

Prepared by:

Tomasz Neugebauer, Digital Projects & Systems Development Librarian

Tim Walsh, Digital Preservation Librarian

Spectrum Format Policy Registry

Version 1.1

October 29, 2019

Introduction

General Introduction

The Spectrum Format Policy Registry (FPR) defines the file format policies implemented by Concordia University Library for Spectrum Research Repository content. Concordia Library always keeps each digital object that has been reviewed by Spectrum editors and made live, in its original file format. For files in some formats, Concordia Library also creates derivative versions in **Preservation Formats**, to be stored alongside original files in preservation storage, and **Access Formats**, to be stored alongside original files in Spectrum. The Format Policy Registry defines which file formats receive this type of treatment.

File uploads to Spectrum may be subject to other policies as well, e.g., limitations on maximum file size.

Preservation Formats have qualities that increase their likelihood of being usable over a longer period of time. These qualities include:

- Wide use and support
- Stable and preferably open specifications
- Uncompressed or lossless compression

When Preservation Formats are created, they are stored alongside original files, either directly in Spectrum or in preservation packages that are only accessible to staff.

Access Formats have qualities that facilitate easier use. These qualities include:

- Wide use and support
- Compact file sizes

In most cases, the access formats are copies of the original files deposited to Spectrum. There are only a few cases for which Concordia Library will distribute an access copy in a file format that differs from the original file. In these cases, Access Formats are created because they are easier to use (e.g. render in a browser) without significantly altering the content or richness of information in a file.

When Access Formats have been made and distributed alongside the original files, the original copy are always available in Spectrum as well.

Levels of Support

Levels of support indicate the Library's ability to maintain the usability of a digital object over time. These assessments are made on the basis of file formats. There are three levels of support: **Basic**, **Watch**, and **Full**.

Basic-Level Support

The minimum level of treatment that all digital objects deposited into Spectrum receive. At this level, the Library preserves the bit-stream (i.e. the 1s and 0s that comprise the code) of a file exactly as-is. Usually, no format migration is performed at this level. This level of support does not necessarily ensure that files will be usable by software available at a future point in time. Any formats not explicitly mentioned in the Format Policy Registry are preserved by the Library at the Basic-level. Below are the assurances that can be granted for file formats that comply with Basic-Level Support:

- Attempt to perform file format identification.
- Reliable and secure storage on backed-up servers.
- Public online access via Spectrum for items out-of-embargo.
- Regular audits of checksums (i.e., fixity check) to ensure that no files have corrupted or changed in any way. This practice ensures that the Library is able to provide an exact copy of original files over time.

Watch-Level Support

File formats at this level of support are those for which the Library is currently only able to offer Basic-level support, but for which we hope to provide Full support in the future. This may be because the formats are common or highly valued or because there is reason to believe that developments in the software industry and digital preservation community will make it easier to perform high-quality batch file format migrations in the future. Below are the assurances that can be granted for file formats that comply with Watch-Level Support, in addition to the Basic-Level Support given to all uploaded files:

- Attempt to perform normalization to a long-term preservation format
- Undertake strategic monitoring of format

Full-Level Support:

File formats at this level of support are those for which the Library has high confidence in their long-term usability, either because the original format is already a preferred Preservation Format, or because the Library consistently and reliably normalizes files in this format to a documented Preservation Format. File formats that benefit from Full-Level Support are those that tend to be more common and in turn hold more stability. These formats are more likely to function cross-platform, be openly documented, and in some cases (e.g., PDF/A), be compliant with ISO standards.. Below are the assurances that can be granted for file formats that comply with Full-Level Support, in addition to the Basic-Level Support given to all uploaded files:

* - Indicates future digital preservation measures to be implemented

- File format identification
- Perform normalization to a long-term preservation format
- Perform validation of file format
- Undertake strategic monitoring of format

Format Policy Registry

File Formats

The following formats are the acceptable and preferred formats by Spectrum document type:

Text

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Archival PDF/A (.pdf)	Preferred	Full	PDF/A	PDF/A
Microsoft Office Open XML Word (.docx)	Preferred	Full	DOCX	DOCX
OpenDocument Text (.odt)	Preferred	Full	ODT	ODT
Standard PDF (.pdf)	Acceptable	Watch	PDF/A	PDF

All other files that identify as “text” mime type are considered “Text” document type with Basic-Level digital preservation support.

Examples: .xml, .csv, .c, .c++, .conf, .cpp, .csh, .css, .doc, .h, .htm, .html, .htmls, .java, .js, .log, .pl, .pm, .py, .rtf, .sgm, .sgml, .tcsh, .text, .tsv, .txt, .zsh.

Spreadsheet

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Microsoft Office Open XML Excel (.xlsx)	Preferred	Full	XLSX	XLSX
OpenDocument Spreadsheet (.ods)	Preferred	Full	ODS	ODS
Comma Separated Values (UTF-8 encoded, .csv)	Preferred	Full	CSV	CSV

Legacy Microsoft Excel files (.xls) will receive Basic-level digital preservation support.

