



Medical
Research
Council

Medical Research Council

Guidance for Data Access Statements



Contents

Introduction	3
What should be included	3
Types of data availability:	3
Examples: Data openly available (includes data about people).....	4
Examples: Controlled access	5
Examples: Data restrictions (limited data sharing).....	6
Examples: Third-party data	7
Examples: Data not held	7
Multiple datasets	8
Useful links	8
References	9
Version control.....	9

Introduction

The [UKRI open access policy](#) came into effect on 1st April 2022. The policy includes a requirement for all in-scope research articles to contain a “data access statement” which “informs readers where the underlying research materials associated with a paper are available, and how the research materials can be accessed” (see paragraph 15 of the policy and Annex 1, s. 9-16).

This document provides best practice examples on how to write a data statement which can be used for publications where no journal or institutional guidance, nor community standards, are available. These examples were collated from MRC-funded publications and external guidance, details of which can be found in the references section of this document.

What should be included

The information provided in the data access statement must include:

- a short description of the underlying data associated with the publication
- where the data are available, a permanent identifier such as a DOI or accession number, or a link to a permanent record for the dataset
- details on how data can be accessed and/or details on any restrictions and/or an explanation for why it can't be shared.

Data¹ here is to be understood as any data or research-relevant digital objects, such as metadata, algorithms, code, software and workflows, used as sources for research or generated as a result of research as reflected in the revised [MRC data sharing policy](#).

Types of data availability:

The following examples have been broken down into five types of data availability which range from completely open data through to data which cannot be shared.

1. Data openly available (includes data about people)
2. Controlled access
3. Data restrictions (limited data sharing)
4. Third-party data
5. Data not held

The type of data availability will influence how the statement should be presented and could include more than one type of data.

¹ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463>

Examples: Data openly available (includes data about people)

Some types of data can be made openly available.

If there are no restrictions on the data, then the data should be deposited with a permanent identifier such as a DOI or accession number, or a link to a permanent record for the dataset included in the statement.

“The data that support the findings of this study, including socio-emotional factor scores and extracted Beta values for significant clusters, are openly available via github at <https://github.com/danakanel>.”²

“Sequence data presented in this study are available on the European Nucleotide Archive (ENA) under project number PRJEB41184 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB41184?show=reads> [61]).”³

“The software developed in this study is available via a Jupyter notebook interface to enable reproduction of the simulations. The notebook, usage guidelines, dependencies, and processed metabolomics data are available via <https://github.com/cwieder/metabolomics-ORA>”⁴

² <https://europepmc.org/article/MED/35178519#free-full-text>

³ <https://europepmc.org/article/MED/34784976#id603330>

⁴ <https://europepmc.org/article/MED/34492007#id196071>

Examples: Controlled access

Some types of data are only accessible from the research team. The process to request access should be made clear in the statement.

“Lothian Birth Cohort 1936 data are not publicly available due to them containing information that could compromise participant consent and confidentiality. Lothian Birth Cohort 1936 data are available on request from the Lothian Birth Cohort Study, University of Edinburgh. If you are interested in working with the Lothian Birth Cohort 1936 data, you must complete a Data Request Form, indicating the variables you wish to access from the Data Dictionaries. Data Dictionaries and Data Request Forms are freely accessible at the following website: <https://www.ed.ac.uk/lothian-birth-cohorts/data-access-collaboration>. Completed forms must then be sent to Dr Simon Cox (ku.ca.de@xoc.nomis) for approval.”⁵

The Swiss Childhood Cancer Registry is the permanent repository of data on childhood cancer cases used in this study. This data cannot be made publicly available for both legal and ethical reasons as this would compromise patient confidentiality and participant privacy. Interested researchers may contact the corresponding author or the Swiss Childhood Cancer Registry (<http://childhoodcancerregistry.ch/>) via its online contact form for further information.”⁶

Further details regarding SABRE can be found at <https://mrc.ukri.org/research/facilities-and-resources-for-researchers/cohort-directory/southall-and-brent-revisited-sabre/>. Because of the sensitive nature of the data collected for this study, data cannot be made publicly available, but requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the MRC Unit for Lifelong Health and Ageing at UCL (ku.ca.lcu@ofnitfiws.AHLCRM).⁷

⁵ <https://europepmc.org/article/MED/35028426#id604610>

⁶ <https://europepmc.org/article/MED/33877514#id689071>

⁷ <https://europepmc.org/article/MED/37368914#id543825>

Examples: Data restrictions (limited data sharing)

Some types of data may have conditions attached that prevent or restrict data sharing.

Reasons for data restrictions

There can be justifiable reasons for restricting access to certain research data. Any restrictions on the data should be explained in the statement. The most common reasons to restrict access to data can be found below:

- **Consent is not in place to allow data sharing**

“The dataset reported in the current article is not available due to lack of informed consent and ethical approval for data sharing’

- **Data Protection issues**

“Patient-level data underlying this article cannot be shared publicly due to data protection requirements.”

