



Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Data Management Support (DMS)
Jürg Brendan Logue

RUTINBESKRIVNING SLU ID: [Skriv numret här]

2022-03-24

Guide to file organisation - Version 1.00

Note that this document will be updated on a regular basis.

Document history			
Version	Date	Changes	Made by
1.00	20210825	First draft of the document	Jürg B. Logue
1.00	20220301	Minor changes	Jürg B. Logue
1.00	20220324	Minor updates	Jürg B. Logue

Should you have any questions about or comments on this guide, please contact Jürg B. Logue (jurg.brendan.logue@slu.se) at SLU's Data Management Support (DMS).

Table of Contents

1. Introduction	4
2. File organisation	4
<i>2.1 Folder structure</i>	<i>4</i>
3. File naming.....	5
<i>3.1 The file naming convention.....</i>	<i>5</i>
<i>3.2 Elements in a file name</i>	<i>6</i>
4. File versioning.....	8
<i>4.1 File version control system</i>	<i>8</i>
5. README file.....	9

1. Introduction

Once you collect, generate, acquire, and, eventually, start processing data things can quickly become disorganised. Having a well-thought-out plan for how to structure and organise data, files, and documents will improve their management tremendously. Thus, it is vital to develop and implement processes and procedures with regard to **file organisation**, **file naming** and **file versioning**, and to do so as early on as possible. This will help to ensure usability, accessibility, traceability, reusability as well as quality and transparency of data, files, and documents.

2. File organisation

One of the essential components of successfully managing data is to establish a file organisation system. Such a system should be logical and hierarchical; it should represent the structure of your work. Doing so helps you organise files into a series of meaningful and useful groups with common properties. Thus, **group files into folders and structure folders hierarchically**.

2.1 Folder structure

When creating a file organisation system, think carefully about how your work will be structured. Often, your work can be divided into a number of meaningful, well-defined areas/categories/tasks (e.g., data, data analysis, results, documentation, etc.). Thus, systematically organise folders hierarchically by meaningful, logic areas/categories/tasks (see *Figure 1*).

Secondly, name folders after the areas/categories/tasks of work to which they relate. Folder names should be logical and concise as well as reflect the content of the respective folder (see *Figure 1*). When naming folders (and subfolders) it is important to follow the same set of rules as for naming files (see 3. *File naming*).

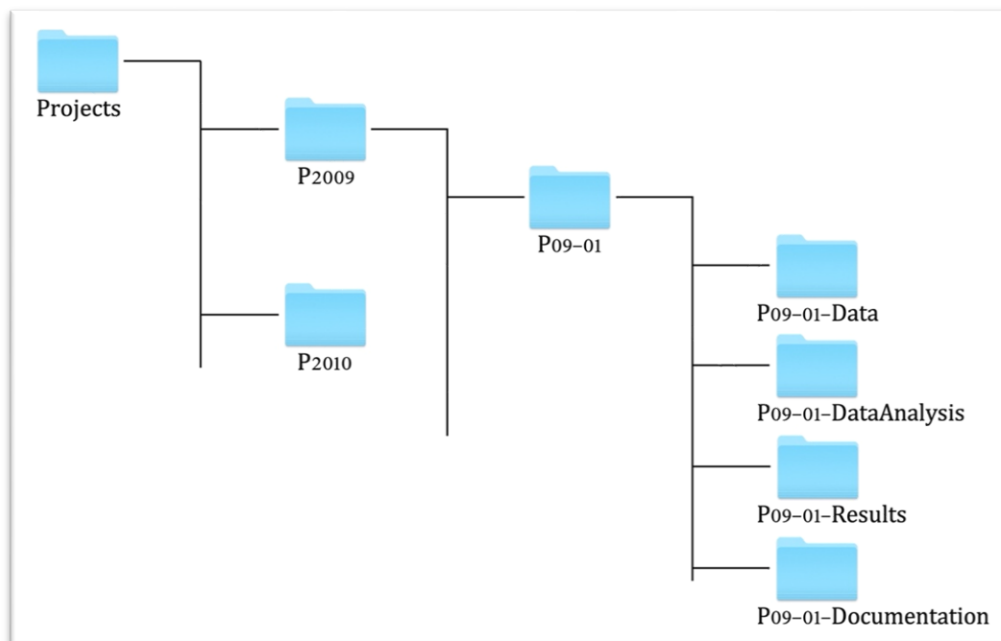


Figure 1 Visualisation of an example of how a file organisation system could look like. The 'Projects' folder contains projects started in 2009 ('P2009') and 2010 ('P2010'). The first, 'P2009', contains the project 'P09-01', which is subdivided into data, data analysis, results, and documentation.

