Best Practices for File-Naming

Just like paper files, electronic files need to be well-organized and labeled correctly so that they are identifiable and accessible by all employees. 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. G.S §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.

Efficient management of electronic records begins with accurate file-naming. 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.

An electronic record is machine-readable, meaning that it 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. 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. For example, the default file name for a Microsoft Word document consists of the first few words written on the first page. Images are frequently labeled 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, this document will frequently use "record" and "file" interchangeably, as the records discussed here are electronic records and, therefore, generally files on a computer.

Keep in mind that 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.

Records will be accessed by others.

As mentioned earlier, good file names are essential to accessibility. 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 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 Vista, Windows XP), or different software (e.g. 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. \ / : * ? " < > | [ ] & $, .

The characters listed above are frequently used 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.

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,

Naming tutorial.doc

would appear as

Naming%20tutorial.doc

if it were available online. 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 e-mail 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

Another difference that is found among operating systems and software is the acceptable length of file names. Some systems allow up to 256 characters, while others allow far fewer.

Rule #3: Err on the side of brevity.

Generally about 25 characters is a sufficient length to capture enough descriptive information for naming a record.

Records will be moved from their original location.

Files are frequently copied to other folders, downloaded, and emailed. 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 works on a project (or staff is in a position to get distracted) 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.

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

Rule #4: The file name should include all necessary descriptive information independent of where it is stored:
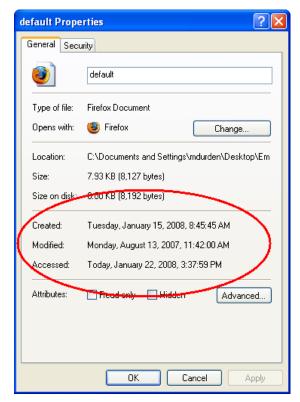
> *Incorrect*: 0001.tif

> *Correct*: wwI_poster_owens_0001.tif

Government records will have an associated retention schedule.

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 dates that do not make much sense to its original creation.

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.

Rule #5: Include dates and format them consistently.

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. 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.

Records will have multiple versions.
A file will frequently have multiple versions, especially when it is created by a workgroup.

Rule #5: To more easily manage drafts and revisions, include a version number on these documents.

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. This can be helpful to quickly identify the most accurate version of the document. When using "FINAL", be sure to use it *instead* of the version number, rather than in addition to it.

Rule #6: Be consistent.

The most important rule of file-naming is to be consistent. Some choices will need to be made about organization that affects 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.

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 for best in those specific environments.

Still have questions? Contact us.

Department of Cultural Resources | Office of Archives and History

Government Records Branch

Mailing Address: 4615 Mail Service Center | Raleigh, NC 27699-4615

Physical Location: 215 N. Blount Street | Raleigh, NC 27601

Phone: 919-807-7350 | **Fax:** 919-715-3627 | **Courier:** 51-81-20

E-mail Address: records@ncmail.net

Additional Resources

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

"Naming Conventions for Electronic Documents." Information Management. Alberta Government.
August 2005. http://www.im.gov.ab.ca/publications/pdf/DocumentNamingConventions.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