

Introduction how to request a DOI for datasets uploaded to the ATTO data portal

Authors: Marcus Guderle

Contact ATTO Data Management Team: attodbm@mpi-mail.mpg.de

Table of content

[1. Overview](#)

[2. FAIR Priciples in the ATTO Data Portal](#)

[3. Data Upload Workflow](#)

[4. DOI registration](#)

1 Overview

The ATTO data portal is the main tool for data sharing within the ATTO consortium and for external researchers. This short tutorial will show you how to request a DOI for datasets uploaded to the ATTO data portal. The first section will give an introduction into the FAIR Principles. In the second chapter shows the workflow of the data upload. Finally, the DOI registration is explained in chapter 3.

2 FAIR Priciples in the ATTO Data Portal

The data management within the ATTO consortium follows the FAIR Guiding Principles for scientific data management and stewardship (Wilkinson et al., <https://www.nature.com/articles/sdata201618>). This means that data should be ...

- **F**indable – good metadata, data are indexed
- **A**ccessible – data are retrievable & metadata are always accessible
- **I**nteroperable – have common metadata standards
- **R**e-usable – richly described, follows community standards & with data usage license

This is implemented by describing the data sets with approx. 80 metadata variables and a data's own dataset ID incl. version number. Furthermore, all metadata can be retrieved without login and the primary data is either publicly available or can be requested. In addition, the metadata follows general metadata standards. For example, the metadata meet the mandatory metadata requirements of DataCite/Max Planck Digital Library (MPDL), which allow an easy publishing process with DOI assignment from DataCite. Table 1 gives an overview of the respective metadata properties required by DataCite. In addition, the FAIR principles are guaranteed by a uniform management of data structures, variables and units of measure.

Data structures contain all variables, which are part of the dataset to be uploaded. Prior to uploading a dataset, the data structure has to be defined in order to have a template for the upload process. Defined data structures can be used and should be used multiple times for long term data. This prevents that the same variable is defined several times with different names and units (e.g. time in seconds and in Universal time). This is essential to enable an extensive and consistent data management as well as exact search results. For this reason, the definition of data structures will be done by the data management team and the following data upload process is mandatory.

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

Figure 1: DataCite mandatory properties for metadata

3 Data Upload Workflow

1. Users create a data sample of their data and send it to the data management (attodbm@bgc-jena.mpg.de).
2. The data management creates a structure for each dataset and defines the variables based on existing entries and future uses. These are part of the data template for the upload.
3. This data template will be sent to the users.
4. Users enter their data in the data template.
5. Together with the data management, the metadata are created and the data will be uploaded to the ATTO Data portal.
6. If the data has been uploaded to the data portal, the respective user rights for the dataset are assigned together with the data management.
7. In connection with point 6 it also results how the data is available. There are the following scenarios: a) Data collected in the context of a (doctoral) thesis will be published latest 6 months after defense. b) Data are freely available for the ATTO Consortium. In both cases it is possible to request the data from

the owner via the "Request Access" button if not freely available. c) Data can be published so that they can be downloaded by interested scientists without registration. In this case it is possible to assign a DOI to the data set via the Max Planck Digital Library/DataCite.

