

Metadata review guidelines

4TU.ResearchData has reviewers who check every submission to ensure the validity of files and metadata and to help make it [Findable, Accessible, Interoperable, and Reusable \(FAIR\)](#). The metadata review is an essential part of our curation process and ensures that the quality requirements of the repository are met.

Process: **Upload** → **review** → **feedback** → **acceptance and completion**

When reviewing the (meta)data, reviewers have a task in increasing the findability and visibility of the data and thus contribute to the reuse of the data.

The following checks are always performed:

- Do the description and data make sense (also when referring to the publication which may not be open access)?
- Do the authors adequately explain its utility to the community?
- Are the protocol/references for generating data adequate?
- Data format (is it standard? potentially re-usable?)
- Is the data well documented?

For the review per field, the reviewer checks to what extent suggestions to improve these fields with even better and clearer information can be made. Comments and suggestions for updates and improvements are returned to the depositor.

All communication with depositors from member institutions takes place via the local front offices.

REVIEW

The first element of the review is checking the **DATA** itself:

- **File formats** (mime types/media types)

Dataset is preferably provided in open, non-proprietary file formats.

List of preferred file formats: https://data.4tu.nl/s/documents/Preferred_File_Formats_2023.pdf

Does the dataset contain files in netCDF/HDF5? Then these files (possibly in consultation with the depositor) are transferred to the OpenDAP server. For this purpose, the files are downloaded and the url to the OpenDAP server is added to the 'Data Link' field. Contact the developer of 4TU.ResearchData to transfer the NetCDF files to OPeNDAP.

If the data contains a zip file or other compressed files containing multiple file formats, the individual formats are listed in the 'Format' field.

- **File organization**

Recommend file organization as single or multiple 'datasets' or 'collections'.

Sometimes data files are uploaded as single datasets while the files can also be merged into 1 dataset item.

In case several datasets belong together, we suggest to the depositor to place them in a collection.

- **README file.**

Has a readme file been provided?

-> YES: does the readme contain sufficient information to enable reuse of the dataset?

-> NO: ask the depositor to provide a readme file.

See also: [Guidelines for creating a README file](#)

- **Privacy aspects**

Do the data files contain personally identifiable, sensitive or inappropriate information?

Data that is personal, confidential or sensitive in nature, is only accepted for archiving when the data has been anonymised so that individuals cannot be identified. Pseudonymised data (with the exception of sensitive personal data and special categories of personal data which cannot be deposited in 4TU.ResearchData) can be shared under restricted access conditions. For survey/interview data and other types of data coming from human participants, an unsigned consent form will be asked.

- **Content**

Does the data match the metadata provided?

Does the data fit within the scope of 4TU.ResearchData? See also: [Data Collection Policy](#). If

NO: then the curator makes suggestions for other data repositories.

The next element of the review is checking the **METADATA** quality (completeness, accuracy and consistency):

An asterisk* indicates the required fields.

The framed texts are the input instructions (help text) on the upload page.

Title*

Give your research a title that is more descriptive than just a file name. This will also help make your items discoverable via search engines such as Google.

The title gives an accurate description of the dataset and should preferably include time indications and geographical references. The title should as far as possible resemble the name under which the research project is known in publications.

If the title entered is unclear, the reviewer will suggest a more specific title.


Authors*

Select co-authors from the list or manually add them by creating a new author record. Hit enter after adding each author and drag and drop names to arrange them in the order you wish.

The author is the researcher who carried out the research. These can be either personal names or institutional names. The name of the depositor is automatically added as author (owner). Co-authors can be derived, or added if not present.

If the author does not have an ORCID, the author is requested to create one. The ORCID is entered in the available space in the 'Author' field.

Example:

Edo Abraham		▼	🗑️
Dimitri Solomatine		▼ ^	🗑️
Ja-Ho Koo	✎	▼ ^	🗑️
Andreja Jonoski 	✎	^	🗑️

We aim to complete author names with ORCIDs where possible. Please note that only author names with a pencil icon to the far right of the name can be changed. The order in which the names are displayed can be adjusted using the up and down arrows.

The reviewer usually compares the authors' names with the names of authors listed in the associated publication (if applicable). If this is different, the uploader is asked whether the co-authors of the publication should also be added to the dataset.

Affiliations of authors are added in the Organizations field.

Description*

Add as much context as possible so that others can interpret your research and reproduce it. Make sure you include methodology, techniques used, and if relevant information about approval for data collection to confirm adherence to legal or ethical requirements. The description should have at least four characters.

