

DEFINITION

A file format is a standard way to encode data as a computer file. It follows a protocol that specifies how bits are used to encode information in a digital storage medium. File formats may be **proprietary** or **open** [1].



DO NOT FORGET

When listing out the data formats you will be using, make sure to include:

- ✓ The necessary **software** to view the data (e.g. SPSS v.3; Microsoft Excel 97-2003);
- ✓ Information about **version control**;
- ✓ If data are stored in one format during collection and analysis, then converted to another for preservation: list out features that may be lost in data **conversion** such as system-specific labels.

IDEALLY...

... prefer file formats that are:

-  **Unencrypted, uncompressed**, commonly **used** by the research community
-  Compliant to an **open, documented** standard: **interoperable** among diverse platforms and applications, fully **published** and available **royalty-free**, fully and independently **implementable** by multiple software providers on **multiple platforms** without any intellectual property [2]

TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
<b>Tabular</b> (extensive metadata)	csv - hdf5	txt - html - tex - fastq - por	
<b>Tabular</b> (minimal metadata)	csv - tab - ods - sql - tsv	xml (if appropriate dtd) - xlsx	xls - xlsb
<b>Textual / Presentation</b>	txt - pdf - odt - odm - tex - md - htm - xml - extxyz - odf	pptx - rtf - docx - pdf (with embedded forms) - eps - ipf	doc - ppt - dvi - ps
<b>Code / Computation</b>	m - r - py - iypnb - rstudio - rmd - NetCDF - aiml	sdd	mat - rdata
<b>Image / Spectroscopy</b>	tif - png - svg - jpeg - fits	jcamp - jpg - jp2 - tif - tiff - pdf - gif - bmp - dm3 - oir - lsm	indd - ait - psd - spc
<b>Audio</b>	flac - wav - ogg - mxl - midi - mei - humdrum	mp3 - aif	
<b>Video</b>	mp4 - mj2 - avi - mkv	ogm - mp4 - WebM	wmv - mov - qt
<b>Geospatial</b>	NetCDF - tabular gis attribute data - shp - shx - dbf - prj - sbx - sbn - postGis - tif - tfw - geoJson	mdb - mif	
<b>3D structures &amp; images</b>	x3d - x3dv - x3db - pdf3D - pov - pdbml	dwg - dxf - pdb	pxp
<b>Other</b>	xml - json - rdf		

WANT MORE?

Check out also these tables:

- [fileformats.archiveteam.org/wiki/Scientific\\_Data\\_formats](https://fileformats.archiveteam.org/wiki/Scientific_Data_formats)
- [loc.gov/preservation/resources/rfs/TOC.html](https://loc.gov/preservation/resources/rfs/TOC.html)
- [archives.gov/records-mgmt/policy/transfer-guidance-tables.html](https://archives.gov/records-mgmt/policy/transfer-guidance-tables.html)
- [nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer](https://nationalarchives.gov.uk/information-management/manage-information/digital-records-transfer/file-formats-transfer)

Credits and sources

[1] [en.wikipedia.org/wiki/File\\_format](https://en.wikipedia.org/wiki/File_format)  
[2] [guides.library.stanford.edu/data-best-practices/format-files](https://guides.library.stanford.edu/data-best-practices/format-files)