

README file for datasets - Best practices and template

Vademecum	1
1. Definition	1
2. Best practices	1
Template.....	3
References	4

Vademecum

1. Definition

A `README` file is a text file that describes a dataset in a directory and explains its reuse. As a form of first-level documentation (easy to write, easy to read), it informs about the content of the other files present in a directory or archive, ex. data, code or software, so that they are correctly interpreted and used.

The `README` file must be understandable by yourself and others in the future. It should be written in a simple (plain) text file, in human-readable format, and contain information about the dataset structure, its content, and its function. There is not one right way to structure a good `README` file, but some best practices exist.

2. Best practices¹

When writing the `README` file:

In general, the earlier you start writing your `README`, the better. At the very least, try to have a `README` file describing the dataset (data, code, etc.) at the time of its publication or archiving.

Structure of the `README` file:

- Name the file simply “`README`”
- Locate the `README` file at the root of the project rather than in a sub-folder
- If needed, create one `README` file per dataset / folder and use the same format for all
- Write it in plain text or markdown: `README.txt`, `README.md`, or simply `README`
- Avoid proprietary formats such as MS Word
- Use of blank lines or dashes to separate the document into paragraphs
- Paragraphs order and title sections are not fixed, simply make it understandable
- Use bullet points or ordered lists instead of long paragraphs
- If needed, create one `README` file per dataset / folder

¹ These lists are non-exhaustive.

Content to include in the README file (strongly recommended elements marked with an *):

- Description of how the content of the directory can be practically used by others, including files naming convention(s), folder(s) structure, contents, and use instructions
- License under which the dataset is released * for reuse by others
- Sources for such standards of terms, labels and keywords, known conventions, taxonomies, and vocabularies of your scientific community
- Dataset name *
- A *General information* section containing: author(s) info *, date of collection (use format) *, geolocation data (use format), funding or sponsorship *
 - Prefer standard formats for dates (ISO 8601), countries (ISO 3166), languages (ISO 639), ...
- A *Sharing / Access information* or *License* section containing: licenses *, terms of use *, citation instructions *, links to related publications, datasets, other research output, URLs in repositories, persistent identifiers...
- A *Data and file(s) overview* or *Data* section containing: files and folders structure description *, file formats *, additional related data, original source if any, dataset version, update description/changelog
- A *Methodological info* or *Preparation* section and *Acknowledgment* section containing: links to publications used as a base for methods, methods for processing data *, necessary software *, necessary instruments, software, hardware and version numbers, parameters or calibration data *, quality assurance process applied, people involved in experiments, surveys, processing, analysis, etc.
- A data-specific section containing information such as: data format *, data dictionary, or data codebook * (variables, units of measures, codes or symbols, abbreviations, semantics, depiction), applied standards

Template

Reorder or adapt (recommended fields are marked with an *)

Download the ready-to-use template here: go.epfl.ch/rdm-readme-template

#####

DATASET TITLE *

#####

General information or Introduction section

- author(s) info (name, affiliation, persistent id) *
- date of collection (use format) *
- geolocation data (use format)
- funding or sponsorship info *

Sharing / Access information or License section

- licenses *
- terms of use *
- citation instructions *
- links to related publications
- links to other research outputs and datasets
- URL in repository
- persistent identifiers

Data and file(s) overview or Data section

- files and folders structure description *
- file formats *
- additional related data
- original source if any
- dataset version, update description/changelog

Methodological info or Preparation section and acknowledgment section

- link to publications used as a base for methods
- methods for processing data *
- technical requirements: necessary instruments and software, hardware and version numbers, parameters or calibration data *
- quality assurance process applied
- people involved in experiments, surveys, processing, analysis etc

Data specific info

- data format *
- data dictionary or data codebook * (variables, units of measures, codes or symbols, abbreviations, semantics, depiction)
- applied standards

References

- 4TU Centre for Research Data. (2017). Guidelines for creating a README file. Retrieved from https://researchdata.4tu.nl/fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf
- Cornell University Readme Template. (n.d.). Retrieved November 13, 2019, from <https://cornell.app.box.com/v/ReadmeTemplate> Desmond, Y.
- Guidelines for Creating a README File. 3. Guide to writing “readme” style metadata | Research Data Management Service Group. (n.d.). Retrieved November 13, 2019, from <https://data.research.cornell.edu/content/readme>
- Library Carpentry: FAIR Data and Software. (n.d.). Retrieved November 13, 2019, from <https://librarycarpentry.org/lc-fair-research/aio.html> Pasquier, G. (n.d.).
- LibGuides: Research Data Management: README.txt. (n.d.). Retrieved November 13, 2019, from <https://libguides.graduateinstitute.ch/rdm/readme> RDC Research Data Canada.
- Guidelines for the deposit and preservation of research data in Canada. (2015). Retrieved from <https://www.rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Canada-2015.pdf>
- README. (2019). In Wikipedia. Retrieved from <https://en.wikipedia.org/w/index.php?title=README&oldid=923233067>
- README Files | Harvard Biomedical Data Management. (n.d.). Retrieved November 13, 2019, from <https://datamanagement.hms.harvard.edu/readme-files>
- README Template—MSU Library | Montana State University. (n.d.). Retrieved November 13, 2019, from <https://www.lib.montana.edu/services/data/toolkit/readme/>
- shashvatshah9. A template to make good README.md for datasets. (n.d.). Retrieved November 13, 2019, from Gist website: <https://gist.github.com/shashvatshah9/5d587605cd087182ccffb46b6cf9e449>
- The Frictionless Data Field Guide. (n.d.). Retrieved November 13, 2019, from <https://frictionlessdata.io/docs/publish-faq/#readme>
- Make a README. (n.d.). Retrieved November 14, 2019, from Make a README website: <https://www.makeareadme.com>