

Metadata review guidelines

4TU.ResearchData has a team of moderators 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 moderators have a task in increasing the findability and visibility of the data and 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 moderator 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/info/fileadmin/user_upload/Documenten/Preferred_File_Formats_2019.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.

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

- **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 moderator 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 (tips) on the upload page.

Title*

Give your research a title that is more descriptive than just a file name. This will also help making your items discoverable via search engines 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 moderator will suggest a more specific title.

Authors*

Select co-authors from the list or manually add them by typing in all mandatory info. 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. Co-authors can be derived, or added if not present.

Because names can be derived and are often linked to profiles in figshare, we are reluctant to improve names to avoid creating duplicate names.

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.

ORCIDs are not shown on the dataset itself, but they are stored. In the near future figshare will also make ORCIDs visible in the datasets.

Affiliations of authors are added in the Organizations field.

Categories*

Discipline(s) or field(s) 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 determined:

- 4TU.ResearchData
- Delft University of Technology
Delft University of Technology Students
- Eindhoven University of Technology
Eindhoven University of Technology Students
Other institutions
- University of Twente
University of Twente Students
Wageningen University and Research

The dataset can be found via the umbrella group '4TU.ResearchData' as well as via the specific group.

Item type*

Defines the content of the item (dataset).

Choice from: 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.

Keyword(s)*

Subject, keyword, classification code, or key phrase describing the resource.
Enter as many keywords as possible that describe the topic of the dataset.
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.

Description*

Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data. Please limit your entry to a single paragraph of text.

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

The moderator 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.

Funding

Search by grant number, name or funder body to add the funding information about your work. Multiple grants are supported.

This field is integrated with Dimensions allowing depositors to look up and extract information from a list. If the grant code is not in the list it can be added as free text.

Resource title

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.

See also: References

Resource DOI

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.

See also: References.

Only the DOI should be provided. Not the full DOI link. E.g: 10.4121/123456.

References

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.

Licence*

Terms and conditions on how the dataset may be used. Our recommended licence is CC0 as it makes your data maximally reusable.

The depositor can choose from different licences for data and software from a menu. The licence '4TU General terms of use' is listed but no longer applies for new dataset uploads. So if this licence has been selected by the depositor, the moderator makes a suggestion for a different licence (preferably CC0).

Publisher*

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

Language

The primary language of the resource.
Choose only when the language is not English.

This is set to 'en' (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

Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country) as accurately as possible.

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

Geolocation longitude

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

Both Geolocation longitude and Geolocation latitude are fields for geographical coordinates.

NOTE: these fields do not result in the display of the location on a map. If these fields are filled in, it is recommended that you alert the uploader.

If a map is desired, the depositor must create and submit a KML file.

Geolocation latitude

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

Format

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.

Data Link

Link to the data that is an integral part of the current item, e.g. an OpenDAP catalog with netCDF files on our OpenDAP server.

Data in netCDF of HDF5 format are 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). If a dataset contains both netCDF and another format, the data files in the other format are shown in the space at the top of the landing page. Example:

<https://doi.org/10.4121/uuid:bd714fbf-1ab1-43e0-93d3-8dbd907a2cf4>

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.

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 necessary, 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".

Depositors can apply the following options of 'restricted access' to their dataset:

Embargo period

Embargo is a period during which a dataset, which will eventually be made public, is unavailable and inaccessible to others.

Apply embargo ×

Embargo period
The embargoed content will become public on the date you select.

Select period... or until specific date... 📅

- 1 month
- 4 months
- 6 months
- 1 year
- 2 years
- Permanent Embargo

Embargo?
the public metadata record.

files are under embargo...

4TU.ResearchData has a maximum embargo period of two years. At the end of the embargo period, the dataset is automatically made available. The landing page shows when the embargo period ends, e.g.:



If an embargo period has been entered, the moderator checks whether the reason for this has been entered, whether the reason is clear and suitable for the purpose.

Permanent embargo

Permanent embargo (or restricted access) can be applied to datasets for which access must be limited. Access can only be granted upon request.

Datasets containing personal data are mostly placed in this category.

NOTE: sensitive personal data and special categories of personal data are NOT accepted for archiving in 4TU.ResearchData.

If an permanent embargo is selected, the moderator checks whether the reason for this has been entered, and whether the contact details of the person/department to whom access to the data can be requested have been added.

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.

FEEDBACK TO AUTHOR/DEPOSITOR

The moderator 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 author/depositor.

The front office colleague sends the email from researchdata@4tu.nl with bcc naar researchdata@4tu.nl (so that the mailbox administrator knows that the mail has been sent out). The front office colleague does sign the email with her/his name.

If no response is received from the depositor/author, 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/author has responded, the moderator processes the enrichment proposals. Once the dataset is published, a confirmation email is sent to the depositor with a recommendation to use the DOI for reference to the dataset.