Research Engagement, Library

University of Reading

RECOMMENDED FILE FORMATS

University of Reading Research Data Archive

Contents

Introduction: file format categories

The University of Reading Research Data Archive will accept any file type that you choose to deposit, but you should wherever possible aim to deposit files that are optimised for long-term preservation and use. For these purposes, the Archive acknowledges three categories of files:

- **Recommended file formats**: standard preservation;
- **Acceptable file formats**: general purpose;
- **Other file formats**: specialist and rarely-used.

Guidance on file formats with examples of recommended and acceptable formats is provided in this document.

Overview: formats for preservation and use

To ensure your data remain accessible and usable in the long term, you should where possible deposit files in standard preservation file formats.

- Preservation formats typically encode information in a way that is software-independent or allows interoperability between systems and applications. These are usually recognised standard or open formats, such as OpenDocument Format (ODF), ASCII text (.txt), comma-separated values (.csv), or XML.

- Proprietary formats are more suitable for preservation if they are widely-used and can be opened in different operating systems and with different programmes or applications: examples include Microsoft Rich Text Format, Excel, and PDF.

- The preferred preservation formats for representations of image, audio and video data are those that encode the information with a lossless algorithm, such as TIFF (.tif), Free Lossless Audio Codec (.flac), and MPEG-4 (.mp4). Unlike lossy formats (e.g. JPEG, .mp3, WMV), lossless formats do not cause data loss when files are saved.

We recommend you deposit your files in the Archive using open, platform-independent or non-proprietary file formats wherever possible.

Preservation formats may not always be the most accessible or usable:

- Open formats may lack the formatting and functionality that allow data to be rendered, manipulated and analysed more effectively.

- Information-rich lossless formats, such as high-resolution image or video, may produce very large files which are not suited for web access or fast processing.

In such cases you may wish to deposit files in more than one format, so that data are optimised for both preservation and use: for example, a richly-formatted Excel file containing manipulated tabular data might be made available alongside a CSV file of the raw numbers; or a large TIFF image file might be deposited for preservation, and a JPEG or other compressed image file made available for web download.

You may need to document your file formats, for example, recording the version of the software used to generate them, and including, if relevant, details of the compression, codec and bit rate used. This information can be entered in the **Data processing and preparation activities** field in the Metadata Record; it can also be recorded in a README.txt file or other documentation file uploaded with your data.

For general information on choosing appropriate formats for preservation and other uses, consult the Further information references at the end of this document.

Recommended file formats

The Archive supports and recommends the following file formats. These formats are optimised for preservation, and you are encouraged to deposit files in these formats (converted from other formats if necessary) wherever possible.

File type	Format name	File extension(s)
Text	Microsoft Powerpoint XML	.pptx
	Microsoft Word XML	.docx
	OpenDocument Presentation	.fodp, .odp
	OpenDocument Text	.fodt, .odt
	Plain Text	.txt, .asc

Recommended File Formats

File type	Format name	File extension(s)
Documentation	Adobe PDF/A	.pdf
	Microsoft Word XML	.docx
	OpenDocument Text	.fodt, .odt
Markup	CSS	.css
	HTML	.htm, .html
	SGML	.sgm, .sgml
	XML	.xml
Tabular	Comma separated values (CSV)	.csv
	Microsoft Excel XML	.xlsx
	OpenDocument Spreadsheet	.fods, .ods
	Tab separated values	.tsv, .tab
Image/graphics	GIF	.gif
	JPEG 2000	.jpxml, .jp3d, .jpf, .jpm, .jpx, .jp2
	PNG	.png
	Postscript	.ps
	Scalable Vector Graphics (SVG)	.svg
	TIFF	.tiff, .tif
Audio	AIFF	.aiff, .aif, .aifc
	Free Lossless Audio Codec (FLAC)	.flac
	WAV	.wav
Video	Motion JPEG2000	.mjp2, .mj2
	MPEG-4	.m4v, .m4r, .m4b, .m4p, .m4a, .mp4
Database	Mineset	.schema, .data
	Minitab syntax and output	.lis, .tj
	R (ascii)	.rdata
	SAS syntax	.sas
	SPSS portable	.por
	SPSS syntax	.sps
	Stata syntax	.do, .dct
	Structured Query Language	.sql
Geospatial	CAD	.dwg
	ESRI Shapefile	.shp, .shx, .dbf, .prj, .sbx, .sbn
	Geo-referenced TIFF	.tif, .tfw

Acceptable file formats: general purpose

The file formats listed below are not considered to be standard preservation formats, for one or more of the following reasons:

- they are proprietary and system-, software- or version-dependent;
- they are lossy (data are lost when compression is applied);
- they are not as common as formats that have been recommended above.

