Best Practices for Selecting Electronic File Formats

Government units, institutions, and individuals create electronic records in a variety of file formats. Over time, these formats may become obsolete or unusable, rendering these electronic records inaccessible. Monitoring and regulating the usage of file formats can help minimize the risk of records loss and reduce costs. For the creation and effective management of electronic records, limit the number of file formats used, review formats on a continuous basis, and be prepared to migrate to more stable and widely adopted formats as needed. These practices are especially important for records that have a permanent retention or that will be transferred to the State Archives at the Wisconsin Historical Society (WHS).

This document provides best practices for selecting and monitoring file formats. It also includes a table of the State Archives' preferred and acceptable formats for permanent retention and transfer.

Criteria for Selecting File Formats

Selected file formats have a great impact on the long-term accessibility of electronic records. The ability to access information in files depends on the ability to store, read, and edit those files. Not all file formats are created equal – formats that have few users, depend on proprietary software to read, or are not well-documented are less likely to be accessible in the future. To help mitigate loss, use the following criteria when selecting file formats for permanent or transferred records.

Open

Open source and non-proprietary formats have openly available specifications, enabling anyone to develop tools to read or edit them. This lowers the chance of these formats becoming inaccessible in the future, making them preferable for records with long or permanent retention. These formats, such as PDF/A (.pdf), JPEG2000 (.jp2), and OpenDocument Text (.odt), often have specifications that are maintained by a community or standards organization.

Popular

Prefer widely-adopted file formats over those that are rarely used. File formats with larger user groups likely will have technical support in the future. Some formats, such as Microsoft Word documents (.doc), are proprietary; but due to their widespread popularity, they are relatively safe and easier to migrate to migrate to preservation worthy, non-proprietary formats.

Documented

File formats with published documentation and standards will be easier to preserve and access in the future. Open formats are often more likely to have publicly available documentation.

Independent

Some file formats depend on specific hardware, operating systems, or software for proper rendering and use. If a file can only be viewed or edited using specific hardware or software, it may become difficult to keep this file accessible over time. Similarly, dynamic content that relies on external data sources for proper rendering will present problems for long-term preservation and accessibility. File formats with fewer external dependencies are more appropriate for longer retention and transfer.

Supports metadata

File formats that have support for metadata are "self-documenting." This means that metadata about how and when a file can be embedded directly in the file and travel with it wherever the file goes. This makes it easier to organize records and keeps the file accessible, even if transferred to another organization.

Lossless

Lossy file formats are those where the data is compressed, or lost, during the encoding process. Lossless file formats do not lose data during encoding. For this reason, lossless formats are often larger and more expensive to store. However, for the purposes of long-term retention and preservation, lossless formats are preferred with lossy file formats more suitable for short-term access purposes. TIFF is an example of a lossless image format, whereas JPEG is a lossy image format, because image data is irretrievably lost due to compression.

No digital rights protection

Some file formats support digital rights management (DRM), which restricts the usage of files. A common example of this includes internet music files that prevent the production of copies or can only be played using specific software. DRM mechanisms can severely inhibit the preservation and access of electronic records. Records scheduled for long-term retention or transfer should not have DRM mechanisms.

Monitoring and Converting File Formats

To ensure that files with long-term retention remain accessible, regularly monitor their formats over time and verify that they are still supported. When formats are no longer

supported, determine whether you will convert the files into supported formats to maintain the information in those files.

While converting files can protect from information loss, the conversion process brings its own risks, and it must be carefully planned. Prior to converting files, consider these three different types of potential loss that can occur during the conversion process:

Data

The file's data is the main component of the information contained within the electronic record. However, file metadata is also at risk of being lost during conversion. In the case of government units, legally, records must be complete and trustworthy, meaning any converted files should retain all their data and metadata after conversion.

Appearance

A file's appearance may also be important to the value of the record. If you convert a Microsoft Word document (.doc) to a Rich Text document (.rtf), for example, you may lose the appearance and structure of the original document. You must consider whether appearance is essential to understanding the record, and whether a loss of appearance would affect the completeness of that record.

