

ReadMe File Preparation Guide

The *ReadMe* file contains contextual information about the deposited dataset and aims to ensure that the data can be correctly interpreted both by the author in the longer term, including after the end of the research project, and by other users who have been granted access to the dataset.

If there are industry-specific standards and requirements for the preparation of a *ReadMe* file, then they should be preferred over the content of the *ReadMe* file recommended here.

In cases where a dataset consists of several contextually closely related files of the same format or files grouped according to their content, it is permissible to create a single *ReadMe* file to describe the datasets. In other cases, it is recommended that you create a separate *ReadMe* file for each file in the dataset.

1. The name of the *ReadMe* file should contain information that:
 - allows it to be identified as *ReadMe* and distinguished from other deposited files – the name must include "readme" or "read-me" or a similar line;
 - clearly indicate which of the added datasets *ReadMe* applies to (if several datasets are deposited together).
2. *ReadMe* should be created as a plain text (.txt) or markdown (.md) format file – the use of proprietary formats such as MS Word should be avoided. The formatting of the text should be such that the information is understandable and easy to understand - not to write all the information in one paragraph, but to separate separate sections, leaving blank lines between them.
3. *ReadMe* text must use UTF-8 encoding.
4. If several datasets with separate *ReadMe* files are deposited together, then in their creation you need to stick to a uniform style and terminology.
5. The naming conventions and designations generally accepted in a particular scientific discipline (for example, for geological or geospatial objects, etc.) must be observed.
6. *ReadMe* complements, but in no way replaces, metadata, which must be entered in as much detail and accurately as possible during the deposit process of the data set.

Recommended content for a ReadMe

General information

1. A brief description of the study and the objectives to be achieved.
2. Brief description of deposited data and their structure.
3. The format of the file, unless it is obvious from the file name.
4. If the dataset consists of several interconnected files, then it is necessary to describe the nature of this obligation.
5. Name, name of the scientific institution represented, *ORCID* and e-mail:
 - a. the principal investigator (or the person responsible for collecting the data),
 - b. co-authors,
 - c. contact person (in case of questions).
6. Date or date range of data collection.
7. Geographical location of the origin of the data.
8. Date of creation of the file (version of the file).
9. Date of modification or addition of the file and nature of the changes made (version history of the file).
10. Keywords that describe the data by nature.
11. Language(s) used.

Metadata on data collection and processing methods

1. A description of the method(s) may also include a link to a resource describing the data collection, processing and quality control procedure.
2. Specific settings of measuring instruments during measurement.
3. Instrument calibration data.
4. Description of the data quality control procedure.
5. Encoding or notations in the data used to highlight low-quality/questionable measurements, jumpers (non-representative values) and missing values where such data are included in the dataset.
6. People who were involved in sampling(s), data processing, analysis.

Specific data information

1. A transcript of the abbreviations and abbreviations used in the data set, a transcript and explanation of the column names for tabular data.
2. Units of measurement(s).
3. Transcript of the designation used to identify the missing data.

Access/sharing

1. License for use of the data and other restrictions on working with the dataset.
2. Links to publications where the relevant dataset is cited or used.

If individual documentation (e.g. code book or detailed metadata) already contains some of the information suggested here, only the most relevant (supplementary) information can be included in the *ReadMe* file.

If the *ReadMe* file is created and regularly updated at the beginning of the project, it is easier to prepare its final version and avoid errors that may occur if all the information is collected only at the end of the project.