Dateiformate

Digitale Objekte werden in drei Versionen gespeichert: als Archivkopie für die Langzeitsicherung, als Nutzungskopie für die wissenschaftliche Arbeit und als Ansichtskopie für die Präsentation im Web. Je nach Medientyp gelten unterschiedliche Anforderungen an Auflösung und Dateiformat. Offene, nicht-proprietäre und verlustfreie Formate sichern langfristige Lesbarkeit und erleichtern die Nachnutzung digitaler Kulturgüter.

- Einstieg: Dateiformate
- Bild
- Text
- Ton
- Video
- 3D
- Weitere Ressourcen zu Datenformaten

Einstieg: Dateiformate

> Digitalisate werden in mehreren Versionen gespeichert, um verschiedene zukünftige Abfragebedürfnisse bedienen zu können: eine für die Archivierung, eine für die Nutzung und eine für die Ansicht des Digitalisats in der Online-Sammlung. Dementsprechend gelten für die **Archivkopie,** die **Nutzungskopie** und die **Ansichtskopie** je nach Medientyp unterschiedliche Anforderungen.

In der Welt der digitalen Archive haben sich Begriffe wie „Archivkopie“, „Nutzungskopie“ und „Ansichtskopie“ durchgesetzt und beziehen sich auf verschiedene digitale Versionen eines Digitalisats.

Archivkopie, Nutzungskopie und Ansichtskopie

- Als **Archivkopie** wird ein Digitalisat bezeichnet, das ein möglichst präzises Äquivalent zum Sammlungsobjekt darstellt. Die Archivkopie und die zugrunde liegenden **Rohdaten** werden bei der Institution gespeichert. Die Kopie wird in einem möglichst langlebigen und damit für die Langzeitsicherung geeigneten Dateiformat gespeichert. Eine Archivkopie soll möglichst unbearbeitet bleiben.
- Eine **Nutzungskopie** ist eine oftmals verkleinerte oder komprimierte Version der Archivkopie, die für den regelmäßigen Gebrauch erstellt wird und mit der wissenschaftlich gearbeitet werden kann. Sie wird Benutzern zur Verfügung gestellt, die das Digitalisat nutzen und bearbeiten möchten. Im Unterschied zur Archivkopie kann sie in leichter zugänglichen Formaten oder Auflösungen vorliegen, um den praktischen Gebrauch zu unterstützen.
- Die **Ansichtskopie** (oder Auslieferungskopie) ist eine niedriger aufgelöste Version des Digitalisats, um einen schnellen Transfer und eine schnelle Anzeige zu ermöglichen, z.B. in einem Webbrowser. Diese Kopie dient hauptsächlich der Präsentation, beispielsweise in Online-Datenbanken oder virtuellen Ausstellungen, und wird für die Veröffentlichung auf Webplattformen bereitgestellt.

Nach dem Erstellen des Digitalisats wird dieses als Archivkopie gespeichert und anschließend nicht mehr verändert. Für das Erstellen der Nutzungskopie (und folglich auch der Ansichtskopie) wird immer nur eine Kopie der Archivkopie verwendet.

Bildanzeige im Kulturpool

- Der Kulturpool generiert für 2D-Digitalisate **Vorschaubilder** auf Basis der übermittelten Ansichtskopie, um schnelle und einheitliche Ergebnisse in der Suche anzuzeigen.
  Die generierten Vorschaubilder sind:
  - Maximale Länge der längeren Bildkante: 400, 800, 1200 Pixel
  - Bildformat: `.webp`

- In der **Vollansicht** werden die **Ansichtskopien der Partnerinstitution** angezeigt. Steht kein IIIF-Manifest zur Verfügung, wird ein statisches Manifest erzeugt, das die Bild-URL(s) der Institution enthält. Damit die Bilder in der Vollansicht angezeigt werden können, müssen die Mediendaten auf den Quell-Servern stets erreichbar sein.
- Für andere Medientypen (Audio, Video, 3D, Text) können Bilder über das Metadatenfeld «edm:object» geliefert werden.

Anforderungen an Dateiformate

Digitalisate sollen das Original so getreu wie technisch möglich wiedergeben und über freie Software & offene Formate darstellbar und editierbar sein. So können (kosten-)effiziente Lösungen für Digitalisierungsvorhaben ermöglicht werden. Die langfristige Bewahrung der vollständigen und ursprünglichen Funktionalität bestimmter Dateiformate ist oftmals nicht praktikabel oder möglich. Die Wahrscheinlichkeit einer erfolgreichen langfristigen Bewahrung von digitalen Inhalten ist höher, wenn die Dateiformate die folgenden Merkmale aufweisen:

- Plattformunabhängigkeit
- nicht-proprietär (vollständig veröffentlicht, einem offenen, dokumentierten Standard folgend und lizenzgebührenfrei verfügbar sowie vollständig und unabhängig von mehreren Softwareanbietern ohne Einschränkungen des geistigen Eigentums an der erforderlichen Technologie)
- keine „verlustbehaftete“ oder proprietäre Komprimierung
- keine eingebetteten Dateien, Programme oder Skripte
- keine vollständige oder teilweise Verschlüsselung
- in der Forschungsgemeinschaft gebräuchlich

Im folgenden Abschnitt finden Sie Dateiformate für die fünf Bereiche **Bild**, **Text**, **Ton**, **Video** und **3D**, die sich in der Arbeit mit Digitalisaten bereits bewährt haben.

