



How to share data related to a publication

SUMMARY

BEFORE YOU START.....	1
THE DATAVERSE: IS IT A PERSONAL DATAVERSE OR A PROJECT DATAVERSE?.....	2
THE STRUCTURE: ONE OR MANY DATASETS?.....	2
THE DATASET: NAMING AND METADATA.....	4
THE RELATED PUBLICATION: IS IT PUBLISHED, UNPUBLISHED OR A PRE-PRINT?.....	4
1. A PUBLISHED PUBLICATION.....	5
2. AN UNPUBLISHED PUBLICATION.....	5
3. PRE-PRINTS.....	6
HOW TO LINK DIFFERENT DATASETS.....	8
LAST NOTICE.....	9
FURTHER HELP.....	9

BEFORE YOU START

- Firstly, consult the [Checklist Prepare your dataset](#) each time you intend to create a dataset and/or a dataverse.
- Remember that well-structured and well-described dataverses and datasets are key for a [FAIR](#) use of research data, for this allows others to access, understand and use your research data more easily.
- Select data that has not been published yet or received a DOI.
- Check the quality of your data using some [useful tools](#) and provide the maximum level of FAIRness in your data to allow its reproducibility and reuse.
- If you have compiled it, please be sure that you are following the same data publication

enlisted in your project's [Data Management Plan](#).

THE DATAVERSE: IS IT A PERSONAL DATAVERSE OR A PROJECT DATAVERSE?

Importantly, think about your data at the main dataverse level since there are some slight changes for uploading data into your personal dataverse or into the dataverse of a project.

- If you are uploading data related to a scientific publication into your **personal dataverse** you can create a new dataset (or, according to your needs, many datasets, as explained in the following paragraph). However, if you are also planning to upload **work data representing the final stages of your research and not related to a scientific publication** to your personal dataverse, you can convey your space differently. Inside your personal dataverse, you can create two dataverses: one for work data which is not related to publications; and one for the data related to your publications. Thus, you will have one dataverse entitled '*Work data for projects*' and a dataverse entitled '*Replication data for publications*'. Use the second dataverse ('*Replication data for publications*') to generate as many datasets as necessary for uploading data associated to your scientific publications.
- If you are uploading data related to a scientific publication into a **dataverse of a project**, then you and your research team should ensure that, inside the dataverse of your project, you have created a dataverse where to **upload and share only the data associated to the publications produced for the research project**. This dataverse should be entitled '*Replication data for publications of "Title of the project"*' and can be used to upload the datasets (as many as necessary) associated to the project's publications. For further information, please, check the guidelines on [How to structure a dataverse for your project](#).

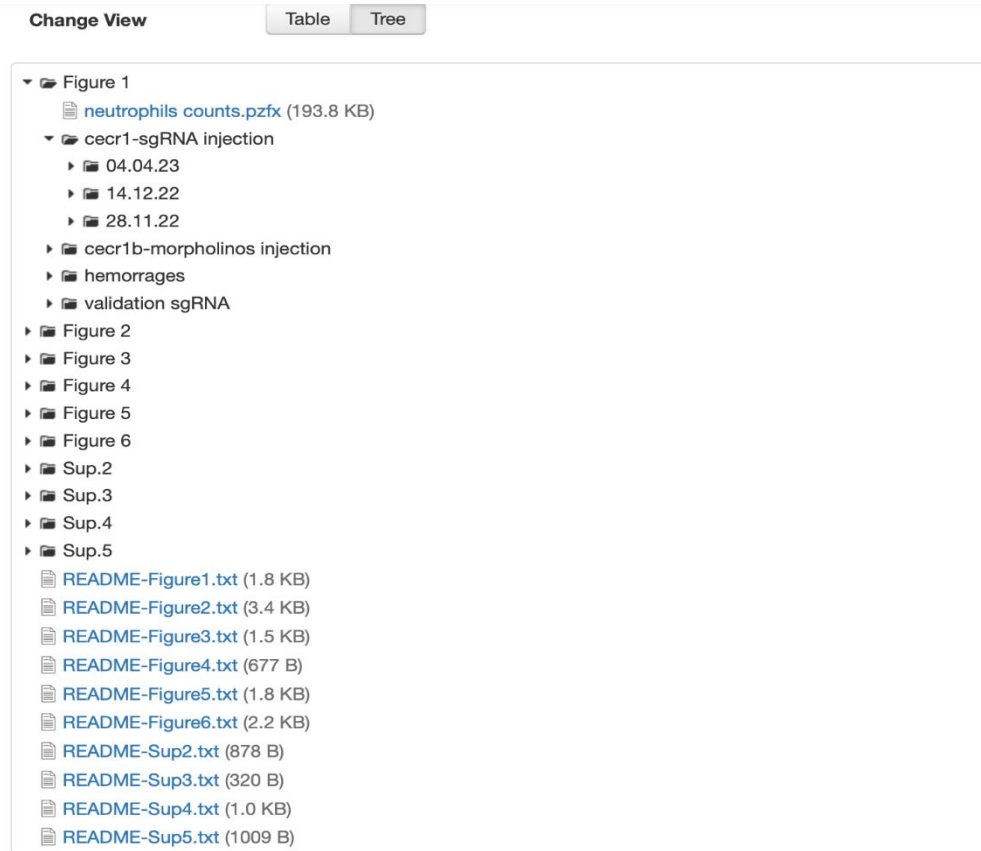
THE STRUCTURE: ONE OR MANY DATASETS?

