

Template Documentation and Data Management Plan

Hassan Haydar

Remember, data include metadata!

Where do I store my files? In my local Git Repo on my MacBook

How and where are my files backed up and archived?

- How do you ensure that files are never without a backup? iCloud and GitHub this ensures if one system had issues I can use the second to continue working.
- How do you ensure that there are back-ups and an archive over the longer term? If I link my local file to GitHub then I am sure I will not lose my data and in every change I do I commit and push them to my GitHub account found online, and also on iCloud I have a automatic change backup which in every step I edit is being locally stored
- How do you implement [version control](#)? It is applied through Git with regular commented commits of the changes, through branches like develop branch which I always use for working on my own code and optimizing, or If I want to test some analysis. I regularly pull and push to sync with my GitHub repo online.

How are my processes documented?

- Where do you document your / the project's "data story" so that others can understand? This includes motivation and reasoning, scope, research questions, hypotheses, tests, and essential references and input (appropriately credited and cited).
I have more than one answer which are:
 1. Through a read.me file which I describe the whole project data info, references, ways that I analyzed my data through and a clear raw data info.
 2. I use commented lines in my code and Markdowns (Jupyter notebook) to explain what I am doing
- Where do you document any required permission to use or share data and other materials?
Generalinfo.txt in which I tell the main rules of using my data with the accessibility restrictions if needed
- Where do you document your exact protocols including all required materials and their sources, models, serial numbers, *etc.*? (Even if you use "the same" protocol, it might be slightly changed in each case to accommodate specific aspects of the samples or dataset, or you might use a different instrument, *etc.*)

How are the integrity and traceability of my data and analyses ensured?

- How do you ensure explicit and reproducible data curation and analysis (ideally using well-documented scripts including environment(s))? I try to implement in each new project a specific `venv` environment in my IDE which has all the same packages installed, which makes it identical in every new project
- How do you ensure the integrity of your raw data, *i.e.*, that raw data are not modified? I save a copy of the raw data to my iCloud, Github repo before starting with any analysis, and then define a new `df` to not overwrite my original data
- How do you make clear which is the current version, which are the raw data, which stage of processing is which, and so on?
I comment each line in my code and when I do changes I specify in the commit message what was done and where, and using branching I always have a develop branch which is for working on and trying new things and when I am sure I merge my changes to the main branch.
- When and how do others check your work for mistakes? Not applicable in my situation till now

- When and how do you check whether someone else can reproduce your work? Send collaboration invitation via GitHub

How do you make your data and documentation FAIR? Store data on a repo on Github which makes it accessible for all who want to discover it, store metadata also on it, and raw materials.

How do you ensure compliance with CARE principles or, more broadly, ensure equitable access and benefit-sharing? setting clear permissions, and making sure data use provides fair benefits and follows ethical and responsible practices.