



The Principles and Practices of Data Sharing

Contents

1.0 Sharing data via data repository	3
1.1 What do people want to do with your data?	3
1.2 Access vs. Re-use.....	3
1.3 Terms of use: defining how data can be used.....	3
1.4 What can you include in terms of use?	3
1.5 What protection do terms of use give you?	4
1.6 Data licensing.....	4
1.7 Creative – Commons (CC) licences.....	4
1.7.1 CC-BY: Creative Commons Attribution Only	5
1.7.2 CC0: Public Domain Dedication	5
1.7.3 Other Creative Commons features.....	5
1.7.4 Why is CC0 recommended (ideal) for data sharing?	5
1.8 Considerations when sharing data via a trustworthy data repository	5
1.9 What is metadata?	6
2.0 Data Publications as an option for sharing and publishing research data	7
2.1 Comparison of typical research journal, typical data journal and typical data repository.....	7
2.2 Process: Publishing a data paper with a data journal	8
2.3 Relevant external data repositories for PEP datasets.....	8
2.4 Relevant data journals for PEP datasets:.....	9
2.5 Considerations when sharing data via a data paper	9
2.6 Data referencing	9
2.7 Metrics for measuring impact of data	10
2.8 Data citation and metrics	10
References	11



1.0 Sharing data via data repository

Data repositories preserve, manage, and provide access to research data in a variety of formats. Data are curated to enable search, discovery, and reuse. There must be sufficient control for the data to be authentic, reliable, accessible and usable on a continuing basis.

1.1 What do people want to do with your data?

- View your data to contextualise your published article.
- Replicate your study.
- Develop their own research based on your data.
- Incorporate your data into a larger dataset or use multiple datasets together.
- Develop commercial products based on your data.

1.2 Access vs. Re-use

When others gain access to your data, they may assume that they are permitted to reuse it. They will only know what they are allowed to do with it if you tell them.

- If your data *can* be reused, you need to clarify whether there are any restrictions on reuse.
- If your data *cannot* be reused, you need to highlight this.

1.3 Terms of use: defining how data can be used

- Terms of use, conditions of access, terms and conditions, copyright statement and licensing information may all be used to convey information about how you want your data to be used.
- Licences and other terms of use may be displayed in the metadata (data about data) of your dataset.
- They provide users with a set of rules or guidelines about how your data can and can't be used.
- Users may implicitly agree to terms of use when accessing or downloading your data from repositories, or they may need to agree explicitly, e.g. to a pop-up.

1.4 What can you include in terms of use?

- Describe the ways that data can be reused.
- Outline the institutions, countries or regions where data can be reused.
- List the types of person who can reuse the data (bona fide researchers versus private citizens).
- Permit (or limit) commercial reuse or profit from the data.
- When terms of use or a licence are stated explicitly alongside data, this will all be done without anyone needing to ask permission directly.

1.5 What protection do terms of use give you?

- Having a licence or terms of use in place doesn't automatically prevent people from misusing your data.
- Terms of use don't have any technical functionality to enforce compliance (like digital rights management (DRM) solutions can).
- If someone breaks the terms of use or ignores your licensing conditions, how will you address this? Maybe, legal action.
- If you are not willing to enforce strict terms of use, then consider a more liberal approach if you want to share your data openly – and receive all of the benefits of open data publication (Increased citation and usage; faster impact; and compliance with open access mandates).

1.6 Data licensing

When you publish your data in a data repository of your choice, a licence agreement will be applied to your data. A licence agreement is a legal arrangement between the creator/depositor of the data set and the data repository, signifying what a user is allowed to do with the data.

Standard licences	Bespoke (customized) licences
<ul style="list-style-type: none"> - Appropriate for most datasets. - Easily understood and widely used. - Do not address very specific circumstances. - Your funder/institution may have a licence which you are required to apply. - The repository you use may only allow you to add a standard licence to your data. 	<ul style="list-style-type: none"> - Time-consuming and will require legal review. - Difficult for other users to interpret. - Not recommended when standard licences are available.

1.7 Creative – Commons (CC) licences

- When sharing data openly, you are very likely to come across Creative-Commons licences.
- They are used by repositories, publishers and institutions (and others) to provide a standard format for licensing content.
- The licences are developed and maintained by an American non-profit called Creative Commons and originally released in 2002.
- They have gone through multiple iterations since then to make them globally applicable – the most recent versions are CC 4.0.
- Licenses are non-revocable – you cannot withdraw copies of work released under Creative Commons at a later date.



