Organize, name and version files

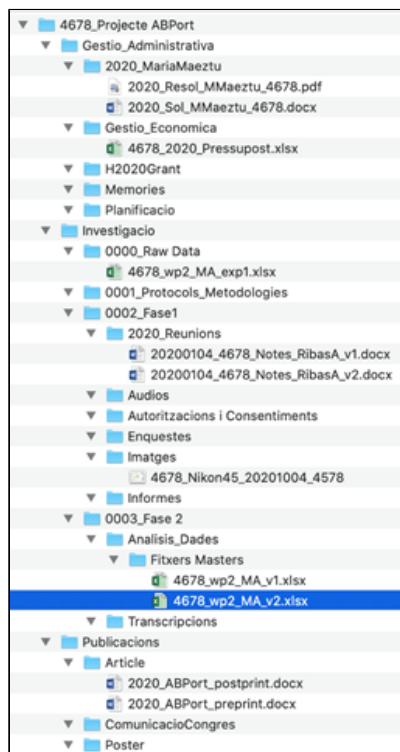| Introduction |
|---|
| When creating or collecting data and working with many files, these files can easily become disorganized. To save time and avoid mistakes, it is important to make decisions about how files will be organized and named, and what version control will be applied whether working individually or in groups.

Having conventions for the structure and organization of research files and data will improve their management and make research data more consistent. If these are also documented, other researchers will be able to understand them and reuse them in the short, medium and long term.

This document provides guidelines for organizing, naming, and versioning files. It is information that many research funding bodies ask to include in the Data Management Plan. |

Organize files

- Save all project documents in the same location, regardless of format or type, or at least link them.
- Choose the location to manage and store files (cloud services, university servers, etc.) based on the type and use you will make of them; not all places are suitable. Even if you work from different locations and devices, establish a single location to archive them all and ensure it complies with the necessary control, access, security, and preservation requirements.
- At the beginning of the research, define a logical folder structure for everyone in the group to use. There is no single criterion for folder organization; it will depend on the type, activities, and stages of your project. Choose what best suits your objectives and working procedures.

A recommended hierarchical structure is from more general to more specific, with 3 or 4 expandable sub-levels if necessary. For example:
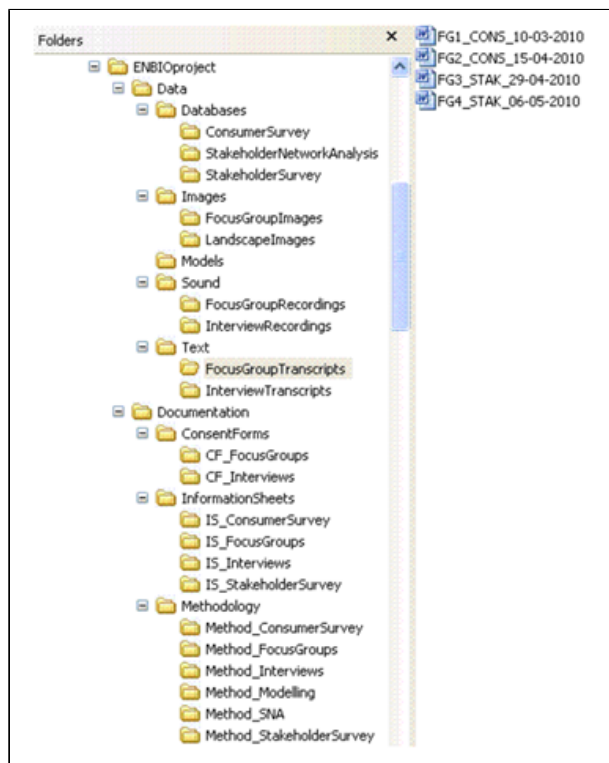
- Use functions as a criterion for creating files: that is, a first level for the project name and, at a second level, folders that group files according to their purpose of use (why the file was created).
- Another option would be a first level for the project name and, at a second level, differentiate between data and documentation, and in the following levels, by type and activities they respond to.