Brief description of the study. Can include the purpose, the methods and the type of data and the result.

The curator can make suggestions to supplement these (e.g. from the abstract of the publication).

You can also refer here for more information, e.g. to the linked publication or thesis. Preferably indicate here for which target group(s) these data may be of interest.

Licence*

Select the appropriate licence from the list. Need to know more? Read our [guidance](#) for more information on the licence types we offer.

The depositor can choose from different licences for data and software from a menu.

NOTE: for software other licenses apply than for data (except CC0). So make sure that the correct license is chosen. All license types are listed at: <https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository/licencing>

It is also important to check in the data files whether a license is mentioned (e.g. in README), or perhaps there is even a separate license file present, and whether it corresponds to the license in the metadata. If the licenses do not match, the uploader must be made aware of this.

The 'Restrictive license' cannot be selected from the list, but is automatically applied to datasets for which 'Restricted access' is selected.

Please note that CC licenses can not be used for datasets with restricted use.

At this point depositors can apply one of the following access conditions to their dataset:

- **Open access**; the default setting in 4TU.ResearchData by which your dataset will be directly accessible to others.
- **Embargoed access**; period during which a dataset is unavailable and not accessible to anyone else. Please enter the end date of the embargo period and add the reason why the files are under embargo.

The metadata form currently offers 2 embargo types: Files-only embargo and Full content embargo.

Embargo type *

Files-only embargo Full content embargo

Please note that files-only embargo is strongly preferred because only the files are inaccessible and the dataset is still findable via the metadata.

NOTE: Embargo is a period during which a dataset, which will eventually be made public, is unavailable and inaccessible to others. The reviewer checks whether the embargo end date and reason has been entered, and whether the reason is clear and suitable for this purpose. 4TU.ResearchData applies a maximum embargo period of two years. At the end of the embargo period, the dataset is automatically made available.

- **Restricted access**; allows you to share your dataset under specific conditions and with users who have been granted access. Please specify the End-User Licence Agreement (EULA) in which you specify the conditions under which you grant users access to your data files, and add the reason why your dataset is restricted.

NOTE: Restricted access can be applied to datasets for which access must be limited. Access can only be granted upon request. Datasets containing personal data (pseudonymized) are mostly placed in this category.

Sensitive personal data and special categories of personal data are NOT accepted for archiving in 4TU.ResearchData. If 'Restricted access' is selected, the reviewer checks whether the reason for this has been entered. The reason must be clear why access to the data is restricted, e.g. because of contracted research, and under what conditions access can be obtained.

Categories*

Select one or multiple categories that this research best correlates with.

Categories are disciplines or fields of research to which the dataset belongs. The available

category options are based on the Australian and New Zealand Standard Research Classification (ANZSRC) Fields of Research (FOR) codes.

Group*

Based on the email address with which the uploader is logged in, the group to which the uploader belongs is automatically set:

Group *

Delft University of Technology

- Delft University of Technology
- Delft University of Technology Students

Eindhoven University of Technology

- Eindhoven University of Technology
- Eindhoven University of Technology Students

University of Twente

- University of Twente
- University of Twente Students

Wageningen University and Research

- Wageningen University and Research

Other institutions

- Other institutions

Publisher

No help text available.

A holder of the data (including archives as appropriate) or institution which submitted the work. Any others may be listed as Organizations.

Default: 4TU.ResearchData

Funding

Select from the list by searching for funder name, grant number or grant name or manually add them by creating a new funding record.

Depositors can look up and extract information from a list. If the funder information is not in the list it can be added by creating a new funding record.

Resource title

Please add the title of the peer-reviewed article/paper or research you want to link back to. This will appear on the public page as the link title. If you add the Resource title, you will need to complete the Resource DOI field.

NOTE: the peer-reviewed article will only be shown on the public landing page if both fields (Resource title and Resource DOI) are filled in.

Resource DOI

Please add the DOI of the peer-reviewed article/paper or research you want to link back to. This is the link destination for the Resource title. If you add the Resource DOI, you will need to complete the Resource title field.

Preferably, only the DOI should be provided. Not the full DOI link. E.g: 10.4121/123456. Once the full DOI link has been provided, e.g. <https://doi.org/10.4121/123456>, it should be checked whether the browser displays the url correctly.

References

Add references (links) to any relevant content or external sources that help describe the resource. All links must be full, valid URLs or DOIs in order to function, eg: <https://doi.org/10.4233/uuid:35022242-b3ba-48d9-b3f6-c6ab20b8cc19>. Hit return to enter new references.

