File naming conventions[1]

Elements that should be considered when developing a file naming strategy (not exhaustive list):

- Project/experiment acronym or ID number or full name;
- Project structure e.g. work package, task, etc. (when appropriate);
- Description of content: type of data (e.g. interviews, images, audio, video, etc.), conditions (e.g. anonymized transcripts, etc.);
- Name of creator (single researcher name/initials) or name of research team/department associated with the data;
- Location/spatial coordinates (when appropriate);
- Date of creation/publication or date range of experiment;
- Version number of file (when appropriate);
- File extension suffix for application-specific files (e.g. .xlsx, .rtf, .mov, etc.).

Best practice in naming files:

- meaningful but brief file names (avoid very long names and use file names to classify types of files);
- date format YYYYMMDD or YYMMDD;
- use leading zeros with sequential numbering system (e.g. use "001, 002, ...010, etc." instead of "1, 2, ...10, etc.");
- avoid dots (**.**) and special characters, such as **~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ' " |**;
- to separate elements in a file name, avoid spaces and instead
    - use dashes (**-**) or underscores (**_**) or both: e.g. “my-file-name.xyz”, “my_file_name.xyz”, or “my-file_name.xyz”
    - do not separate elements: e.g. “myfilename.xyz”
    - use the “camel case” method (first letter of each section of text is capitalized): e.g. “**M**y**F**ile**N**ame.xyz”.
- use consistent naming for all the files connected to one data collection event (e.g. an interview, a focus group session, a laboratory experiment, etc.).

[1] Guidelines based on UK Data Service (https://www.ukdataservice.ac.uk/manage-data/format/organising), CESSDA Training (https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure), and Stanford Libraries (https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming) training material.

Documento redatto dallo staff AlmaDL in data 13/05/2019. AlmaDL è una Unità di Processo dell’Area Biblioteche e Servizi allo Studio (ABIS) dell'Alma Mater Studiorum - Università di Bologna

For example, the audio tape, the transcription and photographs of an interview to a stakeholder held in Bologna:

*MyProject_interview_Stakeholder01_Italy-BO_20180302_**audio.mp3***
*MyProject_interview_Stakeholder01_Italy-BO_20180302_**trans.rtf***
*MyProject_interview_Stakeholder01_Italy-BO_20180302_**image-001.jpg***
*MyProject_interview_Stakeholder01_Italy-BO_20180302_**image-002.jpg***
*MyProject_interview_Stakeholder01_Italy-BO_20180302_**image-003.jpg***
*MyProject_interview_Stakeholder01_Italy-BO_20180302_**image-004.jpg***

- provide documentation about naming convention used (e.g. include in the main folder a "README.txt" file that explains the adopted naming conventions and any abbreviations or codes used)**.**

  E.g. naming convention followed for:

  MyProject_interview_Stakeholder01_Italy-BO_20180302_**audio.mp3
  MyProject_interview_Stakeholder01_Italy-BO_20180302_**trans.rtf
  MyProject_interview_Stakeholder01_Italy-BO_20180302_**image-003.jpg