These are mostly general-purpose formats, and as many are widely used, it is likely that the Archive will be able to preserve them. But preservation of data files in these formats cannot be guaranteed, and they are deposited at your own risk.

If you wish to preserve information, formatting or functionality encoded in files in these formats, you may deposit the original files alongside the same files converted into standard preservation formats.

For instance, you could deposit your SPSS files (.por or .sav) along with standard preservation files in SPSS syntax (.sps) and Plain Text (.txt) data files.

File type	Format name	File extension(s)
Text	Microsoft Powerpoint	.ppt
	Microsoft Project	.mpt, .mpp, .mpx, .mpd
	Microsoft Word	.doc
	WordPerfect	.w51, .wp5, .wp, .wpd
	Rich Text Format	.rtf
Documentation	Microsoft Word	.doc
	Rich Text Format	.rtf
Markup	EAF File	.eaf
	LateX	.ltx, .latex
	PFSX File	.pfsx
	TeX	.tex
	TeX dvi	.dvi
Tabular	Microsoft Excel	.xls
Image/graphics	Bitmap	.ddb, .dib, .bmp
	Encapsulated PostScript (EPS)	.epsi, .epsf, .eps
	JPEG	.jpeg, .jpg, .jpe
	Microsoft Visio	.vsd
	Photo CD	.pcd
	Photoshop	.psd, .pdd
	VTK (Visualisation ToolKit)	.vtu

Recommended File Formats

File type	Format name	File extension(s)
Audio	Audio	.au, .snd
	FMP3	.fp, .fmp, .fp3, .fm
	MPEG Audio	.m4a, .mpa, .abs, .mpega
	RealAudio	.rpm, .ra, .ram
Video	AVI Audio/Video Interleaved Format	.avi
	Flash Video	.f4b, .f4a, .f4p, .f4v, .flv
	MPEG	.mpeg, .mpg, .mpe
	Ogg Vorbis Codec Compressed Multimedia File	.ogg
	Video Quicktime	.qtm, .mov, .qt
Database	DBase, DBF	.dbf
	Microsoft Access	.mdb, .accdb
	SPSS output file	.spv, .spo
	SPSS system file	.sav, .gsav
Archive	BZIP2	.bz2, .bz
	Compressed Archive File	.zip
	GZIP compressed archive file	.gz
	Tarball	.tar

Other file formats: specialist and rarely-used

Some data may be saved in specialist or rarely-used software-dependent file formats that are domain-specific or unique to the instruments or software from which the raw data were generated. These are less likely to be suitable for long-term preservation, and there is a risk that in the future it will not be possible to open the files, if the software becomes obsolete or is not easily available.

Where possible you should convert files to recommended or acceptable formats as listed above. Original and converted files can be deposited together to maximise both current usability and long-term accessibility.

If a file format is not listed in this document, you should ensure your documentation includes a description of the file format and provides information about any software required to render the files.

If you have research data in file formats you are unsure about, need help converting your files to standard preservation formats, wish to propose a file format for inclusion in this list, or have any other enquiries, please contact us at researchdata@reading.ac.uk.

Further information

The National Archives. PRONOM. https://www.nationalarchives.gov.uk/PRONOM/Default.aspx

Online registry of file formats.

UK Data Service. Recommended formats. http://ukdataservice.ac.uk/manage-data/format/recommended-formats

Recommended formats for depositing files in the UK DataService.

Wikipedia. List of file formats. https://en.wikipedia.org/wiki/List_of_file_formats#Tabulated_data

Including common scientific data formats.