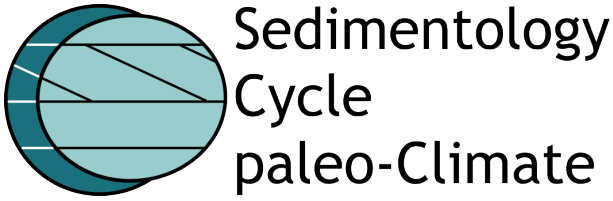


SediCClim Lab Data Management Plan

Anne-Christine da Silva

March 14, 2025

Contents

1. Introduction

This document outlines the data management plan (DMP) for the SediCClim Lab. They are different data/samples that you have accumulated through your project. In this document we will discuss "solid" material management, as well as numeric data management. It provides guidelines on the organization, documentation, storage, access, and preservation of any data associated with geological records, proxy measurements, thin sections, lithological columns, codes, publications, abstract, teaching and so on. Anything you could find interesting could be included upon request.

2. Why a data management plan ?

Open Science and FAIR Data (Findable, Accessible, Interoperable, Reusable) principles are increasingly becoming a research norm in Europe, driving researchers to adopt practices that enhance transparency, reproducibility, and accessibility. Researchers are encouraged to adopt these principles through demands from multiple sources, including journals that require datasets associated with publications to be available for reuse, and research funders who require deliverables related to FAIR Data (e.g., Horizon Europe, Belspo, FNRS, etc.). On a broader scope, the reproducibility crisis, observed over the past decades, has prompted a cultural shift toward greater traceability of published results, demanding higher standards of documentation and transparency in data collection, analysis, and publication. The data management plan (DMP) facilitates the organisation of data management and favours the following.

- **Enhance Research Efficiency**: A well-organized DMP ensures data is systematically stored and easily accessible, saving time and effort in data retrieval and analysis.
- **Ensure Data Integrity and Reproducibility**: Proper documentation and version control help maintain the accuracy and consistency of data, making research results reproducible and reliable.
- **Facilitate Collaboration**: A structured DMP enables smooth data sharing and collaboration among team members and with external researchers, enhancing overall research quality.
- **Compliance with Requirements**: Many funding agencies (including the FNRS) and institutions require a DMP to ensure responsible data handling and long-term preservation of research outputs.
- **Preserve Data for Future Use**: Proper data management ensures that valuable research data is preserved and remains accessible for future studies and researchers. You will always be associated with use of your data.
- **Enhance Data Security**: A DMP includes strategies for secure data storage and backup, protecting against data loss, corruption, and unauthorized access.
- **Support Publication and Dissemination**: Organized data and thorough documentation facilitate the publication process and increase the visibility and impact of research findings.

In other words, it will improve the research quality of the group, improve the duration that your research can "shine" and be useful, you could be associated with more publications, and it could help to prepare you for future professional standards and expectations in academia and industry.

3. Data Types and Formats

3.1. Physical material by geological record

For management of Physical samples, we have an excel sheet (location: XX) which should be filled accordingly. It includes multiple columns. For each record you will have to fill :

- **Country**: Country
- **Section name**: Name of the record
- **Abbréviation**: abbreviation of the record name (e.g. OF for Oued Ferkla)
- **GPS**: GPS coordinates
- **Project**: Project Name
- **Date**: Date of the start of the sampling (yyyymmdd)
- **Worker name**: main worker associated with the data
- **Period**: Period of the record (e.g. Devonian)
- **Stage**: Stage of the record (e.g. Frasnian)
- **Paper DOI**: DOI of the paper associated with the data
- **Amount Spl**: number of samples.
- **Loc smpl**: Location of the samples
- **Loc TS**: Location of the thin sections
- **Loc "Sucre"**: Location of small samples
- **Loc crush**: Location of crushed samples

For the location you can use office numbers on top of its door (ex: 1/22, first floor, office 22), for the dark green shelves downstairs use 0/ shelf number, for the turquoise cabinet for small samples use 0/AST, for the preparation room use 0/XX and for the Compactus in the B18 use B18/XX. If for some reason you move samples elsewhere than ULiège, the file should be updated with a specific new location and contact person.