Bild

> Die Digitalisate in **Bild**, Text, Ton, Video und 3D brauchen unterschiedliche Dateiformate für je Archivkopie, Nutzungskopie und Ansichtskopie.

Bildformate für Rastergrafiken

Für die Archivkopie werden die offenen Formate TIFF v6 oder JPEG2000 Part-1 empfohlen. Falls die Originaldaten in einem Rohformat vorliegen, sollte eine Kopie im unkomprimierten TIFF 6.0-Format exportiert werden, um die langfristige Lesbarkeit in gängiger Software zu gewährleisten. Wenn die verfügbaren Digitalisierungsgeräte eine höhere Auflösung zulassen, soll die höchste Auflösung gewählt werden.

Empfohlene Auflösung

3:2 Bildfeld: ? 20 Megapixel (? 5472 × 3648 px)
1:1 Bildfeld: ? 13 Megapixel (? 3712 × 3712 px)
16:9 Bildfeld: ? 17 Megapixel (? 5568 × 3128 px)
mit 24-Bit True-color (= 3 × 8 Bit, entspricht 224 ? 16,78 Millionen Farben)

Mindestauflösung

Die höchstmögliche Auflösung wird empfohlen, ansonsten **? 4 Megapixel (? 2448 × 1632, 3:2)** und 24-Bit True-Color (3 × 8 Bit, entspricht 224 ? 16,78 Millionen Farben), in begründeten Fällen ? 1 Megapixel.

Pyramid TIFF und JPEG2000 Part-1 sind für eine IIIF-konforme Ansichtskopie notwendig.

Bilddigitalisate sollen IIIF-konform sein, um hochauflösende und zoombare Digitalisate mit einheitlichen Metadaten zu erhalten.

Rastergrafiken			
Zweck	**Erwünschtes Dateiformat**	**Dateiendung**	**MIME-TYPE**

Archivkopie	• TIFF v6
	• JPEG2000 Part-1
		.tiff, .tif
		.jp2, .j2k, .jpf, .jpg2, .jpx, .jpm, .mj2, .mjp2
			`image/tiff`, `image/tiff-fx`
			`image/jp2`, `image/jpx`,
			`image/jpm`, `video/mj2`

Nutzungskopie	• TIFF v6 • JPEG2000 Part-1	*wie oben*	*wie oben*

Ansichtskopie	• JPEG
	• PNG
	• WebP
	• AVIF
		.jpeg, .jpg
		.png
		.webp
		.avif
			`image/jpeg`
			`image/png`
			`image/webp`
			`image/avif`

Formate für gescannten Text

Ein gescannter Text ist ein Foto einer gedruckten Seite, das entweder mit einer Digitalkamera oder einem Scanner erstellt wurde.

Für die Archivkopie werden die offenen Formate TIFF v6, JPEG2000 Part-1, PDF/A-1 und PDF/A-2 empfohlen. Als Basis für Texterkennung (optical character recognition, OCR) werden mindestens 200 bis 300 dpi benötigt. Für zweitönige (bitonal, 1-Bit schwarz-weiß) Aufnahmen werden 300–600 dpi empfohlen, für Graustufenbilder (8-Bit) werden 400 dpi und für Farbaufnahmen (24-bit RGB) 400 dpi empfohlen.

Gescannter Text			
Zweck	**Erwünscht**	**Dateiendung**	**MIME-Type**

Archivkopie	• TIFF v6
	• JPEG2000 Part-1
	• PDF/A-1
	• PNG
		.tiff, .tif
		.jp2, .j2k, .jpf, .jpg2, .jpx, .jpm, .mj2, .mjp2
		.pdf
		.png
			`image/tiff`, `image/tiff-fx`
			`image/jp2`, `image/jpx`, `image/jpm`, `video/mj2`
			`application/pdf`
			`image/png`

Nutzungskopie	• TIFF v6 • JPEG2000 Part-1 • PDF/A-1 • PNG	*wie oben*	*wie oben*

Ansichtskopie	• JPEG
	• PNG
	• WebP
	• AVIF
	• PDF/A-1
		.jpeg, .jpg
		.png
		.webp
		.avif
		.pdf
			`image/jpeg`
			`image/png`
			`image/webp`
			`image/avif`
			`application/pdf`

Bildformate für Vektorgrafiken

Für Vektorgrafiken wird das vom W3C entwickelte Format SVG (Scalable Vector Graphics) unkomprimiert empfohlen.