3. File naming

A file name is a file's principal identifier. Good file names provide useful clues to the content, status and version of a file, uniquely identify a file, and help with classifying and organising files. Good file names, also, facilitate storage as well as increases readability, searchability, and understanding of files. Hence, **when naming files, you should do so in a consistent, logical, meaningful, and predictable manner.**

3.1 The file naming convention

Having a unified process for naming files is nothing else than using a **file naming convention**¹ (or a file naming system). Applying such a convention will help you find what you need, simply and quickly (i.e., in an efficient and effective way).

¹ A file naming convention is a framework for naming files in a way that describes what they contain and how they relate to other files. File naming conventions help staying organised and make identifying files easier. Note, it is strongly recommended to establish a file naming convention prior to collecting files or data in order to prevent a backlog of unorganised content.

Note that for a file naming convention to be of maximal use it is important that a logical file organisation system is in place (see 2. *File organisation* above) and that the convention is used also to name folders (and subfolders) within that system.

A file naming convention may include the following rules²:

1. Keep file names brief but meaningful.
2. Use only characters from the ‘POSIX portable file name character set’ when naming files (“A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h i j k l m n o p q r s t u v w x y z . _ -”).
3. Use (a sufficient number of) elements in a file name for easy searchability and identification (see 3.2 <i>Elements in a file name</i>).
4. Sort elements from general to specific (or from most to least important) (see 3.2 <i>Elements in a file name</i>).
5. Use hyphens (-) to separate elements within a file name.
6. Use capital letters to begin each element as well as delimit words within elements.
7. Include the date when naming files using proper date formatting (according to the format: YYYYMMDD).
8. Include personal initials in the file name when commenting on a document (according to the format: [first name initial][family name initial], “AA” for “Anders Andersson”).
9. Include the version of a document in the file name (especially when creating official documents and documents to be made available on the web) (according to the format: ”v0.10”, ”v0.20”, ”v1.40”, etc.; see 4. <i>File versioning</i>).
10. Include the language a document is written in in the file name (especially when creating official documents and documents to be made available on the web) (according to the ISO 639.1 standard; SV for Swedish, EN for English, etc.).
11. Specify the file format of the document (e.g., ”.docx”, ”.pdf”, ”.xlsx”, etc.).

3.2 Elements in a file name

Using elements in a file name allow for identifying the content of the document without having to open it as well as easy searching of a file (when – for instance – using the computer’s search function). Elements to consider using in a file naming convention may be:

² See the University of Edinburgh for a more comprehensive list of rules comprising file naming conventions: <https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions>.

- description of content,
- project name or number,
- name of creator,
- name of research team/department associated with the data,
- sample,
- date of creation (e.g., 20210825),
- version number,
- modification (e.g., raw/original, processed, analysed),
- location (e.g., Uppsala, Mora),
- type of data (e.g., field, experiment, modelling, interview),
- type of work (e.g., mass spectrometry, sequencing),
- name of funder, client, or contracting body.

You may, further, consider applying names of folders from your file organisation system as elements in a file name. Also, in order to keep file names brief, you may want to use standard and appropriate abbreviations for elements in the file name. In any case, to determine what elements to include and how they should be ordered within a file name, you need to think about how you want to organise, search for, and access files. However, depending on the nature of the document (i.e., working material, official documents, or documents to be made available on the web), different types of elements should be included:

- **Working material**³
 - **[Main folder]-[Name of working material]-[Date]-[Initials].[File format]**
 - Example #1: Methods-MicroscopySOP-20210825.docx
 - Example #2: Project0904-DA-WC-Rin-20210825.csv
 - Example #3: QualityAnalysis-Activity04-20210825-AA.xlsx
 - Example #4: Project0904-Manuscript-20210825-AA.docx
- **Official documents**³
 - **[Main folder]-[Name of document]-[Version]-[Language].[File format]**
 - Example #1: Project0904-Manuscript-v3.10-EN.docx
 - Example #2: Projekt0904-Rapport-v1.00-SV.pdf
- **Documents to be made available on the web**⁴
 - **[Name of document]-[Version]-[Date]-[Language].[File format]**

³ SLU's unit 'Archives, Information Governance and Records' provides an in-depth description of what working material and official documents constitute: <https://internt.slu.se/en/support-services/administrative-support/legal-affairs-data-protection-info-management/info-and-archives-mgmt/manual-research-material/management-preservation/>.