3.2. Digital data

Each of you will have a folder corresponding to your project ID Year_name (for example: 2019_SiluCCarb, 2022_WarmAnoxia, 2023_CarboIce). Furthermore an overarching structure will include data accessible to all (with protocoles, lab publications, teaching ressources, grant proposals, etc.).

- **Geological Record**
  - *Lithological Columns*: Stratigraphic information, descriptions, pictures of the log, csv file to build the log in StratigrapheR, Log at different scales.
  - *Proxy Measurements*: all type of measurements associated with the record (pXRF, MS, Cisotopes, etc.).
  - *Pictures*: picture of the outcrops.
  - *Thin Sections*: Photographs, descriptions, and petrographic analysis results.

  - *Codes*: Scripts, functions, data analysis workflows, and outputs.

- **Publications**: Articles, abstracts, conference papers, software packages, PhD thesis.
- **Reports**: annual reports, grant reports, etc.

- **Teaching Material** any teaching material of interest (e.g. nice field picture, thin section description, field guide, etc.)

- **Protocoles** if you have protocole explaining how to use a device, how to use a package, coding protocole, etc. include it in this section.
- **Applications** Include your applications in a shared file, so everyone can benefit from your experience.

3.3. File Naming Conventions and data format

Having a file convention is important to allow (1) consistent logic of storage and (2) to find something through search engine. So, it is important to use descriptive file names, and to be always consistent (e.g. don't use GRS once and GammaRay for another file). In general avoid using space in your file name, use _ instead or Capital Letter to separate the words (Oued_Ferkla or OuedFerkla). The following characters are usually non valid « * : < > ? / \ | , try to avoid them.

Record $\Longrightarrow$ *Date_Country_Record_Period_Stage.*
Date is the date in this format: yyyymmdd (e.g. 20200512) of the starting of the sampling for a specific record; *Country* of your record; *Record* is the complete name of your record. Avoid using abbreviations for your record name and be consistent in each of your files (e.g. don't say Sallet once and Salet another time). *Period* and *Stage* correspond to the Age of your record. If you don't know the Stage, you can also use the Epoch instead. To avoid to have file names which are too long, use Ordo, Silu, Devo, Carbo, etc. Here is an example of the Devonian Carboniferous boundary record of Chanxhe (so encompassing two different periods).
Ex: 20200915_Belgium_Chanxhe_Devo_Famennian_Carbo_Tournaisian

- **Lithological Column** This folder should include all data about the lithological column, including sample position. If the log exists at multiple scales, include the scales (e.g. 1_10).
  ScanLog for Scan or take pictures of your notebook (Windows Office Lens allows to take successive pictures of your notebook and to directly transfer them as a pdf to OneDrive);
  VectorLog for vectorized log (Corel Draw, Adobe Illustrator, Inkscape). For the vector log, please, include a eps version, it is the most common vector format scale (allowing to open your file with multiple programs), please include also a pdf and the original corel, ai or inkskape files.
  RLog for the code to build Log in StratigrapheR.
  Ex: *20200915_Belgium_Chanxhe_Devo_Famennian_Carbo_Tournaisian_ScanLog*

- **Proxy** - *Date_Country_Record_Period_Stage_Proxy.* Use following codes for your folder and text files: *MS* for Magnetic Susceptiblity, *PXRF* for Portable XRF, *LXRF* for Laboratory XRF, *ICPMS*, *isoCcarb* for Carbon isotope of carbonate, *isoCorg* for Carbon isotopes of Organic Matter, *TOC* for Total Organic Carbon. These are the most common measurements. For other specific measurements, find your own and let me know, so I can include it here. This can be saved as excel files or csv. Your data should be subdivided into ProxyRaw (original file directly coming from device), ProxyDist (ordered, calibrated data in the distance domain), ProxyTime (ordered, calibrated data in the time domain).
  Ex: *20200915_Belgium_Chanxhe_Devo_Famennian_Carbo_Tournaisian_PXRF_Dist*