Secondly, think about the structure of your data: indeed, you can choose to upload data in a single dataset or in many datasets, this will depend on **the complexity you will give to your data files**.

- In a **single dataset** you can upload as many files as you want (within a limit of **5 GB per file**). Anyhow, remember that **the more files you upload, the more you should describe them accurately**: name them in an intuitive and logical way, and make sure you upload a file list and/or a [README file](#) describing data collection methodology amongst other things, this way others will fully understand your data and more easily navigate your files.
- If you are uploading **many files with more structured data related to a scientific publication** you can choose between two options.
 - o The first option is to upload **a compressed ZIP folder inside a single dataset**. Folders can

contain a maximum of **1.000 files**. Dataverse allows to visualize **folders' tree structure** (made of folders and sub-folders), as you can see in the image below

Given that your dataset's structure may be quite elaborated, think carefully about how



to structure the README files in your dataset: you can either create a single and all-comprehensive README file for the entire dataset, or a README file for every folder within your compressed folder. Either way, make sure that you upload the README file(s) at the top level of your compressed folder and that you provide a precise and clear description of the structure of your data files and of the methodology used for collecting and elaborating data.

- The second option for uploading many data files related to a scientific publication is to **divide the structured data files into multiple datasets**. Inside your dataverse, create a **dataverse named 'Replication data for: "Title of your publication"', containing the datasets of the structured data related to the publication**. For instance, if your data is related to a book, you can create a dataset for each chapter within the dataverse *'Replication data for: "Title of the book"'*; if your data is the result of many experiments, you can create a dataset for each experiment within the dataverse *'Replication data for: "Title of your publication"'*; if your data is of different kinds (tables, figures, etc.), you can create a dataset for each type of data within the dataverse *'Replication data for: "Title of your publication"'*; and so on. As mentioned in the previous point, do not forget to upload a README file for each dataset.

Anyhow, you are required to **adequately and uniformly name all the datasets** which you create within

a *'Replication data'* dataverse, as explained in the following paragraph.

THE DATASET: NAMING AND METADATA

When you upload data in your personal dataverse or in a project dataverse, in a single dataset – with many files or with a zip folder – or in many datasets, at the dataset level the basic rules to follow for naming and for metadata are the same.

When naming a single dataset for a publication, use this title: ***'Replication data for: "Title of your publication"'***. If you have created a dataverse containing many datasets which all relate to the same publication, **make sure that the naming of the dataverse and of the datasets conveys the relationship between them**. In this case, the dataverse should be named *'Replication data for: "Title of your publication"'*, whilst the datasets should be entitled as *'Replication data for chapter X/experiment X/tables/figures/etc. of: "Title of your publication"'*. Then, you can use the 'Description' field in the datasets' metadata to explain the basic features and the scope of each dataset as well as how datasets are related to each other and to the same scientific publication. Please, **avoid giving partial** (e.g. only the authors' name/only the publication title) **citation information in the 'Description' field**.

Dataverse provides dedicated [metadata](#) sections for inserting citation information: these are the 'Author' field and, importantly, the 'Related publication' field in the 'Citation metadata' section. **The 'Related publication' field is not mandatory, yet we strongly recommend that you provide the full citation of your publication and its DOI** (if available). If you do not have such information (for instance, because the related publication is still under review or submission), please read the following paragraph for further information. Finally, remember that a single dataset can be related to multiple publications (e.g. two articles in two different journals, or two chapters in the same book) and you should enlist them all. This is why it is important to **provide metadata in the 'Related publication' field for all your datasets**: if you have a single dataset for one publication, or if you have multiple datasets for one publication, or even if you have multiple publications for one dataset do not forget to link all of them to their related publication.

