Report Info

Title:

Report Number:

DataStore Reference ID for Report:

Author 1:

Author 1 Affiliation:

Author 2:

Author 2 Affiliation:

Abstract (required)

*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 (required)

*Shoot for ~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 (required)

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). See the [detailed section in our submission guidelines](https://www.nature.com/sdata/publish/submission-guidelines#sec-5) for advice on writing a transparent and reproducible methods section. 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 Citing Data, 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. Actual analytical code should be provided in Appendices.

Data Records (required)

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 as described below. 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 example Tables below). They should also specify the data output resulting from each data-collection or analytical step, should these form part of the archived record.

Data Quality Evaluation (required)

The Data Quality Evaluation section should present any 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. Wherever possible & appropriate, data quality evaluation should be presented in the context of data standards and quality control procedures as prescribed in the project’s quality assurance planning documentation.

Possible content **may include:**

* Occurrence rates or patterns in data that do not meet established standards or data quality objectives.
* 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
* general discussions of any procedures used to ensure reliable and unbiased data production, such as chain of custody procedures, blinding and randomization, sample tracking systems, etc.
* any other information needed for assessment of technical rigor by reviewers/users

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 (required)

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 (optional)

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.

References (required)

Bibliographic information for any works cited in the above sections, using the standard NPS NR Publication Series referencing style.

In line with emerging [industry-wide standards for data citation](https://www.nature.com/articles/sdata2018259), references to all datasets described or used in the manuscript should be cited in the text and listed in the ‘References’ section in the same manner as a conventional literature reference.

Appendix A. Code Listing (required where appropriate)

Must be included when using R, Python, or SQL to process data.

Appendix B. Analysis Software Session and Version Information (required where appropriate)

Must be included when using R, include when using other tools as appropriate.

Additional Notes

Figures

Figure images should be included in-text near the initial point of reference.

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

### Figure captions 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 (the first sentence) should contain no references of any kind, including to specific figure panels, 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; also 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 text near the initial citation or as an appendix. Legends, where needed, should be included in the Word document. Generally, a Data Publication Report should have fewer than ten tables, but more may be allowed when needed.

Example Data Record Summary Tables

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 |