- **Pictures of the record** I know it's really hard to document all pictures and the link with the record. If you did, that's awesome, explain how. Ideally, the best when is when you take a picture on the field, you write the picture number directly on your notebook just on the side of the stratigraphic interval it has been taken.

- **Pictures of thin sections** Each Picture should include the name of the sample in its name, as well as the magnification and if it's polarized (or in the picture).

- **Codes** The names of your codes should be as self-explanatory as possible and include as much documentation as possible. The title should include when appropriate the RecordName, as well as the type of analysis (Spectral Analysis, Wavelet, TimeOpt, PCA, etc.).

Publications

- **Papers** $\Longrightarrow$ for each specific paper create a folder with Date_Author_AbbreviatedTitle
  Date of the start of the writing of the paper (yyyymmdd), first author (in one word, if different part to the name, such as Da Silva, write is all together (DaSilva), abbreviated Title (in one word).
  Ex: 20220815_Arts_CellonWaveriderSilu

  Include **versioning** in filenames when applicable, as well as the the type (submitted, review, ...). *Ex: 20220815_Arts_CellonWaveriderSilu_SubmittedTextV1.pdf*

  Include also in the folder associated with a specitific paper, a "*Published*" folder, which will include the final published version of the paper (final version of the figures, text, data, codes, etc.).

- **Reports** $\Longrightarrow$ for each report, create a file with a name like Date_Author_Grant.
  Ex: 20230715_Arts_FNRSPDR

- **Abstracts** $\Longrightarrow$ Date_Author_Meeting_Location_Topic
  Ex: 20220425_Arts_EGU2022_Vienna_WaveriderSilu

- **PhD**
  Chapter01_Name_Text
  Chapter01_Name_Figure01

Chapter01_Name_Table01
Chapter01_Name_Code
...

- **Codes & Packages**
  If you published a package or a code, include all associated folders here (while the codes related to specific records should be under the folder associated with this record).

Teaching Material
Includes any teaching material with a reference to the general topic in the file name.

Protocoles
Includes here protocoles on all type of devices, coding, manipulation, field technique, etc. The file name should include the Date of creation or update of the text, as well as the name of the device, technique, code, etc.

Applications
Includes here applications you worte, with the following format: Date_Author_Tool_Goal.
Ex: 20240703_DaSilva_FNRS_PDR_PhDGrant

3.4. File documentations and metadata

Always document what you are doing and in each respective folder include text files with the following information :

- **Structure** - Maintain a *README_FolderStructure* text file in each project directory outlining the data structure, file contents, and usage instructions.

