Alexandra Mills, Special Collections Archivist, Concordia University
Tomasz Neugebauer, Digital Projects & Systems Development Librarian, Concordia University
Tessa Walsh, Digital Preservation Librarian, Concordia University

Special Collections Format Policy Registry

September 5, 2019
Version 1.0

Introduction

Concordia University Library's Special Collections seeks to provide long-term access to the digital materials that compose its holdings. This Format Policy Registry (FPR) defines which file formats are preferred for long-term preservation and access. It covers digital objects received and created by Concordia University Library's Special Collections.

Environmental Scan: Special Collections Needs

Preservation and access are principle responsibilities of the Library's Special Collections. Preserving and providing access to digital holdings presents numerous challenges, including obsolescence, degradation, versioning, and the proliferation of file formats. The majority of digital holdings are a result of digitization initiatives. Special Collections is also committed to preserve and provide access to born-digital materials acquired from individual and organizational donors.

Born-digital

Special Collections acquires materials from private individuals and organizations. Most acquisitions are donations, while some are transfers. Materials are rarely purchased. As a result of how and from whom materials are acquired, Special Collections has little control on which file formats it receives. Limiting which file formats are acquired could negatively impact the acquisition of archival materials. Born-digital materials are usually acquired as part of a larger archival fonds in a variety of file formats.

Digitized

Digitized materials form the bulk of Special Collections' digital holdings. Special Collections digitizes materials for access and preservation. Currently, digitized holdings are primarily audio, video, and photographic materials.

Preservation and Access

Special Collections is committed to the long-term preservation of born-digital and digitized materials. Preservation and access formats should be kept, along with the original documents, to ensure the long-term preservation. Preferably, preservation formats should be stable, uncompressed or losslessly compressed, and openly documented. Access formats should be compact and widely used.

The Special Collections Format Policy Registry defines the file format policies implemented for the preservation and access of Special Collections' digital holdings. Special Collections keeps each digital

object in its original file format. For certain file formats, it will create derivative versions for the purposes of preservation and access. The Special Collections Format Policy Register defines which file formats are migrated and preserved in new formats.

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 managed in Special Collections 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

* Indicates future preservation measure to be implemented.

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:

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

FORMAT POLICY REGISTRY

Text

Format	Acceptable/Preferred	Level of Support	Preservation	Access
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
Microsoft Word (.doc)	Acceptable	Watch	docx*	doc

Basic-level digital preservation support will be provided to other files that identify as "Text" mime type.

Email

Format	Acceptable/Preferred	Level of Support	Preservation	Access
MailBOX (.mbox)	Preferred	Full	mbox	mbox
Personal Storage Table (.pst)	Acceptable	Watch	mbox*	mbox
Email message (.msg)	Acceptable	Watch	msg	msg

Slideshow

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

* Indicates future preservation measure to be implemented.

| Microsoft Office Powerpoint (.ppt) | Acceptable | Watch | pptx* | pdf |

Spreadsheet

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

Images

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
Graphics Interchange Format (.gif)	Acceptable	Watch	gif	gif

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, .ico, .ief, .jfif, .jut, .mcf, .nap, .nif, .pbm, .pct, .pcx, .pgm, .pic, .ppm, .pnm, .qif, .qti, .ras, .rgb, .svf, .turbot, .wbmp, .xbm, .xif, .xpm, .x-png, .xwd.

Audio

Format	Acceptable/Preferred	Level of Support	Preservation Format	Access Format
Waveform Audio File (.wav)	Preferred	full	wav	mp3
Broadcast Waveform Audio File (.bwf)	Preferred	full	wav	mp3
MPEG-3 audio (.mp3)	Preferred	full	wav	mp3
Free Lossless Audio Codec (.flac)	Acceptable	full	wav	mp3

* Indicates future preservation measure to be implemented.

All other files that identify as "audio" mime type are considered "Audio" document type with Basic-Level digital preservation support. Some formats may additionally be normalized to WAV files for preservation by Archivematica, but the Library does not guarantee full support.

Examples: .aif, .aifc, .aiff, .aip, .au, .funk, .gsd, .gsm, .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	Preferred	Full	ffv1 / lcpm / mkv	mp4
MPEG-4 container H264 codec (.mp4)	Preferred	Full	mp4	mp4
QuickTime (.mov)	Acceptable	Watch	ffv1 / lcpm / mkv	mp4

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

* Indicates future preservation measure to be implemented.