Moreover, at the dataset level, ensure that you comply with all the mandatory metadata and with other significant metadata, even if they are not strictly connected to the related publication (such as keywords, funding information, date of collection, data type, eventual software, data source, and the domain-specific metadata), as indicated in [Use Data@UNIMI page](#).

THE RELATED PUBLICATION: IS IT PUBLISHED, UNPUBLISHED OR A PRE-PRINT?

Subsequently, you should consider **what kind of publication is associated with your data, as it will be reflected in the metadata of your dataset(s)**.

1. A PUBLISHED PUBLICATION

If your dataset is associated to a scientific publication which is already published in an academic journal, simply make sure that your publication is appropriately cited in the **‘Related publication’ field of your dataset’s metadata**. Please, ensure that you fill in the metadata with the full citation of the publication with its DOI, so that by acceding to the dataset, users can click on the link and accede to the publication too. Of course, this also means that your dataset will be entitled *‘Replication data for: “Title of your scientific publication”’*, as explained in the instructions in the section above.

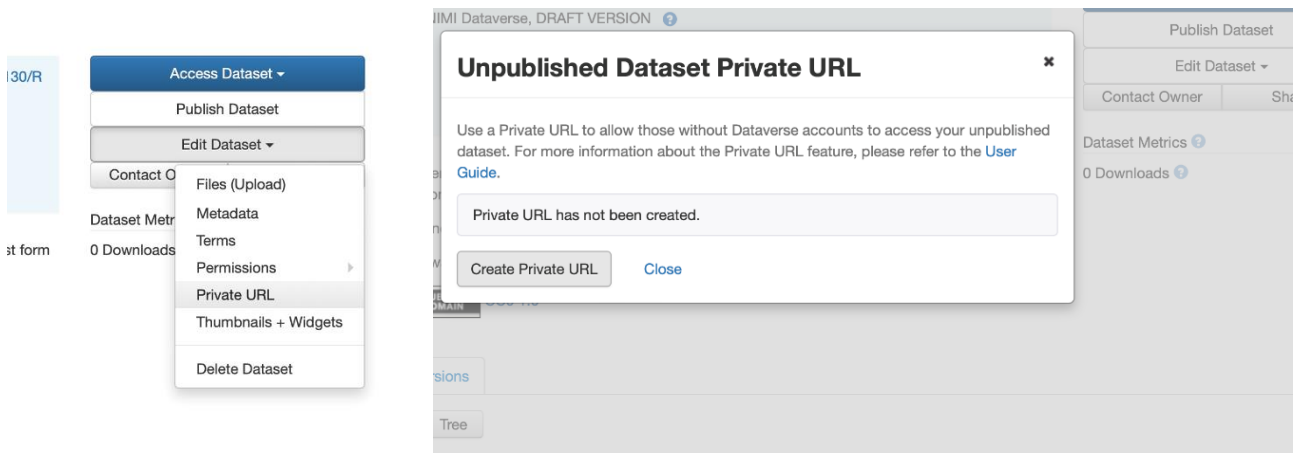
Indeed, if you have created more datasets for a publication – as in previous paragraph – be sure that you fill the ‘Related publication’ field of all your datasets’ metadata with the full citation of the published publication.

2. AN UNPUBLISHED PUBLICATION

If you are uploading data into a dataset which is related to a **scientific publication under submission or review**, you can choose between two options:

- On one hand, you can decide to **publish your dataset whilst keeping accessible your files only under request**. Thus, you can share the DOI of your dataset and grant the specific access to your data files only to other members of your team or other collaborators, or even the reviewers; just remember that, in this case, **only those who have a Data@UNIMI account can ask for your permission to access the files**. Besides, we suggest that you fill in the ‘Terms’ field of the dataset with the statement that the files are restricted because the publication they are related to is under submission/review, and that they will be accessible, accordingly to the chosen license, after the publishing of the related publication. This is a suitable option also in the case of embargoes, since you can indicate in the ‘Terms’ field that your data will be restricted until the end of the embargo. For further information, please [check Data@UNIMI License Management](#).
- On the other hand, you can **keep your dataset related to an unpublished publication in draft form while it is unpublished** and allow reviewers (and all those without a Data@UNIMI account) to access your dataset by using the **‘private URL’** function. This is necessary because an unpublished dataset is given a DOI, but the DOI is automatically activated when the dataset is published; thus, unpublished datasets can be shared only with the ‘private URL’. To obtain it, click on the ‘Edit dataset’ drop-down menu in the sidebar of your dataset, then click on the ‘private URL’ option and then, in the new window which should appear, click on the ‘Create

private URL' button, as in the images below.