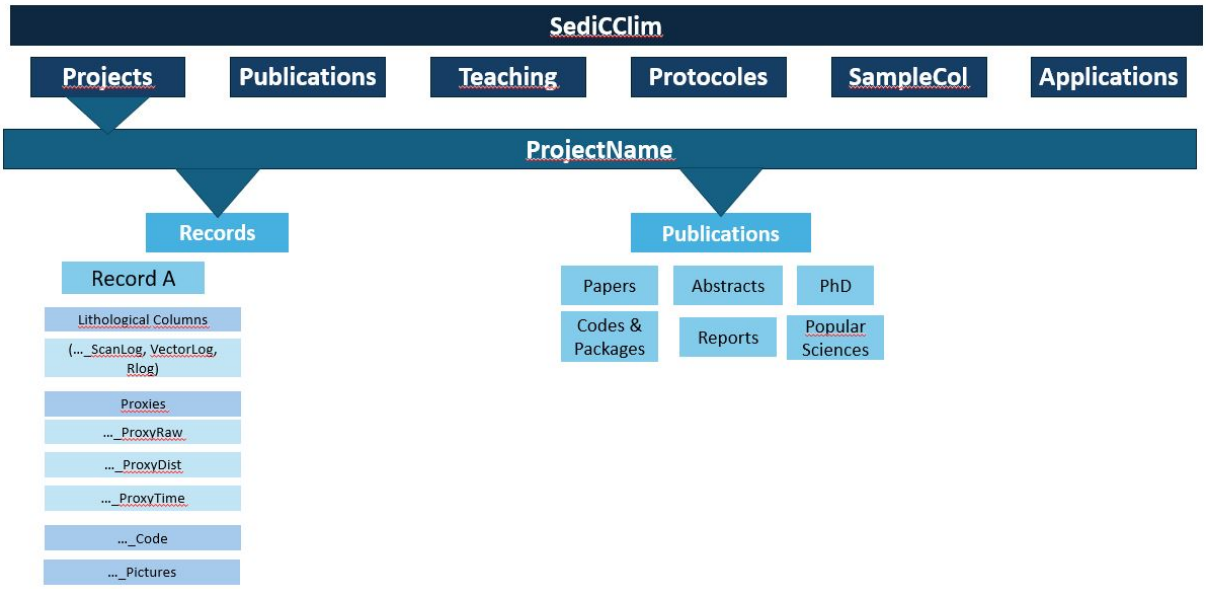
- **Record** - in each specific record directly, include a *README_RecordName* text file with the Name of the record, country, date of sampling, people who sampled, Stratigraphic interval, thickness and a Figure of the global 1p log.

- **Data** - for each type of measured proxy, you should include a *README_RecordName_Proxy* text file which will include the name of the device, location of the device, configuration, calibration (e.g. pXRF, Bruker Tracer i5, 40eV, Calibration Carbonate).

- **Code** - Any interseting info that can help people to navigate in your code (e.g. RecordNameCodeSpectralAnal.R includes all the spectral analysis procedure with WaverideR, ASM and the tuning). Use comments in your R scripts to explain code functionality. Be careful to provide an overview of each script, including inputs, outputs, and dependencies (csv files associated with the codes). Provide a *README_RecordName_codeName* text file, with the explanation of the role of the code, as well as the role of dependencies.

4. Organize Data Structure

On the external drive, there will be a couple of folders accessible to all:

