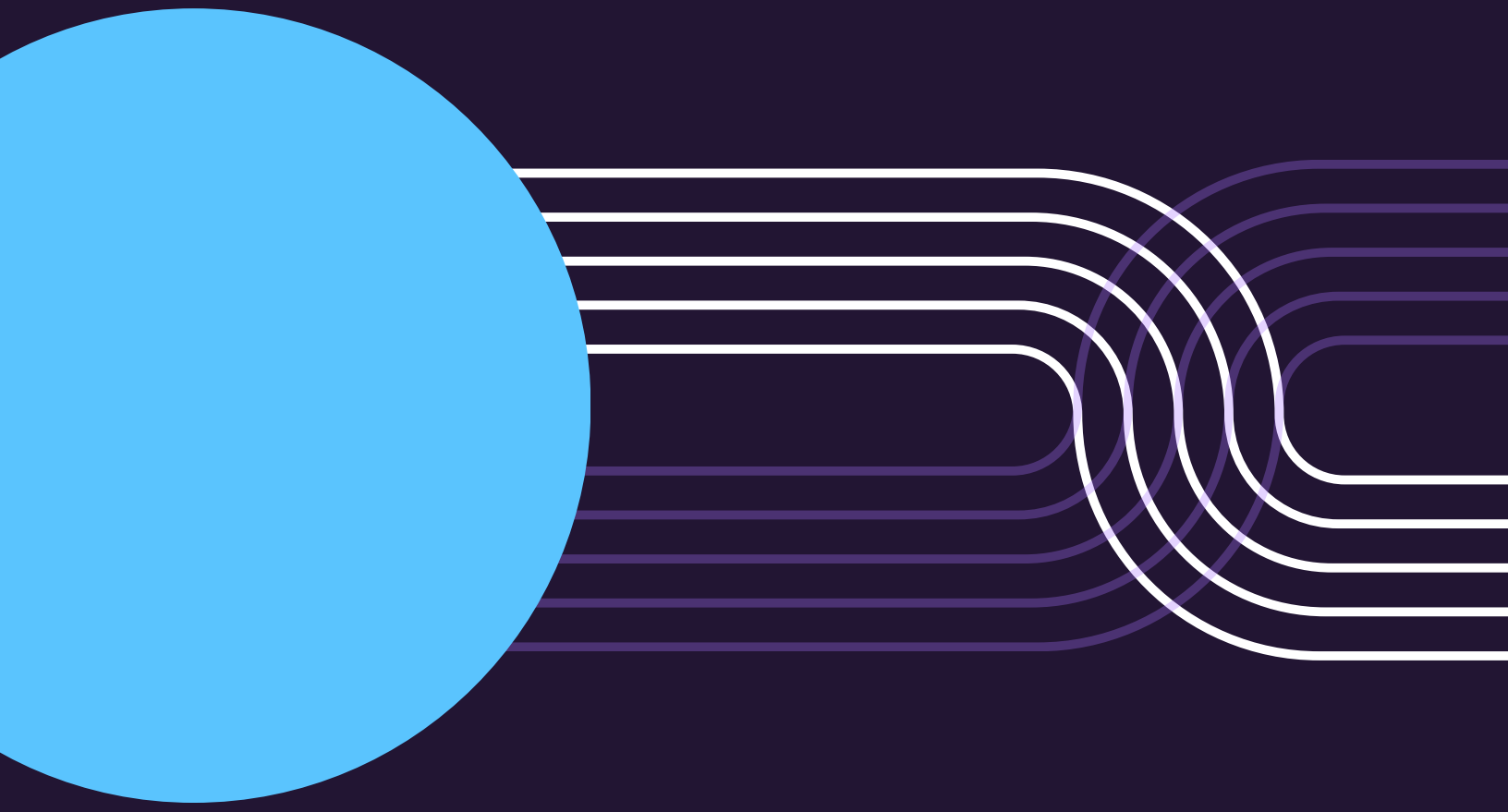
STM Medical Data
Sharing Guidance

September 2026



stm-assoc.org

Introduction

In 2024, STM launched a Task & Finish Group to bring publishers together to create guidance to improve the clarity of data sharing guidance for medical and health science researchers when they publish in academic journals. As first step, workshops were held with publishers and representatives of OpenPharma and their data working group to identify common issues to address.