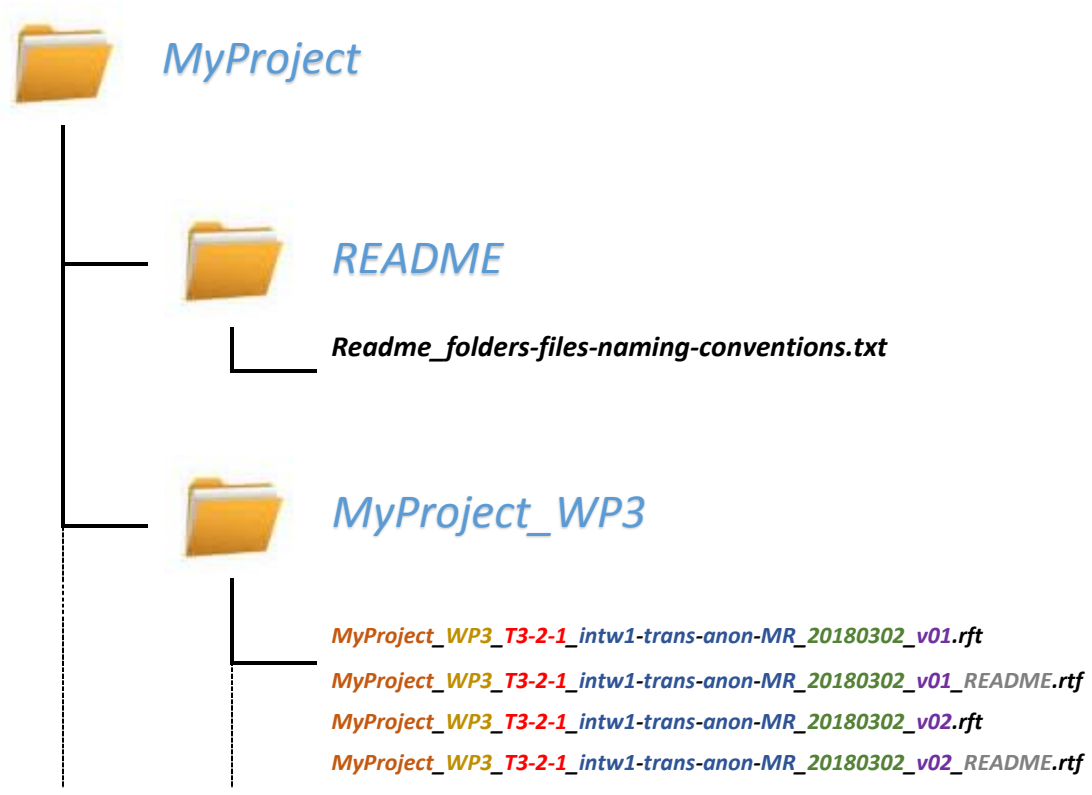
  <project acronym>_<activity>_<subject of study + ID>_<place (country-city code)>_<date of collection (YYYYMMDD)>_<data type + ID>.<file type>

Folder and subfolders structure and naming convention[2]

Folder structure and naming convention depend on how the study is planned and organized. For example, folders can be organized following the project structure based on work packages and tasks (see example 1). Otherwise, in presence of independent data collections, a structure with separate data folders for each collection can be adopted (see example 2). Data and documentation files can also be held in separate folders, with documentation files organized for type of document and research activity (see example 3). However, in all cases background information must never be stored in the file name only.

Example 1

Folders structure based on project work packages and tasks



[2] Guidelines developed starting from UK Data Service (https://www.ukdataservice.ac.uk/manage-data/format/organising), CESSDA Training (https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure), and Stanford Libraries (https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming) training material.

Documento redatto dallo staff AlmaDL in data 13/05/2019. AlmaDL è una Unità di Processo dell'Area Biblioteche e Servizi allo Studio (ABIS) dell'Alma Mater Studiorum - Università di Bologna

In this example, the main folder and the subfolders follow the project structure. The main folder is named with the project acronym, the subfolders with “project acronym_work package number”. It has been adopted also a naming convention for the single files that follow the project structure and provides many information in the filename. The file naming convention adopted is:

<*PROJECT acronym*>_<*WP number*>_<Task number>_<*coverage or other content specifications*>_<*date*>_<version number>.<file extention>