In both cases, do not forget to fill in the 'Related publication' field of your dataset's metadata with the citation of your publication, indicating in the 'Notes' field that the related publication is still unpublished due to submission or revision.

Moreover, **as soon as the related publication is published, you are required to make accessible and publish your dataset**: follow the instructions on the point above (PUBLISHED PUBLICATION) to publish and share your dataset. Relevantly, we strongly recommend updating the 'Related publication' field with the full citation of the published publication and its DOI.

3. PRE-PRINTS

As in above cases, when uploading data related to a pre-print, it is vital to distinguish between published and unpublished pre-prints.

- **Published pre-prints should be treated in the same way as published scientific publications.** Indeed, a 'published pre-print' is any kind of pre-print which has been uploaded and shared – and is thus publicly available - on a publishing server (e.g. pre-print archiving repositories, or other archiving platforms) which should assign a DOI to your pre-print. Clearly, before publishing your pre-print, you should ensure that you are allowed to do it by checking your funder, publisher, and/or journal open access and pre-print policies (here is a [helpful tool](#)). Once you have checked whether you can publish your pre-print and once you have uploaded and shared it to the publishing platform, you can consider it as a published scientific publication as you upload its related data into Dataverse. Thus, your dataset will be entitled as *'Replication data for: "Title of your pre-print"'*. Make sure you provide the dataset's metadata in the '**Related publication**' field, including the full citation of the pre-print that the data uploaded is related to, its DOI and the indication that it is a pre-print. For further information, follow the instructions set out above in the first point (PUBLISHED PUBLICATION). If the published pre-print is, later, followed by a published scientific publication, please, do read the final instructions below.

- **Unpublished pre-prints** can apply to multiple situations: your funder, publisher, and/or journal open access and pre-print policies do not allow you to publicly share your pre-print; your manuscript is under submission, and you are not willing to share the pre-print; your manuscript is not submitted yet; etc. In these cases, data related to an unpublished pre-print can be uploaded and shared on Data@UNIMI on the condition that you clearly report that your **data is referred to an unpublished pre-print in the 'Notes' field**. Similarly, your dataset should be entitled as *'Replication data for: 'Title of your pre-print" pre-print version'*. Please remember that you are required to **promptly update the 'Notes' metadata if the condition of the manuscript/article changes**: from unpublished pre-print to submission, from submission to review. In fact, once the revision process is completed and the manuscript is published the reference of the dataset changes, and you should consider the following instructions.

When the pre-print (both published and unpublished) that your dataset is related to leads to a published scientific publication, you are required to update your dataset as follows:

- If the dataset related to the pre-print remains unmodified or **without substantial changes** and the pre-print becomes a published scientific publication, you can simply **update your dataset's metadata**. If this is the case, edit the dataset's title as *'Replication data for: "New title of your publication"'*. By so doing, the "history" of the publications related to your data will be clear, thus the new citation of the published scientific publication will replace the citation of the pre-print in the 'Related publication' field. Alternatively, you can simply add a second citation, that of the scientific publication, and leave the previous citation of the pre-print. To add another space in the referred metadata, you should click on the + button as shown in the image below.

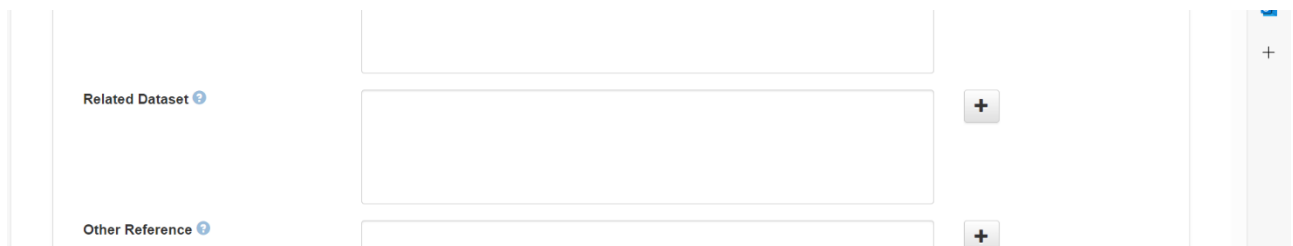
The image shows a user interface for adding related publications. It features a section titled "Related Publication" with a plus icon. Below this, there are two rows of input fields. Each row starts with a "Citation" field. The first row also includes an "Identifier Type" dropdown menu (currently set to "Select..."), an "Identifier" field, and a "URL" field (containing "https://"). To the right of each row are three buttons: a plus sign (+), a minus sign (-), and an "Add" button. The second row only shows the "Citation" and "Identifier" fields and their respective buttons.

