# File Naming and Formats

*Digital Preservation Topical Note 4*

**Digital**Preservation**Coalition**

## Why are file naming and file formats important for digital preservation?
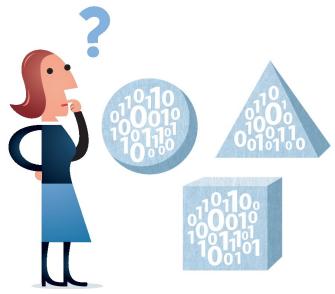
Smart file naming and considered choice of file formats will make a significant difference to the long-term preservation of digital records. Consistently using a standard file name format that follows some basic rules will help ensure files can be found, opened, and preserved for as long as necessary. Choosing file formats that are robust and well-supported will aid the management and maintenance of growing numbers of digital records. Addressing these two issues early in the creation of records is an important step towards successfully preserving digital content.

## What is good practice for naming files?

While each organisation should create its own file naming system, there are some basic rules everyone should follow:

- never use spaces or special characters like !*& % (except for dashes – and underscores _)
- never change the file extension through re-naming
- use a short descriptive file name of content and date
- use the ISO 8601:2004 standard format for dates: YYYYMMDD
- use a consistent method for showing the file versions, such as v1 or ver1

Though not necessary, it can be useful to create file names that begin with the date in the ISO format. For example, meeting minutes might have a file name pattern: '20170314_minutes_monthlystaff.doc' so that they appear in a file directory in chronological order.
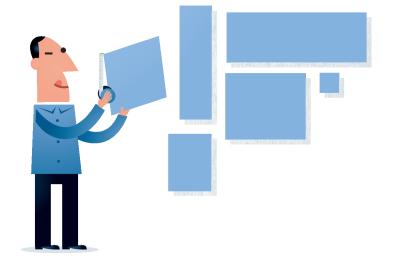


## Why should file names follow these rules?

Ultimately, these rules ensure that computers will be able to read the file correctly and that people working with the files can find and use them. Changing a file extension can create errors that can result in the file becoming unreadable and essentially lost. A descriptive file name is easier for others to find, but even more importantly, in some cases files without a descriptive name might be overwritten. For instance, some devices like scanners and digital cameras automatically assign a file name with a prefix and number, like 'DCS_0001.jpg'. If the file name is not changed, new scanned images or photos uploaded from the device may overwrite the older ones.

## What are the risks involved with file formats?

File formats are like containers for digital content. Like a VHS or a cassette tape, an .mp4 and .wmv hold video and the information needed to play that video. Three main risks associated with file formats are obsolescence, proliferation, and dwindling commercial or community support. While all three issues can be addressed through preservation strategies, it is important to be aware of the them when creating new records.

File Format Obsolescence - when newer generations of software phase out support for older formats. For example, documents created using WordStar (word processing software from the 1980s) cannot be easily opened using modern software programmes.

Proliferation - the rapid creation of new file format types to support new or improved forms of digital content. An organisation can accumulate many different file formats, and versions of those formats, e.g. text documents in different versions of PDF, Word, and Rich Text formats. Tracking and managing all these formats (which ones are at risk, and which software tools can be used for each one) can be a serious challenge.

Lack of Support - if a commercial provider of a proprietary file format discontinues backwards compatibility, older records will become more challenging to access and preserve. Similarly, if a community responsible for an open source file format detects a decrease in use, they may stop supporting it.

## How can risks to file formats be avoided?

While there is a risk that some file formats will become obsolete (see note on What Is Digital Preservation), keeping track of formats and making choices about what formats to use and how to preserve them can prevent data loss. Archivists and digital preservation experts may use one or several of solutions to protect an organisation's records, one of the most important of which is ensuring records are created in robust formats where possible. This usually involves selecting formats that are well-supported (commercially or through an open-source community), that meet standards, that are widely adopted, and that allow users to add metadata (see note on Metadata).

The risks posed to file formats are not as severe as originally feared 10 to 20 years ago, but it is important to be aware of the potential dangers to the long-term lifespan of a record when creating it. Creating records in file formats that are more likely to stand the test of time improve the chances of that record lasting into the future. Most organisations will have guidance on preferred formats to use when creating records that will be preserved, or advice can be sought from archives or records management colleagues. However, the aim of preservation is not to restrict the creation of meaningful, useful records. The best way forward is to find a balance between using new and innovative file formats and choosing preservation strategies to best maintain those records.

**For more information on Digital Preservation visit the DPC Website: https://www.dpconline.org**