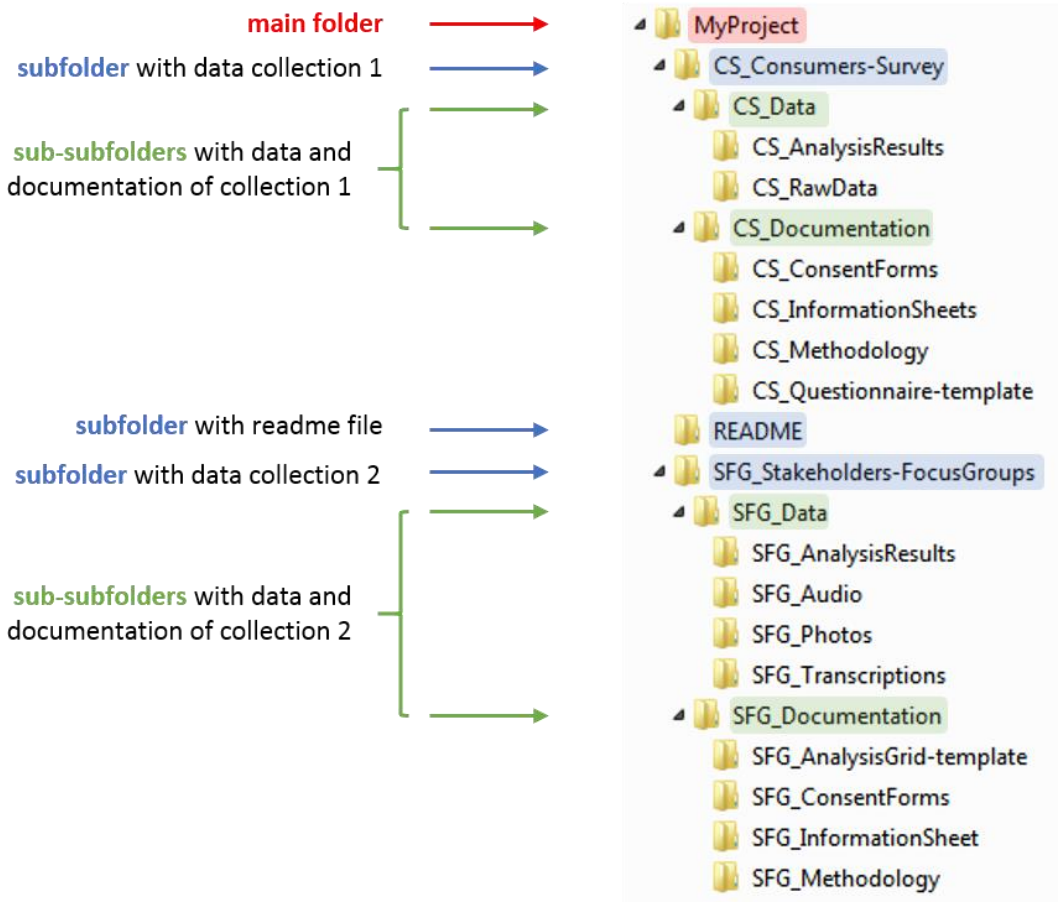
A user aware of this naming convention reading the filenames is informed that:

- the files in the folder “MyProject_WP3” belong to the project “MyProject”, specifically they are generated in WP3, by activities of subtask 3.2.1;
- the files contain anonymized (“anon”) transcripts (“trans”) of an interview (“intw1”) carried out by the researcher Mario Rossi (“MR”);
- the date of collection is 2nd of March, 2018;
- there are two versions of the data, the most recent is the “v02”;
- each data file has associated a file that presents further information about contents, methodologies and associated metadata (“[…]_README.rtf”).

The adopted naming convention, abbreviations and codes (such as “intw”, “trans”, “anon”, “MR”), the date in YYMMDD, and the use of dashes for task numbers must be described in the “README_folders-files-naming-conventions.txt” file, that is in the sub folder “README”.