A small group has taken the workshop outputs and been working to create guidance for publishers utilizing the framework of the RDA endorsed data sharing policy framework (see image on next page). The group identified selected framework features where specific information related to medical and health research would be beneficial.

The aim is to highlight specific information related to sharing medical and health science data that publishers could consider adding or adapting for their data sharing policy information. It may require investment or development to adopt these principles, but the intention is that it will increase the clarity of expectations by authors. This guidance document uses the same heading structure as the RDA policy.

This document should not be interpreted or used to reflect any agreement or understanding among marketplace participants on how they will behave or compete. Each participant remains free to behave and compete as they independently determine. This document does not constitute legal advice.

14 journal research data policy features arranged as 6 policy types (tiers)

Figure 1 - 14 research data policy features arranged as six policy types. Adapted from https://datascience.codata.org/articles/10.5334/dsj-2020-005

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	○	○	○	○	○	○
Exceptions to policy	○	○	○	○	●	●
Embargoes	○	○	○	●	●	●
Supplementary materials	○	○	○	●	●	●
Data repositories	○	○	○	●	●	●
Data citation	○	○	○	○	●	●
Data licensing	○	○	○	○	○	○
Researcher/ author support	●	●	●	●	●	●
Data availability statements		○	●	●	●	●
Data formats and standards				○	○	●
Mandatory data sharing (specific data types)				●	●	●
Mandatory data sharing (all papers)				○	●	●
Peer review of data				○	○	●
Data Management Plans (DMPs)				○	○	○

○ **Provide information**

The text for the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer review process

● **Provide information and action**

The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process

This document offers advisory text that has been drafted following these headers. This has been expanded to include best-practice examples from selected journals or publishers. Any comments and feedback on this guidance will be very welcome.

How to use this document:

- Review the data sharing information that is provided to authors
- Consider whether there is relevant information in this document that could be used to enhance guidance, speaking to authors in terms they recognise
- When making policy update, ensure that all internal stakeholders are aware of the changes and prepared to support authors and answer questions.

Guidance

Data Definition

This part of the policy will help researchers recognise their data and explain what is meant by research data – it is recommended to give examples relevant to the subject areas and consider common research methods used in the area, especially the expectations when publishing clinical trials at their various stages.

For some clinical trials it is common for authors to create a 'data dictionary' for their project that authors will be able to share with journal teams and readers alongside their research paper. It is recommended to consider highlighting this specifically in instructions for authors, and specify whether this is suggested or required.

There are currently discussions underway on how new techniques could be used to allow sensitive data to be shared whilst protecting patients or service users. For example, synthetic data.

For example, BMJ Data Guidelines:

"We encourage you to make available as much of the underlying data from your article as possible (without compromising participant privacy), but at least the minimum data required to reproduce the results presented in the associated article. We consider any files generated by your research as constituting relevant data. This may be raw or processed data. Examples include (but are not limited to):

- Individual-level deidentified patient data
- Survey results
- Interview transcripts
- Statistical code
- Images
- Videos
- Spreadsheets
- Audio files
- Text files
- Imaging and scan files"

Some medical journals provide very detailed guidelines for authors about specific data types. For example, for authors publishing Human Genome Data or Nucleic Acid Sequences there is guidance from National Institutes of Health (https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing) and for an example of how this is represented for a journal see here https://academic.oup.com/nar/pages/data_deposition_and_standardization.

Exceptions to policy

It is recommended that publishers and/or journals provide authors with guidance on what types of data they are encouraged or required to share (see data definition), and what would be an exception to a sharing mandate. When a journal has a mandatory policy, journal instructions will benefit from clearly stating how authors can apply for an exception (e.g. by contacting the Editor or Editorial Office). Examples of exceptions may be helpful but should not imply a lack of flexibility in addressing authors' specific circumstances. The primary reason for exceptions in medical research relates to protecting patients and participants of research projects to ensure their personally identifiable information (PII) are protected and all necessary ethical requirements and guidance are followed. This can be especially relevant where data are related to protected communities or where the number of patients in the sample is small. Journals may wish to decide what is required and what their exceptions are and provide this clarity to authors.

