Data management plan:

Type(s) of data:

Non-technical

SOURCE:

DOCUMENTATION:

SOFTWARE:

STANDARDIZATION:

SHARING:

COLLABORATION:

INPUTS:

REPOSITORIES:

END OF LIFE:

Technical

PROCESSING:

STORAGE:

FORMATS:

SECURITY:

Data management plan for:

Type(s) of data: What particular type of data is this plan covering, and what is the data used for? You can multiple forms for different types of data (e.g. data, software, registry data)

Non-technical

SOURCE: From where do you get data (new data collected, reused, or other)? What are the IPR licenses? Ethical considerations?

DOCUMENTATION: How do you document the data? Is the documentation sufficient to reproduce the data from the inputs, and your work from the data? How will you make it so that you can understand the data in 10 years? Processing scripts and code can be part of the documentation, too (and they should be documented). At minimum, docs should be in a README.txt file along with the data.

SOFTWARE: What software do you use to access data? What methods (e.g. statistical, algorithms) do you use? Can others reproduce these methods?

STANDARDIZATION: How do you ensure that your data is machine-readable, compatible, and linkable with other data (your domain and others)? You should use standard terms and identifiers if available in your field (ontologies). Mention what standards within your field you will follow. If standards don't exist, what will you do instead?

SHARING: Who do you think could use your data (think outside of your domain and of unexpected uses, too)? What will be shared? What license and level of sharing will you use (not shared, by request with conditions, by request, or open and public)? Any privacy concerns? Will there be an embargo period?

COLLABORATION: How will you allow collaboration during processing and analysis? Is collaboration restricted to people in your institution, or can others collaborate as well?   Can someone find, and send improvements back without asking first?

Technical

INPUTS: In what formats do you receive or initially store the data? What's the size? How do you receive it securely? Any immediate processing? Does any raw data go straight to archival in addition to final data?

PROCESSING: How do you do preprocessing? Is this preprocessing reproducible? How do you validate the quality of the data? How do you name files so that they make sense later? How will data be versioned? What are the costs for all data management and how are they covered?

STORAGE: Where will data be stored? Who manages this?

FORMATS: What are your working data formats? Are they standard and open? Ideal formats are open, standard, and have multiple programs that can read them.

SECURITY: How will data be kept secure? Consider: 1) confidentiality, 2) availability (backups), 3) integrity (how do you prevent unnoticed corruption, for example someone editing a master file)?

REPOSITORIES: What repositories are used to share the data? If not in a repository, how will people find the data? In what formats do you share the data? Will people be able to find the data? Will there be persistent identifiers?

END OF LIFE: How do you decide what to delete and what to save? Must any data be deleted? Is the deleted data reproducible? How do you prevent bit rot? (The best solution is to put enough to be reproduced on an archive)

General instructions/advice: Make a copy, remove instructions and fill in your solutions. Fill out as concisely as possible - shorter and more standardized is always better. Where relevant, include names of who is responsible.