

List of EPFL Recommended File Formats

Type of Data	Sub Type	Recommended File Formats for Sharing, Reuse and Long-Term Preservation	Acceptable Formats (up to 10 year)	Not Suitable for Preservation
Dataset	Tabular data with extensive metadata	comma-separated values (CSV) file unicode UTF-8 (.csv) with CSV on the Web descriptive metadata Hierarchical Data Format version 5 HDF5 (.hdf5)	plain text data, ASCII (.txt) Hypertext Mark-up Language (HTML) (.html) LaTeX (.tex) SPSS portable format (.por)	
	Tabular data with minimal metadata	comma-separated values (CSV) unicode UTF-8 file (.csv) tab-delimited file (.tab) OpenDocument Spreadsheet (.ods) Structured Query Language (SQL) dump, preferably from an open tool (PostgreSQL, MariaDB)	eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) MS Excel (.xlsx)	MS Excel (xls, .xlsb)
Text	Textual data	PDF/A (.pdf) plain text, unicode UTF-8 (.txt) Open Document Text (.odt, .odm) LaTeX (.tex) Markdown (.md) HTML (.htm, .html), XHTML 1.0 XML marked-up text (.xml), with specified schema	widely-used proprietary formats, e.g. MS Word (.docx) PowerPoint (.pptx) PDF with embedded forms Rich Text Format (.rtf)	MS Word (.doc) PowerPoint (.ppt)
	Code	Plain text formats (e.g. Matlab/Octave .m, R-project .R, Python .py, and so on) Jupyter Notebook (.iypnb) Rstudio (.rstudio), Rmarkdown (.rmd) NetCDF	Text files for S-plus (.sdd) Matlab .mat, should be saved in HDF format.	Matlab (.mat) R-project (.rdata)
Multimedia	Digital image data	Raster : TIFF version 6 uncompressed (.tif) Portable Network Graphics, PNG uncompressed (.png)  Vector : Scalable Vector Graphics SVG1.1 (.svg)	Raster : JPEG (.jpeg, .jpg), JPEG 2000 (.jp2) TIFF (other versions: .tif, .tiff) Adobe Portable Document Format (PDF/A, PDF) (.pdf) GIF BMP	Vector : Adobe InDesign (.indd) Adobe Illustrator (.ait) Adobe Photoshop (.psd)
	Digital audio data	Free Lossless Audio Codec (FLAC) (.flac) Waveform Audio Format (WAV) (.wav) Vorbis OGG (.ogg)	MPEG-1 Audio Layer 3 (.mp3) Advanced Audio Coding (.mp4) Audio Interchange File Format (AIFF) (.aif)	
	Digital video data	Codecs : MPEG-4 High Profile (.mp4) [Codec : H.264] Motion JPEG 2000 (.mj2), ISO/IEC15444-4  Containers : Audio Video Interleave, uncompressed AVI (.avi) Matroska Multimedia Container (.mkv)	Codecs : OGG Theora Vorbis (.ogg) WebM (.webm)  Containers : Vorbis OGM (.ogm)	Windows Media Video (.wmv) QuickTime (.mov)
Geospatial & CAD	Geospatial data	NetCDF tabular GIS attribute data ESRI Shapefile (essential: .shp, .shx, .dbf and optional: .prj, .sbx, .sbn) PostGIS geo-referenced TIFF (.tif, .tfw) GeoJSON	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data	
	CAD / vector and raster data	CAD data (.dwg) Drawing Interchange Format, AutoCAD (.dxf) Extensible 3D, X3D (.x3d, .x3dv, .x3db) PDF et PDF3D (for documentation)	Good practice would be to : 1. Keep original proprietary design files (for instance .prt and .asm) 2. Keep derived production files (for example .step, .Stl, .iges, .dxf, .drill or .gbr) 3. Generate description files for long term (as an example .pdf or .pdf3d)	
Generic	Generic Data	XML JSON RDF		