Creating a README for Your Dataset

Quick Guide, Version 1, November 2026

What is a README?

A README file is an important document that accompanies your dataset. It provides key project information to ensure your project can be correctly understood, reanalyzed, and reused.

It is good practice to have a README file because it helps preserve the usability of your data for the long-term. It will also continue to describe the dataset even after it has been separated from its original context. Here is a simple template with all minimal elements that we recommend.

How do I create a README?

Creation considerations

We recommend creating your README file at the start of your project. Update your README as soon as project changes are made to prevent information loss.

Any text editor can be used to create a README file. Markdown and TXT are commonly used plain text formats because they're lightweight and non-proprietary (won't rely on specific software), making them durable and good for information preservation.

Store your README in the root directory of your project folder to make the file easily accessible.

Stylistic considerations

Here are some considerations for good README data documentation:

- Be clear and specific when writing about the elements of your project, in addition to using descriptive titles throughout the README sections, including the file itself
- Add relevant dates (e.g. when the README was created and updated) using a standardized format like YYYY-MM-DD or YYYYMMDD (ISO 8601)
- Define acronyms, abbreviations, codes, and/or jargon if used
- Address any research limitations, file quantities, and file versions

Minimum requirements

There are 9 elements that should be included at minimum. These will help users understand the basic details and evaluate if the dataset(s) is relevant to them.

Metadata element	Definition
Author/ creator	The person(s), corporate body(ies), or agency(ies) responsible for creating the dataset, listed as they appear in citations. You may include ORCIDs to properly identify and attribute researchers.
Title	The full, descriptive title of the dataset(s) to clearly convey the subject matter and scope of the data so a user can assess the relevance without opening any files.
Description/ summary	A narrative summary of the dataset's purpose, content, and scope: why the data were collected, what is captured, how the data are structured, and what they might be used for. Write for a broad academic audience to help unfamiliar researchers determine if the data fits their information needs.
Contact information	The name, institutional affiliation, email address, and institutional or project-level address (optional) of the person(s) responsible for answering questions about the dataset.
License information	The full license name and version for how the dataset may be used and shared. Describe any restrictions on use if a standard open license doesn't apply.
DOI/ persistent identifier	A DOI or other persistent identifier assigned to the dataset, to reliably located and cite. If a DOI is not yet assigned at deposit, note it as pending and update the README once available.
Research methodology	A summary of the research methods used to generate the dataset: study design, data collection instruments or procedures, and key processing steps applied to the raw data. This provides users with sufficient context to assess the data's reliability and understand the requirements to replicate the data.
Related publications	Full citations (and DOIs) of journal articles, reports, or other scholarly works that are directly based on or make use of this dataset. This gives analytical and methodological context to the data, along with appropriate credit and citations to the researchers who generated the data.
Software/ tools used	The names and versions of any software, programming languages, specialized tools, or operative system needed to open, read, or reproduce that dataset or its analyses. This information is essential for reproducibility and helps users determine if the data is usable based on the available tools.

Additional resources

Government of Canada – Research Data Management. FAQ: Data retention, deposit, and availability. Link to resource.

Barsky, E., Cevik, Z., Cooper, A., Eber, A., Hu, B., Lui, L. H., & Pratt, I. (2026). README for research data deposits in Borealis: minimum and enhanced metadata elements (Version V1). Borealis. https://doi.org/10.5683/SP3/CJ5X5O

Questions? Contact research.data@ubc.ca

Guide designed by Vanessa Choy & Eugene Barsky, UBC Library.

CC BY