



FILE FORMATS FOR TRANSFER OF ELECTRONIC RECORDS TO THE STATE ARCHIVES OF NORTH CAROLINA

9/10/2012

State Archives of North Carolina

File Formats for Transfer of Electronic Records to the State Archives of North Carolina

STATE ARCHIVES OF NORTH CAROLINA

This is the companion document to *File Formats Guidelines for In-House Preservation and Long-Term Retention*. The current document lists the file formats appropriate for transfer of digital public records to the State Archives of North Carolina (State Archives) for permanent retention. *File Formats Guidelines for In-House Preservation and Long-Term Retention* contains descriptions of all formats classified as “recommended for transfer.” Please refer to that document for more information about the formats listed below. To determine which records should be transferred, please consult your records retention and disposition schedules and contact your records analyst for more information about what records may be transferred to the State Archives for permanent retention.

The following table represents the digital formats that the State Archives of North Carolina accepts for transfer of digital public records. These guidelines organize formats into three categories:

Recommended for transfer: You may transfer records in these formats to the State Archives. File formats that meet the minimum requirements for transfer and long-term retention. In most cases, these are the formats in which the State Archives will maintain files.

Acceptable for transfer under certain circumstances: You need to request permission to transfer records to the State Archives in these formats. These file formats that do not meet the minimum requirements for transfer and long-term retention.

Not recommended for long-term retention: File formats that are not appropriate for long-term retention. Files saved in these formats should not be relied on to last more than five years. Electronic records whose retention periods are over five years should not be stored in these formats.

Type of record	Recommended for transfer	Acceptable for transfer under certain circumstances	Not acceptable for transfer
Word Processing documents	PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A) OpenDocument Text (.odt)	PDF/A-1b (.pdf) (ISO 19005-1 minimally compliant PDF/A) Microsoft® Word Document (.doc) Microsoft® Open XML Document (.docx) Rich Text Format (.rtf)	Corel® WordPerfect® (.wpd) Lotus® WordPro (.lwp) PDF (.pdf)
Plain text documents	Plain Text (.txt) <i>US-ASCII or UTF-8 encoding</i> Comma-separated file (.csv) <i>US-ASCII or UTF-8 encoding</i> Tab-delimited file (.txt) <i>US-ASCII or UTF-8 encoding</i>	Other delimited text files (space-delimited, colon-delimited, etc.) <i>where the delimiting character is not present in the data</i>	
Structural markup text documents	XML (.xml) <i>with DTD/Schema</i> SGML <i>with DTD/Schema</i>		XML without DTD/Schema SGML without DTD/Schema
Spreadsheets	OpenDocument Spreadsheet (.ods) Comma-separated file (.csv) Tab-delimited file (.txt) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A)	Microsoft® Excel® Spreadsheet (.xls) Microsoft® Excel® Open XML Spreadsheet (.xlsx) Other delimited text files (space-delimited, colon-delimited, etc.) <i>where the delimiting character is not present in the data</i>	
Audio	Broadcast WAVE Format LPCM (.wav) WAVE Format LPCM (.wav)	AIFF (uncompressed) (.aif, .aiff) Standard MIDI (.mid, .midi) Windows® Media Audio WMA (.wma) MPEG3 (.mp3) MP4 AAC (.m4a)	Audio CD (Compact Disc Digital Audio system, CDDA, CD-DA) DVD-Audio QuickTime® MP4 AAC Protected (.m4p, .m4b) QuickTime® MP3, iTunes (.mp3) RealAudio® (.rm, .ra) Shorten® (.shn) RIFF-RMID (.rmi) Extended MIDI (.xmi) Module Music Formats, Mods (.mod) SUN Audio, uncompressed (.au) Ogg FLAC (.ogg)

