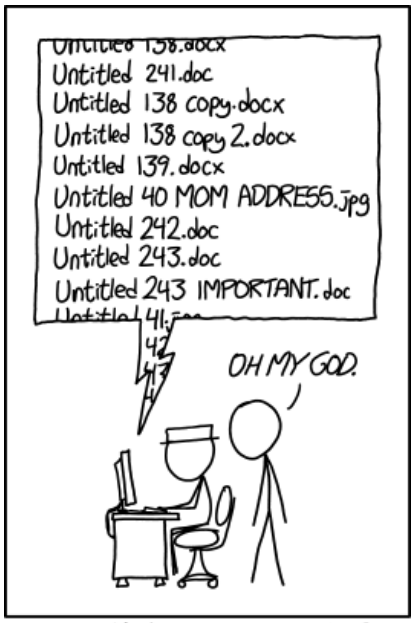
Developing a file naming convention

University of Glasgow

File naming conventions (FNCs) are rules which allow us to title electronic and physical folders, files and records in a consistent and logical way. This ensures that the correct records can be located, identified and retrieved from a filing system in a timely and accurate fashion.

The key to keeping things organized is to develop a file naming system early on in your research process. If you want to keep track of everything and avoid wasting time, you'll need to think about these issues before you start creating or managing files.



https://imgs.xkcd.com/comics/documents.png

Trying to organize and rename files in the midst of a project can be a much more intensive and tedious process.

This guide will take you through the steps involved in developing a file naming convention and explain the components of a good file name. There are also several examples of good and bad file names.

Please note that the information in this guide provides a basic introduction to file naming conventions. A specific file-naming convention that will work in the context of your research should be developed using this guidance as a starting point.

Research groups and those working on collaborative projects should develop and agree their file-naming conventions together.

New PGRs and ECRs joining existing research groups should check to see if the group they've joined already has file-naming conventions, or defined lists of elements for the creation of file naming conventions.

Developing a file-naming convention

Establishing an effective file naming convention is an investment of time and effort. It should be based on your needs as a researcher, and also on those of your colleagues or team. There are no perfect file naming conventions, but there are some basic rules that can help guide you. The following are a mixture of rules, tips and thoughts that will help you develop a meaningful file naming convention:

File (and folder) names should be short(ish) but meaningful

File names should be kept as short as possible while still retaining enough information to disambiguate the contents from other files. Long file names contribute to long file paths which increase the likelihood of error and are more difficult to remember and recognise.

In order to achieve this, you need to find the right balance of elements in your file naming convention. Too few elements and there will be insufficient information to allow you to distinguish between similar files. Too many and the file name becomes unwieldy and could be too long for some

operating systems: e.g. the maximum file path is 259 characters in Windows systems (and 248 characters for folders).

A good way to keep file names short(ish) without compromising on information is to use meaningful abbreviations. If you plan on using abbreviations in any of your file name elements, it is essential to document these – see the section on documentation.

Use relevant elements in your filename to provide description and context

File names should enable the file contents to be easily understood by a human operator reading the file name. The order of the elements should reflect the importance of that information, and the order should be consistent between files.

See the section for examples of elements that might be useful to include.

Avoid special characters and spaces

Some special characters have defined functions in some operating systems, so can cause problems when saving and searching for files. To avoid this issue, don't use special characters and spaces in file names. In addition to this, don't start or end your file names with hyphens or underscores (even though these are acceptable characters in filenames).

E.g. * : \ / < > | ” ? [ ] ; = + & £ $

Repetition and inclusion of redundant words

A common recommendation when developing a file naming convention is to avoid redundancy in file names and file paths as this can increase their length. However, it is important to consider whether your file names need to be meaningful outside of their current folder structure. If files will be moved in the future – shared with others, moved to a repository, moved to a different storage location – their contents still need to be clear.

E.g. /…/GroupMeeting/20041030Minutes.rtf

/…/protocols/BCA_ProteinAssay.rtf

The contents of the above files are perfectly clear in their current contexts. However, if moved or shared, it might not remain clear that the minutes were from a Group Meeting rather than a safety committee meeting or that the file named BCA_ProteinAssay contained a protocol rather than data or results.

Think about how you want to be able to retrieve the files

The way you want to retrieve your files will help to determine the right file naming convention for that type of file:

For example, if the file will be retrieved by date, the date element should appear first. If the file will be retrieved according to a descriptor, that element should appear first.

Dates

If you are using date as an identifier, make sure the format is stable, even over the span of many years.

Dates are a common cause of ambiguity due to the different formats and common usage worldwide.

To avoid any confusion between UK / US (or other) ordering systems, the ISO 8601 date format[1] YYYYMMDD should be used. This can be extended to include time (if this is relevant to your data) by adding the character T followed by the time in ISO format.

