



Bentley Historical Library
Division of Digital Curation
1150 Beal Avenue
Ann Arbor, MI 48109

Sustainable Formats and Conversion Strategies at the Bentley Historical Library

February 14, 2013
Version 1.1

Executive Summary

The Bentley Historical Library is committed to the long-term preservation of and access to its digital collections. Because the library must contend with thousands of potential file formats, Digital Curation Services has adopted a three-tier approach to facilitate the preservation and conversion of digital content:

- Tier 1: Materials produced in sustainable formats will be maintained in their original version.
- Tier 2: Common "at-risk" formats will be converted to preservation-quality file types to retain important features and functionalities.
- Tier 3: All other content will receive basic bit-level preservation.

This document provides further information on the Bentley Historical Library's accepted preservation formats and conversion strategies.

Please see the chart on pp. 3-5 for a list of sustainable preservation formats and at-risk formats that will be subject to conversion.

Tier 1: Preservation of Sustainable Formats

The library has identified a number of sustainable file formats (pp. 3-5) that are widely used and/or nonproprietary, many of which have been recognized as international standards by bodies such as the [International Standards Organization](#) (ISO), [ECMA International](#), and the [Organization for the Advancement of Structured Information Standards](#) (OASIS). The longevity of these formats has furthermore been acknowledged by various peer institutions and experts in the digital curation community, including the Library of Congress's [National Digital Information Infrastructure and Preservation Program](#).

Digital materials stored in these file formats should remain usable to researchers and administrative units at the University of Michigan for the foreseeable future and beyond. The Bentley Historical Library will therefore preserve the original version of content stored in these sustainable formats at the time of accession. Digital Curation Services will monitor community best practices and technological advances in case a migration to alternative preservation formats should prove necessary.

Visit <http://fileinfo.com> to find basic descriptions of file formats or search the [PRONOM Technical Registry](#) for format specifications and more in-depth information.

Tier 2: Conversion of At-Risk Formats

The digital curation community has long acknowledged the disadvantages posed by proprietary formats (for which only specific software may be used) and content encoded with "lossy" compression (i.e. compression that reduces the quality of the data to conserve space). The Bentley Historical Library will therefore convert the most common at-risk formats to preservation-quality sustainable formats. To ensure the authenticity of materials, the original version will be maintained alongside the preservation copy.

See pp. 3-5 for a list of at-risk formats and preservation targets; these strategies reflect the policies and practices of peer institutions as well as the National Digital Information Infrastructure and Preservation Program. Visit the Library of Congress "Sustainability of Digital Formats" site (<http://www.digitalpreservation.gov/formats/index.shtml>) for more information on preservation issues and descriptions of preferred formats.

Tier 3: Bit-Level Preservation of All Other Formats

Because it is infeasible to create conversion plans for the tens of thousands of formats in existence, the Bentley Historical Library will ensure that digital holdings in other formats (i.e. ones not specifically identified in this document) will receive bit-level preservation. The use of integrity checks and regular replacement of storage media (conducted by trusted partners in the University of Michigan Library Information Technology division and Information and Technology Services) will preserve the raw data stored in these files (i.e. the "stream" of 0s and 1s) in its original state. The library concedes that hardware or software obsolescence may reduce the functionality of these files or render them inaccessible. At the same time, the faithful preservation of the content at the bit-level will allow the library to take advantage of future developments in emulation technology.

Tier 1: Preservation of Sustainable Formats	Tier 2: Conversion Strategies for At-Risk Formats	Tier 3: Bit-Level Preservation
Raster Images		
<ul style="list-style-type: none"> • <u>TIFF</u>: Tagged Image Format File • <u>JPEG/JFIF</u>: Joint Photographic Experts Group JPEG Interchange Format File (lossy compression) • <u>JPEG 2000</u>: Joint Photographic Experts Group (lossless compression) • <u>GIF</u>: Graphic Interchange Format • <u>PNG</u>: Portable Network Graphic 	Convert the following to <u>TIFF</u> with ImageMagick: <ul style="list-style-type: none"> • <u>BMP</u>: Windows Bitmap • <u>PSD</u>: Adobe Photoshop Document • <u>PCD</u>: Kodak Photo CD Image • <u>PCT</u>: Apple Picture File • <u>TGA</u>: Targa Graphic 	All others
Raw Camera Images		
Original raw digital camera images will be retained as preservation copies.	Convert the following to <u>JPEG</u> (for access) with ImageMagick: <ul style="list-style-type: none"> • <u>RAW</u>: Raw Image Data File • <u>CR2</u>: Canon Raw Image File • <u>ARW</u>: Sony Digital Camera Image • <u>DCR</u>: Kodak Raw Image File • <u>MRW</u>: Minolta Raw Image File • <u>NEF</u>: Nikon Electronic Format Raw Image File • <u>ORF</u>: Olympus RAW File • <u>PEF</u>: Pentax Electronic File • <u>3FR</u>: Hasselblad 3F Raw Image • <u>X3F</u>: SIGMA X3F Camera Raw File 	

