

# Publishing and preservation steps

As a researcher, you should make reasonable efforts to clean up your data and provide metadata (e.g. authors, title, topic etc.). The metadata should provide a thematic overview of your research data, including information on their development, methods used and legal aspects.

Metadata allow your data to be located via a search of the [Research Collection](#), the [ETH Library search portal](#) and other search engines, and also makes it easier for you, and others, to gain an overview of said data.

Enhanced metadata also include the documentation of your data's thematic context, which is designed to enable its subsequent reuse. The choice of an appropriate data format may also significantly facilitate the reuse and preservation of the data.

The [Step-by-Step Guide on Data Publication for ETH Zurich Researchers](#) supports you with documenting your metadata and datasets and with preparing them for publication in a FAIR data repository.

You should also keep the following in mind:

- Pack the folder structures into ZIP or tar container files.
- Avoid password protection, encryption and compression if possible.
- Ensure that the path length does not exceed 200 characters.
- Avoid special characters in file names.
- Ensure that file extensions are consistent with the file format.

The choice of appropriate file formats will improve the (re)usability of your data and increase the chances of their effective preservation. As a result, it is worth thinking about appropriate file formats at the project's outset, and including these in your Data Management Plan (DMP).

It may also be advisable to convert a specific file format into another format with a longer lifespan after processing (see [File formats for archiving](#)).

Although file formats that offer long-term usability are not required for publication in the repositories of ETH Zurich, problematic formats may significantly impede future use.

All employees who have contributed to a study must give their written consent to the data's publication under the selected specific licence of use for this purpose. Data packages may only be uploaded to repositories by the data producers, or by individuals authorised to do so by the latter.

A large number of subject-specific research data repositories exists in addition to the [Research Collection](#), ETH Zurich's repository.

- An [overview of such repositories](#) has been compiled by PLoS (Public Library of Science).
- If the files are in a format that is well-established in your field of research, we recommend that you publish them in a repository that is experienced in handling this file format. A large number of subject-specific and institutional repositories can also be found in the [re3data index](#).

All data produced in the context of your research at ETH Zurich can be published and archived in the [Research Collection](#). As the data producer, you may select all crucial publication parameters such as [access rights](#) and end user licence independently. After undergoing an internal evaluation by ETH Library staff, your data will be published promptly.

What benefits does the Research Collection offer to data producers?

- Flexible access rights, from "open access" to "access upon request".
- Registration of Digital Object Identifiers (DOIs) and [DOI advance assignment](#)
- Daily download statistics and altmetrics for your published data
- All formats are authorised for upload
- Contents preview for ZIP and tar containers
- Link to GrantIDs of EU and SNSF projects, plus export to the EU portal [OpenAIRE](#)
- Long-term data accessibility
- The Research Collection's so-called "dark archive", the ETH Data Archive is the backbone of data curation. The data you submit to the Research Collection are automatically archived for the long term in the ETH Data Archive (except for data deposited via [Libdrive](#)).

When you submit your data, you can choose who has access rights to the information:

- Open access: freely accessible
- Embargo: free access after an embargo period determined by you
- Members of ETH Zurich: access with [ETH login](#) only
- Selected users: access for a specific group of people
- Closed access: storage for archiving purposes only; access limited to the individuals submitting and to authorised specialists at ETH Library.

In the event that access restrictions are implemented, unauthorised individuals may request access to the data via the "request access" function. The Research Collection team will forward such requests to the respective copyright holders in the event that they are still employed at ETH Zurich.

In any case, your research data's descriptive metadata remain visible and freely accessible in the Research Collection.

The shortest storage period for research data in the Research Collection is ten years. After the chosen storage period has expired, [ETH Zurich University Archives](#) will consult the data producers and decide whether specific data should continue to be stored for historical reasons. For use exceeding ten years, [common, open file formats](#) should be used as far as possible.

The publication or archiving of data is offered by ETH Library as a free service.

All research data and documents published in the [Research Collection](#) receive a Digital Object Identifier (DOI) and can thus be permanently addressed and cited.

We recommend that you cite a research dataset with the following minimum specifications:

*Creator (PublicationYear): Title. Publisher. Identifier*

Example:

Swaminathan, R., Ramya, T., Karthik, C.S. (2013): *Contortrostatin-Reprolysins Domain Structure*. Swiss Institute of Bioinformatics. (Model). <https://doi.org/10.5452/ma-c12zs>

For further information on the correct citation of research data, please consult the [DataCite website](#).

If you wish to refer to a research data set due for publication at a later date in a publication or when submitting a manuscript, you can [reserve a DOI](#) in the Research Collection.