Guide to Open Data

There are many misconceptions around the sharing of research data. We've put together a helpful guide that covers data types and the forms data exists in; shows you that support is available and where you can seek further information; and how data sharing actually establishes and confirms ownership of your data via authorship.
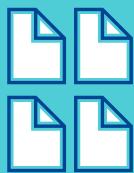
Types of research data

If you aren't sure whether you have any research data, it's important to know that data exists in many different formats: **textual**, **numerical**, **databases**, **geospatial**, **images**, **audio-visual recordings**, **data generated by machines or instruments**, etc. Research data may also include non-digital materials or 'sources', and some non-digital data can be digitized.

In cases where data cannot be easily digitized in a way that maintains its usefulness for you and others, you can still 'share' the data by creating an extensive **metadata record** describing the object, where it is stored, and how to access it. Depositing this metadata record openly in a repository will allow others to find it. You can cite the metadata record in any associated articles (and vice-versa) to establish links between the published work and the dataset.

Where published work genuinely has no data associated with it, it's best to indicate this clearly in a **data availability statement**. This confirms the absence of data to readers, rather than the absence of data sharing.

Data sharing in different subject areas

The practice of data sharing varies considerably from one discipline to another, and this includes how often it is done and how much support there is. However, the key benefits remain the same: **reproducibility**, **credit**, and **potential reuse**.

Open Research Europe endorses the **FAIR Data Principles** and these can be applied to your research data regardless of discipline. There are numerous generalist repositories that accept a wide range of data types in a wide range of formats. You can find guidance on selecting a repository **here**. Across disciplines, data sharing mandates are becoming more common at national, funder, and organizational levels, so it is important to understand these before you start conducting your research project so you can plan how best to meet the requirements.

Open Research Europe

European Commission
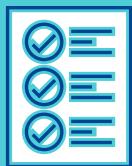
Planning and support with data sharing

To help plan how your data from a project will be shared it is important to create a **data management plan (DMP)**. It is mandatory for Horizon 2020 projects to develop a DMP under article 29.3 of their Grant Agreement, unless projects have opted out of the article. For more information on the requirement and on DMPs see the **Horizon 2020 online manual**. Many institutions have data stewards who can provide expert subject area guidance on how to prepare a DMP and advise on:

- **the handling of research data during and after the end of the project;**

- **what data will be collected, processed and/or generated;**

- **which methodology & standards will be applied;**

- **whether data will be shared/made open access and**

- **how data will be curated & preserved (including after the end of the project).**

Upon publication, the Open Research Europe editorial team can help advise where best to deposit research data and support you in making a data availability statement to maximize the reuse of the data.

Rights to share data

Using a **data management plan** will help make it clear who has the rights to share data, as well as how and when.

Open Research Europe recognizes that openly sharing data may not always be feasible. Exceptions are permitted according to the relevant policy of Horizon 2020. These policies consider the obligation to protect results, confidentiality obligation, security obligations, the obligation to protect personal data, if providing open access would jeopardize the achievement of the main objective of the H2020 project from which the research data derives, and other legitimate constraints. For more information, please see the **Data Guidelines**.

Open Research Europe

European Commission

Sensitive data

Following on from the previous point, if your data is too sensitive to share, you should consider sharing your **metadata**. You can openly publish a description of your data (known as a 'metadata record'). This helps others to discover your data and provides essential information about how the data can be accessed and cited.

For example, you could post a "data codebook" or "data dictionary" in a repository that describes the variables used in your dataset. In this document, you can cite the article in which it appears in order to connect the data description to the paper. Similarly, you can cite the metadata record in your article as part of a **data availability statement**, which should also include the conditions under which your data can be accessed.

Misinterpretation of data

It is important that your data is accompanied by sufficient contextual information to allow others to fully understand your dataset. A data dictionary is a separate file where each variable is defined, including units and ranges, and often includes other useful information for interpreting the dataset. By helping others better understand your data, a data dictionary supports the reuse and reproducibility, and helps to avoid misinterpretation.

Inappropriate reuse of data

Good documentation is key to both stopping and identifying inappropriate use. It is important that your data is supported by **rich metadata** that describes both the purpose of the dataset and any restraints. Where sensitive data is involved, a data use agreement makes clear the terms under which the data can be used.

With this in mind, sharing your research data is likely to have a positive impact and greater reach than you might think. Research data is consumed by a variety of stakeholders beyond researchers, including policymakers and educators. Sharing your data may also help to reduce duplication of work while promoting integrative analysis.

Open Research Europe

European Commission

Claiming priority to results through data sharing

Some researchers feel apprehensive to share data as they think others may take the data and claim priority of the results (sometime known as 'scooping'). However, there is no evidence to support this claim. Instead, data sharing establishes and confirms ownership of your data via authorship. Sharing your data openly increases the opportunity for others to credit your work – where your data is reused by another party, you will receive recognition through a formal data citation, and you may even find new collaborations as a result.

Impact of data sharing on your career

It is true to say that not sharing your data does not hurt your research career, but there are huge benefits to doing so such as increasing your chances of citations and the potential to introduce new collaborations. So, sharing your data now may be valuable to you career in time. In addition, the European Commission believes that data sharing will help to facilitate more evidence-based policies, better solutions to societal challenges such as climate change, and innovation in science, healthcare and business practices.

Another worry is often that sharing data now may impact on your ability to publish later on. However, Open Research Europe agrees with the majority of publishers that the novelty of a research article is not undermined by data sharing. This means that sharing data early-on should have no impact on your ability to publish a research article later. In addition, there is evidence to suggest that publications associated with a shared dataset have a citation advantage.

Data sharing for commercial innovation and industry applied research

Data that is generated from projects that focus on innovation in a commercial and/or industrial setting can also be shared on Open Research Europe. However, in some circumstances it is realized that due to the nature of this research it may not be possible as there is an obligation to protect the confidentiality and security of certain results.In this instance, an extensive metadata record describing the research, where it is stored, and how to access can be deposited openly in a repository and cited in the **data availability statement**.

Open Research Europe

European Commission