

File Naming

Summary

A file name is the chief identifier for a record. In the world of electronic records, the record's file name provides metadata that places the record in context with other records, records series, and records retention schedules. In most organizations, the policy for naming a file (and hence a record) is left to individuals or to groups of individuals (e.g., departments, committees). Consider establishing an agency-wide file naming policy that complements your electronic records management strategy.

Consistently named records foster collaboration based on mutual understanding of how to name files and use file names (including the file name metadata). Consistently named records also help you to meet your legal requirements. Legally, your records must be trustworthy, complete, accessible, legally admissible in court, and durable for as long as your approved records retention schedules require. Records that are consistently and logically named are easier to manage to meet these requirements.

In other words, with each staff member consistently naming electronic records, another staff member will be able to look at a record's file name and use the information in the record's file name to recognize the contents and characteristics of the record and to make decisions about the record. For example, a staff member could see that "HF0035broch96/97P.pdf" is a brochure about a House bill (HF0035) in the 1996/1997 session that is available to the public.

Legal Framework

For more information on the legal framework you must consider when developing a file naming policy, refer to the *Introduction* and Appendix D of the *Trustworthy Information Systems Handbook*. Also review the requirements of:

- Official Records Act (Minnesota Statutes, Chapter 15.17) (available at: <<http://www.revisor.leg.state.mn.us/stats/15/17.html>>), which mandates that government agencies must keep records to maintain their accountability and specifies that the medium must enable the records to be permanent. It further stipulates that you can copy a record and that the copy, if trustworthy, will be legally admissible in court.
- Records Management Act (Minnesota Statutes, Chapter 138.17) (available at: <<http://www.revisor.leg.state.mn.us/stats/138/17.html>>), which establishes the Records Disposition Panel to oversee the orderly disposition of records using approved records retention schedules.
- Minnesota Government Data Practices Act (MGDPA) (Minnesota Statutes, Chapter 13) (available at: <<http://www.revisor.leg.state.mn.us/stats/13/>>), mandates that government records should be accessible to the public unless categorized as not-public by the state legislature.

- Uniform Electronic Transactions Act (UETA) (Minnesota Statutes, Chapter 325L, available at <<https://www.revisor.leg.state.mn.us/statutes/?id=325L>>) and Electronic Signatures in Global and National Commerce (E-Sign), a federal law (available at: <<http://thomas.loc.gov/cgi-bin/query/z?c106:S.761:>>). Both UETA and E-Sign address the issues of the legal admissibility of electronic records created in a trustworthy manner and the application of the paper-oriented legal system to electronic records.

Key Concepts

As you develop your file naming policy, you will need to be familiar with the following:

- Differences among file names, file paths, and addresses
- Common file name elements
- General challenges in file naming
- Internet file naming protocols
- Domain names
- URL protocols
- Challenges in Internet-based file naming
- General file naming issues

Differences Among File Names, File Paths, and Addresses

A *file name* is the name of the file as it stands alone. The *file path* shows the location of the file. For example, the file “CommitteeAMinutes021401.doc” might be stored in a series of nested directories for all committees as:

“X:Committees/CommitteeA/Minutes/2001/February/CommitteeAMinutes021401.doc.” An

address describes the location of a file delivered on the Internet. For example, a map of a public park named Smith Park might have the following address: “<http://www.parks.org/smith.html>.”

Common File Name Elements

When developing your file naming policy, you may wish to include some of the following common elements:

- Version number (e.g., version 1 [v1, vers1])
- Date of creation (e.g., February 24, 2001 [022401, 02_24_01])
- Name of creator (e.g., Rupert B. Smith [RBSmith, RBS])
- Description of content (e.g., media kit [medkit, mk])
- Name of intended audience (e.g., general public [pub])
- Name of group associated with the record (e.g., Committee ABC [CommABC])
- Release date (e.g., released on June 11, 2001 at 8:00 a.m. central time [61101_0800CT])
- Publication date (e.g., published on December 24, 2003 [pub122403])
- Project number (e.g., project number 739 [PN739])
- Department number (e.g., Department 140 [Dept140])
- Records series (e.g., SeriesX)

General Challenges in File Naming

As you develop your policy, you will encounter the following challenges in file naming:

- *Version control.* You will need to determine how and whether to indicate the version of the record. Some organizations put current and obsolete drafts in different electronic file folders without altering the file name. However, when these records are moved from the active electronic file folder to another storage area, identical file names may conflict and cause confusion.
- *Uniqueness.* To avoid file names conflicting when they are moved from one location to another, each record's file name should be unique and independent from its location. For example, if letters are simply named with the word *letter* and the date, they are not independent from location because they could fit into any records series that contains letters, and all letters sent on that date would have the same file name.
- *Persistence over time.* File names should outlast the records creator who originally named the file. With good stakeholder and staff input, and training, you should be able to develop file names that make sense to staff members once the file creators are no longer available.

