

hbz LZV-Dateiformatliste

Version 02, Stand 10.1.2020

Die Langzeitstabilität von Datei- bzw. Datenformaten ist von mehreren Faktoren abhängig. Sie dienen als Grundlage der vom hbz vorgenommenen Einschätzungen.

Die vorliegende Dateiformatliste stellt den "State of the Art" dar und wird stetig weiterentwickelt. Die Liste ist als Modell zu verstehen. Grundsätzlich ist im jeweiligen Einzelfall zu entscheiden, ob die Durchführung einer Migration sinnvoll und entsprechend eines vorab definierten Qualitätsanspruchs durchführbar ist.

Die Liste enthält derzeit gut 50 Formate. Die Bewertungen beruhen auf den Einschätzungen der KOST (KaD), der Library of Congress (LoC) und IANUS sowie Veröffentlichungen der ETH Zürich.

Bisher nicht aufgeführte Formate werden bei Bedarf und soweit möglich analysiert und der Liste hinzugefügt.

Die Grade der Unterstützung sind in folgenden Kategorien zusammengefasst:

- Dunkelgrün**: Dateiformat wird derzeit als langzeitstabiles Format eingeschätzt.
- Hellgrün**: Dateiformat gilt derzeit als nicht langzeitstabil. Mindestens ein Migrationspfad ist bekannt und in der Rosetta-Infrastruktur des hbz durchführbar (Plugin gestützt).
- Gelb**: Dateiformat gilt derzeit als nicht langzeitstabil. In der Rosetta-Infrastruktur des hbz existiert (derzeit) kein Migrationstool. Mindestens ein Migrationspfad ist bekannt.
- Orange**: Dateiformat gilt derzeit als nicht langzeitstabil. Es ist aktuell kein Migrationspfad bekannt.
- Rot**: Bestimmte Eigenschaften dieser Dateiformate widersprechen Anforderungen an langzeitstabile Formate und/oder verhindern eine Formatmigration (z.B. Verschlüsselung, dynamische Inhalte/Funktionen).

Formatbezeichnung	Dateinamen- Erweiterung / Grad der Unterstützung	PUID - PRONOM Unique Identifier	Informationsquelle	Bemerkungen Migration Tool (→ Zielformat)
3D-Anwendung				
COLLADA Digital Asset Exchange	*.dae	fmt/1209	IANUS	"Collada"
Wavefront OBJ	*.obj	fmt/1210	IANUS	Wavefront OBJ (referenzierte Materialdatei: *.mtl)
Polygon File Format	*.ply	fmt/831	IANUS	
Extensible 3D	*.x3d	fmt/579, 580, 581, 582	IANUS	
Audio				
MP4	*.m4a	fmt/596	ETH	MP4 (Apple Lossless Audio Codec) Zahlreiche Open Source Konverter (→ .wav) Sofern keine unkomprimierte Repräsentation der Information (z.B. als *.wav) vorliegt, sollte in Erwägung gezogen werden das *.mp3 unverändert zu archivieren. MP3toWaveMigrationTool (→ .wav)
MPEG 1/2 Audio Layer 3	*.mp3	fmt/134	ETH, KOST	

Formatbezeichnung	Dateinamen- Erweiterung / Grad der Unterstützung	PUID - PRONOM Unique Identifier	Informationsquelle	Bemerkungen Migration Tool (→ Zielformat)
Ogg Vorbis Codec Compressed Multimedia File	*.ogg	fmt/203	hbz	Quelloffenes Container-Format (inkl. Zugriff auf RAW-Daten) Zahlreiche Open Source Konverter (→ .wav)
Waveform Audio	*.wav	fmt/141	KOST, ETH	Linear Pulse Code Modulation (LPCM)

Bild/Rastergrafik

Windows Bitmap	*.bmp	fmt/114-119	ETH	
Digital Negative Format	*.dng	fmt/152, 436-438, 730	ETH	Rohdatenformat (inkl. Zugriff auf RAW-Daten) Zahlreiche Open Source Konverter (→ .tif)
Graphics Interchange Format	*.gif	fmt/4	ETH	GIF89a ImageMagick (→ .tiff)
JPEG 2000 part 1	*.jp2	x-fmt/392	KOST	JPEG 2000, verlustfreie Komprimierung
JPEG	*.jpg	fmt/41-44, 645; x-fmt/390, 391, 398	KOST, ETH	Sofern keine unkomprimierte Repräsentation der Information (z.B. als *.tif) vorliegt, sollte das *.jpg unverändert gespeichert werden. ImageMagick, VIPS (→ .tiff)
Open Microscopy Environment - Tagged Image File Format	*.ome.tiff	n.a.	hbz	https://docs.openmicroscopy.org/bio-formats/5.9.1/
Portable Network Graphics	*.png	fmt/11-13	ETH	ImageMagick; VIPS (→ .tiff)
Tagged Image File Format	*.tif	353; x-fmt/387, 388, 399	ETH, KOST	

GIS (Geoinformationssystem)

Geography Markup Language	*.gml	x-fmt/227, fmt/1047	hbz
---------------------------	-------	---------------------	-----

Officeformate

Microsoft Word	*.doc	x-fmt/273-276; fmt/37-40	hbz	Binäres Microsoft-Office-Format, tw. heute schon obsolet LibreOffice (→ .pdf)
Microsoft Word Macros	*.docm	fmt/523	hbz	Macro enabled Microsoft Word Document OOXML
Microsoft Word XML	*.docx	fmt/412	KOST	XML-basiertes Office Open Format; International genormter Standard (ECMA-376, Norm ISO/IEC 29500), nicht quelloffen LibreOffice (→ .pdf)
Microsoft Word XML (Encrypted); Microsoft PowerPoint XML (Encrypted); Microsoft Excel XML (Encrypted)	*.docx *.pptx *.xlsx	fmt/494	hbz	Microsoft Office Encrypted Document