Authors should only agree to share data that they have collected and prepared themselves. Any third-party data they have used should be linked, cited and, if online, then the persistent identifier provided. For more information, see the section below on data licensing.

The authors should use the Data Availability Statement (DAS) to include the key information about their data, how it can be accessed, or whether it is subject to an exception and is not available.

For example, PLOS Medicine:

"Data sharing should never compromise participant privacy. It is therefore not appropriate to publicly share personally identifiable data on human research participants. The following are examples of data that should not be shared:

- Name, initials, physical address
- Internet protocol (IP) address
- Specific dates (birth dates, death dates, examination dates, etc.)
- Contact information such as phone number or email address
- Location data"

Data that are not directly identifying may also be inappropriate to share, as in combination they can become identifying. For example, data collected from a small group of participants, vulnerable populations, or private groups should not be shared if they involve indirect identifiers (such as sex, ethnicity, location, etc.) that may risk the identification of study participants.

It may also be useful to state other sensitive data types that may be inappropriate to share, depending on the journal scope - for example identifying photos and images.

Peer Review of Data

It is recommended that publishers clearly outline the policy of the journal with respect to peer review of data. The policy may need to be reflected in multiple places, such as general peer review guidance as well as journal specific instructions to authors and reviewers. Authors should be asked to make data available to reviewers, editors, and journal teams, either as standard or on request in accordance with their Institutional

Review Board (IRB) approval. Authors should be asked to provide details for reviewers on how to access supporting data at submission along with any links/ passwords etc., where data is made available for review via controlled access, and consider suitable formats to enable reviewers to conduct this work. Some repositories (e.g. Figshare, Dryad, Open Science Framework) provide 'private peer review links' to enable anonymised access to data to enable peer review. Reviewer guidance and author information should be clear that peer reviewers, editors, and journal teams must agree to maintain confidentiality.

Publishers may want to consider being clear on the expectations for reviewers if they have access to the data. As per the ICMJE policy, clinical trials that began enrolling participants on or after 1st January 2019, authors must include details of the data sharing plan and reviewers should be asked to check and confirm that the authors have adhered to their data sharing plan.

Depending on the peer review model of the journal (single/double anonymised), there may need to be extra instructions for authors on how to share their data whilst maintaining the integrity of an anonymised peer review process. This may require authors to share data and generate a 'peer review link' to allow reviewers to access the data without seeing the identity of the researcher who uploaded it.

PLoS provides the following helpful guide for peer reviewers outlining its expectations regarding peer review of data: Peer Review Data - PLOS.

Data Repositories

While generalist repositories (repositories without a specific subject area focus e.g. GREI repositories) may be suitable for some data types, some medical and health science research projects may require more specialist repositories to be used. For industry projects some companies are Clinical Study Sponsors for clinicalstudydatarequest.com which provides controlled access to research data - only employees of Clinical Study Sponsors can upload data to this repository however.

The National Institutes of Health provide a curated list of relevant domain specific repositories which may be useful to direct researchers to if they want to store data alongside similar data sets - Open Domain-Specific Data Sharing Repositories

Controlled access data repositories should be allowed for storing sensitive data as long as a route to guide readers to gain access is provided via the DAS. Institutional repositories may also be encouraged or even required for use for some researchers, and may offer a way for data to be shared without fee for the researcher.

If a publisher or journal has a recommended repositories list, it is recommended to review this list to ensure there are repositories that support relevant data types and authors who will submit to the journal, whilst ensuring repositories support data sharing best practice (able to mint persistent identifiers, support data versioning, long term preservation etc.). This may need to be reflected in multiple locations, such as at the publisher and journal level.

For example: The New England Journal of Medicine:

"Microarray Studies
Data obtained by microarray must be submitted to a repository such as the Gene Expression Omnibus or ArrayExpress before manuscript submission. Raw and transformed datasets for each microarray experiment must be provided through the repository, and the accession number for each experiment or series must be provided in the Methods section of a manuscript. If data are password-protected, the user name and password must be provided in both the cover letter and Methods section of the manuscript at the time of submission. A criterion of publication is full access to the relevant datasets through a publicly accessible repository." (From https://www.nejm.org/about-nejm/editorial-policies, March 2025).