⁴ SLU provides instructions on how to publish documents on the web (only in Swedish): <https://internt.slu.se/riktat/administrativa-roller/webbpublicerare/webborganisation/publicera-dokument-pa-webben/>. SLU, further, recommends adhering to DIGG's guidelines on how to provide documents with a clear file name (only in Swedish): <https://webbriktlinjer.se/riktlinjer/9-ge-dokument-tydliga-filnamn/>.

- Example #1: QualityGuide-v2.00-20210825-EN.pdf
- Example #2: FomaZoonos-Rapport-v1.00-20210825-SV.docx

4. File versioning

Versioning is the process of managing different versions of files over time. Versioning can help differentiate between versions of and track changes in files. A working version control system also ensures that you always work with the most updated versions of your files⁵.

4.1 File version control system

A suitable version control system (or version control strategy) depends on whether files are used by single or multiple users, in one or more locations, and whether versions across users or locations need to be synchronised or not.

Version control can be done through one or more of the following manual and/or automated strategies:

- **recording the date in the file name** (manual) (e.g., Methods-MicroscopySOP-20210825.docx),
- **including the version in the file name** (manual) (e.g., Project0904-Manuscript-v3.10-EN.docx),
- **using a version control table documenting the file's history** (manual) (see *Figure 2*),
- **using the software's own version control application or add-ons** (automated) (e.g., XLTools or Colectica for Microsoft Excel),
- **using a version control software** (automated) (e.g., Git, GitHub, GitLab, Subversion, ArcGIS, electronic lab notebooks [ELN]),
- **using file-sharing services** (automated) (e.g., Microsoft SharePoint, Microsoft OneDrive, Google Docs, Dropbox).

⁵ See the Australian National Data Service (ANDS) for further information about why data versioning is important: [https://www.ands.org.au/working-with-data/data-management/data-versioning?ct=t\(andsUP_13SEP_2016\)](https://www.ands.org.au/working-with-data/data-management/data-versioning?ct=t(andsUP_13SEP_2016)).

Version control table – document history			
Version	Date / Last amended	Changes / Notes	Made by / Responsible
1.00	20210825	First draft	Anders Andersson
1.20	20210830	Minor revisions	Bertil Bertilsson
2.00	20210831	Substantive revision	Carl Carlsson

Figure 2. Example of how a version control table could look like.

Version numbering in file names or version control tables can be through discrete or continuous numbering, depending on minor or major revisions:

- Major versions are identified by integers (e.g., v.4.00; a major version or larger version represents a milestone, such as a document that is made available or submitted for archiving),
- Smaller versions are identified by decimal numbers (e.g., v.4.20; a smaller version is an ongoing document or material under development).

Note that work material that in the future leads to an official document (e.g., a strategy document, internal action plan, etc.) must contain a version control table. For more complex projects, laboratory information management systems (LIMS) or document management systems (DMS) may be considered. In any case, even if you are tracking changes and using version control, **always keep a copy of the original, unedited file/data safe and available.**

5. README file

Projects often develop over the course of many years and can include several colleagues/collaborators. To ensure that both the file organisation system, the file naming convention, and the file version control system are understood by everyone and months or years after they were initially drawn up, they should be documented in detail. The simplest way to do so is to create a README file. This README file should be kept with your files (preferably highest up in the file organisation system it documents).

With regard to the file organisation system, a README file should include a description of the system itself that is a description of the main folders (i.e., the kind of files that should be maintained in what folders; see Figure 3) and the system's structure. This will help you and others recognise which files belong into which folders.

Data	The main folder 'Data' shall contain the project's data files. It contains the 2 subfolders 'RawData' and 'ProcessedData'. These again contain subfolders reflecting the data's methodological origin.
DataAnalysis	The main folder 'DataAnalysis' shall contain working material generated by the various analytical procedures applied. Each analytical procedure is, moreover, accompanied by a file documenting exactly what analytical steps were undertaken during that procedure.
Results	The main folder 'Results' shall contain the final results that shall go into a report or publication.

Figure 3. Example of how a description of the main folders in the README file could look like.

When it comes to the file naming convention, a README file should include a description of the rules adopted, the elements used, and any abbreviations applied. This will help you and others to adhere to the convention.

Lastly, a README file should also include a description of the strategy or strategies applied when it comes to file version control. Describe how version control is executed, which strategies are used in connection with what files/documents.