measurements reported in the dataset | Data assimilation procedure | Method to generate output data | dataFile2 |

**Depositing your data to an appropriate repository**

Your *Scientific Data* manuscript will not be sent to review unless the dataset(s) described therein have been deposited in an appropriate public repository ([please see our list of recommended repositories](http://www.nature.com/sdata/policies/repositories)). Should a specific repository not be available for your field or data-type, or should the repository of your choice not permit confidential peer-review, you may upload your data to one of our [recommended generalist repositories](https://www.nature.com/sdata/policies/repositories#general). Integrated submission systems are available for both figshare and Dryad.