

Organization and naming of research data files

Rendez-vous de l'info scientifique

0. Data Organization

Organizing research data files is time-consuming and may appear daunting at first. However, it is worthwhile in the long run both for individual and collaborative work since well-structured and well-named data are easier to locate and identify.



- A. Invest time up front to save time later**
- B. Consider who will need to access the data when setting up your organization system**
- C. Be precise but realistic in the choices you make (organization, naming convention, etc.) and stick to your decisions**

1. File structures – hierarchical method

The hierarchical system is a familiar system widely used in almost all types of systematic classification. The tree structure is particularly suitable for representing the structure of information. In addition, similar items are usually stored together.



However, conceiving a hierarchical system can be surprisingly difficult: folders and subfolders sometimes multiply, leading to a large number of clicks before reaching the desired document. One is sometimes tempted to put a file in several places and thus versions of the same document can multiply. Finally, the reorganization of an obsolete tree or the creation or modification of a tree afterwards takes a lot of time and efforts.

- A. Avoid overlapping categories:** subfolders with the same name will become obscure when taken out of context.
- B. Weight the depth and width of the tree:** Avoid endless lists of files inside a folder or "Russian doll" subfolders.
- C. Associate metadata with your files** (e.g. README.txt)
- D. Think broadly and anticipate sorting files/folders** (e.g. YYYYMMDD for dates or 01 instead of 1)
- E. Do not multiply copies of a file:** use shortcuts pointing to it
- F. Consider using a dedicated software for managing references** (e.g. EndNote or Zotero)

2. File naming

The naming conventions make it possible to bring logic and coherence within the hierarchical structure. In concrete terms, this means defining in advance the syntax that will govern the names of the files. A naming convention must be:

A. Descriptive

A file's name is easier to read and allows the related file to be easily findable if it is composed with elements such as:

- A unique identifier (ex. Name of the project or #Financement in the file's name)
- The conditions of data production (instrument, producer, etc.)
- The reference of the experiment or the version (sequential)
- A date



B. Consistent and uniformly applied

You need to maintain order. For example, always use the same pattern for dates (YYYYMMDD or YYYY-MM-DD, etc.) or the same syntax for the names of the files. The choice of dates or syntax pattern is free, but they must be used systematically.

C. Good practices for naming files:

- × **Avoid special or accented characters:** & , * % # ; * () ! @ \$ ^ ~ ' { } [] ? < > -
- × **Avoid spaces** (use _ - or « camel case » writing instead)
- × **Avoid periods** (reserve them for the extension indicating the type of file (.docx, .csv, etc.)).
- × **Avoid lengthy names** (since they will not be displayed completely)
- × **Avoid vague or too broad titles** (if taken out of context, they will become obscure or difficult to understand)
- ✓ **Choose an unambiguous date format that allows automatic chronological sorting:** YYYYMMDD or YYYY-MM-DD
- ✓ **Choose numbering that allows you to process and sort large volumes of items** (001 and not 1)

Examples:

- × Présentation_étudeDMP@HUG_fév.18.ppt
- × Resultats|enquête|v.corr
- ✓ 201802_DMP_HUGpresentation.ppt
- ✓ 2018_QualiteVie_EnqueteGE_Results_v02.csv

Here is an example of a naming convention:

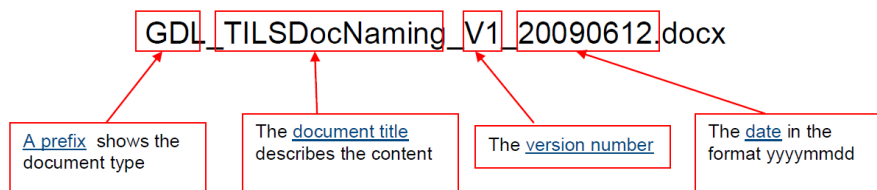


Figure 1 Queensland University of Technology Naming Convention.
<https://www.library.qut.edu.au/about/management/documents/QUTTILSDocNamingConvention.pdf>

3. File versioning

Since the number of versions of a file may increase rapidly, it is necessary to distinguish them easily, even at a later date.

Here are some tips on this subject:

- A. Number files as follows:**
 - a. Major changes: v02, v03
 - b. Minor changes: v02-01, v02-02
- B. Avoid putting the "final" label or reserve it for the publicly released version**
- C. Keep your source files in a separate folder**
- D. Create an archive folder to keep the old files, if necessary.**
- E. Ask yourself which intermediate versions you really need to keep**

4. Sources and Resources

- MIT Libraries Data Management Services. (2017) *Data Management: File Organization*. <https://libraries.mit.edu/data-management/services/workshops/> CC-BY-NC
- EDINA and Data Library, University of Edinburgh. (s.d.) *Research Data MANTRA [online course, Organising Data module]* <https://mantra.ed.ac.uk/>. CC-BY.
- CORNELL University RESEARCH DATA MANAGEMENT SERVICE GROUP (s.d.) *Guide to writing "readme" style metadata* <https://data.research.cornell.edu/content/readme> CC-BY

