README File Checklist

Make your dataset easy to use and understand

A good README file is like a map that helps others navigate your data. Use this checklist to make your README clear, complete, and user-friendly, enabling others to make the most of your hard work. This checklist provides general guidance but may need to be adjusted to fit the specific requirements of your research.

GENERAL INFORMATION

- [ ] **Title of Dataset*:** Clearly state the name of the dataset.
- [ ] **Author(s)/Contributor(s)*:** Include names, affiliations, corresponding author and its contact information.
- [ ] **Date of Creation/Release*:** Specify when the dataset was created or made publicly available.
- [ ] **Dataset Version:** Indicate the version number (e.g., v1.0).
- [ ] **DOI or Persistent Identifier*:** Provide a unique identifier if available (depend on the repository).
- [ ] **Funding Information*:** Mention any funding sources or grant numbers.
- [ ] **Acknowledgments:** List individuals or organizations that contributed.

DATASET OVERVIEW

- [ ] **Summary/Abstract*:** Provide a brief description of the dataset.
- [ ] **Purpose/Objective*:** Explain why the dataset was created and its intended use.
- [ ] **Directory Structure*:** Describe the organisation (list) of files and folders
- [ ] **Dataset Description*:** Describe clearly the content of dataset including folders and file with their file format
- [ ] **Data Dictionary:** Provide a data dictionary or table describing each variable/column.

METHODOLOGY

- [ ] **Ethical Approvals*:** Mention any IRB or ethical board approvals if applicable.
- [ ] **Data Collection Methods*:** Explain how the data was collected, including tools, devices and protocols.
- [ ] **Data Processing*:** Describe any cleaning, transformation, or preprocessing steps.
- [ ] **Data Analysis*:** Specify the software and their versions used to analyze the datasets.

USAGE AND ACCESS

- [ ] **Licence*:** Specify the licence under which the dataset is shared (e.g., CC BY, MIT).
- [ ] **Access Restrictions:** Indicate if there are restrictions (e.g., embargoes, required permissions).
- [ ] **Usage Instructions*:** Provide guidance on how to open and use the files.
- [ ] **Citing the Dataset*:** Include a citation format or example.
- [ ] **Related Publications/Protocol:** Cite any papers, articles, protocols or posters related to the dataset.

*These steps should always be included.

What would you want to know if you were encountering this dataset for the first time?

Think as a Consumer of your Data not the Producer!

Anticipate questions users might ask and address them upfront.
Remember, a well-documented dataset is more likely to be reused and cited!

Funding: MRC grant MR/X009726/1

THE UNIVERSITY of EDINBURGH

Contact us: *bio_rdm@ed.ac.uk*