E.g. 20200430T062833 (YYMMDD-T-hhmmss)

https://imgs.xkcd.com/comics/iso_8601.png

Sequential numbering

When developing a file naming convention, try to plan for the longer-term. If your project is going to last for three years, you may end up with thousands of files.

If you're using a sequential numbering system, try to estimate the number of files you might accumulate and use leading zeros to make sure files sort in sequential order.

For example, if you anticipate 750 data files in an experiment or project, lead your number with two zeros, starting at 001.

If you anticipate 7500 data files, lead your number with three zeros, starting at 0001.

Version numbers

The version number of a record should be indicated in its file name by the inclusion of 'V' followed the version number and, where applicable, 'Draft' or 'Final'.

Some records go through a number of versions. For example, they start out as working drafts, become consultation drafts and finish with a final draft, which may then be reviewed and updated at a later date. It is important to be able to differentiate between these various drafts by giving them each their own number.

Where a version number is applicable, it should always appear in the file name of the record so that the most recent version can be easily identified and retrieved.

E.g. Org_Hier_2002_v2.xls

Beware software or tools which autosave your work (e.g. newer versions of Microsoft Office). If the autosave function is turned on, versioning has to be done consciously. To do this, save a new version of the file, prior to editing, and rename this to the version sequence. Then go ahead and make the changes in the new version.

[1] https://www.iso.org/iso-8601-date-and-time-format.html

Document and share your file naming convention

Having developed a file naming convention for your work, take the time to document it thoroughly. This will make it easy for you to use and understand, and also make it straightforward for others to understand and navigate the contents of your files (at the very least, your supervisor or PI will need to be able to do this).

You might want to include this documentation in a readme.txt file in the main shared folder. You will also need to provide an explanation of the file naming convention when you deposit data files in a repository for long-term storage and sharing.

If you are using abbreviations in your file names, it is particularly important to have a list of these available for clarity and consistency.

An example of simple file name documentation can be found in the examples section.

Anatomy of a file name

File names are comprised of a series of elements, clearly delimited from one another.

Elements are descriptive subunits which can be combined to form a useful filename. Examples of elements can be found in the table below:

Element
Date or date range
Researcher's name or initials
Project or experiment name or acronym
Location / spatial coordinates
Version number of file
Type of data
Language
Conditions
File extension
Sample name
Analysis type
Interviewee name / pseudonym
Instrument / equipment identifier
Brief descriptor

There are two main ways in which elements can be distinguished from one another in a file name:

With underscores: eg 20200423_expt004_repeat2.xls

With CamelCase: eg 20200423Expt004Repeat2.xls

Both have their uses:

Underscores make reading the filename easier for humans and clearly delimit elements in a machine-readable way.

CamelCase saves on characters where the length of the filename is becoming problematic, however, some file systems are case sensitive and some are not, so relying on case to differentiate filename elements could be risky.

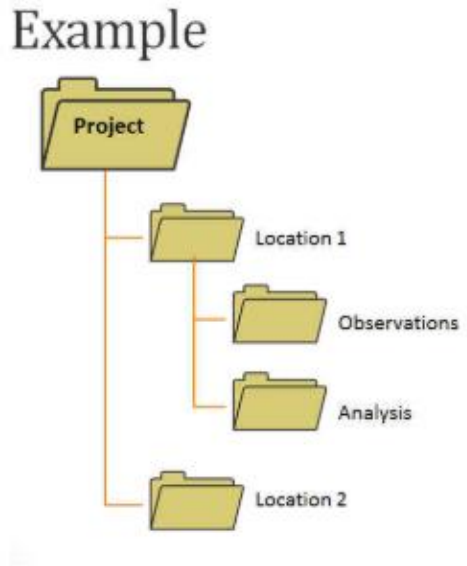
File organisation

As for file naming, with folder naming consistency is key. Organize files in a way that makes sense within the context of your project but would also make sense to someone who was not intimately familiar with your project. Your supervisor or PI may need to be able to navigate your files and folder structure at some point in the future, without reference to your knowledge of how you arranged things.

How files are nested in directories can be dependent on the number of files you are working with, and what aspect of those files is most important for analysing or re-using the information in them.

For instance, if you have hundreds of thousands of image files collected over many years from many different locations, you may want to organize first by year, then month, then location. You could also organize them entirely by date and include the location in the filename. Alternatively, organize by location, and only include the date in the filename.

A hierarchical structure which can be easily adapted to a range of situations is illustrated below:



As with file naming, it is important to develop a folder structure which makes sense with your project and data.

Batch renaming

You may already have a lot of data collected for your project and wish to organize and rename these files for easier data management. Alternatively, you may be collecting files from an automated instrument which does not allow you to pre-set filenames.

