

# U-M Library Digital Repository Services Registered Formats and Support Levels

The U-M Library (the “Library”) is committed to providing long-term access to all deposited content by applying best practices for data management and digital preservation while also acknowledging the complexities involved in preserving digital information. The Library commits to preserving the content in the form it is originally deposited and, for some formats, will preserve the content, structure and functionality of the files through migration or other preservation strategies. In addition, the Library will provide basic services including secure storage, backup, management, fixity-checks, and periodic refreshment by copying the data to new storage media.

At the outset, the Library will provide three levels of preservation support for specific file formats. We have determined these support levels by applying a set of evaluation criteria including prevalence of the file format in the marketplace, whether the format is proprietary, the availability of tools for emulation or migration and the availability of local resources to take specific preservation actions. The Library will undertake appropriate format monitoring and provide adequate staffing and other resources to support the services offered at each level. Over time, our ability to provide full preservation support for more formats is likely to grow as additional tools and techniques are developed. To assist content creators in saving and depositing documents that meet the level of quality necessary for full information capture and the highest degree of preservability over time, the Library is developing a set of specification and format best-practice guidelines for common content types.

The Library provides three levels of support for various submission file formats.

## Level 1

The Library will provide its **highest level** of preservation support, making its best effort to maintain the content, structure and functionality in the future. This service level is currently provided only for formats that are both publicly documented and widely used, giving us a high degree of confidence in our preservation commitment, making it more likely that tools will exist or be developed to undertake preservation actions, and that those actions will result in an understood and controlled transformation or migration. The content may also be normalized (transformed to another stable format) to provide additional assurance that the information content is preserved. Finally, the content will be preserved as originally deposited to ensure the original bitstream is always available. TIFF is an example of a Level 1-supported format, as its specifications are publicly available, it is well-supported and widely deployed.

## Level 2

The Library will make **limited** efforts to maintain the usability of the file as well as preserving it as submitted (bit-level preservation). The format will be monitored and may be transformed when significant risk to access is imminent but it is likely to be difficult to predict or control the consequences of any transformation or migration on content, structure or functionality. The file may also be transformed to a more preservable format to ensure that the information content is not lost, even if some structure and functionality are sacrificed. This level of support is generally applied to proprietary formats that are widely used, where there is substantial commercial interest in maintaining access to files saved in the format, and therefore tools will likely be available to migrate them to successor formats (e.g., Microsoft Word).

## Level 3

The Library provides basic preservation of the file (bitstream) and associated metadata **as-is** with no active effort made to monitor the format and associated risks or to normalize, transform or migrate the file to another format. Files may be openable and/or readable by future applications, but there is no guarantee that the content, structure, or functionality will be preserved. This service level usually applies to files written in highly specialized, proprietary formats, often usable only in a single software environment, formats no longer widely utilized, and/or formats about which little information is publicly available. PhotoCD is an example of a format that would receive Level 3 support in the Repository. Any format not yet reviewed and evaluated will also receive Level 3 service on deposit. A higher level may be assigned after format review takes place.

The following chart summarizes the primary preservation services that the Library will provide at the various service levels:

Feature	Level 1	Level 2	Level 3
Persistent identifier that will always point to the object and/or its metadata	•	•	•
Provenance records and other preservation metadata to support accessibility and management over time	•	•	•
Secure storage and backup	•	•	•

Periodic refreshment to new storage media	•	•	•
Fixity checks using proven checksum methods	•	•	•
Storage in a trusted preservable format (making a normalized version, if necessary)	•	for some formats	
Strategic monitoring of format	•	•	
Migration to succeeding format upon obsolescence	•		

The three levels of preservation commitment are made at the individual file level. Complex content items comprised of multiple files in various formats will need additional evaluation to determine whether the operational relationships between the files can be maintained. If the original relationships are documented externally in metadata, that information will be preserved in any case. In addition, executables and some files that rely on a specific hardware/software environment will require additional evaluation because not only the format but the access environment must be considered in making a preservation determination. Because of the collection policy for accepting finished work, we discourage submissions in some "working" formats, such as Photoshop and Final Cut Pro, since we will be unable to offer the highest level of preservation support for them. If you have material to submit that falls into any of these categories, please contact [digitalpreservation@umich.edu](mailto:digitalpreservation@umich.edu) prior to making your submission.

