

Smithsonian Data Management Best Practices

Choosing File Formats (for Preservation *and* Access)

Support for a particular file format now does not guarantee readability in the future. If your data is in a proprietary format, or is dependent on specialty software or hardware to read it, consider converting it to one of the more sustainable file types below and deposit your data in both its original format plus an equivalent sustainable format.

In general, save data for archiving in a non-proprietary, platform-independent, unencrypted, lossless, uncompressed, commonly used file format.

If you need to store multiple files in one container, choose an **uncompressed TAR, GZIP, or ZIP**.

File Type	Prefer	Avoid
Text	PDF/A-1 or PDF/A-2, OpenDocument Text Format (.odt), Plain Text (.txt - using ASCII or Unicode encoding)	WordPerfect, Microsoft Word
Structural markup	SGML with DTD, XML with a DTD	SGML or XML without a DTD
Tabular data	comma (.csv) or tab-delimited (.txt or .tab) file, OpenDocument Format Spreadsheet (.ods)	
Database	comma or tab delimited flat file(s) (.csv or .txt), XML with DTD, JSON, OpenDocument Format Spreadsheet (.ods)	MSAccess, FileMaker Pro
Geospatial*	GeoTIFF, Geographic Markup Language (GML) or GML in JP2, Keyhole Markup Language (KML)	
Audio	Broadcast Wave Format (BWF – an extension of WAVE), Free lossless codec of the Ogg project (FLAC), Audio Interchange File Format (AIFF), MP3 (uncompressed), Microsoft Wave (WAV)	
Video	AVI, Quicktime (MOV), Windows Media Video (WMV), MPEG4 or MPEG-2 Video, Material Exchange Format – lossless (MXF)	Raw, Adobe Flash
Still image	TIFF, JPEG2000, PNG, PDF/A	Camera RAW
Presentation	PDF/A, PDF	MS PowerPoint, Prezi

*for Geospatial files, ESRI Shapefiles and ESRI ARC files are acceptable, but not preferred, per NARA.

References

(accessed 2018-02-23) Wikipedia. List of file formats. https://en.wikipedia.org/wiki/List_of_file_formats

2014. NARA Records Management Regulations, Policy, and Guidance Appendix A: Tables of File Formats <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>

(accessed 2018). Library of Congress. Recommended Formats Statement. <http://www.loc.gov/preservation/resources/rfs/data.html>

2016. Recommended Preservation Formats for Electronic Records. <https://siarchives.si.edu/what-we-do/digital-curation/recommended-preservation-formats-electronic-records>