File Naming Guidelines for Digital Materials

"It is impossible to accurately predict all of the situations in which a file might be used. Therefore, in the interest of preserving access to digital files, we choose file name components that are least likely to cause a problem in any environment. File names should provide context and be easily understood by humans and computers, now and in the future."

From Association of College & Research Libraries TechConnect blog
http://acrl.ala.org/techconnect/?p=2607

Why is file naming important?

- Facilitates access and retrieval
- Provides version control
- Provides sorting control
- Prevents accidental overwriting
- Aids system compliance and portability

When assigning filenames, it is important to consider the following:

- What naming elements to use and in what order
- Controlled vocabularies for those elements, including abbreviations
- What constitutes a new version
- Documentation of naming conventions for each project

Overall, whatever file naming convention you develop, be sure to apply it consistently throughout a project.

Operational Considerations

Length

Thankfully, the days of 8.3 are behind us, but brevity is still a valuable quality to have in a filename. The Windows operating system imposes a maximum file path length of 260 characters, which includes all the components of a path, such as the server name, drive, folders and the file name and extension. As networks and storage continue to grow and layer, you can see that it quickly adds up.

Blank example of 89 characters–

\\server\sharedrive\foldername\subfoldername\subsubfoldername\filename_desc_date_vers.ext

Specific example of 113 characters-

\\SRC-Douvan\SRO\projects\ACL5\1 - Management\Budget\2011_Aug\08-0025R03 ACL Wave 5 Workscope Transmittal Memo.doc

Characters

Use only alpha numeric characters for your filenames. Avoid punctuation characters altogether, with the exception of hyphens(-) and underscores(_). This is because some characters such as \ / : * ? " < > | are reserved for use by the operating system itself.

Spaces

There are a number of reasons not to use spaces:

- Not all operating systems tolerate spaces equally.
- Command line and scripts need to escape spaces, which creates extra effort for programmers and can cause inaccuracies.
- URLs need spaces to be escaped or replaced, such as with %, which is even more unreadable than no space.

Even if you think your files won't be acted on in this way, you don't really know for sure. For instance, many preservation activities could require scripts and URL protocols.

Naming Elements

These elements are to be considered for accuracy in retrieval of files, readability, sorting capabilities, and version control. Place the elements in the most logical order based on retrieval methods. For example, use the date first on files that are time specific or reoccurring, and use the title for files that are more subject specific.

Title

This is purely descriptive and free form. Think of the intended audience for retrieving this file and how they might search for it. Keep it brief, but meaningful.

Version

This is one of the most important components of a filename to consider. It will be shaped by the number of people working with the file, the editing software, and file system being used. Often times it is the only way to know which file is the most current. Always use some way to denote what version it is, whether it is a working document or a final document, etc. Possibilities include:

\meetings\minutes_20120120_v01
\meetings\minutes_20120120_rev01
\meetings\minutes_20120120_working
\meetings\minutes_20120120_final (don't use final unless you mean it!)

It is helpful to put the version information at the end of the filename so as not to obscure more important retrieval information. If you anticipate more than ten versions of a document, be sure to use a zero as placeholder.

Author

If there is more than one person working on a file, a last name and first initial can be a helpful identifying component along with the version.

Date

Relying exclusively on file properties for this information may cause problems because an operating system, software program, or user may act on a file in unpredictable ways. Indexing, opening, or unintentional modifications will change the date and time properties of a file.

Dates are useful for sorting purposes. If there is a large volume of files that rely on chronological organization, the date element is necessary.

The standard format for dates should always be **YYYYMMDD** with or without a separating character (**YYYY_MM_DD**). You can use just **YYYY-MM**, or **YYYY** or **YYYY-YYYY**, but keep it consistent. See ISO 8601 for more information.

Document type

Files may fall into a specific type of document, i.e. email, memo, minutes, schedule, agenda, budget, report, etc. If so, including this element in the name may facilitate sorting as well as retrieval.

Unique identifier

If you are working with a large number of similar files, sometimes a unique identifier is the only way to easily and efficiently describe the file. Be sure to document how the identifier is constructed. For instance:

4110240-SS.pdf

4	110	240	S	S
Grade	School ID	Individual student	season	type

Include the controlled vocabularies (the directory of school names and seasons, etc.) with your documentation.

Other

You can use numbers as prefixes to sort folders, if that is important:

8/15/2013 filename_guidelines_v4.docx

01 Project Management
02 Sampling
etc….