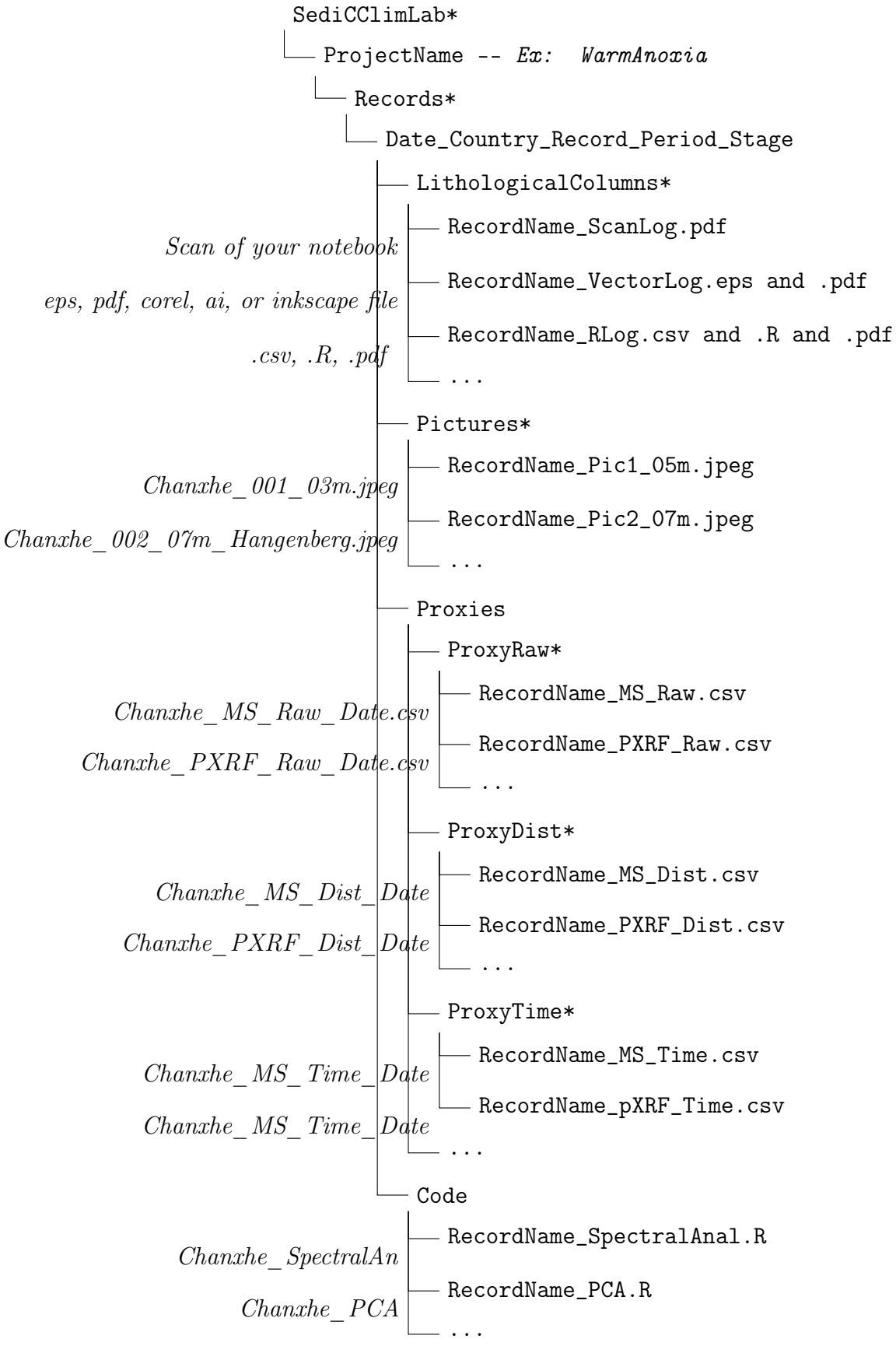


- *Applications*: You can include here your applications with the following format: Date_Author_Tool_Goal. I will also include applications related to your project and submitted applications. We can all learn and save time from other applications
- *Protocoles*: include all programs and protocoles related to each device available in the lab and abroad.
- *Publications*: subdivided into SediCClim publications and Publications for all papers
- *SampleCollections*: will include the excel file to organize Research Samples, Teaching Samples and Thin Section samples.
- *Teaching*: Teaching material

Furthermore, each of your project will have its own folder divided in 3 folders: Protocoles - Publications - Records

4.1. Directory Structure for Each record

A folder followed by * means that the name of the folder cannot be changed.


SediCClimLab*
ProjectName -- Ex: WarmAnoxia
Records*
Date_Country_Record_Period_Stage
LithologicalColumns*
RecordName_ScanLog.pdf
Scan of your notebook
RecordName_VectorLog.eps and .pdf
eps, pdf, corel, ai, or inkscape file
RecordName_RLog.csv and .R and .pdf
.csv, .R, .pdf
...
Pictures*
RecordName_Pic1_05m.jpeg
Chanxhe_001_03m.jpeg
RecordName_Pic2_07m.jpeg
Chanxhe_002_07m_Hangenberg.jpeg
...
Proxies
ProxyRaw*
RecordName_MS_Raw.csv
Chanxhe_MS_Raw_Date.csv
RecordName_PXRF_Raw.csv
Chanxhe_PXRF_Raw_Date.csv
...
ProxyDist*
RecordName_MS_Dist.csv
Chanxhe_MS_Dist_Date
RecordName_PXRF_Dist.csv
Chanxhe_PXRF_Dist_Date
...
ProxyTime*
RecordName_MS_Time.csv
Chanxhe_MS_Time_Date
RecordName_pXRF_Time.csv
Chanxhe_MS_Time_Date
...
Code
RecordName_SpectralAnal.R
Chanxhe_SpectralAn
RecordName_PCA.R
Chanxhe_PCA
...

