Data Organization

Midi de l'info scientifique

0. Data Organization

The organization of data files takes time and seems complicated, but it is really effective for both individual and collaborative work.

A. **Spending a little time upfront, can save a lot of time later on**
B. **Be realistic: strike a balance between doing too much and too little**
C. **There's no single right way to do it; establish a system that works for you**
D. **Think about who your system needs to work for: Just you? You and your lab group? Collaborators?**

1. File structures – hierarchical method

The hierarchical system, used almost in all kinds of systematic classification, is a familiar and widely used system. The tree structure is particularly suited to represent the structure of information. Similar items are stored together and subfolders can function as task lists. However, the establishment of a hierarchical system can be surprisingly difficult; it is necessary to find the right balance between the width and the depth of the tree, an element can only be in one place and the reorganization of an obsolete tree takes an enormous amount of time.
Here are some tips and best practices to create a hierarchical system and find your files more easily:

A. **Avoid overlapping categories**
B. **Don't let your folders get too big**
C. **Don't let your structures go too deep**
D. **Order dates beginning with the year to enable sorting by date** (e.g., YYYYMMDD)
E. **Embed metadata in your files (if possible)**
F. **Add shortcuts to files within other relevant folders**
G. **Consider reference management software such as EndNote, Zotero, Mendeley**

2. File Naming

It is important to respect certain naming conventions, designed to create logic and consistency within a structure.
A naming convention should be:

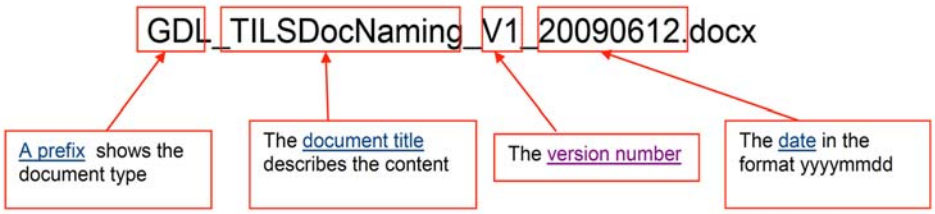
A. **Descriptive**
   Consider including:
   - Unique identifier (ie. Project Name or Grant # in folder name)
   - Project or research data name
   - Conditions (Lab instrument, Solvent, Temperature, etc.)
   - Run of experiment (sequential)
   - Date (in file properties too)
   - Version #

B. **Consistent**
   You need to maintain order. For example, when naming a date, always use the same pattern (YYYYMMDD or MMDDYYYY, etc.). The choice of the pattern is free, but its use must be systematic.
   You must also include the same information (DateProjectID or TimeDate, etc) each time. Once again, the choice of naming pattern is free, but its use must be systematic.

There are some best practices for naming files, which you can find in the table below:

Best Practice	Example
Limit the file name to 32 characters (preferably less!)	32CharactersLooksExactlyLikeThis.csv
When using sequential numbering, **use leading zeros** to allow for multi-digit versions For a sequence of 1-10: 01-10 For a sequence of 1-100: 001-010-100	**NO** ProjID_1.csv ProjID_12.csv **YES** ProjID_01.csv ProjID_12.csv
Don't use special characters & , * % # ; * ( ) ! @$ ^ ~ ' { } [ ] ? < > -	**NO** name&date@location.doc
Use only one period and use it before the file extension	**NO** name.date.doc **YES** name_date.doc
Avoid using generic data file names that may conflict when moved from one location to another	**NO** MyData.csv **YES** ProjID_date.csv

Finally, according to the TILS Document Naming Convention, the naming of a document should follow this convention:



3. File versioning

The number of different versions of a document can quickly increase within a research project. The naming and organization of these versions is very important in order to avoid loss or confusion.
Here are some tips on how to best manage document versions:

A. **Avoid imprecise "final" labels**
B. **Put older versions in a separate folder**
C. **Save new versions**
D. **Establish a consistent convention**
E. **Document your convention**
F. **Consider your version control needs**

4. The 5 C's to organize your files

The files organization can finally be summarized in five words:

A. **Clear**
B. **Concise**
C. **Consistent**
D. **Correct**
E. **Conform**