Relationships

Relationships within or between files can also be lost during conversion. If you convert a Microsoft Excel spreadsheet (.xls) to a comma-separated file (.csv), for example, you may lose spreadsheet formulas that determine certain values. You must determine if the loss of these relationships would affect the completeness of the record.

File Formats for Transfer and Preservation

The table below is organized by type of file (word processing documents, audio, presentations, etc). For each file type, common file formats are listed as optimal, acceptable, and unacceptable for long-term retention and transfer to the Wisconsin Historical Society. Please note that this table is not exhaustive – if you have questions about a format not listed, contact WHS staff.

Optimal formats

These formats meet all requirements for long-term retention and preservation. You may transfer files in these formats to the State Archives. These also represent the optimal formats for permanent retention in government units.

Acceptable formats

These formats meet some of the requirements for long-term retention and preservation. You may transfer files in these formats to the State Archives. If these files are scheduled for permanent retention in a government unit, the unit should consider converting them to an optimal format in consultation with IT staff and/or the State Archives.

Unacceptable formats

These formats are not appropriate for transfer or long-term retention, as they cannot be relied on to last more than five years. Many proprietary formats created using legacy or less common proprietary software programs are unacceptable for transfer or long-term retention. For government units, electronic records whose retention periods are over five years should not be stored in these formats. If you have records scheduled for transfer to the State Archives or for permanent retention in any of these formats, contact your IT staff and/or State Archives staff to discuss options for conversion.

Table last updated February 2026

Record type	**Optimal formats**	**Acceptable formats**	**Unacceptable formats**
Word processing documents	PDF/A-1a (.pdf) OpenDocument Text (.odt)	Microsoft Word document (.doc) Microsoft Open XML Document (.docx) Rich Text Format (.rtf)	Corel WordPerfect (.wpd) Lotus WordPro (.lwp)
Plain text documents	Plain text (.txt) Comma-separated file (.csv) Tab-delimited file (.txt)		
Spreadsheets	OpenDocument Spreadsheet (.ods) Comma-separated file (.csv) Tab-delimited file (.txt) PDF/A-1a (.pdf)	Microsoft Excel Spreadsheet (.xls) Microsoft Excel Open XML Spreadsheet (.xlsx)	
Raster images * Raster files store images as a collection of pixels and cannot be scaled without distortion	TIFF (.tif, .tiff) JPG 2000 (.jp2) JPEG (.jpg, .jpeg) PNG (.png) GIF (.gif)		RAW (.raw, various) Adobe Photoshop (.psd)
Databases	Comma-separated file (.csv) Tab-delimited file (.txt) Structured Query Language (.sql)	Microsoft Access (.accdb, .mdb) dBase Format (.dbf)	
Presentations	OpenDocument Presentation (.odp)	Microsoft PowerPoint Presentation (.ppt)	

	PDF/A-1a (.pdf)	Microsoft Open XML PowerPoint Presentation (.pptx)	
Email	Microsoft Outlook Personal Storage Table (.pst)	Microsoft Outlook Message (.msg)	
Audio	Audio Interchange File Format (.aif, .aiff) WAVE Format (.wav)	Windows Media Audio (.wma) MPEG3 (.mp3) MP4 AAC (.m4a)	Audio CD DVD-Audio MP4 AAC Protected (.m4p, .m4b)
Video	AVI, lossless (.avi)	AVI, lossy (.avi) MPEG-4 (.mp4) MPEG-2 (.mp2) MOV (.mov) WMV (.wmv)	DVD-Video VOB (VIDEO_TS, AUDIO_TS) Blu-ray Disc Adobe Flash (.fla, .swf)

Additional Resources

The Library of Congress Recommended Formats Statement

The Library of Congress Sustainability of Digital Formats website

The Digital Preservation Coalition's Technology Watch Publications

The Digital Preservation Coalition's Digital Preservation Handbook

The Digital Preservation Coalition's Global 'Bit List' of Digital Materials