1.7.1 CC-BY: Creative Commons Attribution Only

- The most permissive Creative Commons licence.
- It allows reuse of your data by anyone, for any purpose, as long as you receive acknowledgement and an attribution for your work (e.g. a citation).
- However, it is not appropriate for work from which you intend to gain commercially, either now or in the future as you have no control over who will reuse your work.
- This licence closely aligns with Open Access principles and is generally applied for Open Access publications.

1.7.2 CC0: Public Domain Dedication

- Technically not a licence, CC0 signifies that you have released your data into the Public Domain – you no longer claim your copyright over it, and anyone can use it for any purpose.
- CC0 does not require any attribution or acknowledgement for you, but good practice is that you will still receive an attribution.

1.7.3 Other Creative Commons features

Creative Commons also supports other licence features, which can be mixed and matched (as long as they don't conflict with each other).

These include:

- **SA – Share-alike** – the data can be used for any purpose as long as any derivatives are licensed under the same terms.
- **NC – Non-commercial** – the data can be used for any non-commercial purpose.
- **ND – No derivatives** – the data can be shared and used but no new derivatives can be produced from it.

1.7.4 Why is CC0 recommended (ideal) for data sharing?

- Does not limit or restrict reuse in any way.
- Maximises potential reuse and audience for your data.
- No ambiguity.
- Acknowledges technical barriers to attribution.

1.8 Considerations when sharing data via a trustworthy data repository

- Data are discoverable independently of any research articles.
- Data are more likely to be accompanied by relevant metadata facilitating interpretation.
- Data are more likely to be accompanied by terms of use (data licence) facilitating reuse.
- Data will be preserved beyond the funding that initially generated the data.

- Preparing data for deposit is often time consuming.
- Data may not be linked to the relevant research article.
- You must identify a suitable trustworthy data repository and prepare data for sharing according to its requirements.

1.9 What is metadata?

Metadata is often described as “data about data”. Data that defines and describes the characteristics of other data, used to improve both business and technical understanding of data and data-related processes. It usually adds some context or additional descriptive information to the data you're looking at.

Examples of metadata:

- Metadata describing the dataset (title)
- Metadata about the dataset's creators
- Metadata describing the contents of the dataset
- Metadata about the dataset's related publications.

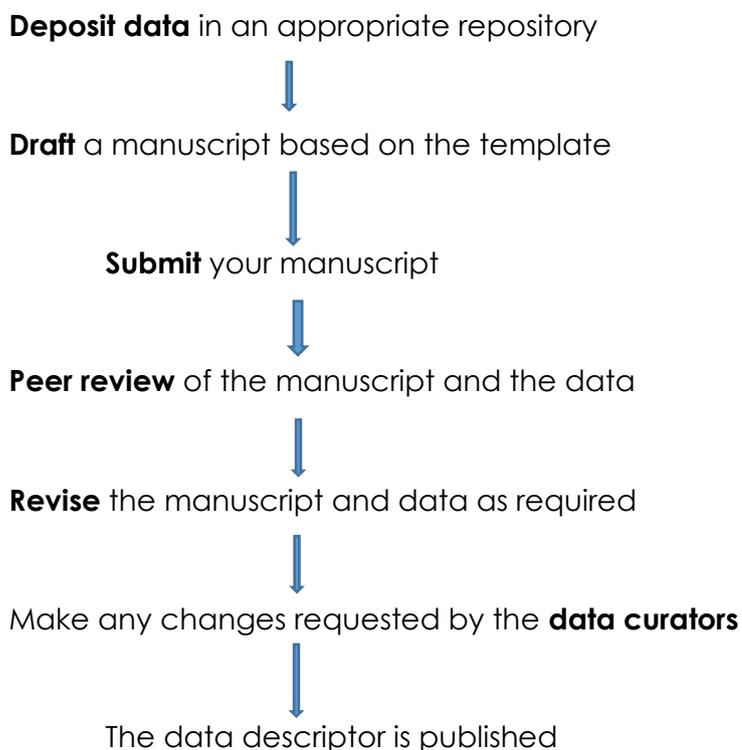
2.0 Data Publications as an option for sharing and publishing research data

Apart from publishing data in a data repository, data can also be published in form of a data paper/data article in a data journal. Below is a comparison of data journal and data repository to research journal:

2.1 Comparison of typical research journal, typical data journal and typical data repository

Typical research journal	Typical data journal	Typical data repository
Dataset checking may be available as an additional paid for service	Data curation and checking may be offered as part of the normal data journal workflow	Data files may be reviewed by repository staff for <ul style="list-style-type: none"> •Technical integrity of data •Violation of repository conditions of use
Reviewers may be allowed to request to see data	Data peer reviewed	May facilitate peer review of data prior to it being publicly available
Data referencing may or may not be allowed	Journal staff typically check if data have been correctly cited	May provide guidance to data users on how data should be cited
Final output is a citable scholarly article	Final output is a citable data article	Final output is a citable dataset

2.2 Process: Publishing a data paper with a data journal



2.3 Relevant external data repositories for PEP datasets

Repository name	Information on fees/costs	Size limits
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated
figshare	100 GB free per <i>Scientific Data</i> manuscript. Additional fees apply for larger datasets	1 TB per dataset
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded
Zenodo	Donations towards sustainability encouraged	50 GB per dataset
Mendeley Data	Contact repository for datasets over 10 GB	10 GB per dataset

2.4 Relevant data journals for PEP datasets:

Data journal	Article processing charges
Journal of Open Humanities Data	£ 100 per data paper
Scientific Data	100 GB free per <i>Scientific Data</i> manuscript. Additional fees apply for larger datasets
Elsevier Data in Brief	Not stated
Journal of Big Data	£1060.00/\$1570.00/€1290.00 for each article

2.5 Considerations when sharing data via a data paper

- Data are discoverable from the data paper *and* from the data archive.
- The quality of your data is going to be paramount for a data paper.
- Data should be technically reliable and suitable for use by others – ensuring this happens is time consuming.
- Flaws in study design (e.g. lack of suitable controls) are likely to result in rejection.
- Presence of relevant metadata may be checked by peer reviewers and editors.
- Appropriateness of terms of use may be checked by editors.
- Data (along with full contextual detail) is preserved for the long term.
- Data producers receive academic credit for their work.
- The effort taken to draft a data paper is likely to be similar to that of drafting a traditional research article.

2.6 Data referencing

Citation of data functions in the same way as a citation of a book, monograph, conference paper, journal article or any other material included in a reference list.

Authors or Creators >> Title >>Publisher (Repository) >>Persistent identifier >>Publication year

Murigi, M. & Kariuki, P. Household socioeconomic data from Murang'a County, Kenya, surveyed in 2018. figshare <https://doi.org/10.6084/m9.figshare.11961363> (2020).

Properly archived datasets will be given a unique, persistent identifier by the repository.

2.7 Metrics for measuring impact of data

- Number of times data have been viewed
- Number of times data have been downloaded
- Number of times data have been cited in a scholarly publication
- Number of newspaper articles written about a dataset
- Number of policies referencing/making use of a dataset

2.8 Data citation and metrics

- **Supports reproducibility**, providing evidence for the claims in the paper and provenance tracking of the outputs and material used in the paper.
- There is evidence that the inclusion of data citations in published papers **increases the citation rate of the paper** as a whole.
- Citing data supports the idea that **research data are as important as research publications**.
- Citing data provides **credit** to those who created or generated the dataset.
- When datasets are cited, it is possible to create **metrics to track re-use**.

References

- Creative Commons Australia. (2014). *What are the benefits of offering and using creative commons licences?* Retrieved August 2020, from Creative Commons Australia:
http://creativecommons.org.au/content/Benefits_of_CC_08.pdf.
- Hrynaszkiewicz, I., Birukou, A., Astell M., Swaminathan, S., Kenall, A., & Khodiyar, V. (2017). Standardising and Harmonising Research Data Policy in Scholarly Publishing. *International Journal of Digital Curation*, 12 (1), 65.
- Hrynaszkiewicz, I. (2016). Promoting research data sharing at Springer Nature. Retrieved August 2020, from
<http://blogs.nature.com/ofschemesandmemes/promoting-researchdata-sharing-at-springer-nature>
- Naughton, L., & Kernohan, D. (2016). Making sense of journal research data policies. *Insights the UKSG Journal*, 29(1), 84–89.
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175.
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. *PloS One*, 11(1), e0146695.
- Solon, O. (2011). *Creative Commons 101: An introduction to CC licences*. Retrieved August 2020, from Wired: <http://www.wired.co.uk/news/archive/2011-12/13/creative-commons-101>