Type of record	Recommended for long-term retention	Acceptable for long-term retention	Not recommended for long-term retention
Digital Video	AVI, full frame (uncompressed), WAVE PCM audio (.avi)	AVI, containing H.264/MPEG-4 AVC (lossy) ¹ (.avi) MPEG-4, containing H.264/MPEG-4 AVC (lossy) (.mp4) MPEG-2, containing H.262/MPEG-2 (lossy) (.mp2) MOV, containing H.264/MPEG-4 AVC (lossy) (.mov) ASF, containing WMV (lossy) (.wmv) MXF, containing Motion JPEG 2000 ² (lossless) (.mxf) Ogg, containing Theora (lossy) (.ogg)	DVD-Video VOB (VIDEO_TS, AUDIO_TS) Blu-ray Disc™ HCAM® Digital VHS (D-VHS) DVCam®
Raster Images	TIFF (.tif, .tiff) <i>uncompressed</i> JPG 2000 (.jp2)	JPEG (.jpg, .jpeg) PNG (.png) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A) GIF (.gif)	RAW (.raw, various) Adobe® Photoshop® (.psd) Kodak PhotoCD Encapsulated PostScript (.eps) FlashPix™ (.fpx) PDF (.pdf)
Vector Images (*See below for geospatial vector sets.)	Scalable Vector Graphics 1.1 (.svg) AutoCad Drawing Interchange Format (.dxf) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A)		Adobe Illustrator (.ai) Corel® Draw CDR (.cdr) Micrografx Draw DRW (.dwr) Windows Metafile WMF (.wmf, .emf) Standard for the Exchange of Product Model Data STEP (.stp) Computer Graphics Metafile DXF (.dxf) AutoCAD Drawing Format (.dwg)
Databases	Software Independent Archiving of Relational Databases (SIARD) Delimited Flat File (Plain Text) with DDL	Microsoft® Access® (.accdb) Microsoft® Access® (.mdb) dBase Format (.dbf)	
Presentations	OpenDocument Presentation (.odp) PDF/A-1a (.pdf) (ISO 19005-1 compliant PDF/A) <i>for presentations without animation</i>		Microsoft® PowerPoint® Presentation (.ppt) Microsoft® Open XML PowerPoint Presentation (.pptx)

¹ One of the H.264/MPEG-4 AVC profiles is sometimes described as lossless: MPEG-4 AVC High 4:4:4 Profile. For more information, see the Notes section of “MPEG-4, Advanced Video Coding, High 4:4:4 Profile,” *Sustainability of Digital Formats: Planning for Library of Congress Collections*, <http://digitalpreservation.gov/formats/fdd/fdd000218.shtml> (accessed 5/16/2012).

² See “MXF File, OP1a, Lossless JPEG 2000 in Generic Container,” *Sustainability of Digital Formats: Planning for Library of Congress Collections*, <http://digitalpreservation.gov/formats/fdd/fdd000206.shtml> (accessed 5/16/2012) and “Motion JPEG 2000 jp2 File Format,” *Sustainability of Digital Formats: Planning for Library of Congress Collections*, <http://digitalpreservation.gov/formats/fdd/fdd000127.shtml> (accessed 5/16/2012).

Type of record	Recommended for long-term retention	Acceptable for long-term retention	Not recommended for long-term retention
Email	<p>The State Archives currently has systems in place to collect email accounts (including emails, email folders, and calendars) of state employees subject to E.O. 18.</p> <p>Microsoft Outlook Personal Storage Table (.pst)</p>	<p>Multiple emails MBOX, MIME (.mbx, .mbox)</p> <p><i>Individual email messages saved in any of the following formats:</i> Email Message, MIME (.eml, .txt), with email header Plain Text (.txt), with email header Rich Text (.rtf), with email header PDF/A-1 a (.pdf) (ISO 19005-1 compliant PDF/A), with email header HTML (.html), with email header Microsoft® Outlook® Message (.msg), with email header</p>	<p><i>Individual email messages saved in any of the following formats:</i> Apple® Mail (.emlx), with or without email header Plain Text (.txt) without email header Rich Text (.rtf) without email header PDF/A-1 a (.pdf) (ISO 19005-1 compliant PDF/A), without email header PDF, with or without email header HTML (.html) without email header</p>
Websites / Social Media	<p>Note: The State Archives is currently collecting state agency webpages, Twitter, YouTube, Flickr, and other online content through an automated web-crawling tool. To submit a new website to the web archive, please contact webarchives@ncdcr.gov .</p> <p>Web Archive (.war, .war)</p>		
Geospatial Vector Data	<p>Note: Extensive documentation covering proper preparation of geospatial data and current North Carolina geospatial preservation standards and workflows can be found at http://geomapp.net. Currently, the State Archives of North Carolina collects geospatial vector layers that have been superseded in the North Carolina GIS clearinghouse, NC OneMap, as shapefiles containing seven files with the same filename prefix and the following extensions:</p> <ul style="list-style-type: none"> Main file (.shp) Index file (.shx) Database file (.dbf) Projection file (.prj) Shapefile spatial index file (.sbn) Shapefile spatial index file (.sbx) Geospatial metadata file (.shp.xml) 		