

Increasing Findability: Document Naming Conventions

Sofia Empel, MLIS, CRM, CDIA+¹

© 2010 Sofia Empel. All rights reserved.

Keywords: naming conventions; findability; electronic documents; e-docs; information retrieval; search, browse.

Even a well defined taxonomy in a shared data system² can not compensate for poorly named files that must be opened and then viewed to determine content; furthermore, documents that are named uniformly allow an additional point of access through the “search” function of the operating system or application. Standardized names also allow the user to “browse” through a list in order to find the target document or group of relevant documents. An effective means of improving the findability of electronic documents, or e-docs, is through the use of naming conventions (Gingrande 2004; Inmon, O’Neil and Fryman 2008). Well-defined naming conventions increase the findability of e-docs and support information retrieval techniques such as search and browse.

There are other reasons to standardize document names. Naming conventions can keep track of document versions which are difficult to control without versioning software. Names also affect the logical sequence of the sorting and ordering of documents. For example, “nothing” comes before something— “policy” will sort before “policy brief”—and numbers sort from the left to right so “zero filling” (see Table 1) may be necessary (Alberta Government 2005).

Table 1: Numerical Sorting

Sorting without zero-filling	Sorting with zero-filling
300	0300
3000	0320
3005	0330
320	3000
330	3005

¹ Sofia Empel can be contacted by email at sofia@recordsupdate.com.

² Taxonomy in a shared data system refers to the hierarchical structure of electronic folders on an organization’s shared drive system.

Further, naming conventions not only aid in the identification of e-docs for discovery and legal hold orders, but the date in the file name also allows e-records to be associated with an organization's retention schedules, or the rules that determine when records can be destroyed in accordance with legal, regulatory, statutory, fiscal, and operational requirements.

A consensus on the required format of standard terms and commonly used names should be determined by a group of interested users from across business units (Bedford and Morelli 2006). The "main" name of the e-doc should be logical and consistent with the business unit's terminology (Kremer, Kolbe and Brenner 2005). For example in the legal environment, e-doc names may include descriptions such as client name, matter number, or type of document. Additionally, a case e-folder contains e-docs from the same case, so the name of an e-doc will be more meaningful than the name of the case folder (Cunningham and Montana 2006). Another example might be the naming of trademark e-docs: the registration number from the trademark office might be more useful in identifying each e-doc rather than simply the type of document. Since users most often classify information according to "taskonomy," or the task at hand, this should also be considered when determining a naming convention (Dougherty and Keller 1982).

Consistent standards should be determined for the following components of an e-doc name: use of a specified separator; use of abbreviations; naming elements and their order; and the definition of a new version number. Name length which includes the total path must also be taken into account, particularly if an e-doc will be moved to a different folder at any point in the future. Names of e-docs should make sense to both the author and to the other users. The document's name should be descriptive enough to identify it without viewing its content. Finally, naming conventions provide not only unique identifiers for each digital document, but they also permit file exchange if standards for cross-platform compatibility are observed (Riecks 2009).

Table 2 below lists the nine best practices for developing naming conventions for e-docs in shared data systems and the explanations of why these rules are important (Alberta Government 2005; Department of Cultural Resources 2008). Adhering to these best practices can be very effective in helping the user find the target or relevant group of documents through either the search feature or by browsing through a list of documents. The findability of documents improves both user access and organizational compliance and for these reasons naming conventions are well-worth the user's extra effort.

Table 2: Document Naming Conventions

Rules	Explanations
Always include the date as first 8 characters of the file name	The date should be in standard international notation with no separators as follows: "YYYYMMDD". The date can be used to apply the company's retention schedule.
Example: 20091202	
Limit file names to: A - Z a - z 0 - 9 underscore (_) period (.)	These characters provide maximum interoperability between operating and storage systems. Other forms of punctuation may have different uses in various systems causing confusion.
Example: 20091202_Information	
No Special Characters	The following characters are used for specific tasks in an electronic environment and should not be used in file names: \\/: * ? < > [] & , .
Incorrectly named file would appear on Web as: 20091202%20Information	
Use a single period (.) as a separator followed by the 3 letter extension at the end of the filename	Operating systems either partially or wholly rely on the extension to determine which application opens the file. This format ensures backward compatibility with some legacy systems. Some email applications consider a file with 2 periods to be a virus or a worm so your attachment will not be sent to the recipient.
Example: 20091202_Information.doc	
Use underscores (_) instead of periods or spaces as a separator	Periods are used for file extensions and spaces are translated to "%20" in the Web environment. Spaces in file names can cause broken links.
Example: 20091202_Information_Retrieval.doc	
Approximately 25 characters is sufficient length	Some systems allow up to 256 characters for file names; however, the length should only be enough to be descriptive.
File names should be independent of folder names	Files are frequently copied, moved, or downloaded to other folders; therefore, the file name should be independent of the folder.
Include version number after the main file name	Version names should be listed as "V01", "V02", etc. An exception is the word "FINAL" with no version number to easily identify the last version.
Example: 20091202_Information_Retrieval_V01.doc	
Be Consistent	Use rules determined by the business unit including what abbreviations to use.

Works Cited

- Alberta Government. 2005. *Naming Conventions for Electronic Documents*. Information Management, Branch Government and Program Support Services: Edmonton, Canada.
- Bedford, Diane and Jeff Morelli. 2006. Introducing information management into the workplace: a case study in the implementation of business classification file plans from Sector Skills Development Agency. *Records Management Journal* 16(3): 169-75.
- Cunningham, George and John Montana. 2006. *The Lawyers Guide to Records Management and Retention*. American Bar Association: Chicago, Illinois.
- Department of Cultural Resources. 2008. *Best practices for File-Naming*. Government Records Branch of North Carolina: Raleigh, North Carolina. Revised May 7, 2008. Available at http://www.records.ncdcr.gov/erecords/filenaming_20080508_final.pdf.
- Dougherty and Keller. 1982. Taskonomy: a practical approach to knowledge structures. *American Ethnologist* 9(4): 763-74.
- Gingrande, Arthur. 2004. The value of indexing. *Edoc*. Available at <http://www.aiim.org/ResourceCenter/Archive/Article.aspx?ID=27718>.
- Inmon, William, Bonnie O'Neil and Lowell Fryman. 2008. *Business Metadata*. Morgan Kaufman Publishers: Burlington, MA.
- Kremmer, Stefan, Lutz Kolbe and Walter Brenner. 2005. Towards a procedure model in terminology management. *Journal of Documentation*. 61(2): 281-295.
- Riecks, David. Recommendations for limitations on image file naming. Available at http://www.controlledvocabulary.com/imagetdatabases/filename_limits.html. Accessed December 1, 2009.