

Recommended Data Formats for Preservation Purposes in the Florida Digital Archive

This table is intended to help Florida university administrators develop guidelines for preparing and submitting files to the Florida Digital Archive. It is not intended to suggest that these file formats are allowable formats for ETDs or any other official publication of any Florida university. (See the notes following this table.) Source: FCLA, Aug. 2008.

Media	High Confidence Level	Medium Confidence Level	Low Confidence Level
Text	<ul style="list-style-type: none"> - Plain text (encoding: US-ASCII, UTF-8, UTF-16 with BOM) - XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema and character encoding explicitly specified) - PDF/A-1 (ISO 19005-1) (*.pdf) 	<ul style="list-style-type: none"> - Cascading Style Sheets (*.css) - DTD (*.dtd) - Plain text (ISO 8859-1 encoding) - PDF (*.pdf) (embedded fonts) - Rich Text Format 1.x (*.rtf) - HTML (include a DOCTYPE declaration) - SGML (*.sgml) - Open Office (*.sxw/*.odt) - OOXML (ISO/IEC DIS 29500) (*.docx) 	<ul style="list-style-type: none"> - PDF (*.pdf) (encrypted) - Microsoft Word (*.doc) - WordPerfect (*.wpd) - DVI (*.dvi) - All other text formats not listed here
Raster Image	<ul style="list-style-type: none"> - TIFF (uncompressed) - JPEG2000 (lossless) (*.jp2) - PNG (*.png) 	<ul style="list-style-type: none"> - BMP (*.bmp) - JPEG/JFIF (*.jpg) - JPEG2000 (lossy) (*.jp2) - TIFF (compressed) - GIF (*.gif) - Digital Negative DNG (*.dng) 	<ul style="list-style-type: none"> - MrSID (*.sid) - TIFF (in Planar format) - FlashPix (*.fpx) - PhotoShop (*.psd) - RAW - JPEG 2000 Part 2 (*.jpf, *.jpx) - All other raster image formats not listed here
Vector Graphics	<ul style="list-style-type: none"> - SVG (no Java script binding) (*.svg) 	<ul style="list-style-type: none"> - Computer Graphic Metafile (CGM, WebCGM) (*.cgm) 	<ul style="list-style-type: none"> - Encapsulated Postscript (EPS) - Macromedia Flash (*.swf) - All other vector image formats not listed here
Audio	<ul style="list-style-type: none"> - AIFF (PCM) (*.aif, *.aiff) - WAV (PCM) (*.wav) 	<ul style="list-style-type: none"> - SUN Audio (uncompressed) (*.au) - Standard MIDI (*.mid, *.midi) - Ogg Vorbis (*.ogg) - Free Lossless Audio Codec (*.flac) - Advance Audio Coding (*.mp4, *.m4a, *.aac) - MP3 (MPEG-1/2, Layer 3) (*.mp3) 	<ul style="list-style-type: none"> - AIFC (compressed) (*.aifc) - NeXT SND (*.snd) - RealNetworks 'Real Audio' (*.ra, *.rm, *.ram) - Windows Media Audio (*.wma) - Protected AAC (*.m4p) - WAV (compressed) (*.wav) - All other audio formats not listed here

Media	High Confidence Level	Medium Confidence Level	Low Confidence Level
Video	- Motion JPEG 2000 (ISO/IEC 15444-4) (*.mj2) - AVI (uncompressed, motion JPEG) (*.avi) - QuickTime Movie (uncompressed, motion JPEG) (*.mov)	- Ogg Theora (*.ogg) - MPEG-1, MPEG-2 (*.mpg, *.mpeg, wrapped in AVI, MOV) - MPEG-4 (H.263, H.264) (*.mp4, wrapped in AVI, MOV)	- AVI (others) (*.avi) - QuickTime Movie (others) (*.mov) - RealNetworks 'Real Video' (*.rv) - Windows Media Video (*.wmv) - All other video formats not listed here
Spreadsheet/ Database	- Comma Separated Values (*.csv) - Delimited Text (*.txt) - SQL DDL	- DBF (*.dbf) - OpenOffice (*.sxc/*.ods) - OOXML (ISO/IEC DIS 29500) (*.xlsx)	- Excel (*.xls) - All other spreadsheet/ database formats not listed here
Virtual Reality	- X3D (*.x3d)	- VRML (*.vrl, *.vrm) - U3D (Universal 3D file format)	- All other virtual reality formats not listed here
Computer Programs		- Computer program source code (*.c, *.c++, *.java, *.js, *.jsp, *.php, *.pl, etc.)	- Compiled / Executable files (EXE, *.class, COM, DLL, BIN, DRV, OVL, SYS, PIF)
Presentation		- OpenOffice (*.sxi/*.odp) - OOXML (ISO/IEC DIS 29500) (*.pptx)	- PowerPoint (*.ppt) - All other presentation formats not listed here

Notes:

- We do not intend to provide full preservation for formats listed under Low Confidence Level.
- UMI ProQuest lists file formats that they will accept for theses and dissertations on their website (http://www.proquest.com/products_umi/dissertations/UMI_PreparingYourManuscriptGuide.pdf). Some of the formats UMI ProQuest accepts are listed in the “Medium” or “Low” confidence level in the above table. Because UMI ProQuest currently does not provide full preservation, this should not represent a conflict for any universities in deciding on acceptable ETD file formats.
- No files with viruses will be accepted (refers especially to these file formats: DOC, XLS, MDB, PPT, ZIP, EXE). Please scan your files for viruses with up-to-date virus scanners before transmitting your files to the Florida Digital Archive.
- We do not intend to provide full preservation for any files fully or partially encrypted.
- We do not intend to provide full preservation for any files fully or partially password-protected.
- We do not intend to provide full preservation for any files fully or partially compressed with a proprietary compression algorithm.
- We do not intend to provide full preservation for any files fully or partially encoded with unembedded proprietary fonts.
- We do not intend to provide full preservation for any files produced with Digital Right Management controls.

- It is permissible to archive files in the unencrypted ZIP (*.zip) format for transmission to the Florida Digital Archive. Upon the receipt at FCLA, those ZIP files will be restored to their original formats and archived accordingly.
- As a general rule, use platform-independent, vendor-independent, non-proprietary, stable, open and well-supported formats.