Example 2

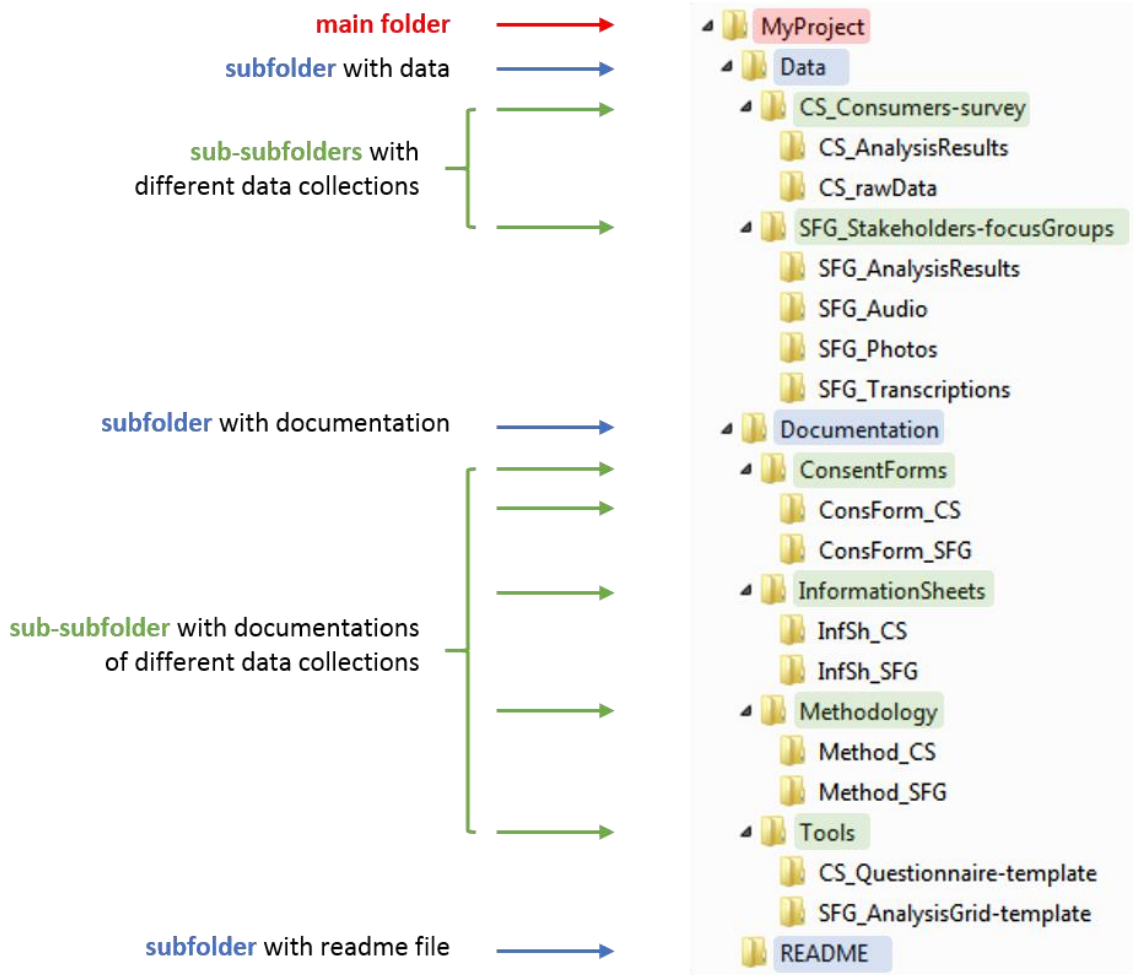
Folders structure organized as independent data collections



In this example, the main folder is named with the project acronym, data from different collections is saved in two specific subfolders. For each data collection there are sub-subfolders containing the data itself and the documentation, respectively. The adopted naming convention, abbreviations, codes, date, file naming and the structure of folders and folder contents must be described in a “README” file, saved in the sub folder “README”.

Example 3

Folders structure with separated data and documentation



In this example, the main folder is named with the project acronym, data from different collections is saved in subfolder "Data" and organized in sub-subfolders. The same structure is applied to the data documentation. The adopted naming convention, abbreviations, codes, date, file naming and the structure of folders and folder contents must be described in a "README" file, saved in the sub folder "README".