Creating a README for your dataset

A README is portable, durable way to provide information to other researchers about how to use your dataset

What is a README?

A readme is a guide to your dataset and is usually a plain text file to maximize its usability and long-term preservation potential. The purpose of a readme is to assist other researchers to understand your dataset, its contents, provenance, licensing and how interact with it. Readme files are generally named README, readme.txt or read-me.md and are included as component of a dataset.

What does a README do?

A readme complements but does not replace the metadata that data repositories (e.g. Dataverse or FRDR) ask you to provide when you deposit your data. The best practice is to record information in both the repository's metadata and the readme. The repository's metadata will support findability within and between data repositories while the readme is portable and continues to describe the dataset after it has been separated from its original context. In all cases, you should use any conventions appropriate to your discipline to record the information about your dataset.

How do I Create a REAMDE?

The Content

Core elements of any readme include:

⇒ Contact information for the researcher(s)
⇒ The use license for your data (unless that is included in a separate file)
⇒ The context of your data collection (the goal of your research)
⇒ Your data collection methods (protocols, sampling, instruments, coverage, etc.)
⇒ The structure of files
⇒ Naming conventions for files, if applicable
⇒ Your sources used
⇒ Your quality assurance work (data validation, checking)
⇒ Any data manipulations or modifications
⇒ Data confidentiality and permissions
⇒ The names of labels and variables
⇒ Explanations of codes and classifications

Cornell University's Guide to writing "readme" style metadata provides a thorough description of the content that you should consider including in your readme file. It also provides a template readme.txt file for you to work from.

The Style

How you write your readme is as important as the information you include. Always remember to be as clear as possible. The following are some best practices related to data documentation:

⇒ Don't use jargon
⇒ Define terms and acronyms
⇒ State limitations
⇒ Use descriptive titles
⇒ Be specific and quantify
⇒ Use appropriate and specific keywords
⇒ Make it machine readable (avoid symbols)

The Process

Document your work as you go, so you don't lose track of any details. If you wait until the end of your project, you might already have lost or forgotten valuable information.

You can create a readme using any text editor (e.g. TextEdit, Notepad++, Atom.io, Sublime Text) or word processor (e.g. Word, LibreOffice). Save your readme as UTF-8 encoded text. Using plain text helps preserve your information because it relies on durable, open standards rather than proprietary formats. If you're using GitHub as a repository for your data, your readme should be written using Markdown syntax (readme.md).

Additional Resources

- UBC Library DataGuide. https://bit.ly/3eTjHVP
- Cornell University. Guide to writing "readme" style metadata. https://bit.ly/2W4t9xa
- The Graduate Institute Geneva. Readme.txt. https://bit.ly/3aH6AUx
- GitHub. Mastering Markdown. https://bit.ly/2y7c4dZ

Questions? Contact research.data@ubc.ca