

README for Data Deposit – Template Guide

Why do I need a README file for my data deposit?

A README is a text-based file that includes information about the content and structure of the files it accompanies. A README for data deposit will include information about the dataset, including administrative information, file naming conventions, folder organization, the relationship between files, research methodology, data processes, and information about data access and reuse. When depositing data in a repository, a README provides detailed contextual information that allows users to navigate, understand, and reuse the data.

I'm already including information about my dataset as metadata in the data repository - do I still need a README?

Yes, it is still a good idea to include a README with your dataset. Some of the information in your README may be duplicated elsewhere, such as in the repository metadata or in your data documentation/data files. This is not a problem, as these things serve different purposes. Metadata allows users to search for and find your dataset and to determine if the dataset meets their needs. Including a README with your data deposit means that when users download the dataset they will have a local copy of this important contextual information in one, easy-to-identify file.

How do I include a README in my deposit?

- Create a README by downloading the [README for Data Deposit Template](#) or by opening a blank text document.
- Save your file in the top-level folder of your dataset so it can be easily found by users. The template is formatted so that it can be saved as a simple text file (.txt) or as a Markdown (.md) file. Saving your README as a .md file may be useful if you're familiar with Markdown and you're using a data repository that supports Markdown formatting preview.
- Make sure the file name includes "README" so it can be easily identified. Depending on the system you're using, you may want to add a leading 0 in the file name (e.g., "0_README") to ensure it shows up first in your list of files. We recommend this practice if you're depositing your data in U of T Dataverse.

Use the guidance below to understand what information to include in each section. Note that some sections may not be relevant for your dataset.



README file sections and fields

Quick Links

Section: General information	3
Field: Title	3
Field: Author(s)	3
Field: Dataset contact	3
Field: Description	3
Field: Datasets cited	4
Field: Data collection date/range	4
Field: Data collection geographic location.....	4
Field: Language	4
Field: Dataset contributors	4
Field: Funding	5
Field: Keyword(s).....	5
Section: Sharing and access information	5
Field: Licenses	5
Field: Suggested dataset citation	5
Field: Terms of access.....	6
Section: File information	6
Field: File name(s) and extension(s).....	6
Field: File naming conventions.....	7
Section: Methodological information.....	7
Field: Methods used to collect or generate data	7
Field: Methods used for data processing	7
Field: Quality-assurance procedures performed on the data	7
Section: Versioning.....	7
Field: Version number	7
Field: Version updates	8
Section: Additional information	8



Field: Data documentation	8
Field: Related publications	8
Field: Related datasets	8
Field: Notes.....	9

Section: General information

Field: Title

What to include: The title of the dataset, as listed in the repository.

Additional context and considerations: Give your dataset a clear and descriptive name that allows users to understand what the data is. Avoid vague language (e.g., “Thesis data”; “TDM project research data”) that would make it difficult for future users to search for or understand what this dataset is.

Field: Author(s)

What to include: The name and contact information of all authors. This should include anyone included in the citation for the dataset.

Additional context and considerations: If there are people who have contributed to the dataset in some way (e.g., data collection, analysis, data cleaning, etc.) but are not listed as authors, you can include them in the [Dataset contributor](#) section of the README file.

Field: Dataset contact

What to include: The name and contact information for the person who can be contacted about the dataset.

Field: Description

What to include: A summary describing the purpose, nature, and scope of the dataset.

Additional context and considerations: A good description should describe the content and context of the dataset and provide users with enough information to determine if the dataset is relevant to their needs. This description will not be the same as your journal article abstract - while they may contain some overlapping information, your dataset description should focus on describing your data. Some elements you may want to include in your description include:

- Dates of collection
- Geographical location



- Type of data
- Other dataset(s) used in creating the dataset
- Overview of the method of analysis
- Data processing tools used
- Other relevant information

Field: Datasets cited

What to include: The citation(s) for any third-party dataset(s) used to create this dataset.

Additional context and considerations: Including this information ensures that the original dataset creators receive proper credit for their work and enhances the transparency of your dataset. Note that if you're reusing data from other sources, you need to abide by their licensing terms and/or seek permission to use and share derivatives created with their data. Some licenses require that anything derived from the data must be shared under the same open license, which will determine what license is most appropriate to apply to your dataset. Please [contact us](#) if you have questions about licensing.

Field: Data collection date/range

What to include: The date the data was collected. This might be a single date or a date range.

Additional context and considerations: You can also note if your dataset contains files representing different dates of collection/creation.

Field: Data collection geographic location

What to include: The latitude and longitude, or city/region, province/state, and country where the data was collected.

Additional context and considerations: You can also note if your dataset contains files representing different or multiple geographies.

Field: Language

What to include: The language used in the dataset.

Field: Dataset contributors

What to include: A list of all individuals and organizations that contributed to the dataset's creation. This may include people who are not listed as authors, but who supported the creation of the dataset through data collection, analysis, curation, etc.

Additional context and considerations: Including this information enhances the transparency of your dataset and allows you to credit individuals for their work.

Field: Funding

What to include: Information about the funding sources supporting the creation of this dataset.

Additional context and considerations: Including this information increases the transparency of your dataset. This may help others understand the context in which your research was conducted and help the funder, publisher, institution, or project partners track outputs.

Field: Keyword(s)

What to include: Keyword(s) that represent the research areas covered by the dataset.

Additional context and considerations: Including keywords can help users search for and discover your data and can make it easier to find other datasets that use the same terms. Separate each keyword with a comma.

Section: Sharing and access information

Field: Licenses

What to include: The license applied to this dataset.