- *Access and ease of use.* The policy should be simple and straightforward. A simple policy will help staff members logically and easily name records and help ensure that records are accessible to staff members and/or to the public (as determined by the MGDPA). A simple policy will be more consistently used, resulting in records that are consistently named, and thus easier to organize and access.
- *Ease of administration.* The policy should work with your computer infrastructure, so that you can monitor policy compliance, manage records and records series, gather metadata, and perform other administrative tasks easily and in compliance with all legal requirements. For example, if all the records in a specific records series are easily identifiable by file name, they will be easier to gather and manage.
- *Scalability.* Consider how scalable your file naming policy needs to be. For example, if you want to include the project number, don't limit your project numbers to two digits, or you can only have ninety-nine projects.

Internet File Naming Protocols

Several file naming protocols are currently in use on the Internet. They all fall under the category of Uniform Resource Identifiers (URIs). *URIs* are short text strings that identify resources (e.g., documents, images, electronic mailboxes) on the Internet. These text strings commonly appear in the address window of web browsers. The first part of a URI specifies the *transfer protocol* in use (the method for transmitting the file from one device to another, such as hypertext transfer protocol [HTTP]). The second part specifies the address, often including the domain name, of the file.

Within the broad grouping of URIs are:

- *Uniform Resource Locators (URLs).* URLs are specific schemes that allow browsers and other software to access resources on the Internet. URLs indicate the resource's location (e.g., address and name).
- *Persistent Uniform Resource Locators (PURLs).* PURLs are functionally URLs, but act as an intermediary for the URL of a web site by redirecting the browser to a PURL server instead of the actual URL. The PURL server associates the PURL with the real URL and returns the URL to the viewer's browser.
- *Uniform Resource Names (URNs).* URNs are designed to serve as persistent, location-independent resource identifiers. Some overlap exists between URLs and URNs, and some URNs are PURLs. URNs are intended to overcome the problem of persistence and location-independence by providing a long-term identifier for resources. URNs use a resolution service to enable a web browser to use the URN to find the URL location for the resource.

Domain Names

Common practice is to include the domain name in the URI. A domain name, such as “microsoft.com,” is nearly always a part of the URL, because URLs identify resources by location. Review the resources in the Annotated List of Resources for more information on domain names. Domain names are administered by the Domain Name System.

URL Protocols

You will encounter several common types of URL transfer protocols, including:

- *Hypertext Transfer Protocol (HTTP)*. This is the most common type of URL protocol accessed on the Internet (e.g., <http://www.mnhs.org>).
- *File Transfer Protocol (FTP)*. This protocol type is commonly used to transfer large files via the Internet (e.g., <ftp://ftp.mm.com>).
- *Gopher*. This protocol was used primarily in academic and governmental settings, and is rarely used today.
- *News*. This protocol accesses newsgroups (e.g., <news:rec:knitting>).
- *Telnet*. This protocol allows users to control the activity on another computer and participate in interactive sites for such activities as games, live chats, and text information exchange.
- *Mailto*. This protocol is for e-mail exchange.

Challenges in Internet-Based File Naming

Naming for the Internet is particularly challenging. File names should meet your general criteria, especially for uniqueness, independence from location, and persistence over time. The file names should persist even if you move the files to another server, reorganize your web site, or use another software program or method for producing your web pages.

A carefully constructed policy for naming Internet-delivered files will ensure that:

- *Your web site links stay live*. Links contain embedded information about the location of the resource being linked to. Moving files from one server to another may result in dead links. If you develop a policy that builds in persistence and location-independence, you should be able to avoid this problem.
- *You can more easily manage your web site records*. Because the file names are independent of location, you can be assured that you will be able to find records if they are reorganized. For example, if a department within your agency reorganizes its web pages and moves some files to another server, as long as the file names of the records are independent of location, you can still efficiently manage and archive them.

- *Your Internet-accessed files mesh with your other electronic files.* By integrating your file naming policy with that used for other electronic records, your public records will remain accessible as long as necessary and not-public information will be protected as appropriate.