Complete the directory of folders by establishing conventions for naming folders, subfolders, and files. Folder names should reflect content, not the person or team that created it.

- Manage different versions of files carefully, following the recommendations in point 4.
- Save the original data (raw data) of the research carefully:
  - Create an exclusive folder configured as "read-only" for raw data.
  - Use copies of the original data to work with them and avoid losing information.
  - Since files can be continuously updated from different devices, create a folder called "master files" to store the final files. The files you work with should be treated as temporary copies. Synchronize them or update the master files periodically or after each change.
  - Periodically back up the folder with the master files in different locations.
- Assign metadata and create explanatory documentation about the data, as it is very useful for understanding and interpreting them without having to open the files. Doing so adds value: it facilitates management, allows understanding the context in which the data was generated, finding them, and reusing them. There are three types of metadata and documentation: administrative, technical, and descriptive.
- When creating a file, use standard formats to facilitate access for all project members, their preservation, and, later, their dissemination and reuse. Preferably use open, non-proprietary formats compatible with different platforms and widely used. On the page "*Format Recommendations*," you can expand on this information.
- Control information security and access to folders and files throughout the project, as it is essential to comply with current legislation on data protection and copyright. Establish who, when, and how can access each folder and/or file. Therefore:
  - Agree in a document the access and use conditions for all project members.
  - Encrypt the folder, subfolder, or file so that only those who know the code can access it.
  - Establish security measures where all files are stored: use antivirus and firewall programs, secure networks, continuous software updates, periodic backups, and access passwords.
- Determine which files to keep in the long term and for how long. This will depend on the project's purpose and how the data will be used. With the folder structure created, it will be easier for you to select the files to keep: you will preserve the "master files" and "original data" folders forever. On the other hand, documents that are drafts, duplicates for individual use, or temporary files can be deleted when they are updated or are no longer useful.

```
▼ 📁 4678_Projecte ABPort
   ▼ 📁 Gestio_Administrativa
      ▼ 📁 2020_MariaMaeztu
           📄 2020_Resol_MMaeztu_4678.pdf
           📄 2020_Sol_MMaeztu_4678.docx
      ▼ 📁 Gestio_Economica
           📄 4678_2020_Pressupost.xlsx
      ▼ 📁 H2020Grant
      ▼ 📁 Memories
      ▼ 📁 Planificacio
   ▼ 📁 Investigacio
      ▼ 📁 0000_Raw Data
           📄 4678_wp2_MA_exp1.xlsx
      ▼ 📁 0001_Protocols_Metodologies
      ▼ 📁 0002_Fase1
         ▼ 📁 2020_Reunions
              📄 20200104_4678_Notes_RibasA_v1.docx
              📄 20200104_4678_Notes_RibasA_v2.docx
         ▼ 📁 Audios
         ▼ 📁 Autoritzacions i Consentiments
         ▼ 📁 Enquestes
         ▼ 📁 Imatges
              📷 4678_Nikon45_20201004_4578
         ▼ 📁 Informes
      ▼ 📁 0003_Fase 2
         ▼ 📁 Analisis_Dades
            ▼ 📁 Fitxers Masters
                 📄 4678_wp2_MA_v1.xlsx
                 📄 4678_wp2_MA_v2.xlsx
         ▼ 📁 Transcripcions
   ▼ 📁 Publicacions
      ▼ 📁 Article
           📄 2020_ABPort_postprint.docx
           📄 2020_ABPort_preprint.docx
      ▼ 📁 ComunicacioCongres
      ▼ 📁 Poster
```

