Open Data

How to Write a
Data Availability Statement

Submitting to F1000Research? (Great choice, by the way). Just remember to include a Data Availability Statement at the end of your article.

This is a short section of text which tells the reader how, where, and under what conditions the data associated with your research can be accessed and reused. In line with our **Open Data Policy**, failure to provide this data openly is likely to result in your submission being rejected.

What types of data does the Data Availability Statement cover?

Your Statement should reference **all data associated with your article**, together with details of any software used to process results. Your Data Availability Statement could include up to four sections, depending on the types of data used. If these apply to your research, you must include them in your Statement:

1 Source Data

This is data you have not collected yourself as part of your study, but has been used for analysis. This includes data which has been obtained from a third party.

2 Underlying Data

This is data that you have collected or produced as part of your study. This needs to be uploaded to an approved online repository.

3 Extended Data

These are additional materials that support the key claims made in your article, but are not required to follow the study design and analysis. Examples of extended data include questionnaires and supporting images or tables. This data needs to be uploaded to an approved online repository.

4 Reporting Guidelines

Where mandatory reporting guidelines apply, a copy of the relevant guidelines must be uploaded to an approved repository.

We recommend that all data (except third party) should be uploaded to an approved online repository together as a single dataset, which can then be clearly cited throughout your article, and in the Data Availability Statement. If you have multiple data deposits, they will all need to be listed as individual datasets. Visit our **Data Guidelines** to find a list of F1000Research-approved repositories.

For repository-hosted data, your Data Availability Statement must include:

· Repository name
· Title of the dataset
· DOI
· List of all data items (including the full file name, and a description of its contents)
· Data license

Here's an example:

Repository: Manually annotated miRNA-disease and miRNA-gene interaction corpora. https://doi.org/10.5256/repository.4591.d34639.

This project contains the following underlying data:
· Data file 1. (Description of data.)
· Data file 2. (Description of data.)

Data are available under the terms of the **Creative Commons Zero "No rights reserved" data waiver** (CC0 1.0 Public domain dedication).

Visit our **Data Guidelines** for more specific information on what details to include for each of our approved repositories, as this can vary. For example, some Data Availability Statements will need to include URLs, Accession Numbers, Study IDs, or Routes of Access, depending on the repository which hosts the data.

What about third party data?
If your data belongs to a third party, it won't necessarily be stored in an open repository. In this case, authors should describe exactly how they gained access to the data, so that readers can do the same.

What if my article doesn't have any data?
If there is no data associated with your article, you should still include a Data Availability Statement which makes this clear to readers. In this case, the Statement should read:
"No data are associated with this article"

If all associated data is presented within the article itself, the Statement should read:
"All data underlying the results are available as part of the article and no additional source data are required"

What about software?
If you are describing new software, you must make the source code available on a Version Control System (VCS) such as GitHub or BitBucket. We will also need a permanent, archived version of the source code at the time of publication to be uploaded to an approved repository. You must also provide details of the license under which the software can be used.

Still got questions?
Get in touch! **Our Editorial team** are on hand to advise you on what to include in your Data Availability Statement.

F1000Research

CHOOSE **open**

Open Data **Appendix**

Still not sure what a Data Availability Statement looks like, or what to include? Here are a few examples of different Statements from published articles on F1000Research.

If you're still not sure what to includes in your Statement, **get in touch with our Editorial team** before submission and they'll be happy to advise you.

Using subheadings for different data types

This **Research Article** has source and underlying data available from different locations, so the authors have split out their Data Availability Statement using subheadings to make this clear.

Data availability

Source data

Structures of natural compounds were downloaded from the ZINC Database.

Crystal structures of COVID-19 main protease were downloaded from the Protein Data Bank, accession numbers 6LU7 (in complex with N3) and 6Y7M (with OEW).

Extended data

Harvard Dataverse: Replication Data for: Computational screening for potential drug candidates against SARS-CoV-2 main protease. https://doi.org/10.7910/DVN/GYFXA0[67].

This project contains the following extended data:

- 2D interaction maps of all OEW pharmacophore-like ligands (PNG).

- 2D interaction maps of all Remdesivir pharmacophore-like ligands (PNG).

- 2D interaction maps of all Hydroxychloroquine pharmacophore-like ligands (PNG).

- 2D interaction maps of all N3 pharmacophore-like ligands (PNG).

Extended data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Third party data

This **Research Article** uses third party source data owned by the Demographic and Health Surveys (DHS) Program. Because the authors do not own this data, they have made it clear how readers can gain access to the dataset.

Data availability

Source data

The data for this study is owned by the DHS Program. The Individual Recode datasets for the PDHS 2006–07, 2012–13 and 2017–2018 were used for this study and can be obtained here: https://www.dhsprogram.com/data/available-datasets.cfm?ctryid=31

The electronic data is available from the DHS Program under its terms of use. Before downloading the data, users must register as a DHS user for reasons laid out on the DHS Program website and dataset access is only granted for legitimate research purposes.

Sensitive data

This **Data Note** describes a dataset containing sensitive details of intensive care patients, which cannot be uploaded to an open repository for ethical reasons. They have explained this in their Data Availability Statement and included instructions on how readers can apply for access to this dataset.

Data availability

Underlying data

The sensitive nature of these data means that they are only available internally to UHB staff for the purposes of clinical audit and service evaluation activities via the CAG guidelines. For external researchers, ethical approval may be obtained via formal application to the NHS Integrated Research Application System (IRAS) for a specific research project. The IRAS website (www.myresearchproject.org.uk) has full instructions; however, interested parties are advised to contact the corresponding author (christopher.bourdeaux@uhbristol.nhs.uk) to discuss the application.

Software

This **Software Tool Article** includes a Software Availability Statement so that readers can easily find and access the software described in this article.

Data availability

All data underlying the results are available as part of the article and no additional source data are required.

Software availability

Software available from: http://bioconductor.org/packages/HDCytoData

Source code available from: https://github.com/lmweber/HDCytoData

Archived source code at time of publication: https://doi.org/10.5281/zenodo.3551051[26]

Licence: MIT License

Want to see more examples?

The good news is there are thousands of articles on F1000Research, so grab a cup of coffee and **browse the archive** to see more Data Availability Statements in action.