

PLOS Data Policy Summary

What does the PLOS Data Policy state?

The PLoS Data Policy¹ states that authors wishing to publish a paper in one of their journals must ensure that:

- Data that underpins their research findings is made available to others in an appropriate form
- A Data Availability Statement which outlines how data can be located is provided in the paper itself.

This requirement came into force at the start of March 2014. Additional guidance published on the PLOS blog² clarifies several aspects of the policy, addresses common questions and provides examples.

What data is covered by the PLOS requirement?

PLOS define the 'minimal dataset' that should be made available as:

- Data "*used to reach the conclusions drawn in the manuscript*" which is necessary to understand, validate, or replicate the work. E.g. tabular data that underpins a graph. Other project data that is not part of the analysis described in the paper *is not* covered by the policy.
- Documentation necessary to understand the data content and the methods applied to produce it.

If the paper describes the creation of open source software, the following resources should be made available:

- Source code and, where appropriate, executable code
- Documentation explaining how to set-up and configure the software
- A test dataset with associated control parameter settings

What file format should be used to store data?

Data should be made available in an open format, such as CSV or tab-delimited text, whenever possible. PLOS reserve the right to request specific data types on a case-by-case basis. Guidance on recommended file formats is available through the UK Data Archive³.

How will PLoS monitor application of the policy?

PLOS will monitor the policy using two methods:

- PLOS editors/reviewers will evaluate information provided in the Data Availability statement on submission. If the author refuses to share data described in the paper and cannot provide a satisfactory reason for it to be withheld, the paper will be rejected.
- PLOS encourage researchers to contact them if they are unable to obtain an associated dataset. The PLoS Policy does not state the activities that will take place following this event. However, they may choose to retract the paper in some cases.

¹ <http://www.plosone.org/static/policies#sharing>

² <http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/> and <http://blogs.plos.org/everyone/2014/03/08/plos-new-data-policy-public-access-data/>

³ UK Data Archive File Formats table: <http://www.data-archive.ac.uk/create-manage/format/formats-table>

What methods may be used to share data?

Decisions on data sharing methods should be made on a case-by-case basis. PLOS suggest various methods that may be used to provide data access:

- **Publication through PLOS:** Data files of 10MB or less can be published alongside the paper as ‘supporting information files’. If the combined files are over 10MB in size, they should be compressed.
- **Publication through a data repository:** The preferred approach is to publish data through an appropriate institutional, disciplinary, or general data repository. E.g. ClinicalTrials.gov for clinical data and GenBank for gene sequences⁴. Authors who have datasets too large for sharing through a repository are advised to contact the relevant journal for advice.
- **Ad hoc sharing via an appropriate Data Access group:** In cases where data cannot be submitted to a data repository, the availability statement should indicate how the requester may contact the appropriate committee/group to request access (e.g. an email address or URL). See ‘Available on request’ below.

What access status should be applied?

Data may be made available under different conditions:

- **Available for anonymous download:** PLOS encourage authors to make their data available for free download, where possible. See Dryad (<http://datadryad.org/>) for examples of this access method.
- **Available on request:** If data cannot be made available for download, a request access approach may be appropriate. In this scenario, a user must submit an access request, indicating their intended use. The access request is evaluated by a data access group and a decision made on whether to provide the data.

What about secondary data owned by a 3rd party?

PLOS do not expect authors to breach third party rights agreements. If data is owned by a third party that does not allow it to be shared, the Data Availability statement should state the data source and how access may be obtained (e.g. “data available through email application to the Clinical Trial Unit at University of Oxford”).

What about research that contains personal information?

PLOS do not expect personal information to be made publicly available. They suggest various methods of balancing the need to protect participants’ privacy with data sharing expectations. These include: restricting sensitive segments of the dataset, de-identifying content and applying restrictive licence conditions.

What licence should be assigned to data?

PLOS encourage authors to licence data under a Creative Commons Attribution licence (<https://creativecommons.org/>). However, they appear willing to discuss alternative licences with authors, if required. The LSHTM RDM website provides guidance on choosing an appropriate data licence⁵.

What is the School doing to support LSHTM researchers in this area?

- The recently launched Research Data Management website (<http://www.lshtm.ac.uk/research/researchdataman/>) offers guidance on data management and sharing.
- The RDM Support Service offers advice to individuals, departments and research groups on data management and sharing. Email researchdatamanagement@lshtm.ac.uk to book an appointment.
- An LSHTM Research Data repository is currently being developed by the RDM Support Service. This will offer functionality necessary to: (1) assign DOIs to datasets (2) perform long-term curation and (3) deliver publicly available data and handle data requests, as required.

⁴ The LSHTM RDM site offers a list of possible data repositories
http://www.lshtm.ac.uk/research/researchdataman/share/locate_datarepository.html

⁵ http://www.lshtm.ac.uk/research/researchdataman/share/choose_licence.html