Embargoes

In many cases, data from pharmaceutical led medical research may be subject to stricter embargoes than academic research, and it may not be possible to share the data before, for example, a medicine is approved. Journals may have policies that require data to be made available on publication, which do not always align with pharmaceutical embargoes. It is therefore helpful for journals and publishers to clearly outline whether embargoes are allowed, and under what circumstances. The author guidance may benefit from including information for authors specifying how they can contact the journal to discuss embargoes to their data in the event these contradict the data policy of the journal. This may also link to the 'exceptions' section.

Authors could be invited to share any information relating to embargoes for their data in their data availability statement, and this could be included in the guidance for the DAS.

For example, from Nature Medicine:

"A condition of publication in a Nature Portfolio journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript."

Data Availability Statements

Most journals will already include information in guidance to authors relating to any requirements for data availability statements (DAS). Where journals are publishing medical data, additional information can be provided to authors to outline in what ways authors can make their data available, and what requirements the journal has relating to medical data sharing. If the authors are including email addresses as a route to access data, the addresses may benefit from being for a laboratory, library, department or other shared account rather than an individual.

The DAS is the public facing summary of all the key information about the data from the study. To meet best practice, it is recommended that all key information as outlined elsewhere in the policy (licensing, route to access, PIDs, repository location, embargo), is included in this statement for long term storage and preservation.

For further guidance, see the Belmont Forum Data Accessibility Statement and Policy: https://www.belmontforum.org/wp-content/uploads/2019/10/Data_Accessibility_Statement.pdf, or the Research Data Policy Framework from the Research Data Alliance: https://datascience.codata.org/articles/10.5334/dsj-2020-005

For example, from The Lancet:

"From September 21, 2020, all submitted research Articles must contain a data sharing statement, to be included at the end of the manuscript. Data sharing statements must include:

- Whether data collected for the study, including individual participant data and a data dictionary defining each field in the set, will be made available to others ("undecided" is not an acceptable answer);
- What data will be made available (deidentified participant data, participant data with identifiers, data dictionary, or other specified data set);
- Whether additional, related documents will be available (eg, study protocol, statistical analysis plan, informed consent form);
- When these data will be available (beginning and end date, or "with publication", as applicable);
- Where the data will be made available (including complete URLs or email addresses if relevant);
- By what access criteria data will be shared (including with whom, for what types of analyses, by what mechanism – eg, with or without investigator support, after approval of a proposal, with a signed data access agreement - or any additional restrictions). See table for examples. Clinical trials that begin enrolling participants on or after Jan 1, 2019, must include a data sharing plan in the trial's registration. If the data sharing plan changes after registration, this should be reflected in the statement submitted and published, and updated in the registry record."

Data Formats and Standards

Depending on the kinds of data being generated, a journal or publisher may need to be specific about the formats used to store and share data - reminding authors that reviewers and readers may be accessing the files shared.

In general, publishers may wish to encourage authors to use file formats which promote access and reuse - rather than proprietary formats which reviewers/readers may not be able to make use of without specific programmes.

For example, from BMJ data guidelines:

To enable reuse and enhance reproducibility, all data should be shared using the sources file in which they were originally generated, for example:

- Images should be provided as .png, .jpg, .eps, etc.
- Text files should be provided as .txt, .doc, .rft, etc.
- Spreadsheets should be provided as .csv, .xls, .tsv, etc.
- Videos should be provided as .mp4, .avi, .wav, etc.
- Imaging and scan files should be provided in .img, .dcm, hdr, etc.

Other relevant file formats, such as audio content, could also be specified.

Data Licensing

It is recommended that authors include information on how the data they have used is licensed, to ensure that readers understand how the datasets can be reused. There may need to be separate guidance on how to treat data that authors have created themselves (author generated), and data that they have reused (third-party data) from other sources (created by other researchers, available from databases, governments or other projects).

