

# Guideline for Long-Term Archiving in Edmond

## 1 Goal of the Guideline

Supporting users in documenting their data with the aim of making the data available for reuse beyond the ten-year period.

## 2 Process of long term archiving

During Upload:

1. User uploads the research data in one dataset.
  - a. Documentation of the data according to the criteria below.
2. User informs the Edmond team about the desired long term archiving of the dataset.
3. If necessary, the Edmond team advises the user on suitable options for long-term archiving within and/or outside the MPG.
  - a. If applicable, completion of a consent form regarding any necessary data conversions by the Edmond team.

After the ten-year period:

1. The Edmond team will check suitable filing locations upon request.
2. The Edmond team transfers the data to a suitable archive upon request. The archive stores the data according to its specifications.

## 3 Criteria for consideration regarding the data for long-term archiving within the MPG

Edmond ensures retention for at least ten years in accordance with GSP requirements. If researchers wish to retain data beyond the specified ten-year period, the Edmond team will transfer the data to an appropriate archiving solution.

The following recommendations should be considered:

- Documentation of the data set regarding provenance and other post-use aspects as comprehensive as possible.
  - o Store Readme file
  - o Filled out "Description field"
  - o Usage of "Tags"
- Edmond users are encouraged to add domain-specific metadata in the provided metadata schemas.
  - o Additional information on data conventions.
  - o If custom software is required to read the data, it should be included in the data set.
- Set license terms.
- Use of data types that are as widespread as possible.
  - o Favoring of open data standards, for further information see <https://rdm.mpdl.mpg.de/before-research/file-formats/>
- Avoidance of proprietary data types.
  - o If the data must be stored proprietary, it is recommended to additionally store the data in a standard data format, as far as this is possible.
  - o Depending on the volume of data, it may also be necessary to offer a copy in a non-proprietary format.
- Consideration of data types explicitly designed for longevity.

## 4 Criteria for subsequent use

If the user wants to facilitate the subsequent use of their data, the following criteria should be met:

- As a basic rule, as much metadata as possible should be made available for others to be able to understand and reuse the data in an appropriate form.
- Orientation towards the requirements of the FAIR principles.
- Storing information about provenance, e.g. in the "Description field" or in a Readme file.
  - o Provenance description (beyond Study type)
    - Underlying data conventions
    - Information on method and experiment setup
    - Information on experiment equipment
- Compliance with basic data management principles already during the production of the data.
  - o Notes on data management plans <https://rdm.mpg.de/before-research/data-management-plans/>
  - o Notes on data quality <https://rdm.mpg.de/before-research/data-quality/>

## 5 Contact

You can always contact the Edmond support team via [edmond@mpdl.mpg.de](mailto:edmond@mpdl.mpg.de).

*Last edit by 22 February 2023.*