

---

# Data sharing

---

Sharing research data is one of the basic building blocks of open science. Data sharing makes validating research results easier and the published research becomes more transparent and credible. If the data are published under an [open license](#), they can be used by other researchers and also the wider public. Contrary to widespread assumptions, sharing research data does not mean that everything needs to be open to everybody. The general idea is that data should be **As open as possible, as closed as necessary**.

In this section, you will find a general introduction to data sharing (which data to share, why, and when), information on possible ways to share research data, tips for data sharing, recommendation for citing research data and myths about data sharing.

## Why should you share your data?

These days, a number of **publishers and research funders require the sharing of underlying data** to published studies. However, even if it is not a publisher's or a research funder's requirement, there are many reasons why research data should be as open as possible – it benefits you, as well as the wider research community.

- Opening your research data boosts the robustness of your research as it allows others to replicate the results
- Enhancing your reputation as an honest and careful researcher and thus increasing your impact
- Increased citation rates – your data can be cited as well! Moreover, research suggests that sharing your research data may increase the citation rate of the associated publication (e.g., [Piwowar et al. 2007](#), [Colavizza et al. 2020](#))
- Compliance with a funder's or a publisher's policy
- Effective use of resources - the same data do not have to be recreated
- Sharing research data may speed up the research process
- Possibility of establishing new cooperation between the authors of the data and their users
- Combining research data from multiple sources may lead to new findings
- Encouraging citizen science
- Reducing academic fraud

## Which data to share?

In general, it is considered good scientific practice to share **all data which are required for replicating the published study**. Apart from the collected or generated data, this might also include a computer script or software used for their processing or analysis, lab notebooks, field notes, codebooks, and other materials. It is also important to share the **documentation** (metadata) along with the dataset to provide additional information about the data such as who the author is, when and where the data were collected, methodology of data collection and processing, and other information that will help the users interpret your data.

Besides sharing underlying data, you can also share **standalone datasets**, especially if it is a larger dataset that does not relate to just a single publication. Such datasets can be valuable to researchers within your field and beyond.

Depending on disciplinary practices, context or the type of the data collected, it might be preferable to share either **raw data**, or **processed data**, which may have been sorted, cleaned, or annotated in some way, or even both types. In both cases, the script for and/or the description of the method of processing should also be shared.

It is also important to remember that **not all data can be shared**. Sharing might be restricted, for example, in cases where data sharing could infringe on intellectual property rights, the right to privacy and personal data protection, or the right to protection of trade secrets, state security, or other legitimate interests of the beneficiary (e.g., commercial use). Even if the data cannot be shared, it is recommended that you deposit your data in a trusted storage (e.g., a repository) for long-term preservation and that you share at least a metadata record.

## When to share data?

Research data can be shared at **any stage of the research process**. It is a common practice to share data **along with the publication** they relate to.

Some researchers may decide to open up the whole research process and share data even before the related publication is published. This can be done, for example, through a platform like [Open Science Framework](#) (OSF), which also enables [preregistrations](#) that you can use if you have further research projects planned, using the shared dataset. The advantage of this approach lies in the transparency (and so enhancing the credibility of your research) and the opportunity to identify potential errors in study design or data analysis prior to submitting your article to a journal.

## Useful Resources

Colavizza, Giovanni et al. 2020. The citation advantage of linking publications to research data. *PLOS ONE* 15(4). e0230416. <https://doi.org/10.1371/journal.pone.0230416>

Drachen, Thea M., Olle Ellegaard, Asger V. Larsen & Søren B. F. Dorch. 2016. Sharing data increases citations. *LIBER Quarterly* 26(2). 67–82. <http://doi.org/10.18352/lq.10149>

Piwowar, Heather A., Roger S. Day & Douglas B. Fridsma. 2007. Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3): e308. <https://doi.org/10.1371/journal.pone.0000308>

Open Data Institute: [Open Data Essentials](#) - e-learning course