

PubData Dateiformat-Policy

Richtlinie zur Übernahme, Aufbewahrung und Veröffentlichung von Dateiformaten in PubData, dem Datenservice des Medien- und Informationszentrums (MIZ) der Leuphana

Version 1.0, Stand: 01.12.2021

Das MIZ setzt sich das Ziel, Forschungsdaten und Materialien so zu sichern und zu veröffentlichen, dass diese Datenbestände ohne technische Einschränkungen und bestenfalls kostenfrei langfristig nutzbar sind. Um eine langfristige Verwendbarkeit zu garantieren, werden entsprechende Sicherungs- und Aufbewahrungsverfahren durchgeführt (siehe Leuphana Archivierungsrichtlinie Forschungsdaten).

Eine wichtige Grundvoraussetzung zur technischen Aufbewahrung und Nachnutzbarkeit ist die Wahl geeigneter Dateiformate. Je nach verwendeter Software sind einige Dateiformate besser geeignet. Ausschlaggebend für die langfristige Sicherung sind Aspekte wie die **Eignung zur Langzeitaufbewahrung**, **Kompatibilität** zu anderen Programmen, **verlustfreie Konvertierung**.

Im Sinne von Nachnutzung und Open Science sind außerdem **offene, nicht-proprietäre** oder zumindest **gängige, weit verbreitete Dateiformate** zu bevorzugen.

Für die Übernahme, Sicherung und Veröffentlichung von Datenbeständen werden diese Kriterien zu Rate gezogen und in der PubData Collection-Policy Forschungsdaten verankert. Das bedeutet, dass übermittelte Daten im Rahmen einer technischen Eingangsprüfung auf Eignung der Dateiformate für die Aufbewahrung, Sicherung und bestenfalls Veröffentlichung geprüft werden.

Das MIZ behält sich hierbei vor, Datenbestände mit für genannte Zwecke ungeeigneten Dateiformaten vorerst abzulehnen. Eine solche Entscheidung wird allerdings stets in Absprache mit der datengebenden Partei und nach Erüierung von Alternativlösungen getroffen. Denken Sie aber dennoch rechtzeitig bei der Wahl Ihrer Software und der Dateiformate an die daraus resultierenden Konsequenzen bezüglich Aufbewahrung und ggf. Veröffentlichung im Sinne guter wissenschaftlicher Praxis (weitere Informationen finden Sie [hier](#)).

Die im Dokument referenzierten Leuphana-Richtlinien und Policies befinden sich im Downloadbereich auf der [Leitlinienseite von PubData](#).

Im Folgenden eine kurze Übersicht samt Empfehlungen zu Dateiformaten nach Formattyp:

Formattyp	Empfohlene Formate	Akzeptierte Formate	Nicht empfehlenswert
Text	ANSI X3.4/ECMA-6/US-ASCII (7-bit); EPUB (unverschlüsselt) (*.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 mit BOM; SGML; XML und XML-basierte Markup-Formate; 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 mit eingebetteten Schriftarten oder Verschlüsselung; PDF/A-3; DVI (*.dvi); Microsoft Word (*.doc); WordPerfect (*.wpd); Google Docs
Bild	JPEG; PNG; SVG; TIFF	Bitmap oder BMP (*.bmp); Computer Graphic Metafile (*.cgm); Digital Negative DNG	Encapsulated Postscript (*.eps, *.epsf, *.ps); FlashPix (*.fpx); GIF; JPEG 2000 Part 2 (*.jpf, *.jpx); Macromedia

		(*.dng); GIF (*.gif); Google WebP (*.webp); JPEG 2000; JPEG/JFIF (*.jpg); JPEG2000 (lossy) (*.jp2); Drawing Interchange Format (*.dxf)	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 (komprimiert) (*.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)

Präsentation	Abhängig vom Dateiinhalt	PDF-A; Microsoft Powerpoint OOXML (.pptx); Open Office (.sxi and .odp)	Microsoft Powerpoint (.pptx)
Strukturierte Daten	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); Komprimiertes Matlab
E-Mail	MBOX; Internet Message Format (EML)	Email Account XML Schema	Microsoft Outlook Item (*.msg); Personal Storage Table (*.pst); Outlook für Mac (*.olm)
Webarchivformat	Mozilla Archive Format (MAFF) (*.mar); Web Archive File (WARC) (*.warc)	-	-
Archiv-Dateiformat	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)
Sonstige	XD3 (unbekannte oder unbestimmte Dateierweiterung)	VRML; U3D; Computer Pgm Source Code (*.c, *.c++, *.java, *.js, *.jsp, *.php, *.pl, etc)	Compiled oder Executable Files (EXE, *.class, COM, DLL, BIN, DRV, OVL, SYS, PIF)