NOTE: in the case of a peer-reviewed article, preferably use the Resource title / Resource DOI field for better visibility on the public landing page.

If no link to a publication is indicated, the moderator can search for it or ask the depositor/author for it.

This can also be added later after publication of the dataset by the depositor himself.

DOI reservation

Upon publication, a DOI will be assigned to your dataset. If you need to have the DOI before publication, you can use this value.

This field is automatically populated with the DOI registered for the dataset, but only becomes active when the dataset is published.

Language*

Select a language from the list.

The primary language of the resource. This is set to 'English' by default and can be adjusted.

Time coverage

Indicate the dates to which the data refer. Enter the year, or beginning and ending dates.

Period to which the data relates.

Format: year or starting year-end year, or free format e.g. for a geological period, or a series of discrete points in time (separated by a comma+space). If there are dates in it, then in format 2020-09-13, and with start- and end dates both separate with a "/", e.g. 2019-12-25/2020-01-05. Also possible on monthly level 2019-12/2020-01, or year 2019/2020.

Geolocation

When your dataset covers geographical data, fill in the following section.

- **Name**

The geographic area to which the data refer (e.g. municipality, town/city, region, country).

Geographical area in which the data has been collected. Place, region, country or other geographical unit.

- **Longitude**

Geographic longitude in decimal degrees, East is positive, West is negative. Values: -180 to 180.

Both Longitude and Latitude are fields for geographical coordinates.

NOTE: if coordinates are filled in, a link 'view on openstreetmap' will automatically appear on the public landing page.

- **Latitude**

Geographic latitude in decimal degrees, North is positive, South is negative. Values: -90 to 90.

File format(s)

Format of the data file(s), e.g. the media type like image/jpeg or a description like "g-zipped shape files".

This field contains the mime types of migrated datasets. For new datasets it is advisable to indicate the individual file formats here in case the dataset consists of a zip file or other

compressed files.

Link to extra data that belongs to the current item, e.g. an OpenDAP catalog with netCDF files on our OpenDAP server.

Data Link

Data in netCDF or HDF5 format is stored on a separate OpenDAP server. Data in other formats are stored on the regular server. The field 'Data Link' is used to refer to an external location (e.g. OpenDAP). Example: <https://doi.org/10.4121/18614102.v1>. The link to the OpenDAP server is shown in the DATA section.

Derived From

URL of the dataset from which the current item was derived.

Same As

URL of exactly the same data. This can be a copy in another location or an alternative URL that redirects to the current item.

Keywords*

Add keywords that will help make your research more discoverable. Use a semicolon (;) or hit return after each keyword you enter.

Keywords with which the dataset can be found. Keywords can often be taken from the Description, or from the Abstract of the corresponding publication. Keywords can also be derived from the name of the faculty or department.

Organizations

Name of the organizations that contributed to the creation of the resource. Separate different names with a semicolon (;)

Make the suggestion to depositors to include the faculty name and, if applicable, the research department as an Organization.

NOTE: previously we used the Contributors field for this. So the Contributor field may still appear in previously published datasets. When these datasets are changed and then accepted again, the Contributor field is automatically removed. When reviewing modified datasets, make sure that the names of Contributors are transferred to 'Organizations' or in case of personal names to 'Description'. In this case, add at the end of the Description (new line): "Contributors: person1,

person2”.

File type

Based on the depositor's choice of 'File deposit' or 'Software deposit', the file type is automatically set to 'Dataset' or 'Software'.

Dataset: group of data files, usually numeric or encoded, related to a specific topic and collected for a specific purpose.

Software: set of programs, procedures, code, scripts, etc. to run a computer system or application.

FEEDBACK TO DEPOSITOR

The reviewer formulates the suggestions for improvement in such a way that the front office colleague can easily adopt the suggestions and comments in the mail to the depositor.

If no response is received from the depositor, then reminders are sent after 2 weeks, 4 weeks and 6 weeks.

If no response has been received after 5 weeks, it will be examined whether the suggestions are such that they really need to be approved by the depositor or whether we can publish the dataset without approval.

In the final reminder (after 6 weeks) we indicate:

- 1) that we will publish the dataset with our suggestions if no response is received to this reminder, or
- 2) that we will decline the dataset if no response is received to this reminder.

ACCEPTANCE AND COMPLETION

Acceptance

Once the depositor has responded, the reviewer processes the enrichment proposals. Once the dataset is published, an automatic email confirmation is sent to the depositor with a recommendation to use the DOI for reference to the dataset.