## Registered Formats and Support Levels

*This list of formats and support levels will be regularly reviewed and updated based on our growing experience with digital preservation and the emergence of new formats and standards.*

Under review as of June 11, 2020

TEXT AND PAGE DESCRIPTION FORMATS ( <a href="#">Best Practices For Creating Quality PDFs</a> )				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
PDF/A	.pdf	application/pdf	Level 1	<a href="#">Best Practices</a>  Files not created per the "Best Practices" receive Level 2 support, and may be migrated to PDF/A
Plain Text UTF-8 (Unicode)	.txt	text/plain; charset=UTF-8	Level 1	
Plain Text ANSI X3.4/ECMA-6/US-ASCII (7-bit)	.txt	text/plain; charset=US-ASCII	Level 1	
SGML	.sgm, .sgml	application/sgml	Level 1	Requires that DTD is deposited with SGML file and that SGML file parses against it
XML	.xml	text/xml	Level 1 / Level 2	Level 1 requires that DTD/schema is deposited with XML file and that XML file parses against it; Level 2 assumes no DTD/schema but

				that XML file is well-formed
HTML	.html, .htm	text/html	Level 2	Requires HTML 4.0 or 4.01 validated markup and CSS files(s), if referenced, be deposited with document
LaTeX	.latex	application/x-latex	Level 2 / Level 3	Level 2 requires that referenced style files and/or embedded items be deposited with document
Postscript	.ps	application/ps	Level 2	
Rich Text	.rtf	text/richtext	Level 2	
TeX	.tex	application/x-tex	Level 2 / Level 3	Level 2 requires that referenced style files and/or embedded items be deposited with document
Plain Text ISO 8859-x (8-bit)	.txt	text/plain; charset=ISO-8859-x	Level 2	
Plain Text; all other encodings (including, but not	.txt	text/plain	Level 3	

limited to ISO 646 national variants)				
---------------------------------------	--	--	--	--

COMMON DESKTOP SOFTWARE FORMATS ( <a href="#">Best Practices</a> )				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
Microsoft Word	.doc	application/msword	Level 2	Requires that macros be disabled
Microsoft PowerPoint	.ppt	application/vnd.ms-powerpoint	Level 2	Requires that macros, animation and other effects be disabled
Microsoft Excel	.xls	application/vnd.ms-excel	Level 2	Requires that macros be disabled. (See also <a href="#">Best Practices for Datasets</a> )

IMAGE FILE FORMATS ( <a href="#">Best Practices</a> )				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
JPEG	.jpg	image/jpeg	Level 1	
TIFF	.tiff	image/tiff	Level 1	
JPEG 2000			Level 2	Level 1 support expected as more tools become available
PNG	.png	image/png	Level 2	

BMP	.bmp	image/x-ms-bmp	Level 3	
GIF	.gif	image/gif	Level 3	
Photo CD	.pcd	image/x-photo-cd	Level 3	
Photoshop	.psd	application/x-photoshop	Level 3	

AUDIO ( <a href="#">Best Practices</a> )				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
AIFF	.aif, .aiff	audio/aiff, +	Level 1	
Wave	.wav	audio/x-wav or audio/wav	Level 1	
Audio/Basic	.au, .snd	audio/basic	Level 2	
MPEG audio	.mp3	audio/mpeg, audio/mp3	Level 2	
AAC_M4A	m4a, .mp4	audio/m4a, audio/mp4	Level 2	
Real Audio	.ra, .rm, .ram	audio/vnd.rn-realaudio	Level 3	
Windows Media Audio	.wma	audio/x-ms-wma	Level 3	

VIDEO ( <a href="#">Best Practices</a> )				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
AVI	.avi	video/avi, video/msvideo, video/x-msvideo +	Level 2	
Quicktime	.mov	video/quicktime, video/x-quicktime	Level 2	
MPEG-1 MPEG-2	.mp1 .mp2	video/mpeg video/mpeg2	Level 2	Many variants possible; preservation level not yet established
MPEG-4	.mp4	video/mp4	Level 2	Many variants possible; preservation level not yet established
Windows Media Video	.wmv	video/x-ms-wmv	Level 3	

OTHER/MISCELLANEOUS				
Format	File Extension	Mime Type	Support Level	Qualifying Factors/Notes
ZIP/tar	.zip, .gz, tar.gz	application/zip; application/x-gzip	Level 1; see "Qualifying Factors/Notes"	ZIP or tar files are only as good as their contents; all best practices above still apply. (See also <a href="#">Best</a>



				<a href="#">practices for producing ZIP and tar files)</a>
--	--	--	--	--