Image

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Tagged Image File Format (.tiff)	Preferred	Full	TIFF	JPG
Joint Photographic Experts Group JPEG (.jpg)	Acceptable	Full	TIFF	JPG
Portable Network Graphics (.png)	Acceptable	Full	PNG	PNG

All other files that identify as “image” mime type are considered “Image” document type with Basic-Level digital preservation support. Some formats may additionally be normalized to TIFF files for preservation by Archivematica, but the Library does not guarantee full support. Examples: .bmp, .dxf, .fif, .flo, .fpx, .g3, .gif, .ico, .ief, .jfif, .jut, .mcf, .nap, .nif, .pbm, .pct, .pcx, .p gm, .pic, .ppm, .pnm, .qif, .qti, .ras, .rgb, .svf, .turbot, .wbmp, .xbm, .xif, .xpm, .x-png, .xwd.

Slideshow

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Microsoft Office Open XML Powerpoint (.pptx)	Preferred	Full	PPTX	PPTX
OpenDocument Presentation Document (.odp)	Preferred	Full	ODP	ODP

Legacy Microsoft Powerpoint files (.ppt) will receive Basic-level digital preservation support.

Audio

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
MPEG-3 audio (.mp3)	Preferred	Full	WAVE(LPCM)	MP4 & OGG
Waveform Audio File (.wav)	Preferred	Full	WAVE(LPCM)	MP4 & OGG
Broadcast Waveform Audio File (.bwf)	Preferred	Full	WAVE(LPCM)	MP4 & OGG
Free Lossless Audio Codec (.flac)	Acceptable	Watch	WAVE(LPCM)	MP4 & OGG

Audio files are migrated to MP4 and OGG for **access format**:

- [Audio.ogg](#) (Format: Vorbis | Bitrate: 96K | Sampling Rate:44.1 kH | Container: OGG)
- [Audio.mp4](#) (Format: AAC (Advanced Audio Codec) | Bitrate: 96K | Sampling Rate: 44.1 kH | Container: MPEG-4)

All other files that identify as “audio” mime type are considered “Audio” document type with Basic-Level digital preservation support. Examples: .aif, .aifc, .aiff, .aip, .au, .funk, .gsd, .gsm, it, .jam, .kar, .la, .lam, .lma, .midi, .mid, .mod, .mp2, .ra, .ram, .rm, .rmi, .rmm, .rmp, .s3m, .sid, .snd, .tsi, .tsp, .voc, .vox, .vqe, .vql, .xm.

[Video](#)

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Matroska container FFV1 codec (.mkv)	Preferred	Full	FFV1/LPCM in MKV	MP4 & OGG
MPEG-4 container H264 codec (.mp4)	Preferred	Full	MP4	MP4 & OGG
QuickTime (.mov)	Acceptable	Watch	FFV1/LPCM in MKV	MP4 & OGG

Video files are migrated to MP4 and OGG for **access format**:

- [Video.ogg](#) (Video Format: Theora | Audio Format: Vorbis| Video Bitrate: 500 KB/s | Video Size: 640*360 (16:9) | Video Frame Rate: 10 FPS | Audio Sampling Rate: 44.1 kHz | Audio Bitrate: 96 kb/s | Container: OGG)
- [Video.mp4](#) (Video Format: AVC (Advanced Video Codec) | Audio Format: AAC (Advanced Audio Codec) | Video Bitrate: 500 KB/s | Video Size: 640*360 (16:9) | Video Frame Rate: 10 FPS | Audio Sampling Rate: 44.1 kHz | Audio Bitrate: 96 kb/s | Container: MPEG-4)

All other files that identify as “video” mime type are considered “Video” document type with Basic-Level digital preservation support. Some formats may additionally be normalized to FFV1/LPCM in MKV files for preservation by Archivematica, but the Library does not guarantee full support.

Examples: .afl, .asf, .asx, .avi, .avs, .dif, .dl, .dv, .flv, .fli, .fmf, .gl, .isu, .mjpg, .lsf, .qt, .rv, .scm, .vdo, .viv, vos, .xdr., .wmv

[Archive](#)

Format	Acceptable/Preferred	Level of Support
ZIP (.zip)	Acceptable	Basic-level
TAR Archive file (.tar)	Acceptable	Basic-level
GZIP compressed file (.gz/.tgz/.tar.gz)	Acceptable	Basic-level

[Other](#)

Document Typeset

Format	Acceptable/Preferred	Level of Support
LaTeX (.tex)	Acceptable	Basic-Level

Geospatial data

Format	Acceptable/Preferred	Level of Support
ESRI Shapefile (.shp, .shx, .dbf)	Preferred	Basic-Level
CAD data (.dwg)	Preferred	Basic-Level
Keyhole Mark-up Language (.kml)	Preferred	Basic-Level
ESRI Geodatabase format (.mdb)	Acceptable	Basic-Level
MapInfo Interchange Format (.mif)	Acceptable	Basic-Level