When researchers are depositing their own data online, repositories may offer a range of licenses which range from very permissive to more strict. For this author-generated data, journals may require the application of specific licences, for example CC0 or CC-BY. It is recommended that publishers be clear about the licenses that they accept (especially if there is more than one policy type), and how authors should contact the journal to discuss exceptions if they are unable or unwilling to comply. Some commercial data owners may have specific licensing requirements based on their investment in the data and want to retain ownership of it for future commercial use. Journals may need to have flexibility in their policies to allow for this and should be clear as to which party has the responsibility for ensuring licensing policies have been followed (especially if it is the author).

In some cases, datasets that the author has used in their study include data that they did not create (third-party data). Examples include: Health Data Research UK, the Norwegian Mother, Father and Child Cohort Study (MoBa), or National Vital Statistics System (NVSS) - National Cardiovascular Disease Surveillance Data from the CCDC.The existing license on the data should be stated, as well as how the author gained access and reuse permissions for the data. Some data sources and repositories may have specific guidance on how to refer to the data within manuscripts and how editors/reviewers can gain/request access.

This may also be appropriate for third party tools, for example the Epsworth Sleepiness Scale, which is only free to use in certain situations and authors need to obtain a license. See more here: https://epworthsleepinessscale.com/licenses/#:~:text=The%20ESS%20and%20the%20ESS%2DCHAD%20are%20both%20subject%20to%20copyright.

Alongside any data policy, it is recommended that each journal or publisher should decide if there is any flexibility on data licensing in extenuating circumstances and make this explicit to authors, pre-submission. A publisher or journal may wish to encourage potential authors to submit a pre-submission enquiry to clarify whether alternative licenses for their dataset would be compliant with the policy.

For example, from BMJ Open:

"Data should not be shared in any way that could compromise participant anonymity or privacy, and data should not be shared if that would require the authors to break any laws or licensing agreements. If the data used were licensed from a third party, the data availability statement should explain how to obtain a licence for that data." (From https://authors.bmj.com/policies/data-sharing, March 2025).

FAIR Data Principles

If a journal or publisher asks authors to comply with the FAIR principles, it is recommended to be specific about how this relates to tangible actions they (the authors) can take in preparing and managing their data. If making a decision to promote FAIR, this information may need to be reflected in different areas of a publisher's or journal's data sharing policy and guidance information.

The FAIR Data Principles provide a framework to support the sharing of data which is "Findable, Accessible, Interoperable and Reusable." The Principles can help to guide the use of appropriate data repositories, persistent links, licensing information and metadata. They can be a useful way to support the sharing of useful and reusable datasets, even when the data cannot be made publicly available or "open".

Authors unfamiliar with the framework may assume that FAIR data is openly available or public. Publishers and journals should be clear about the journal's requirements and what is expected of authors. Ensure that the guidance is clear so clinical data / human data does not need to be shared openly in order to be made FAIR.

For example, from Health Open Research:

"Health Open Research endorses the FAIR Data Principles as a framework to promote the broadest reuse of research data.

Additional, practical guidance can be found on the GoFAIR website.
For research software, please consult the FAIR4RS Principles.
Findable
Findable data should be easy for both humans and machines to find.
Findable data should be easy for both humans and machines to find.
Findable data requires that:

F1. (Meta)data are assigned a globally unique and persistent identifier.
F2. Data are described with rich metadata (defined by R1 below).
F3. Metadata clearly and explicitly include the identifier of the data they describe.
F4. (Meta)data are registered or indexed in a searchable resource.
The best way to achieve Findable data is by:

- Depositing your dataset into a recognized data repository which assigns globally unique persistent identifiers (such as DOIs).
- Add as much contextual information (metadata) as possible when depositing your dataset into the repository." (From: https://healthopenresearch.org/for-authors/data-guidelines/#fairdata, March 2025)

Acknowledgements

The authors of this document would like to acknowledge OpenPharma for the work in creating and distributing the original survey which highlighted the problem that this work is looking to address.

We would also like to thank both the publisher and pharma company representatives who attended and contributed to the workshops in the early stages of this process; the STM Open Research committee members; Hylke Koers at the STM Association for support and guidance; and Katie Eve (Elsevier) and Lauren Cadwallader (PLOS) who gave feedback on drafts of this document.

Created by:
Matthew Cannon - Taylor & Francis
Rebecca Taylor-Grant - Taylor & Francis
Rhiannon Meaden - Oxford University Press