

Writing a Data Availability Statement: Expert Guidelines & Illustrations

Author

Dr. Shweta Murudkar

Post Url

<https://www.enago.com/academy/writing-data-availability-statement/>



Research thrives on the legacy of nourished and deeply rooted literature. While building on this strong knowledge base, researchers challenge themselves with goals set higher in the quest for reaching the skies! A strong foundation or knowledge base puts us in a solid position to continue this journey of scientific advancements. How do you accomplish this mammoth task? As a responsible researcher, it is important to allow your readers access to all your research data. This is where the “Data Availability Statement” comes into picture. Most journals today mandate a Data Availability Statement, where authors provide all the necessary information to reproduce the study.

What is a Data Availability Statement?

A Data Availability Statement, also commonly referred to as Data Access Statement tells the reader whether the data created, procured or used for a research project can be accessed. If yes, how and where? These [statements are important because](#) they support reuse, validation and citation of the reported work. They allow the readers an opportunity to verify the legitimacy of your paper's underlying data. Other researchers who wish to reuse the information can validate your findings and conduct further research. Additionally, it helps in improving the chances of the work being properly cited. Moreover, a complete Data Availability Statement increases transparency, allows compliance with journal policies, promotes trust, and encourages good scientific practice.

What Type of Data Should Be Included in a Data Availability Statement?

Based on the types of data reported in your study, there are primarily four sections that should be included in your Data Availability Statement.

Source data: This includes the data that the authors have not collected but used for analysis in the reported study. In other words, it is the data that are obtained from the third party. Authors should remember that the third party data cannot be distributed legally. Therefore, authors must consider including the following information in their Data Availability Statement:

- A description of the dataset and the third-party source
- Publishable contact details (email address or an individuals' ORCID identifier)
- Verification of permission to use the dataset
- All the conditions under which reuse is permitted

Underlying Data: This includes data that the author has produced or collected during the course of your investigation. Authors must upload this data in an authentic online repository.

Extended Data: This includes additional materials such as questionnaires and supporting images or tables that support the key claims made in the manuscript but are not essential to the main text. These must be approved from an online repository.

Reporting Guidelines: A copy of relevant guidelines and checklists must be uploaded to an approved repository wherever applicable.

Types of Data Availability Statements

Your data availability statement should describe the complete process required to reproduce the study results. It should also include links to all the supporting documents and contact information of the accountable data manager. Several organizations and

publishers such as [Hindawi](#), [PLOS ONE](#), and Springer Nature provide guidance for drafting data availability statements.

If instructions regarding how to write an effective Data Availability Statement are not specified by journals, here are some templates that you might find handy. Furthermore, you can tailor these to your requirements.

Illustration of Data Availability Statement Types:

- 1. Datasets available in a public (general, institutional, or discipline-specific) or a funder-mandated repository that assigns persistent identifiers to the dataset**

Examples:

All the data for growth temperatures of prokaryotes created or used during the study are openly available from the TEMPURA Archive Center at <http://togodb.org/db/tempura>

Protein sequence data that support the findings of this research have been deposited in GenBank with the accession code KP27898039 [*hyperlink the code*]

An actual example from [BMC Biology](#)

All data generated or analysed during this study are included in this published article (and its additional files). The ChIP-seq data have been deposited in the Gene Expression Omnibus database [GEO:GSE83330] (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=avylwwgwvtmrlmz&acc=GSE83330>). Requests for material should be made to the corresponding authors.

- 2. Datasets published in the literature**

Examples:

Datasets used for this study are included in *White et al.* (2019)

- 3. Datasets derived from public resources and made available in the article**

Examples:

Data analyzed in the current research were a re-investigation of existing data, which are openly available in the [*repository name*] at [*DOI*].

An actual example from [Microbiome Journal](#)

Sequence files and metadata for all samples used in this study have been deposited in Figshare (<http://dx.doi.org/10.6084/m9.figshare.687155>). A full record of all statistical analysis is included as Additional file 5, and was created using the knitr package in R. Original R scripts are available in GitHub (https://github.com/jfmeadow/Meadow_etal_Surfaces).

4. Data sets that are not publicly available or are restricted

Examples:

Owing to propriety nature, supporting data cannot be made available openly. Further information about the data and access conditions is available at [*repository name*] at [*DOI*].

Due to privacy and ethical concerns, neither the data nor its source code can be made available to the public.

Due to confidentiality agreements, data can be made available subject to a non-disclosure agreement. For further information, you may contact [*data manager contact information*].

Due to the sensitive nature of the data, information created during and/or analysed during the current study is available from the corresponding author [*name and contact information*] on reasonable request to bona fide researchers.

An example from [Palgrave Communications](#)

The datasets analyzed during the current study are available from the Web of Science repository, owned by Thomson Reuters (<http://scientific.thomson.com/isi/>) but restrictions apply to the availability of these data, which were used under license from Thomson Reuters, and so are not publicly available. Data are however available from the authors upon reasonable request and permission of Thomson Reuters. The downloading scripts used in the study are available in the Dataverse repository: <http://dx.doi.org/10.7910/DVN/MCXTHF>

5. No valid data repositories identified by the authors

Examples:

The authors of this study were unable to find a valid data repository for the data produced in this study. Supporting documents are available from [*data manager contact information*] at [*institution*].

The quantitative data and XYZ model simulations upon which the research is based are too large to transfer or archive. Therefore, we are providing all the information required to replicate the simulations here: [*model version, code, compilation script, condition files, etc.*] / [*DOI or permanent URL*]

6. Data sharing not applicable (for theory based articles or review articles)

Example: No datasets were analyzed or generated during the course of the current study.

Let us know how this article helped you in formulating Data Availability Statements. You can also visit our [Q&A forum](#) for frequently asked questions related to different aspects of research writing and publishing answered by our team that comprises subject-matter experts, eminent researchers, and publication experts.

Cite this article

Dr. Shweta Murudkar, Writing a Data Availability Statement: Expert Guidelines & Illustrations. Enago Academy. 2021/07/21. <https://www.enago.com/academy/writing-data-availability-statement/>