

---

# Repositories and Data Journals

---

Like with scientific publications, there are multiple ways how to publish your data:

1. Supplementary Materials to your research article
2. Repository
3. Data journal

## Supplementary Materials to a research article

Some publishers offer the option of attaching **Supplementary Materials** to an article. In this way, authors can add, for example, raw data or code to their article, and the publisher then publishes the materials online and adds a link to the materials to your published article. If you decide for this option, keep in mind that unless the publication is Open Access, the **copyrights are usually transferred to the publisher**. Moreover, this option would not allow readers to reuse the data, or cite it independently of the main publication.

## Data repositories

The best way to preserve your data - whether you decide to share them or not - is to deposit them in a **data repository**. To increase the impact of your data, you should deposit your data in a **subject specific repository**. The advantages of subject specific repositories are that they bring together researchers from a particular scientific discipline who share their data with one another, and they are also usually better equipped to meet the needs of the community, for example, when it comes to the type of data the researchers share. An example of a subject specific repository is [LINDAT/CLARIAH-CZ](#) for linguistic data and tools, which is developed at the Institute of Formal and Applied Linguistics at MFF UK.

When choosing a suitable subject specific repository, it is a good idea to use a repository which is already established for your research domain - you can ask your colleagues where they deposit their data, or use the international **registry of data repositories** [re3data.org](#) (Registry of Research Data Repositories).

If you cannot find a suitable subject specific repository, you can deposit your data in a general-purpose repository, which store data of all scientific disciplines. The most commonly used general-purpose repositories are [Zenodo](#), [Figshare](#), or [Dryad](#). The [Generalist Repository Comparison Chart](#) can help you select a general repository.

When choosing a suitable repository, check the following as well:

- Does the repository give your submitted dataset a **persistent and unique identifier** (e.g., DOI)? Thanks to a persistent identifier, your data are more easily findable and citable.
- Is the repository certified as a **'trusted data repository'**? If the repository is certified, it is more likely that your data will be well looked after.
- Does the repository enable **open access** to your data? If you decide to share your data, this is key information.
- Does the repository **license** your data? Does it offer clear terms and conditions for data reuse? It is important that others know what they can and cannot do with your data.
- Does the repository provide a **landing page** for your dataset with **metadata**? Metadata will help others find your data, tell what they are and how to cite them.
- Does the repository enable **versioning**? If you update your dataset, you can upload it as a new version of the original dataset. The new dataset is given its own identifier and users can easily find out what is the latest version or which version was used in a particular study.

You can easily check some of these information at [re3data.org](#). Each entry, in the upper right corner, includes a series of pictograms which tell you, for example, whether the repository uses persistent identifiers, whether it is open or certified. You can find more information in the detailed entry or on the website of the repository.

**LINDAT/CLARIN repository**

Subject(s) Humanities Linguistics Artificial Intelligence, Image and Language Processing Humanities and Social Sciences Computer Science  
Computer Science, Electrical and System Engineering Engineering Sciences

Content type(s) Standard office documents Audiovisual data Plain text Structured text Archived data Source code

Country Czech Republic European Union

LINDAT/CLARIN is designed as a Czech "node" of Clarin ERIC (Common Language Resources and Technology Infrastructure). It also supports the goals of the META-NET language technology network. Both networks aim at collection, annotation, development and free sharing of language data and basic technologies between institutions and individuals both in science and in all types of research. The Clarin ERIC infrastructural project is more focused on humanities, while META-NET aims at the development of language technologies and applications. The data stored in the repository are already being used in scientific publications in the Czech Republic.

## Data journals

Data journals **mirror the scientific publication model** and promote the publication of data papers, which can be defined as "a scholarly publication of a searchable metadata document describing a particular online accessible dataset, or a group of datasets, published in accordance to the standard academic practices" (definition by [Chavan & Penev, 2011](#) ). Some journals are dedicated to publishing data papers only (mind that different publishers may call this type of publication differently, e.g., database article, or data descriptor), others are mixed and publish data papers alongside traditional research articles. Data paper is therefore analogous to traditional research papers, can be cited and can be reported in OBD (the system used at Charles University for registration of creative activities).

Different data journals may have different policies but the general aim is for the data paper to **describe the dataset**, provide information on what the data are, how and where they were created and so on, and the paper should also contain a **link back to the dataset** (ideally via a persistent identifier such as a DOI). Data papers should not include scientific analyses made by using the data. Generally, the publishers should not host the data; instead, the **dataset should be deposited in a trusted open access repository** so that even if the paper might have restricted access, the dataset would still be available.

Most data journals also perform **peer review**, however, there may be differences in terms of what the reviewers are requested to assess and whether the peer review process is open or not.

Listed below are some examples of data journals and you can find further examples [here](#) .

- [Scientific data](#) - mainly natural science disciplines
- [Earth System Science Data](#) - geosciences
- [Journal of Open Archaeology Data](#) - archaeology
- [Biodiversity Data Journal](#) - biodiversity
- [GigaScience](#) - life and biomedical sciences
- [Journal of Open Research Software](#) - research software

## Useful Resources

Chavan, V., Penev, L. 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12, S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>

Gould, Julie. 2014. How to publish your data in a data journal. Naturejobs Blog. <http://blogs.nature.com/naturejobs/2014/12/04/how-to-publish-your-data-in-a-data-journal/>

Stall, Shelley, Maryann E. Martone, et al. 2020. [Generalist Repository Comparison Chart](#) . Zenodo. <http://doi.org/10.5281/zenodo.3946720>