

Data sharing: A primer from UKRN

What is data sharing?

Data sharing is the process of making [research data](#) available for wider dissemination. The **FAIR principles** state data and metadata should be: **findable** (easy to find for both humans and computers); **accessible** (accompanied by clear instructions for access and authorisation); **interoperable** (compatible with other data and/or tools); and **reusable** (suitably described to allow further use).

Why should we share data?

There are many positive reasons to share data. Sharing data opens opportunities for scientific enquiry through the promotion of innovation via new data uses and collaboration. It maximises transparency and helps ensure the reliability of the scientific record.

Might people find mistakes in my work?

Sometimes mistakes are found in shared work, but finding mistakes is good - it means we no longer believe wrong things. Everyone makes mistakes, and [genuine errors are unlikely to be damaging to your work or career](#).

Might people use open data to scoop my ongoing research?

Scooping almost never happens, but if you're worried about it you can *embargo* data while sharing so that you have exclusive access to it for a few months or years.

How do I share data?

What can be shared with whom?

Each field has its own conventions, challenges, and solutions. The tips in this section are general, and focus on human data (but may also be relevant where for example commercial, security or legal implications exist).

It's important to plan for sharing of research data from the start of your project. Most human research data can be shared ethically and legally with the appropriate use of informed **consent**, **anonymisation**, **rights management**, and **access control**. Within a single research project, some aspects of the data may be able to be shared completely openly, while others may need to be protected (e.g. by an end-user agreement).

Human data

Data protection laws (e.g. GDPR) govern the processing of [personal data](#), but they do not apply to anonymised data, and there are some [exemptions](#) for scientific studies. Be aware that some seemingly [anonymised data](#) can be [re-identified](#). Your institution or funder may have additional rules regarding data management (e.g. requiring a [Data Management Plan](#)).

Consent

✓ Inform participants about what data will be stored, how it will be preserved, how it is likely to be used in the long-term, and how confidentiality will be protected. The UK Data Service has a bank of [example sharing-friendly forms](#) for different types of data.

✗ Ensure consent forms **do not** promise to destroy the data or promise that the data will only be seen or accessed by the research team.

Anonymisation

- ✓ Consider procedures for [anonymisation](#) at the beginning of your project in conjunction with the planning of what consent you will seek from participants.
- ✓ Work with anonymized data whenever possible. This minimizes the risk of leaking personally identifiable information and ensures that anonymisation does not distort the data.
- ✓ Remove direct and indirect identifiers from [quantitative](#) and [qualitative](#) data.
- ✓ Where data cannot be de-identified sufficiently, consider imposing [access controls](#) instead.

Access Control

Where data are sensitive, safeguarding or shielding subsets of data may be appropriate. Most large repositories allow managing access through an [End User Licence](#) which stipulates conditions for access, including agreement not to attempt to re-identify individuals.

The data itself

To enable shared data to be used effectively, they should be [understandable to humans and machines](#), which means conforming to standards. [Meta-data standards](#) facilitate long-term use and integration; some also apply to [specific kinds of data](#). The DCC keeps a [list of standards](#) you can check. [Roche et al \(2015\)](#) provide several generally-applicable recommendations for making data complete and reusable, including:

- ✓ Use [clear and detailed data descriptions and metadata](#) to increase usability. As a result, where data relate to an accompanying research paper, it should be transparent which parts of the dataset refer to which claims in the paper.
- ✓ Try to make the data accessible on their own terms, independent of another paper, especially where additional supplementary data can be provided.
- ✓ Ensure missing data or data exclusions are fully annotated.
- ✓ Provide as much unprocessed data as possible so users can “rebuild” information.
- ✓ Include analysis code and processing scripts where feasible.
- ✗ Avoid proprietary formats / file types as the **only** way to access data. Where suitable, convert the files to open or standard formats before storing data in a repository.
- ✓ Consider asking friends and colleagues to review the data submission you are preparing to ensure quality control, especially if the dataset will not be peer reviewed.

Sharing the data

1. [Find a repository](#) that is appropriate for your dataset. Make sure the repository has the capabilities you need: Is a general or specific repository more suitable? What are the file size limits? How long will the data be available? Do you need to manage sharing permissions or set up an embargo period? Will you get a [DOI](#) or similar [persistent identifier](#)? Check the terms and conditions for a commitment to retain your data. You can search a list of repositories by discipline at <https://www.re3data.org/>
2. Provide reuse guidance: [decide on an appropriate license](#) for your data; be clear about **how you would like the dataset cited**. Some repositories will do this automatically, while others will allow you to customise everything.
3. Share the persistent URL! Let people know about the resource you've just created - in a paper, at a conference, on social media.

Data should be **as open as possible** and **as closed as necessary**

Other resources

Check out institutional help (e.g. the Library) for advice, guidance and experience.

Digital Curation Centre: Resources and information <http://www.dcc.ac.uk/resources>

FAIRsharing: Resource on data and metadata standards <https://fairsharing.org/>

MANTRA: Free online course on digital data for researchers <https://mantra.edina.ac.uk/>

UK Data Archive: Social, economic and population data <https://www.data-archive.ac.uk/>

Caldicott Principles: For sharing medical data <https://www.ukcgic.uk/manual/principles>

OpenAIRE: Resource on ethical and legal issues <https://www.openaire.eu/guides>

FOSTER Open Science: Free data protection and ethics course

<https://www.fosteropenscience.eu/learning/data-protection-and-ethics/#/>

Cessda Training: Data management guide

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>

Reviewed by Chris Gorgolewski (0000-0003-3321-7583), Olivier Klein (0000-0003-2737-8049), Frederick Verbruggen (0000-0002-7958-0719), and Tony Weidberg (0000-0002-5158-307X). Edited by Emma L Henderson (0000-0002-5396-2321) and Jackie Thompson (0000-0003-2851-3636)

Conflict of Interest Statement: Reviewer Chris Gorgolewski works for Google Dataset Search (<https://g.co/DatasetSearch>)

This work is licensed under [CC BY 3.0](https://creativecommons.org/licenses/by/3.0/)