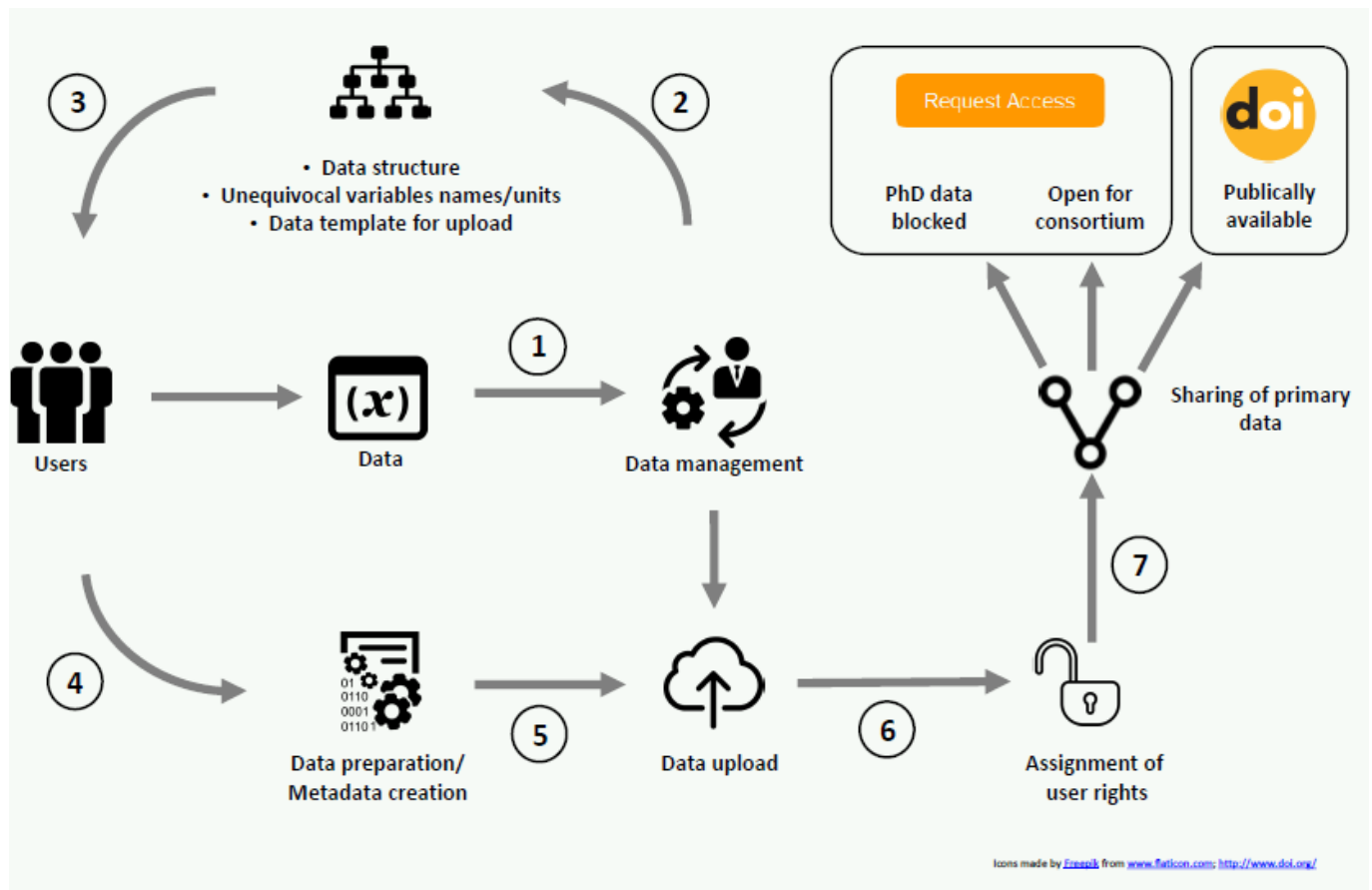


Figure 2: Detailed workflow for data uploads to the ATTO Data Portal

4 DOI registration

Currently, DOI registration for datasets is done manually by the data management team. Therefore, we ask all consortium members to contact the data management team (attodbm@bgc-jena.mpg.de) in case datasets should be provided with a DOI. A prerequisite is of course that datasets have been uploaded to the ATTO data portal. Please check the metadata and primary data for accuracy.

The data management team will assign a DOI to the respective dataset via the Max Planck Digital Library/DataCite, which has the following structure:

10.17871/atto.DatasetID.VersionNr.VersionID

Since the DOI refers to the respective dataset with a specific dataset ID, version number and version ID, it is necessary to lock the respective dataset after assigning the DOI, so that a possible change of the dataset is prevented. Since any change - both to the metadata and to the primary data - will increment the three numbers mentioned above. If the case arises that after assigning the DOI an error becomes apparent in the dataset, the dataset must be uploaded again and will be assigned a new DOI, which is following the rules of DOI.

If you have any questions or require a DOI, please contact the data management team via attodbm@mpi-mail.mpg.de.