

Edinburgh DataShare: Recommended File Formats

To ensure access and usability of your data to the broadest audience into the long term, the Data Library encourages you to deposit standard preservation file formats which we are confident we can preserve.

The digital preservation community recommends these formats because, either:

- they encode the information in a way that is software independent or allow interoperability between systems and applications. Often, these formats are a recognised standard, or published in an open format. Some of these file formats might be proprietary but can be opened in different operating systems and with different programmes or applications.

or

- the information is encoded with a lossless algorithm and for that reason there is no data loss when files are 'saved as' and stored in these formats.

Where possible, we recommend you deposit your files in Edinburgh DataShare using open, platform independent or nonproprietary file formats. Some of the most common standard preservation formats use extensions such as: .csv, .xml, .pdf, .tif, .txt, .jp2, .mp4, .html, .flac.

Here is a list of file formats that Edinburgh DataShare supports and recommends:

Format name	File extension(s)
Adobe PDF /A	.pdf
XML	.xml, .sgml
Text	.txt, .asc, .sts
HTML	.htm, .html
CSS	.css
GIF	.gif
PNG	.png
TIFF	.tiff, .tif
JPEG 2000	.jpxml, .jp3d, .jpf, .jpm, .jpx, .jp2
Scalable Vector Graphics (SVG)	.svg
AIFF	.aiff, .aif, .aifc
WAV	.wav
Free Lossless Audio Codec (FLAC)	.flac
Postscript	.ps
Structured Query Language	.sql
Comma separated values (CSV)	.csv
Tab separated values	.tsv, .tab
MPEG-4	.m4v, .m4r, .m4b, .m4p, .m4a, .mp4
Motion JPEG2000	.mjp2, .mj2
OpenDocument - text	.fodt, .odt

Format name	File extension(s)
OpenDocument - spreadsheet	.fods, .ods
OpenDocument - presentation	.fodp, .odp
SAS syntax	.sas
SPSS syntax	.sps
Stata syntax	.do, .dct
Minitab syntax and output	.lis, .tj
R (ascii)	.rdata
Mineset	.schema, .data
Microsoft Powerpoint XML	.pptx
Microsoft Word XML	.docx
Microsoft Excel XML	.xlsx

Other selected file formats

File formats such as the ones listed below have been deposited in the repository but are not considered standard preservation formats because they are either proprietary and system, software or version dependent, are considered lossy (data are lost when compression is applied) or are not as common as the ones mentioned above. Most of these formats are widely used and it is likely we will be able to preserve them, but cannot guarantee it. If you have files in these formats, you may deposit them in Edinburgh DataShare. We recommend you deposit original files in these formats along with the same data files converted into standard preservation formats. For instance, you could deposit the .tex and .bib files used to generate your documents along with your final LaTeX output in .pdf (PDF/A) format, or deposit your SPSS files (.por or .sav) along with standard preservation files in .sps (SPSS syntax) and .txt (Plain Text data file).

Format name	File extension(s)
Microsoft Word	.doc
Microsoft Powerpoint	.ppt
Microsoft Excel	.xls
JPEG	.jpeg, .jpg, .jpe
Audio	.au, .snd
MPEG	.mpeg, .mpg, .mpe
RTF	.rtf
Microsoft Visio	.vsd
FMP3	.fp, .fmp, .fp3, .fm
BMP	.ddb, .dib, .bmp
Photoshop	.psd, .pdd
Video Quicktime	.qtm, .mov, .qt
MPEG Audio	.m4a, .mpa, .abs, .mpega
Microsoft Project	.mpt, .mpp, .mpx, .mpd
Mathematica	.nbp, .nb
LateX	.ltx, .latex
TeX	.tex
TeX dvi	.dvi
SGML	.sgm, .sgml
WordPerfect	.w51, .wp5, .wp, .wpd
RealAudio	.rpm, .ra, .ram
Photo CD	.pcd
NetCDF Network Common Data Form	.nc

Format name	File extension(s)
Compressed Archive File	.zip
DBase, DBF	.dbf
Microsoft Access	.mdb
SPSS portable file	.por
SPSS system file	.sav, .gsav
PFSX File	.pfsx
EAF File	.eaf
TextGrid	.textgrid, .TextGrid
SPSS output file	.spv, .spo
Flash Video	.f4b, .f4a, .f4p, .f4v, .flv
ML source code file	.ml
LAB	.lab
MTRANS file	.mtr
PSC	.psc
PITCH File	.pitch, .PITCH
PitchTier File	.pitchtier, .PITCHTIER
RESULTSMFC File	.RESULTMFC
FLT	.flt
AVI Audio/Video Interleaved Format	.avi
Ogg Vorbis Codec Compressed Multimedia File	.ogg
NifTi	.img, .hdr, .nii
Encapsulated PostScript (EPS)	.epsi, .epsf, .eps
VTK (Visualisation ToolKit)	.vtu
HDF (Hierarchical Data Format)	.he4, .h5, .hdf4, .h4, .hdf, .he5, .hdf
GZIP compressed archive file	.gz
BZIP2	.bz2, .bz
Tarball	.tar
BED	.bed
bedGraph	.bg

Unknown files

When you upload a file format in Edinburgh DataShare which the system has never received before, the file will be tagged as “unknown.” The file you have submitted might be a fairly standard format in your discipline, but has been created with a proprietary or open software and saved as a proprietary or open file format which DataShare has never encountered before. The Data Library will look into these formats, contact you if more information about them is required, and evaluate them for inclusion in the repository. We recommend you also consider exporting files which Edinburgh DataShare does not recognise into supported file formats (standard preservation) and deposit them along with your original files. The danger of storing your research data only in proprietary or software dependent formats is that you and others might not be able to open the files in the future when using a newer version of the same software or with other free, open or proprietary software packages.

If you have research data in file formats you are unsure about, need help converting your files to standard preservation formats, or simply want to discuss your needs with us, please contact us at datalib@ed.ac.uk.