

# **PubData File Format Policy**

## **Guideline for the submission, storage and publication of file formats in PubData, the data service of the Media and Information Centre (MIZ), Leuphana**

Version: 1.0, Date: 2021-12-01

The MIZ aims to archive and publish research data and materials in a commonly and scientifically accepted way ensuring reuse and replication of the research results in the long term without technical restrictions and, at best, free of charge. In order to guarantee long-term usability, appropriate backup and storage procedures are carried out according to the official Leuphana Archiving Directive for Research Data (Leuphana Archivierungsrichtlinie Forschungsdaten).

The choice of suitable file formats is one of the most important requirements for technical preservation and reusability of data. Depending on the software used for the creation and processing of the data, some file formats are more appropriate for long-term preservation than others. The key aspects are suitability for long-term storage, compatibility with other programmes and lossless conversion. According to reusability in the sense of Open Science, it is crucial to prefer open, non-proprietary or at least common, widely used file formats.

These criteria are applied during the submission, storage and the publication of data collections as they are documented and anchored in the PubData Collection Policy Research Data. This implies a technical file format check for submitted data by MIZ as an essential part of the data ingest check and as a mandatory prerequisite for the transfer into the Leuphana archiving system, the preservation, storage and, at best, their publication.

The MIZ reserves the right to reject data sets including file formats that are unsuitable for the named purposes. However, such a decision will always be made in consultation with the data depositing party and after exploring alternative solutions. Nevertheless, when choosing your software and file formats, please consider in advance the resulting consequences with regard to storage and, if applicable, publication in terms of good scientific practice (in accordance with the Leuphana Guideline for Safeguarding Good Scientific Practice [-> more details](#)).

The Leuphana guidelines and policies referenced in this document can be found in the download area at the [PubData Guidelines Overview](#).

In the upcoming table, you can find a brief overview of recommended and accepted file formats by format type:

Format type	Recommended	Accepted	Not recommended
Text	ANSI X3.4/ECMA-6/ US-ASCII (7-bit); EPUB (unencrypted) (*.epub); Open Office (*.sxw/*.odt); PDF/UA; PDF/A-1; PDF/A-1a; PDF/A-1b; PDF/A-2; PDF/A; Plain text: UTF-8, UTF-16 with BOM; SGML; XML and XML-based Markup formats; HTML; Cascading Style Sheets (*.css)	PDF; Postscript; Rich Text Format 1.x (*.rtf); SGML (*.sgml); DTD (*.dtd); LaTeX; TeX; OOXML (ISO/IEC DIS 29500) (*.docx)	PDF with embedded fonts or encryption; PDF/A-3; DVI (*.dvi); Microsoft Word (*.doc); WordPerfect (*.wpd); Google Docs
Image	JPEG; PNG; SVG; TIFF	Bitmap or BMP (*.bmp); Computer Graphic Metafile (*.cgm); Digital Negative DNG	Encapsulated Postscript (*.eps, *.epsf, *.ps); FlashPix (*.fpx); GIF; JPEG 2000 Part 2 (*.jpf, *.jpx);

		(* .dng); GIF (* .gif); Google WebP (* .webp); JPEG 2000; JPEG/JFIF (* .jpg); JPEG2000 (lossy) (* .j2); Drawing Interchange Format (* .dxf)	Macromedia Flash; MrSID (* .sid); Photo CD (* .pcd); Photoshop (* .psd, * .psb [Large doc format], * .acv, * .atf [Photoshop Curve File]); RAW; TIFF (Planarformat)
Audio	AES3 (* .aes); AIFF (* .aif, * .aiff); Broadcast WAV (BWF); WAV	Advance Audio Coding (* .mp4, * .m4a, * .aac); Free Lossless Audio Codec (* .flac); MP3 (MPEG-1/2, Layer 3)(* .mp3); Ogg Vorbis (* .ogg); Standard MIDI (* .mid, * .midi); SUN Audio/Basic	AIFC (compressed) (* .aifc); NeXT SND (* .snd); Protected AAC (* .m4p); RealNetworks 'Real Audio' (* .ra, * .rm, * .ram); WAV; Windows Media Audio (* .wma)
Video	AVI (Audio Video Interleaved) (* .avi); Motion JPEG 2000 (ISO/IEC 15444-4); Motion JPEG 2000 (* .mj2); QuickTime Movie (* .mov); MPEG-4; FFV1	DV; MPEG-1; MPEG-2; Matroska Multimedia Container (* .mkv)	Ogg Theora (* .ogg); RealNetworks 'Real Video' (* .rv, * .rm); Windows Media Video (* .wmv); Flash Video (* .flv); Macromedia Flash (* .swf); Advanced Authoring Format Object (* .aaf); Material Exchange Format (* .mfx); Digital Cinema Initiative Distribution Master (DCDM); Apple ProRes; Cinepak; DivX; Digital Moving Picture Exchange (DPX)
Presentation	Depends on the file content	PDF-A; Microsoft Powerpoint	Microsoft Powerpoint (.pptx)

		OOXML (.pptx); Open Office (.sxi and .odp)	
Structured Data	CSV; TXT; SQL DDL	Microsoft Excel (.xlsx); OOXML; Open Office; DBF; JSON; R; SPSS (.por and .sav); SAS; Stata; MatLab (*mat) in HDF Format	Microsoft Excel (.xls); compressed Matlab
E-Mail	MBOX; Internet Message Format (EML)	Email Account XML Schema	Microsoft Outlook Item (*msg); Personal Storage Table (*pst); Outlook for Mac (*olm)
Web Archive Format	Mozilla Archive Format (MAFF) (*mar); Web Archive File (WARC) (*.warc)	-	-
Archive File Format	ZIP (*.zip); Tape Archive (*tar); gzip (*gz);	Unix Archiver (*a, *.ar); CPIO (*.cpio); lzip (*.lz); lzma (*.lzma); 7z (*.7z); bzip2 (*.bz2)	Shell archive (*shar); lzop (*lzo); 7zX (*s7z); RAR (*rar); FreeArc (*arc); Cabinet (*cab); extensible Archive format (*xar)
Other	XD3 (unknown or undefined file extensions)	VRML; U3D; Computer Pgm Source Code (*c, *.c++, *.java, *.js, *.jsp, *.php, *.pl, etc)	Compiled or Executable Files (EXE, *.class, COM, DLL, BIN, DRV, OVL, SYS, PIF)