Best Practices for File-Naming

July 2019
Version 2.0

Contents

Purpose

Just like paper files, electronic files need to be well-organized and labeled correctly so that they are identifiable and accessible by all employees. Efficient management of electronic records begins with accurate file-naming. This is especially important for government offices in order to comply with legal requirements to ensure the availability, integrity, accessibility, and, if appropriate, the confidentiality of public records. General Statute §132-8.1 provides that each agency shall establish and maintain an active, continuing program for the economical and efficient management of records of that agency.[1]

This document is intended to be used primarily by individuals creating records on a daily basis. The rules discussed below are best practices and therefore highly recommended; however, every rule may not be relevant to every office. Regardless, this document should provide a foundation for developing a consistent and easy to use file-naming standard to be implemented in any office. The following file-naming convention incorporates best practices in electronic records management and information technology. This document explores the general characteristics of records, how records are used and referenced, and the file-naming rules that should be applied to all electronic records.

Default File Naming Explained

An electronic record is machine-readable, a series of 1s and 0s and requires hardware and software to be accessed and read. Organization is especially important so that these records can be found and retrieved. Electronic records include documents, spreadsheets, databases, images, video, and audio, as well as other file types. If not managed, a computer assigns a unique name for these files when saved, but these names do not provide a context for the file, nor are they logical from a human-readable perspective. For example, the default file name for a Microsoft Word document consists of the first few words written on the first page. Imaging devices automatically name images with sequential numbers. These types of file names do not promote accessibility and ease of identification. For ease of explanation, the examples in these guidelines will focus primarily on documents; however, the standards outlined below apply to all file formats. In addition, these guidelines frequently use “record” and “file” interchangeably, as the records discussed here are electronic records and, therefore, generally files on a computer.

Records Will Be Accessed by Others

Many offices utilize a network server to store files so that they are accessible from multiple locations by various people. This requires that file names (as well as folder structures) make sense to more than just the files’ creator. A file name should be clear to everyone in the department or agency in which the file was created. A record should be distinguishable from files with similar subjects as well as different versions of the same file.

When other individuals access a record, they may be using different operating systems (Microsoft Windows, Mac OSX), different versions of the system (e.g. Windows 7, Windows 10), or different software (e.g.

[1] https://www.ncleg.net/EnactedLegislation/Statutes/PDF/BySection/Chapter_132/GS_132-8.1.pdf

Microsoft Word, OpenOffice.org Writer, Notepad); therefore, it is important to follow rules that will allow a file to be recognized in as many different environments as possible.

Rule #1: Avoid using special characters in a file name such as \ / : * ? " < > | [ ] & $, .

Computers and other electronic devices frequently use the characters listed above for specific tasks in an electronic environment. For example, a forward slash is used to identify folder levels in Microsoft products, while Mac operating systems use the colon. Periods are used in front of file-name extensions to denote file formats such as .jpg and .doc; using them in a file name could result in lost files or errors. Using these reserved characters can cause the system to misinterpret or be unable to read the file.

Rule #2: Use underscores instead of periods or spaces

As mentioned above, periods already have a specific function in a file name, which is to tell the computer program where the file-name extension begins. Spaces are frequently translated in a web environment to be read as "%20". For example, if it were available online,

Naming tutorial.doc

would appear as

Naming%20tutorial.doc

This alteration can cause confusion in identifying the actual file name. Spaces in file names can also cause broken links, because word processing tools like Microsoft Word, and email clients like Microsoft Outlook, recognize spaces as an opportunity to move to another line. Therefore, a link to

\\Ah1\Intranet\ar\naming conventions

could become

\\Ah1\Intranet\ar\naming
conventions

Rule #3: Err on the side of brevity

Different operating systems and software have different limits on the acceptable length of file names and file paths. Some systems allow up to 256 characters, while others allow far fewer. Keeping file names short can help prevent any future issues. Generally about 25-30 characters is a sufficient length to capture enough descriptive information for naming a record.

Rule #4: The file name should include all necessary descriptive information independent of where it is stored

Files are frequently copied to other folders, downloaded, and emailed. They may be migrated to newer storage, or moved permanently if transferred to the archive. It is important to ensure that the file name, independent of the folder where the original file lives, is sufficiently descriptive. Electronic records are usually organized in a series of folders. For example:

World_War_I\Posters\Owens\0001.tif

While this is a very organized way of storing records, it is only efficient as long as the files stay in their original folders in their original context. When multiple staff work on a project or if staff are multitasking, it is very easy to misfile a document. As soon as "0001.tif" is copied to another folder, or emailed to an agency, the context provided by the folders in which the document is nested is lost. Context is particularly important in legal situations because it provides authenticity and trustworthiness of the record. Losing the context of a record can possibly compromise its trustworthiness and therefore its validity or admissibility for court proceedings. Additionally, if the file has the same name as another file on the system in a different folder, once it is pulled out, the computer could mistake it for the same file and write over one of the files.

