Guidelines for Writing Useful Readme Files

Before submitting data to the MSD-LIVE data repository you should write a "readme" file that can be uploaded along with the dataset itself. Readme files are meant to summarize the key features of the dataset in a short and digestible way. They are crucial to promoting data reuse and can save you a lot of time answering questions from potential downstream users who may access your dataset. This document provides a list of suggested content and best practices for your readme file. The suggestions and associated template follow the guidelines from Cornell's research data management group: https://data.research.cornell.edu/content/readme. You can access their readme template here: https://cornell.app.box.com/v/ReadmeTemplate.

Best Practices

- Create a separate readme file for each unique type of data file in your dataset. For example, if you had a dataset consisting of climate forcing files and model output files you would create two readme documents - one for each unique type of data. Clearly name each readme file so that it is clear which data type it is associated with (e.g., climate_forcing_files_readme.txt).
- Document important information in the readme as the research is progressing.
- Create the readme as a plain text (i.e., .txt or .md) file so that it can be opened on any system and does not require licensed software to be viewed. Save it in the root of the directory it describes.
- Avoid jargon and define acronyms, but also follow accepted scientific conventions for your discipline.
- Follow the 'Golden Rule': Put yourself in the shoes of a downstream user and think about what information you would want to know if accessing someone else's dataset.

Recommended General Content

- Descriptive title of the dataset, the same as the name you are giving to the record in MSD-LIVE
- Data location: A MSD-LIVE DOI or URL to the landing page
- Contact information for the data creator or curator and the project PI:
    - Name, Institution, Email
- Date on which the dataset was downloaded, generated, or committed to the repository
- Information about the geographic location where the data was collected
- Licenses or restrictions placed on the data
- Description of methods for data collection or generation, including links to the appropriate code repositories for the processing code when possible
- Description of methods used for data processing, including links to the appropriate code repositories for the processing code when possible
- Citations for the research papers in which data was generated or used

Recommended Data-Specific Content

- File name and a short description of what data each file or group of files contains
- Date the file was created
- A description of the file's basic data structure and the relationship between files if it is not obvious
- A list of all the variables included in each data file and their associated units (e.g., wind speed [m/s], temperature [°C], relative humidity [%], etc.)
- Definitions for codes used in the data, including symbols used to record missing data (e.g., NA)