Vector Images		
<ul style="list-style-type: none"> • <u>SVG</u>: Scalable Vector Graphics File 	Convert the following to <u>SVG</u> with Inkscape: <ul style="list-style-type: none"> • <u>AI</u>: Adobe Illustrator • <u>WMF</u>: Windows Metafile • <u>EMF</u>: Enhanced Metafile Convert the following to <u>PDF</u> with Ghostscript: <ul style="list-style-type: none"> • <u>PS</u>: PostScript • <u>EPS</u>: Encapsulated PostScript 	All others
Audio Files		
<ul style="list-style-type: none"> • <u>MIDI</u>: Musical Instrument Digital Interface File • <u>XMF</u>: Extensible Music File • <u>WAV</u>: Waveform Audio File Format • <u>AIFF</u>: Audio Interchange File Format • <u>MP3</u>: Moving Picture Experts Group Layer 3 compression • <u>OGG</u>: Ogg Vorbis Audio File • <u>FLAC</u>: Free Lossless Audio Codec File 	Convert the following to <u>WAV</u> with ffmpeg: <ul style="list-style-type: none"> • WMA: Windows Media Audio • RA/RM: Real Audio • SND: Apple Sound File • AU: Sun Audio File 	All others

Tier 1: Preservation of Sustainable Formats	Tier 2: Conversion Strategies for At-Risk Formats	Tier 3: Bit-Level Preservation
Video Files		
<ul style="list-style-type: none"> • <u>MPEG-1/2</u>: Moving Picture Experts Group • <u>AVI</u>: Audio Video Interleave File (uncompressed) • <u>MOV</u>: QuickTime Movie (uncompressed) • <u>MP4</u>: Moving Picture Experts Group (with H.264 encoding) • <u>MJ2</u>: Motion JPEG 2000 • <u>MXF</u>: Material Exchange Format File (uncompressed) • <u>DV</u>: Digital Video File (non-proprietary) 	Convert the following to <u>MP4</u> (with <u>H.264 encoding</u>) with ffmpeg: <ul style="list-style-type: none"> • <u>FLV</u>: Flash Video • <u>WMV</u>: Windows Media Video • <u>RV/RM</u>: Real Video 	All others
Office Documents and Text Files		
<ul style="list-style-type: none"> • <u>DOCX</u>: MS Word Open XML Document • <u>XLSX</u>: MS Excel Open XML Document • <u>PPTX</u>: PowerPoint Open XML Document • <u>PDF/A</u>: Portable Document Format (Archival) • <u>TXT</u>: Plain Text File • <u>RTF</u>: Rich Text Format File • <u>XML</u>: Extensible Markup Language Data File • <u>CSV</u>: Comma Separated Values File • <u>TSV</u>: Tab Separated Values File 	Convert the following to <u>Office Open XML</u> with Microsoft Office File Converter (OFC): <ul style="list-style-type: none"> • <u>DOC</u>: MS Word Document • <u>XLS</u>: MS Excel Document • <u>PPT</u>: PowerPoint Document Convert PDF to <u>PDF/A</u> with Ghostscript.	All others
Email		
<ul style="list-style-type: none"> • <u>MBOX</u>: Mailbox File 	Convert the following to <u>MBOX</u> with Aid4Mail: <ul style="list-style-type: none"> • <u>EML</u>: Email Message • <u>PST</u>: Outlook Personal Information Store File • Eudora mail, etc. (40 total) 	All others