For instance, if the following files were pulled out of their appropriate folders, they would appear to be the same file:

World_War_I\Posters\Owens\0001.tif

World_War_I\Posters\RedCross\0001.tif

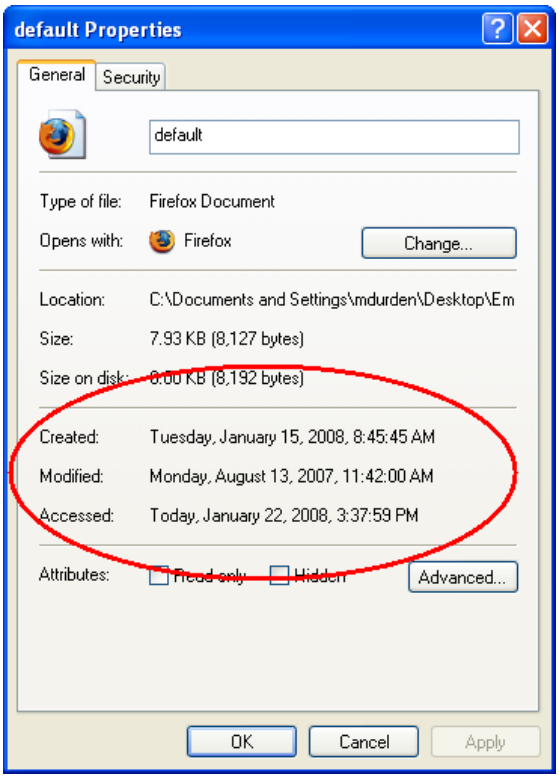
Incorrect: 0001.tif

Correct: wwI_poster_owens_0001.tif

Rule #5: Include dates and format them consistently

The point of organizing an agency's electronic records is to enable accessibility not only by current users, but by future users as well. Records retention schedules are applied to electronic records just as they are to paper records. Having the relevant date associated with the file is essential. Though many operating systems store this information with the file, as users move the file among folders and computers and as the file is re-saved as revisions are made, those dates change. A file could have a "created on" date that does not accurately reflect when they were created.

For example, the image to the right displays the properties for a file named 'default'. The date modified appears to be before the date created. This disparity appeared when the file was copied from a server to the desktop of a PC.



The best way to prevent confusion is to embed the relevant date (the date that the file was created or revised) in the file name itself.

Some workgroups might find it more useful to have the date at the start of the file name, while others might prefer it at the end. Either way, it is a useful sorting tool when the files are organized. Just be sure to keep it consistent.

The best way to list the date is based on an international standard – ISO 8601.[2] ISO 8601 specifies numeric representations of date and time to be used in electronic format. The international standard date notation is:

YYYY_MM_DD or YYYYMMDD

YYYY is the year, MM is the month of the year between 01 (January) and 12 (December), and DD is the day of the month between 01 and 31. For example, January 5, 2008 is written as 2008_01_05 or 20080105.

This format allows ease of sorting and comparing files by date and prevents confusion with other date formats (especially in other formats that use just two digits for the year).

For example, this document could be named:

filenaming_20080507

to reflect that this draft was last edited on May 5, 2008.

Rule #6: Include a version number on documents to manage drafts and revisions more easily

A file will frequently have multiple versions, especially when it is created by a workgroup. Specifying the version of a file can help you quickly identify the most accurate or most definitive version of the document.

The easiest way to do this is to use the letter "v" to represent "version number." Then, "v01, v02, v03" can be added as needed to a file and the main file name can stay the same. This is much more effective than other common additions like "update," "new," "old," etc. An exception to this rule is using "FINAL" to indicate the final version of the document. When using "FINAL", be sure to use it *instead* of the version number, rather than in addition to it.

Rule #7: Be consistent.

The most important rule of file naming is to be consistent. Some choices will need to be made about organization that affect the entire workgroup – where to include the date, what abbreviations to use, etc. Regardless of what the group decides, it is only effective if everyone follows the rules consistently.

There Will Be Exceptions

One notable consideration is the batch-scanning process. This process typically relies on a program that sets its own parameters on the file names allowed. If this is the case, take advantage of the folder hierarchy and, when possible, apply the rules outlined here to folder names. Avoid spaces and special characters. Be consistent throughout the project; consider developing a file-naming standard for all batch-scanning projects.

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

Remember: this document is not going to apply absolutely to every situation; it should be used as a guide to encourage discussion in offices and workgroups to develop file-naming practices that work best in those specific environments.

Additional Resources

Controlled Vocabulary. "**Recommendations for Limitations on Image Filenaming."** http://www.controlledvocabulary.com/imagedatabases/filename_limits.html

"Document Naming Conventions." Information Management Branch. Alberta Government. February 2018. https://www.alberta.ca/assets/documents/im-naming-conventions.pdf

Kuhn, M. "A summary of the international standard date and time notation." http://www.cl.cam.ac.uk/~mgk25/iso-time.html

Minnesota Historical Society. "File Naming." March 2004. http://www.mnhs.org/preserve/records/electronicrecords/docs_pdfs/erfnaming.pdf