- If during the review of the manuscript/article you are asked to **apply significant changes to your data** (and therefore add/remove/modify files of your dataset) you should **consider creating a new dataset with the reviewed data**. In this situation, you will have two datasets: one dataset will be entitled *'Replication data for: "Title of your publication"'* and will contain

the reviewed data and the full citation of the scientific publication in the 'Related publication' field. The second dataset will be entitled '*Replication data for: "Title of your publication" pre-print version*' and will contain the unreviewed data and the full citation of the published pre-print in the 'Related publication' field (or the indication of the unpublished pre-print in the 'Notes' field). Most importantly, make sure that **the new dataset with reviewed data is linked to the old one containing the unreviewed data related to the pre-print**. For details on how to link different datasets, please read the paragraph below.

HOW TO LINK DIFFERENT DATASETS

If necessary, you can link different datasets to each other. This can be useful in the above-mentioned situation, as well as in the case of datasets related to different scientific publications but referred to the same subject. **The link between datasets can be obtained by compiling the 'Related dataset' field in the datasets' metadata**, as shown in the image below



The image shows a portion of a web form. On the left, there are two labels: 'Related Dataset' and 'Other Reference', each followed by a small blue circular icon with a question mark. To the right of these labels are three empty text input fields stacked vertically. To the right of the first and second input fields are small grey square buttons with a white plus sign. To the right of the third input field is a vertical grey bar with a white plus sign at the top.

The 'related dataset' field can be freely filled in by the author with the necessary information. As such, you can insert all the information necessary for ensuring that the datasets are exhaustively described in relation to each other. For example, you can insert a dataset's title and the DOI. Furthermore, to further motivate why you are linking two datasets, you can also indicate which dataset is related to the pre-print and which dataset is related to the final publication, and, if you prefer, you can elaborate on the content differences between the two datasets (e.g. the final dataset has 2 more files, or, in the second dataset I have reviewed the data in table 1).

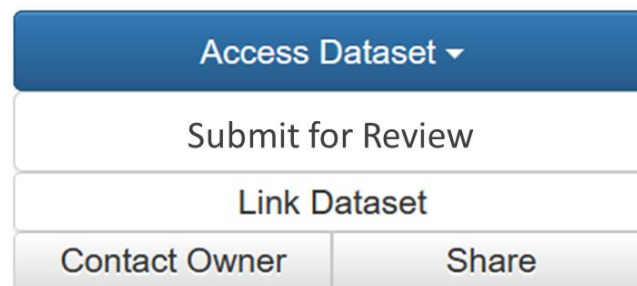
Moreover, you can indicate multiple related datasets by clicking on the + button next to the referred metadata, as in the image above.

Finally, it is also possible to link your entire Data@UNIMI to a dataset that exists outside of your Dataverse collection, so it appears in your dataverse collection's list of contents without actually being in that dataverse collection; vice versa, you can link other users' datasets to your dataverse collection. Indeed, it can be a useful tool for datasets related to scientific publications written by many researchers or produced by a research project that you want to link to your personal dataverse. For further information, consult Data@UNIMI [User guide](#).

LAST NOTICE

Please, keep in mind that once your dataset is unpublished you are able to make all the necessary changes. Once you have defined the structure of your dataverse, created your dataset(s), exhaustively filled in metadata sections, uploaded data files and documentation, defined license and terms of access to your dataset(s) and files, and checked the quality of your data, you should submit your dataset(s) for publication.

Just click on 'Submit for review' button in the left menù of your dataset, so that Data@UNIMI support staff will automatically receive a notification, review your dataset and, eventually, publish it.



Once your dataset(s) is published, you will still be able to edit it (description, number of files, related publication, title, and so on), creating an updated version, which will automatically replace the previous one. New versions can consist in a **minor version** (e.g. from version 1.0 to version 1.1) or in a **major version** (e.g. from version 1.1 to version 2.0). Data@UNIMI will keep track of the history of the changes made and of the new versions.

FURTHER HELP

- Didn't find what you are looking for? Check the [guide](#) or the UNIMI research data management [website](#).
- Still have some questions? Contact [Data@UNIMI support](mailto:Data@UNIMI).