Source: own preparation

Source: *https://www. ukdataservice.ac.uk /manage-data /format/organising*

Name files

- Name the files based on items such as project name or number, research group/team name, measurement type, topic, creation date, version number, etc.
- Names must be descriptive, clear and short, maximum 25 characters.
- They must show the relevance of each document without having to open them.
- Don't use words like: draft, presentation, or spreadsheet. Use names and/or codes from the most generic to the most specific to facilitate easy sorting within folders.
- Use only characters from the sets A-Z, a-z, 0-9, hyphen, or underscore.
- Avoid using special characters such as &% $ # : or)
- If you use sequential numbering, use leading zeros to ensure that the files are ordered correctly: 0001, 0002 ... 1001, etc. and not: 1, 2 ... 1001, etc.
- Make sure the file names are independent of their location to avoid problems if files are moved.
- For digital photos and scanned images, change each system-generated file number to more meaningful names.
- Create a log file to explain the coded elements.
- Thus all collaborators will have the information to establish the codes.

ⓘ **Examples**

- NTI_wp2_MA_exp1.csv: project, work package within the project, measurement type, measurement test identifier
- MicroSensor_NTI023_20201031.xls: description of content, project number, date following the international standard

- NameExperiment_NameInstrument_DataScreenshot_ImageID.tif

Create a new version when changes occur in the structure, content, or status of the resource. In the case of research data, create a new version of a dataset when you process, correct, or add additional data. Version control allows tracking changes associated with dynamic data.

A file version can be identified by:

- The date recorded in the file name or within the file.
- Version numbering in the file name (v1, v2, v3.3).
- Version description in the file name or within the file (draft, final).
- File history, version control table, or notes included in a file, where versions, dates, authors, and details of changes to the file are recorded.
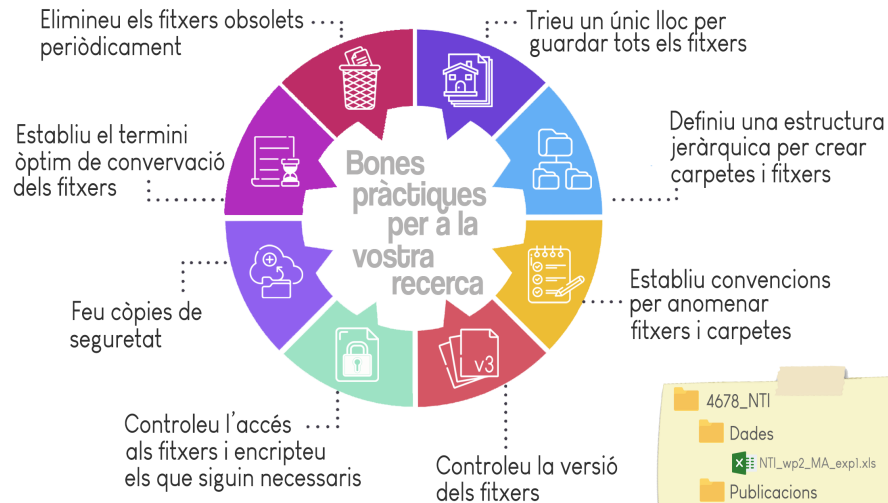
Version control can also be accomplished through:

- The use of version control software, such as GIT or Subversion (SVN).
- The use of file-sharing services like Dropbox, Google Docs, or Amazon S3.

ⓘ **Exemples**

- 20200104_ProjectA_Ex1Test2_RibasA_v1.xlsx
- 20200104_ProjectA_MeetingNotes_RibasA_v2.1.docx

Organitzeu els fitxers del projecte

Elimineu els fitxers obsolets periòdicament

Establiu el termini òptim de convervació dels fitxers

Feu còpies de seguretat

Controleu l'accés als fitxers i encripteu els que siguin necessaris

Bones pràctiques per a la vostra recerca

Trieu un únic lloc per guardar tots els fitxers

Definiu una estructura jeràrquica per crear carpetes i fitxers

Establiu convencions per anomenar fitxers i carpetes

Controleu la versió dels fitxers

4678_NTI
  Dades
    NTI_wp2_MA_exp1.xls
  Publicacions
    2020_NTI_Results_v1.doc
    2020_NTI_Results_v2.doc

Per a més informació: http://hdl.handle.net/2072/362618

Grup de treball de Suport a la Recerca del CSUC