Additional context and considerations: Including a license lets users know how to use your data, under what conditions, and what they can do with any derivative products they create using your data. Common examples of licenses that can be applied to data include [Creative Commons](#) and [Open Data Commons](#).

Some repositories (like U of T Dataverse) require you to apply a license to your dataset. The license you include in your README must match the license in the data repository.

If your data includes or was created using third-party data, it's important to be aware of any associated licenses. Some licenses require that anything derived from the data must be shared under the same open license, which will determine what license is most appropriate to apply to your dataset.

For more information about research data licensing, see:

- [How to License Research Data](#)
- [Creative Commons License Chooser](#)

Field: Suggested dataset citation

What to include: The preferred citation that others should use when citing the dataset.

Additional context and considerations: Including a preferred citation lets users know how they should credit you if they use your dataset. If you're depositing the dataset into a data repository, the repository may provide a recommended citation format. It is also helpful to note the citation style so users can easily identify and update it if needed. For more information on how to cite data go to [Citing Data](#) or [contact us](#).

Field: Terms of access

What to include: Any information about the access conditions associated with the dataset. This can include who can access and use the data, their responsibilities towards the data, and whether they need to sign an agreement (e.g., a data use agreement) before accessing the data.

Additional context and considerations: Including access conditions ensures users are aware of any conditions, requirements, and responsibilities associated with accessing and using this data. For example, if your dataset contains sensitive or confidential data, you may decide that potential users must complete a data use agreement that outlines how the data can and cannot be used, and their responsibilities with regards to safely working with and storing the data. Please note that if you are working with human subject data - including de-identified data - you will need consent from your participants and approval for the Research Ethics Board to share data. [Contact us](#) for more information.

Section: File information

Field: File name(s) and extension(s)

What to include: A list of all the files in the dataset, including a short description of the file.

Additional context and considerations: Providing a description of file contents will help users navigate and identify your files and can help users ensure they have the full dataset. You may want to include the following information in your description:

- Type of file (e.g., is it data, documentation, code, etc.)
- Brief description of content
- Any important relationships with other files in your dataset. This might include:
 - Relationships between data documentation (e.g., a codebook or data dictionary) and the file it describes
 - Relationships between code and the file it runs on
 - Relationships between files in a series (e.g., a series of monthly data files)

Example: TDMP_Survey_2017_codebook_EN.txt: English codebook for the 2017 survey data (TDMP_SurveyData_2017.csv).

Field: File naming conventions

What to include: The file naming conventions used to name your files.

Additional context and considerations: Including your file naming conventions will help users identify relevant files and understand what they are without opening them. This is particularly useful if your dataset contains many files. For more information about creating a file naming convention, go to [Organizing Your Data \(File naming\)](#).

Section: Methodological information

Field: Methods used to collect or generate data

What to include: A brief description of methods used for collecting or generating data. Include links to relevant publications or documentation with experimental design or protocols used in data collection.

Field: Methods used for data processing

What to include: Information about any software or instruments used to process the data. This may include the software name, version, system requirements and developer.

Additional context and considerations: Including this information allows users to understand how you processed/analyzed the data, increases transparency, and supports reproducibility and replicability. This may include the name and version of any tools and/or software used.

Field: Quality-assurance procedures performed on the data

What to include: The description of any quality-assurance procedures performed on the data.

Additional context and considerations: Some examples of quality-assurance procedures may include protocols for inputting/processing data, data validation, methods for cross-checking and verifying data, etc. Including this information supports data transparency and increases the trustworthiness of your data.

Section: Versioning

Field: Version number

What to include: The version number of the dataset.

Additional context and considerations: Use digits (e.g., V2.0; V3.0 etc.) to indicate major version updates and decimals for minor updates (e.g., V2.1, V2.2, etc.). Note that if you update your dataset, you will also need to update the version number in your README.

Field: Version updates

What to include: An overview of what has changed between versions.

Additional context and considerations: Including a log of version updates allows users to see the difference between versions and understand what has changed. You can use a numbered list to outline the changes one-by-one.

Section: Additional information

Field: Data documentation

What to include: Any additional information required to understand, interpret, and reuse your data (e.g., explanation of the data variables in data files).

Additional context and considerations: If your dataset already contains data documentation files you may not need to include this information in your README. However, these files should be included and described in the “File Information” section.

If your dataset does not include data documentation files, you may want to include this information in your README so users can properly understand, interpret, and reuse your data. This may include a description of your data variables, the number of variables and rows in a spreadsheet, coding structures, and any other relevant information.

For more information and examples of data documentation, go to [Data Documentation](#).

Field: Related publications

What to include: Citations and links to publications and outputs related to the dataset.

Additional context and considerations: Including links and persistent identifiers (e.g. DOI) for related publications can enhance the discoverability of your work and showcase the application of your dataset.

Field: Related datasets

What to include: Citations and links to any related datasets.

Additional context and considerations: This may include datasets that are part of the same series (e.g., data that is collected annually) or project. Including links to related datasets can enhance the discoverability of your work and allow users to discover other relevant datasets. It's strongly recommended to include a persistent identifier (PID), such as a [DOI](#), in your citation.

Field: Notes

What to include: Any additional notes that may be helpful for understanding and using the datasets.

Other Resources

Some content in this document was based on and modified from:

- [Guide to writing “readme” style metadata – Cornell Data Services](#)
- [Preparing Your Data - Federated Research Data Repository](#)
- [UA Research Data Repository Policies & Documentation](#)

For more information visit the [University of Toronto Libraries RDM website](#) or [contact us](#).

Last updated: 2024-11-14

README for Data Deposit – Template Guide © 2024 by University of Toronto Libraries RDM Committee is licensed under CC BY-NC 4.0