- **Legal (such as patents)**

“We are currently unable to provide the source code at this stage, as part of the algorithm is being used for ultrasound imaging applications, which is being considered for a patent application for medical imaging and diagnosis.”

- **Data can't be shared due to commercial or other sensitivities**

“Data on household socio-demographics were purchased from a private company and cannot be shared due to contractual agreements.”

The data for this study is not available for data sharing. This is because consent is not in place to share the data beyond the original study.
The data for this study consists of interview transcripts of participants that contain potentially identifying and sensitive information. The data cannot be shared publicly due to concerns of participant confidentiality and ethics requirements. Participants consented to the study with the understanding that only de-identified quotations would be made public, not the entirety of the transcripts.
Identifiable patient data cannot be accessed in concordance with the NHS Code of Confidentiality and the signed patient consent form.

Examples: Third-party data

Some types of data are used but not generated by researchers. Access is on application and might be via a Trusted Research Environment or third-party infrastructure.

The patient data used in this study are supplied from Clinical Practice Research Datalink (CPRD; www.cprd.com) but restrictions apply to the availability of these data, which were obtained under licence from the UK Medicines and Healthcare products Regulatory Agency, and so are not publicly available. For re-using these data, an application must be made directly to CPRD. Instructions for how to submit an application and the conditions under which access will be granted are explained at <https://www.cprd.com/research-applications>.⁸

The de-identified data used in this study is available via the CVD-COVID-UK consortium coordinated by BHF Data Science Centre for accredited researchers working on approved projects in the NHS Digital trusted research environment. For information on the access and approvals process, please email ku.ca.kurdh@csdfhb or see <https://web.www.healthdatagateway.org/dataset/7e5f0247-f033-4f98-aed3-3d7422b9dc6d>.⁹

Data are available from the UK Biobank (<https://www.ukbiobank.ac.uk/>) for researchers who meet the criteria for access to de-identified UK Biobank data.¹⁰

Examples: Data not held

A data access statement should be included even when there is no data associated with the study, and an explanation given as to why no data was referenced to produce the publication.

Note: If there is any underlying data that has already been published then this should be included in the statement.

Examples

“Data sharing not applicable to this article as no datasets were generated or analysed during the current study.”

“No data were created or analysed in this study.”

⁸ <https://europepmc.org/article/MED/34056137#id698742>

⁹ <https://europepmc.org/article/MED/35192597#id509436>

¹⁰ <https://europepmc.org/article/MED/37061546#id676649>

Multiple datasets

Many research publications will be associated with more than one dataset which may have different types of availability. We recommend that each data set is presented in the same format in a list. For example, datasets can be accessed from the same place so would be efficient to only include one link for both datasets. An example template is listed below:

“[DATASET A] and [DATASET B] generated and/or analysed during the study are publicly available from [REPOSITORY NAME] at [DOI/REPOSITORY NAME].”

If there is more than one type of data availability, then this can be presented in the following format:

“[DATASET A] generated during and/or analysed during the current study is available in [REPOSITORY NAME] at [DOI/PERSISTENT WEB LINK].

[DATASET B] presented in this article is not readily available because [REASONS WHY DATA IS NOT AVAILABLE]. Requests to access the datasets should be directed to [URL].”

Feedback

This guidance should be read alongside [MRC's Data Sharing Policy](#). This guidance will be revised on a regular basis and if you have any questions or feedback about this guidance, please email: mrcdata.sharing@mrc.ukri.org

Useful links

- [UKRI open access policy](#)
- [UKRI Open Access Policy: frequently asked questions](#)
- [MRC's Data Sharing Policy](#).
- [MRC Data management and sharing guidance](#)

References

Publishers

<https://www.nature.com/nature/editorial-policies/reporting-standards#availability-of-data>

<https://authorsS.bmj.com/policies/data-sharing/>

<https://journals.plos.org/plosone/s/data-availability>

<https://reviewer.elifesciences.org/author-guide/data-availability>

<https://www.frontiersin.org/guidelines/policies-and-publication-ethics>

<https://scientific-publishing.webshop.elsevier.com/manuscript-preparation/data-availability-statement/>

Funders

<https://wellcomeopenresearch.org/for-authors/data-guidelines#opendata>

Universities

<https://library.bath.ac.uk/research-data/archiving-and-sharing/data-access-statements>

<https://www.library.manchester.ac.uk/services/research/research-data-management/sharing/data-access-statements/>

<http://www.bristol.ac.uk/media-library/sites/research/documents/lib-res-sup/rdm/guidance-documents/sharing/Data%20Access%20Statements%20-%20biomedicine%20and%20health%20science.pdf>

<https://www.strath.ac.uk/research/researchdatamanagementsharing/dataaccessstatements/>

Version control

<u>Version Number</u>	<u>Revision Date</u>	<u>Author(s)</u>	<u>Summary of changes</u>
<u>1.0</u>	<u>July 2024</u>	<u>Michael Cherrington</u> <u>Anisha Ahmed</u> <u>Alex Bailey</u> <u>Lou Chandler</u> <u>Geraldine Clement-Stoneham</u>	<u>New guidance created</u>