Template Documentation and Data Management Plan

Suvam Kumar Das

Remember, data include metadata!

Where do I store my files?

- Data should be stored on secure cloud services or version-controlled platforms like GitHub or GitLab.

How and where are my files backed up and archived?

- All project files (raw data, processed data, scripts, and documentation) are stored on the institutional project server (e.g., CSCS project space). Files are never kept only on a local machine.
- Automatic server backups run daily, and important datasets are manually copied to a separate "Backup" directory. Final versions of data and scripts are archived in a dedicated "Archive" folder with read-only permissions for long-term preservation.
- Version control is implemented using Git/GitHub/GitLab. Commits are documented clearly, and
- major versions are tagged. Folder naming and timestamps ensure traceability across versions.

How are my processes documented?

- A running project log ("Project_Log.md") documents the data story, including motivation, scope, research questions, hypotheses, key references, and decisions made during analysis.
- Any required permissions or licenses for datasets or external materials are stored in a "Permissions" folder.
- Protocols, software settings, hardware models, and any changes made during the project are described in "Protocol.md." All code is annotated to explain analytical choices, parameter settings, and environment details (modules, Python environments, SLURM scripts, etc.)

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

- Raw data are stored in a read-only "raw" folder and never edited directly. All processing steps are performed using scripts, ensuring reproducibility. The directory structure distinguishes raw, processed, analysis, and results layers clearly.
- Data transformations are logged step-by-step through scripts or notebooks. Intermediate outputs are verified regularly. The analysis is re-run in a clean environment to confirm reproducibility. Peer checks or supervisor reviews help identify mistakes early.

How do you make your data and documentation FAIR?

Findable: Clear file names, metadata, and README files improve discoverability.
Accessible: Data stored on shared institutional servers and archived in long-term storage when allowed.
Interoperable: Standard file formats (CSV, TXT, XYZ, JSON, NetCDF) support compatibility.
Reusable: Documentation, metadata, code comments, and licensing support future reuse.

How do you ensure compliance with CARE principles or, more broadly, ensure equitable access and benefit-sharing?

If data involve communities or restricted sources, the CARE principles guide usage: collective benefit, authority to control, responsible handling, and ethical practice. If not applicable, this is acknowledged while still ensuring equitable and ethical data management.