General File Naming Issues

General issues to consider as you develop a file naming policy include:

- *Determining what metadata to collect.* You will need to decide what metadata to collect and include in file names. Collection and use of metadata in file names will help ensure the long-term usefulness of your records and help you to meet legal requirements for accessibility (for public records) and accountability, as well as protect not-public records.
- *Universal retrieval.* Ensure that the staff and the public (as appropriate) can access your files. Legally, public records must be accessible. Standard file names allow users to find records efficiently.
- *Determining the official copy.* Determine which file is the “official” copy. As part of your web content management (see the *Web Content Management* guidelines), you should include in your policy which web site files are official records, and which version of the electronic file is the official record. Including an indicator of official record status in a file name may be useful for this purpose. The inclusion of this parameter in your policy will help you meet your legal requirements to capture records as set forth by the Official Records Act. The inclusion of this designation may also make administration of your web site records easier.
- *Determining file naming boundaries.* Pay close attention to the freedom you give staff members (and outside vendors) in naming files. Provide guidelines and training on file naming. You will not be able to manage every electronic record’s file name, so you will need to rely on staff members and vendors to name files in compliance with your policy. By providing guidelines and training, you can maximize policy compliance in a way that meets your operational and legal requirements.
- *Relationship to and connection with paper records.* Determine how the names of your electronic records relate to the names of paper files you have stored. Because electronic records may be part of records series that include paper records, the file naming policy for electronic records should fit logically with your paper records naming. For example, a letter published on a web site might be part of a records series that includes additional paper documents in a file folder. By ensuring that the electronic records’ and the paper records’ file names mesh, you can more easily manage the records series.

Key Issues to Consider

Now that you are familiar with some of the basic concepts of file naming, you can use the questions below to discuss how they relate to your agency.

Pay special attention to the questions posed by the legal framework, including the need for public accessibility, as appropriate. Consider your current and future activities and records to help determine the components of a file naming policy that will work now and in the future. For example, you may currently publish official statements or press releases on paper, but in the future, you may publish such records on the web.

Discussion Questions

- Who will use the file naming policy to name files? What policy will make sense to each group?
- Who will need access to these records? Are there different groups with different needs (e.g., the public, internal users)? How will people “think of” this record (e.g., “I need to find a copy of XYZ.doc.” or “I need information about legislation passed in 2002/2003.”)?
- Will the records move location (e.g., from one server to another, from a server to long-term storage)? How will these changes affect file naming?
- What style issues are important? For example, how should the record names appear in print?
- How does file type affect file name? Does our software or computer system limit the number of digits in the file name?
- What types of electronic records will we name?
- How will staff members and the public access and open files in the short-term and long-term? What limitations do these systems have for file naming?

Annotated List of Resources

Primary Resources

Cool URIs Don't Change. In: *Style Guide for Online Hypertext*. Cambridge, MA: World Wide Web Consortium (W3C), 1998.

<<http://www.w3.org/Provider/Style/URI>>

This section of the complete style guide discusses the file naming concepts for the World Wide Web to ensure the accuracy of links and the longevity of the names.

Naming and Addressing: URIs, URLs,....

<<http://www.w3.org/Addressing>>

These web pages describe the relationship of URIs, URLs, and URNs. The pages also provide links and other information about other file naming topics for the web, such as metadata, markup languages, events, and history.

Webopedia

<<http://webopedia.internet.com>>

This comprehensive online encyclopedia for the information technology community provides an easy-to-understand, searchable database of terms and topics, including entries on file names and file formats.

PURL

<<http://purl.oclc.org/>>

The OCLC PURL Service provides a comprehensive introduction to the subject of PURLs. Available from this web site are Frequently Asked Questions on PURLs, introductions to the subject, and the opportunity to create and modify a PURL.

Additional Resources

Identifiers for Digital Resources. Washington, D.C.: Library of Congress, National Digital Library Program, 1996.

<<http://memory.loc.gov/ammem/award/docs/identifiers.html>>

These web pages describe the desirable characteristics for file naming for digital records. For illustrative purposes, the pages use the American Memory Collection as a case study for a file naming scheme.

Mims, J. "Files Control." In *Records Management: A Practical Guide*. Washington, D.C.: International City County Management Agency, 1996: 73–84.

This chapter on file management discusses such topics as filing systems, filing system creation, filing system maintenance, and filing system equipment. This chapter also offers information on troubleshooting file system control. The content focuses primarily on

paper systems, but the management principles apply across all media.

Minnesota Historical Society, State Archives Department. *Trustworthy Information Systems Handbook*. Version 4, July 2002.

<<http://www.mnhs.org/preserve/records/tis/tis.html>>

This handbook provides an overview for all stakeholders involved in government electronic records management. Topics center around ensuring accountability to elected officials and citizens by developing systems that create reliable and authentic information and records. The handbook outlines the characteristics that define trustworthy information, offers a methodology for ensuring trustworthiness, and provides a series of worksheets and tools for evaluating and refining system design and documentation.