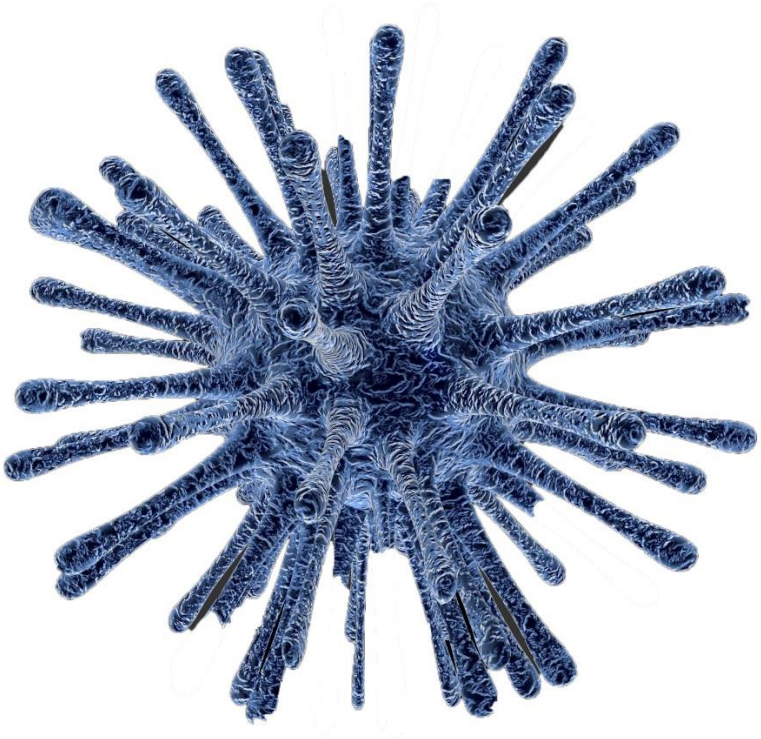
Meeting funder expectations on data sharing in Biomedicine and Health Science

Data Access Statements

Version 2.1 October 2018



University of Bristol

Research Data Service

Image: Virus Infected Cells, Pixabay, Public Domain

University of BRISTOL

DATA ACCESS STATEMENTS

UKRI (formerly RCUK) and other national and international funding bodies now require grant holders to provide a 'Data Access Statement' in each of their published articles. These should tell readers how they can access sufficient underlying data to evidence the published claims (or alternatively, why readers cannot access evidence). The nature of these statements will vary depending on where your data is. Please note a simple direction to interested parties to 'contact the author' is no longer considered sufficient. The remainder of this document provides examples of acceptable Data Access Statements for the following situations:

Contents

UNDERLYING DATA PUBLISHED AS SUPPLEMENTARY MATERIAL

Underlying data can be published in a variety of ways depending on the conventions of your field and the journal in which you are publishing your article. In this example, the authors have included a number of figures within the text of the paper, but the raw data required to validate the study's findings is too long to be included in this way.

As the publisher allows for extensive supplementary information to accompany an article, the authors have therefore deposited this supporting data as a supplement. Their data access statement makes this clear and fulfils the funder's expectations regarding data availability.

Data access statement

"All relevant data are within the paper and its Supporting Information files."[i]

UNDERLYING DATA PUBLISHED IN AN EXTERNAL DATA REPOSITORY

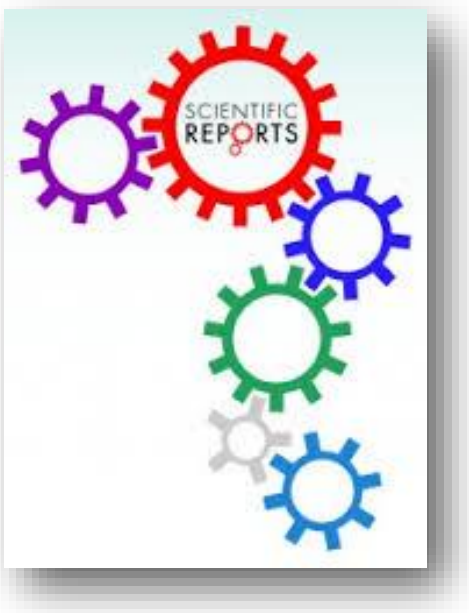
If your publisher is unable to host your data, a suitable external repository may be able to provide this service; these are usually discipline-specific and free of charge.

In this example, the authors have included some data as a supplement to the paper (microscopy images, structural diagrams and graphs), but the underlying protein structure data required to validate their findings is not suitable for sharing in this manner and has been deposited in the Protein Data Bank (PDB) instead. This is an online repository for information about the 3D structures of biological molecules. PDB have provided the authors with an accession number for their data; this has been cited in the data access statement in order to satisfy funder expectations.

Data access statement

"The structure solved has been deposited to the Protein Data Bank (www.pdb.org) and assigned the following accession number: 4WIQ."[ii]

UNDERLYING DATA PUBLISHED IN THE DATA.BRIS REPOSITORY



If there are no suitable external repositories, you can deposit your data in the University of Bristol's data repository, data.bris. In the following example, the authors have included some aggregated data in graphs within the text of the article, but the full meta-analysis results are too complex to present in this way.

The authors have therefore deposited the underlying meta-analysis data in the data.bris repository where the dataset is issued a DOI, which is then cited in the body of their article in order to meet funder expectations.

