



# UNIVERSITY *of* ROCHESTER

## URRR (University of Rochester Research Repository)

### Preferred File Formats

#### Introduction

This brief guide overviews the preferred file formats and specifications for materials submitted to the UR Research Repository, as well as some common software used to view, render, or otherwise interact with those formats.

When deciding on format preferences, URRR errs on the nexus of faithful, preservable and accessible, where “faithful” means the file is as the creator originally intended or presented; where “preservable” means the file will likely be usable well into the future; and where “accessible” means the file will be usable by the greatest number of users across as wide a spectrum of contexts as possible.

Ideally, users will submit files that exist at the aforementioned nexus, but this is sometimes difficult or impossible depending on the source material. Therefore, **URRR administrators will usually accept any format** and attempt, when appropriate and desirable, to make more preservation-friendly, accessible file derivatives.

#### Preferred File Formats

File Type	Preferred Format	Software	More Info
<b>Word Processing</b>	PDF/A (.pdf); Open Document Format (.odf); MS Word Open XML (.docx)	Adobe Acrobat Reader, MS Word, Apache OpenOffice	Word processing documents are any text-based works that rely on formatting. The listed formats are widely accepted and versatile. Avoid embedding audio or video files in PDFs.

<b>Text</b>	Plain UTF-8 text (.txt); plain text with markdown formatting	Notepad, MS Word, Apache Open Office	Plain text is the most widely accepted way to format and distribute textual data, with or without formatting notation.
<b>Data &amp; Datasets</b>	Delimited flat file (.csv, .txt)	MS Excel, Notepad	This is specific to tabular data and includes tab- and comma-separated formats. Some datasets cannot be represented so simply, and we can accept most proprietary formats (e.g., ACCDB, SAS).
<b>Raster &amp; Vector Images</b>	PNG (.png); TIFF (.tif); JPEG2000 (.jp2); Scaled Vector Graphics (.svg)	Adobe Photoshop and/or Illustrator, MS Paint, IrfanView	For images, we accept whatever format in which the work was originally exported but prefer uncompressed losslessly compressed formats like those listed.
<b>Audio</b>	WAVE (.wav); AIFF (.aiff); FLAC (.flac)	VLC, Windows Media Play	For audio files, we accept whatever format in which the work was originally exported but prefer uncompressed or losslessly compressed formats like those listed.
<b>Video</b>	AVI (.avi); MKV (.mkv), Motion JPEG2000 (.mj2), MP4 (.mp4, .m4v)	VLC, Windows Media Player, Adobe Premiere	For video files we accept whatever format in which the work was originally exported but prefer uncompressed or losslessly compressed formats like those listed (*MP4 video is compressed, but universally accepted).
<b>Presentations/ Slide Decks</b>	MS PowerPoint Open XML (.pptx); PDF/A (.pdf)	Adobe Acrobat Reader, MS PowerPoint	Like word processing documents, we prefer the most accessible versions that preserve formatting and are faithful to their original presentation.

<b>Spreadsheets</b>	CSV (.csv); MS Excel Open XML (.xlsx)	MS Excel, Notepad	Like tabular datasets, spreadsheets are best preserved as comma-separated value files or MS Excel Open XML files without multiple worksheets.
---------------------	---------------------------------------	-------------------	---

For more information on file formats like their accessibility features, predicted longevity, and uses, check out The Library of Congress's format sustainability website:

<https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>

If you have questions regarding your submission and whether your files are acceptable, please feel free to [contact us](#).

## Glossary

**Delimited flat file:** a file wherein data fields and their contents are separated – or delimited – by commas, pipes, or even tabs. These are simple and interoperable ways to represent tabular data, like those in spreadsheets.

**Lossless compression:** compression standard that makes files smaller and more easily transmittable but retains all data and fidelity in the process. Examples of lossless compression formats include FLAC for audio, Motion JP2000 for video, and JPEG 2000 for images.

**Lossy compression:** compression standard that makes files smaller and more easily transmittable but sacrifices data and fidelity in the process. Examples of lossy compression formats include h.264 for video, Ogg Vorbis for audio, and JPEG for images.

**Raster images:** images based on thousands of pixels, or squares. Examples include scanned photographs, digital camera images, and screenshots. File types for raster images include JPEG, TIFF, and PNG.

**Uncompressed files:** files that receive no compression before export from their native environments. Such files are often used for archiving and are not recommended for common use or transmission through a web-based repository. Examples of uncompressed files include PCM for audio, YUV for video, and BMP for images.

**UTF-8:** a character encoding set based on Unicode that constitutes encoding for almost all operating systems, electronic documents, and web-based text. It is backwards compatible with earlier character sets like ASCII.

**Vector images:** images based on points along a line and the equations that connect them. These are typically used in graphic design and include things like Adobe Illustrator projects or print-ready graphic documents. File types for raster images include EPS, SVG, and even PDF.