4.2. Directory Structure for Publications

```
SediCClimLab
└── ProjectName
    ├── Papers
    │   └── Date_Author_RecordName_Topic
    │       ├── Submitted and Reviews
    │       │   ├── Date_SubmittedText.docx
    │       │   ├── Date_SubmittedFigure1.eps
    │       │   ├── Date_Review1_ReviewerComments1.pdf
    │       │   ├── Date_ResponseToReview1.docx
    │       │   ├── Date_Rev1_Figures.eps
    │       │   └── ...
    │       └── Published
    │           ├── Year_Author_Journal_Topic.pdf
    │           ├── Year_Author_Journal_Topic_SupplMat1.pdf
    │           └── ...
    ├── Abstracts
    │   ├── Date_Author_Meeting_Location_Keywords.pdf
    │   └── ...
    └── PhD
        ├── Chapter1.pdf
        └── ...
```

5. Data Storage and Backup

5.1. Storage

- *Individual External drive*; in which you can include all your data in any format.
- *SediCClim Drive all*; in which you can include all your data in any format under your name.
- *SediCClim Drive DMP*; in which you can include all your data in the DMP format under your project name.
- *Individual OneDrive*; in which you can include all your data in any format.

- *SediCClim Teams* in which you will also have storage space to include your data formatted accordingly to this DMP.

- *NASS* - Research Unit of Geology has connected drives with different units, allowing multiple backups. This external drive is only accessible to me, however I will also save your data (saved according to the DMP) there each month from SediCClim Drive DMP.

For the individual solutions (external drive or OneDrive), you can use one or both.

5.2. Backup Strategy

- If you don't use OneDrive, implement automatic daily backups on your drive. For this Synckback (2BrightSparks) is a good tool. Regularly test backup restoration procedures.Store backups in geographically separated locations.

- Save all your data (any format) on the external drive of the Lab (SediCClimDrive) at least each Month.

- Save all your data on the external drive of the Lab (SediCClimDMP) at least each month.

6. Sharing

- Use DOIs for data sets to ensure persistent and citable references.
- Share data via institutional repositories or data sharing platforms (e.g., myOrbi, Zenodo, Github, Dataverse).

Here is a view of the structure and convention names together, + where you should document the files/folders.

Physical Data - Fill the excel file

Country	Country of Sampling
Record name	Name of the record
Abbr. Record	Abbreviation of the record name
GPS	GPS coordinates
Date	Date of the start of the sampling (Year_Month_Day)
Project Name	
Worker name	
Age	Period
Stage	Stage
Paper DOI	When published
Amount Spls	Amount of samples
Thickness	Thickness of the record
Comments	Any comment you find acceptable
Loc TS	Location of Thin sections
Loc spl	Location of Samples
Loc Sucre	Location of "Sucres"
Loc Geoc	Location of geochemical Samples

Digital Data

Structure — Common files — : Publications (only for published material), Teaching, Protocoles, SampleCollection and applications. — Project files —: Records (Litholog, Proxies -Raw, distance and time-, Codes, Pictures), Publications (papers, abstracts, Packages, PhD).

Name convention - use date (YYYYMMDD) as much as possible, and be consistent with your names.

Documentation - make Readme files documenting the tree structure and the location of your files. Document your proxies (device, conditions, etc.) and your codes.