Research Data Management: Guide to writing “readme” type metadata

A readme file provides information about a data file and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data. Standards-based metadata is generally preferable, but where no appropriate standard exists, writing “readme” style metadata is an appropriate strategy.

- **Create one readme file for each data file, whenever possible.** It is also appropriate to describe a "dataset" that has multiple, related, identically formatted files, or files that are logically grouped together for use (e.g. a collection of Matlab scripts). When appropriate, also describe the file structure that holds the related data files.
- **Name the readme so that it is easily associated with the data file(s) it describes.**
- **Write your readme document as a plain text file**, avoiding proprietary formats such as MS Word whenever possible. Format the readme document so it is easy to understand (e.g. separate important pieces of information with blank lines, rather than having all the information in one long paragraph).
- **Format multiple readme files identically.** Present the information in the same order, using the same terminology.
- **Follow the scientific conventions for your discipline for taxonomic, geospatial and geologic names and keywords.** Whenever possible, use terms from standardized taxonomies and vocabularies

Recommended content

Recommended minimum content for data re-use is in **bold**.

Introductory information

1. **For each filename, a short description of what data it contains**
2. Format of the file if not obvious from the file name
3. If the data set includes multiple files that relate to one another, the relationship between the files or a description of the file structure that holds them
4. **Name/institution/address/email information for**
    - **Principal investigator (or person responsible for collecting the data)**
    - Associate or co-investigators

    - Contact person for questions
5. **Date of data collection (can be a single date, or a range)**
6. **Information about geographic location of data collection**
7. **Date that the file was created**
8. Date(s) that the file(s) was updated and the nature of the update(s), if applicable
9. Keywords used to describe the data topic
10. Language information

Methodological information

1. **Method description, links or references to publications or other documentation containing experimental design or protocols used in data collection**
2. Any instrument-specific information needed to understand or interpret the data
3. Standards and calibration information, if appropriate
4. Describe any quality-assurance procedures performed on the data
5. Definitions of codes or symbols used to note or characterize low quality/questionable/outliers that people should be aware of
6. People involved with sample collection, processing, analysis and/or submission

Data-specific information

1. **Full names and definitions (spell out abbreviated words) of column headings for tabular data**
2. **Units of measurement**
3. **Definitions for codes or symbols used to record missing data**
4. **Specialized formats or abbreviations used**

Sharing/Access information

1. Licences or restrictions placed on the data
2. Links to publications that cite or use the data
3. Links to publicly accessible locations of the data
4. Recommended citation for the data
5. Information about funding sources that supported the collection of the data

Acknowledgements

These guidelines have been adapted from Cornell University's Guide to writing "readme" style metadata: http://data.research.cornell.edu/content/readme