

Data Formats for Long-Term Digital Preservation

The Data Format Confidence Table below lists the most secure and sustainable ways to preserve electronic public records that have a long-term retention period (more than 15 years). The formats and corresponding confidence levels represent the Kentucky Department for Libraries and Archives (KDLA) recommendation for long-term digital preservation. The “primary preservation” option is preferred for permanent records being transferred to the Kentucky State Digital Archives (KSDA), though the Electronic Records Branch can accept any format. Agencies are free to use other formats (including those not listed) for active business use if they meet with state-approved standards and architecture. While some systems may use proprietary formats (GIS or CAD, e.g.), please ensure the agency can export the records into one of the below formats for long-term or permanent retention.

Table of Contents:

Factors Determining Confidence Level	1
Agency Considerations Regardless of Format	2
Data Format Confidence Table	2
Resources	3

Factors Determining Confidence Level

The confidence levels identified in the table consist of Primary Preservation (high-confidence) and Secondary Preservation (low-to-mid confidence) levels. These confidence levels are determined by a combination of factors including:

1. **Disclosure.** Degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content. What is most significant is not approval by a recognized standards body, but the existence of complete documentation.
2. **Adoption.** Degree to which the format is already used by the primary creators, disseminators, or users of information resources. This includes use as a master format, for delivery to end users, and as a means of interchange between systems.

3. **Transparency.** Degree to which the digital representation is open to direct analysis with basic tools, such as human readability using a text-only editor.
4. **Self-documentation.** Self-documenting digital objects contain basic descriptive, technical, and other administrative metadata.
5. **External Dependencies.** Degree to which a particular format depends on specific or proprietary hardware, operating system, or software for rendering or access and the predicted complexity of dealing with those dependencies in future technical environments.
6. **Impact of Patents.** Degree to which the ability of archival institutions to sustain content in a format will be inhibited by patents.
7. **Technical Protection Mechanisms.** Implementation of mechanisms such as encryption that prevent the preservation of content by a trusted repository. This could also include Digital Rights Management (DRM), though less common in government.

Agency Considerations Regardless of Format

1. Fully or partially encrypted files must be unencrypted prior to transfer to KDLA.
2. Password protected files must be opened with protections removed prior to transfer to KDLA.
3. Any files produced with Digital Right Management controls must have all controls removed prior to transfer.
4. In general, use platform independent, vendor independent, nonproprietary, stable, open, and well-supported formats.
5. Contact Electronic Records Branch staff for any file format or electronic records management questions at Electronic.Records@ky.gov.

Data Format Confidence Table

Media	Primary Preservation Format (preferred)	Secondary Preservation Format (acceptable)
Audio	BWF-Broadcast WAV (.wav is the extension)	WAV, AIFF, FLAC
CAD	PDF/A, PDF/E, PDF with original file	Original
Database Management Systems (DBMS)	Original	CSV or XML with schema
Email Messages or Accounts	MSG, PDF	PST, MBOX

GIS, Geospatial and Non-GIS Cartographic	Original file with PDF or TIFF	
Graphics	TIFF (uncompressed), SVG 1.1 (for vector graphics)	PDF
Images	TIFF (uncompressed)	JPG, DNG, PNG, JP2, PSD, GIF, Camera Raw (i.e. .nef, .crw, .arw)
Other	Consult the KDLA Electronic Records Branch	
Presentations	PDF/A, PDF (PDFs should include speaker notes)	Original
Software	Consult the KDLA Electronic Records Branch	
Spreadsheet Applications or Structured Data	CSV, PDF/A, PDF (must capture entire workbook - macros disabled)	Tab-delimited TXT, XML, XLS, XLSX
Text or Word Processing Applications	PDF/A, PDF/UA, PDF	RFT (text), TXT, XML with schema
Video	Motion JPEG 2000 (.mj2), MOV, AVI	MPEG-4
Websites and Social Media Sites	WARC	ARC IA, WACZ, CDX, Files from Content Management System

Resources

Smithsonian Institution Archives Digital Preservation Formats

- <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>

LOC Recommended Formats Statement 2025-2026

- <https://www.loc.gov/preservation/resources/rfs/format-pref-summary.html>