Data access statement

"Cotinine GWAS meta-analysis summary results (doi: 10.5523/bris.182rhz19hg3lz1172a7yfcap9v) accessible in full at https://data.bris.ac.uk/data/dataset/182rhz19hg3lz1172a7yfcap9v."[iii]

UNDERLYING DATA PUBLISHED AS A RESTRICTED ACCESS DATASET

If your dataset contains information that you cannot share openly for ethical or commercial reasons, you can still deposit it in a repository – you can apply restrictions on who can access the data and for what purpose. An embargo can also be applied if temporary access restrictions are required.

In the following example, the authors have not obtained consent to openly share data from the research participants, so have deposited their data in data.bris as a restricted dataset. The authors have chosen to give a full explanation of what this entails in their data access statement, but this level of detail is not required to satisfy funder expectations – a simple statement that the data is available to bona fide researchers subject to a data access agreement will suffice.

Data access statement

"It is the authors' intention to share their underpinning research data in order to maximise reuse and evidence their findings. At the time these data were generated, participants were not asked for their permission to share data beyond the immediate project team. The Faculty of Science Ethics Committee at the University of Bristol is the responsible ethics committee overseeing the data. The data will be deposited at the University of Bristol Research Data Repository (data.bris.ac.uk/data) where once published, they will be assigned the following doi: 10.5523/bris.8jnf6kn5dj6c1ky1nrgopc3r0. A metadata record will be published openly by the repository and this record will clearly state how data can be accessed by bona fide researchers. Requests for access will be directed to the Research Data team at Bristol, who will assess the motives of potential data re-users before granting access to the data. No authentic request for access will be refused and re-users will not be charged for any part of this process."[iv]

A STUDY USING THIRD PARTY DATA

If you are performing a secondary analysis of existing data belonging to someone else, you may not be able to publish the data at all. However, you should still cite the underlying data so that your readers will know how to access it if they choose.

In this example, the authors have used data from ALSPAC (Avon Longitudinal Study of Parents and Children) and are not permitted to re-publish that data. However, they have cited the original dataset in their data access statement and therefore satisfied the funder expectation on data sharing.

Data access statement

"Data used for this submission will be made available on request to the ALSPAC executive committee (alspac-exec@bristol.ac.uk). The ALSPAC data management plan (available here: http://www.bristol.ac.uk/alspac/researchers/data-access/) describes in detail the policy regarding data sharing, which is through a system of managed open access."[v]

A REVIEW PAPER WITH NO PRIMARY DATA

Data accessibility statements can be difficult to structure for review papers, as whilst they are clearly based on data, the data have not been generated during the course of the work underpinning the review. Instead, the supporting data are contained within the papers cited in the body of the review, and it would be pointless to include references to these in both the reference list and in data access statement. Instead, the authors have simply stated that no new data were generated, which fully satisfies funder and publisher expectations regarding data availability.

Data access statement

"This systematic review draws on published research only. No new data were collected or generated in undertaking this study."[vi]

A STUDY WITH NO UNDERLYING DATA

Finally, due to the nature of the study or the discipline some papers will not have any underlying data at all. In these cases, it is important to state explicitly that there is no underlying data.

In this example, the authors describe a series of mathematical theorems and proofs. All the information required to verify their findings is included within the paper, and there is no underlying 'data' as such.

The authors have simply stated this in the acknowledgements section and fully satisfied the funding body's expectation on research data management.

Data access statement

"No data were created during this study."[vii]

FURTHER INFORMATION

For help structuring your data access statement, or if your data does not fit into any of these categories (for example if you are unable to share your data) please contact the Research Data Service (data-bris@bristol.ac.uk).

---

[i] Example taken from: Gilbert R, et al. (2015) Incorporating Known Genetic Variants Does Not Improve the Accuracy of PSA Testing to Identify High Risk Prostate Cancer on Biopsy. PLoS ONE 10(10): e0136735. doi: 10.1371/journal.pone.0136735

[ii] Example taken from: Bozzi M, et al. (2015) The Structure of the T190M Mutant of Murine α-Dystroglycan at High Resolution: Insight into the Molecular Basis of a Primary Dystroglycanopathy. PLoS ONE 10(5): e0124277. doi:10.1371/journal.pone.0124277

[iii] Example taken from: Ware JJ, et al. (2016) Genome-Wide Meta-Analysis of Cotinine Levels in Cigarette Smokers Identifies Locus at 4q13.2. Sci Rep. Feb 1;6:20092. doi: 10.1038/srep20092

[iv] Example taken from: Leonards U, Fennell JG, Oliva G, Drake A, Redmill DW (2015) Treacherous Pavements: Paving Slab Patterns Modify Intended Walking Directions. PLoS ONE 10(6): e0130034. doi:10.1371/journal.pone.0130034

[v] Example taken from: Hammerton G, et al. (2015) Association between Maternal Depression Symptoms across the First Eleven Years of Their Child's Life and Subsequent Offspring Suicidal Ideation. PLoS ONE 10(7): e0131885. doi:10.1371/journal.pone.0131885

[vi] Loades ME, Sheils EA, Crawley E. "Treatment for paediatric chronic fatigue syndrome or myalgic encephalomyelitis (CFS/ME) and comorbid depression: a systematic review" BMJ Open 2016;6:e012271 doi:10.1136/bmjopen-2016-012271

[vii] Example taken from: M. R. Atkin et al. "Random Matrix Ensembles with Singularities and a Hierarchy of Painleve III Equations". International Mathematics Research Notices, 2016, 2016 (8), pp 2320–2375. doi:10.1093/imrn/rnv195