If you have too many files to rename them all by hand, try one of the following applications for renaming your files:

- Bulk Rename Utility (Windows, free)
- Renamer 6 (Mac)

You can also bulk rename and manipulate files by scripting in the programming language of your choice, using PowerShell (Windows) or the Finder (Mac).

Some final thoughts on file naming

Collaborations

If you are working on a collaborative project, make sure all collaborators are using the same principles to organize and name files!

Sensitive data

If you are working on sensitive data, don't include personal or identifying information in your file names. It is, however, useful to include an element in the file name which allows you to identify that there is identifying information in the file contents. This makes accidental sharing of personal information much less likely.

Future stability

Keep in mind that files can be moved and, without the inherited folder structure, important descriptive information about the contents could be lost. Consider whether a filename would be meaningful outside of your chosen directory structure, and if not, how important the loss of that context would be, e.g., if the date a file was created is important, include it in the filename rather than just the directory name.

One size does not fit all

A single file naming convention may not suit the needs of all the data or files you will collect or create during the course of your research. It is perfectly reasonable to develop separate conventions for different types of files or data – just make sure you document them appropriately.

File extensions

The file extension should match the file format used. Do not change the extension code!

Examples of file-naming conventions

Example 1: File naming convention for instrumental datasets in the Centre for Environmental Data Archive (CEDA):

instrument_[location|platform]_YYYYMMDD[hh][mm][ss][_extra][_cor#].ext

- `instrument` - is the instrument name (full or shortened) or model name
- `location|platform` - is the location name (full or shortened) or the name of the platform on which the instrument is deployed
- `YYYYMMDD` - is the date on which measurements were taken, e.g. 20151118
- `[hh][mm][ss]` - is the time of day specified (optional).
- `[_extra]` - additional code to define different range resolutions etc.
- `[_cor#]` - the file is a corrected version of a previously released file
- `.ext` - will normally be .nc (NetCDF) or .na (NASA Ames)

Example 2: Institute of Museum and Library Services file naming convention oral history interviews in the Oral History in the Digital Age project:

interviewer_interviewee_[status]_[part#]_[date].format

Examples:
boyd_johnson_pres_01_20120801.wav
macdowell_benberry_mez_02_20120801.h264

[status]
Raw = capture format
Pres = preservation copy
Mez = mezzanine copy (working copy)
Ed01 = edited version #1
Dis = dissemination copy, e.g. for web

In this example, the status was the part that the researchers felt needed documenting – there are five different options for this element, each denoting a different version of the same recording.

Example 3: Stanford University Libraries case studies in file naming

STANFORD UNIVERSITY LIBRARIES

Stanford Libraries have shared two case studies in file naming:

The first is an example of an organised and thorough method used by a research group to name a large set of image files: https://library.stanford.edu/research/data-management-services/case-studies/case-study-file-naming-done-well

The second contains real-life examples of the problems that can arise when good file naming choices are not made: https://library.stanford.edu/research/data-management-services/data-best-practices/case-study-file-naming

Example 4: Example of a digital photo file naming convention

Professional photographers also use file naming conventions to organise their photos. A photographer may take thousands photos in a single shoot, and they do not depend on file names produced by their camera, or rely on folder structures. Rather, they typically use a file naming convention, such as: [Date] – [place or event] – [number] – [comment].

Examples:

- 2011.11.11-kampala-riot-000001.tiff
- 2011.11.11-kampala-riot-000002.tiff
- 2011.11.11-kampala-riot-000003.tiff
- 2011.11.11-kampala-riot-000004.tiff
- 2011.11.11-kampala-riot-000004-cropped-slider660x510.jpg

As you can see, the photos above relate to riots that took place in Kampala on 11 November 2011. They were shot in TIFF format. The last photo is derivative of the previous one: it's an image cropped for the slider of the NGO's website. Even if there are tens of thousands of photos in the same folder, it's easy to filter for "kampala" and "riot".

Photography software like Adobe Lightroom or Adobe Photoshop allows you to batch rename files as above.

References and acknowledgements

This guide has drawn on information from a range of similar guides from a wide range of institutions. These guides contain many more examples of good and bad file naming practice and are recommended for further or alternative reading.

Purdue University Libraries and School of Information Studies 'File naming conventions'
https://guides.lib.purdue.edu/c.php?g=353013&p=2378293

Stanford University Libraries 'Best practices for file naming'
https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming

University of Edinburgh 'Naming conventions'
https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions

Smithsonian Institute 'Smithsonian Data Management Best Practices'
https://library.si.edu/sites/default/files/tutorial/pdf/filenamingorganizing20180227.pdf

Princeton University Library 'File naming'
https://libguides.princeton.edu/c.php?g=102546&p=930626

Huridocs 'File naming conventions: why you want them and how to create them'
https://www.huridocs.org/2016/07/file-naming-conventions-why-you-want-them-and-how-to-create-them/