Formatbezeichnung	Dateinamen- Erweiterung / Grad der Unterstützung	PUID - PRONOM Unique Identifier	Informationsquelle	Bemerkungen Migration Tool (→ Zielformat)
OpenDocument- Zeichnung	*.odg	fmt/139, 296, 297	ETH	XML-basiertes OpenDocument Format; International genormter, quelloffener Standard (Norm ISO/IEC 26300) LibreOffice (→ .pdf)
OpenDocument- Präsentation	*.odp	fmt/138, 292, 293	ETH, KOST	XML-basiertes OpenDocument Format; International genormter, quelloffener Standard (Norm ISO/IEC 26300) LibreOffice (→ .pdf)
OpenDocument- Tabellenvorlage	*.ods	fmt/137, 294, 295	ETH	XML-basiertes OpenDocument Format; International genormter, quelloffener Standard (Norm ISO/IEC 26300) LibreOffice (→ .pdf)
OpenDocument-Text	*.odt	fmt/136, 290, 291	ETH, KOST	XML-basiertes OpenDocument Format; International genormter, quelloffener Standard (Norm ISO/IEC 26300) LibreOffice (→ .pdf)
Microsoft PowerPoint	*.ppt	x-fmt/88; fmt/125, 126, 179, 181	ETH	Binäres Microsoft-Office-Format, tw. heute schon obsolet LibreOffice (→ .pdf)
Microsoft PowerPoint XML	*.pptx	fmt/215	ETH, KOST	XML-basiertes Office Open Format; International genormter Standard (ECMA-376, Norm ISO/IEC 29500), nicht quelloffen LibreOffice (→ .pdf)
Microsoft Excel	*.xls	fmt/55- 57, 59, 61, 62, 175-177	KOST	Binäres Microsoft-Office-Format, tw. heute schon obsolet LibreOffice (→ .pdf)
Microsoft Excel XML	*.xlsx	fmt/214	ETH	XML-basiertes Office Open Format; International genormter Standard (ECMA-376, Norm ISO/IEC 29500), nicht quelloffen LibreOffice (→ .pdf)

PDF (Portable Document Format)

Acrobat PDF 1.4 - Portable Document Format	*.pdf (PDF 1.4)	fmt/18	IANUS, KOST
Acrobat PDF/A - Portable Document Format 1a	*.pdf (PDF/A- 1a)	fmt/95	IANUS, KOST
Acrobat PDF/A - Portable Document Format 1b	*.pdf (PDF/A- 1b)	fmt/354	IANUS, KOST
Acrobat PDF/A - Portable Document Format 2a	*.pdf (PDF/A- 2a)	fmt/476	IANUS, KOST
Acrobat PDF/A - Portable Document Format 2b	*.pdf (PDF/A- 2b)	fmt/477	IANUS, KOST

Formatbezeichnung	Dateinamen- Erweiterung / Grad der Unterstützung	PUID - PRONOM Unique Identifier	Informationsquelle	Bemerkungen Migration Tool (→ Zielformat)
Acrobat PDF/A - Portable Document Format 2u	*.pdf (PDF/A- 2u)	fmt/478	IANUS, KOST	
Acrobat PDF/A - Portable Document Format 3	*.pdf (PDF/A- 3)	fmt/479- 481		

Anwendungsunabhängiges textbasiertes Format

Character-Separated Values	*.csv	X-fmt/18	ETH	
Hypertext Markup Language	*.html	fmt/96- 100, 471;	ETH, IANUS, KOST	Hypertext, Auszeichnungssprache
Markdown	*.md	fmt/1149	hbz	Auszeichnungssprache
Standard Generalized Markup Language	*.sgml	fmt/195	IANUS	Auszeichnungssprache
Text file	*.txt	x-fmt/111	ETH, IANUS	
Extensible Hypertext Markup Language	*.xhtml	fmt/102, 103	IANUS, KOST	Hypertext, Auszeichnungssprache
Extensible Markup Language	*.xml	fmt/101	ETH, IANUS, LoC	Auszeichnungssprache

Vektorgrafik

Scalable Vector Graphics	*.svg	fmt/91, 92	ETH	zu beachten: SVG ohne JavaScript binding
-----------------------------	-------	------------	-----	---

Video

Motion JPEG 2000	*.mj2, *.mjp2	fmt/337	ETH, KOST	
Matroska Multimedia Container (FF video codec 1)	*.mkv (FFV1)	fmt/569	ETH, KOST	Videocodec, das für die Archivierung entwickelt wurde
Moving Pictures Experts Group	*.mpg, *.mpeg (MPEG-1)	x-fmt/385; fmt/649	ETH	MPEG-1 Handbrake, ffmpeg (→ Container- Formate .avi oder .mov)
Moving Pictures Experts Group	*.mpg, *.mpeg (MPEG-2)	x-fmt/386; fmt/640	ETH	MPEG-2, H.262 Handbrake, ffmpeg (→ Container- Formate .avi oder .mov)
Moving Pictures Experts Group	*.mpg, *.mpeg (MPEG-4)	fmt/199	ETH, KOST	MPEG-4 Part 10, AVC, H.264 Handbrake, ffmpeg (→ Container- Formate .avi, .mov oder .mkv)

Webarchivierung

Internet Archive	*.arc	x-fmt/219, fmt/410	hbz	
Web ARChive	*.warc	fmt/289	KOST	