Vektorgrafiken

Zweck	Erwünschtes Dateiformat	Dateiendung	MIME-Type
Archivkopie	• SVG	.svg	`image/svg+xml`
Nutzungskopie	• SVG	*wie oben*	*wie oben*
Ansichtskopie	• SVG	*wie oben*	wie oben

Linktipps

- SELFHTML: Übersicht über MIME-Typen für Bilder
- IANA: Liste aller MIME-Typen für Bilder
- Angela Kailus: Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten. 10. Empfehlungen für Dateiformate
- ArcGIS: Imagery formats and performance
- Europeana Knowledge Base: ContentTier 2-4: Image type
- International Image Interoperability Framework (IIIF)
- plroit: Skyreach – A JPEG2000 Transcoding Library

Text

> Die Digitalisate in Bild, **Text**, Ton, Video und 3D brauchen unterschiedliche Dateiformate für je Archivkopie, Nutzungskopie und Ansichtskopie.

Dateiformate für Text

Für den Archivmaster werden die offenen Formate Unicode Text, PDF/A-1, PDF/A-2, ODF, HTML empfohlen. Wenn eine Auswahl für die Kodierung möglich ist, wählen Sie bitte **UTF-8 ohne Byte Order Mark (BOM).**

Textformate

Zweck	Erwünschtes Dateiformat
Archivkopie	• Unicode Text • PDF/A-1 • PDF/A-2 • ODF • HTML
Nutzungskopie	• Unicode Text • PDF/A-1 • PDF/A-2 • ODF • HTML
Ansichtskopie	• PDF • Unicode Text

Linktipps

- SELFHTML: Übersicht über MIME-Typen für Text
- IANA: Liste aller MIME-Typen für Text
- Angela Kailus: Handreichung für ein FAIRES Management kulturwissenschaftlicher Forschungsdaten. 10. Empfehlungen für Dateiformate
- asprise: Accurate Real-time Receipt OCR: Blog » EU European Union receipt OCR API

Ton

> Die Digitalisate in Bild, Text, **Ton**, Video und 3D brauchen unterschiedliche Dateiformate für je Archivkopie, Nutzungskopie und Ansichtskopie.

Für die Archivkopie werden die offenen verlustfreien Formate LPCM (Linear Pulse Code Modulation) in einem WAV-Container oder FLAC empfohlen. Eine Quantisierungsauflösung & Abtastrate (s*ample rate*) von mindestens 24 Bit/48 kHz.

Bei neu digitalisierten analogen Quellen werden 96 kHz Abtastrate empfohlen.

Audioformate

Zweck	**Erwünschtes Dateiformat**	**Quantisierungsauflösur *ample rate*)**
Archivkopie	• LPCM (in WAV-Container) • FLAC	• 24 Bit/48 kHz, 24 B
Nutzungskopie	• LPCM (in WAV) • FLAC • MP3	• 16 Bit/48 kHz • mind. 192 kbit/s fü
Ansichtskopie	• MP3	• 128 kbit/s

WAV Dateien mit erweiterten Metadaten werden auch „BWF“ (Broadcast Wave Format) genannt. Im Zweifelsfall sind BWFs reinen WAVs vorzuziehen.

Linktipps

- Kulturpool Forum: Audio- und Videoformate (Anmeldung nötig)
- SELFHTML: Übersicht über MIME-Typen für Ton
- IANA: Liste aller MIME-Typen für Ton
- Angela Kailus: Handreichung für ein FAIRES Management kulturwissenschaftlicher Forschungsdaten. 10. Empfehlungen für Dateiformate
- Reto Kromer: Digitalisierung von Ton

Video

> Die Digitalisate in Bild, Text, Ton, **Video** und 3D brauchen unterschiedliche Dateiformate für je Archivkopie, Nutzungskopie und Ansichtskopie.

Für die Archivkopie wird die folgende Kombination von offenen verlustfreien AV-Formaten empfohlen: Container: Matroska (.mkv), Video-Codec: FFV1. Als Audioformat unkomprimiertes LPCM (Linear Pulse Code Modulation) mit 48 kHz, 16 oder 24 Bit. (LPCM wird umgangssprachlich oft auch „WAV“ genannt, da dies meistens das Format in WAV-Audiodateien ist.)

Dateiformat und Codec sind Teil der freien Codec-Sammlung „libavcodec“ des Projektes FFmpeg. Je nach Quellmedium können unterschiedliche Farbmodelle (z. B. RGB, YUV) und Subsampling (z. B. 4:2:2, 4:2:0, 4:4:4) vorkommen.

Videoformate

Zweck	Container	Video-Codec	Audio-Codec
Archivkopie	Matroska (.mkv)	FFV1	PCM (linear) 48 kHz/ 16 oder 24 Bit
Nutzungskopie	Matroska (.mkv)	MPEG-2, AVC/H.264[*1]	AAC
Ansichtskopie	WebM, MP4	VP9, AVC/H.264[*1], AV1	Opus

[*1]Lizenzkosten bei nicht freien und öffentlichen Inhalten: Via Licensing Alliance LLC via-la.com

Linktipps

- Kulturpool Forum: Audio- und Videoformate (Anmeldung nötig)
- SELFHTML: Übersicht über MIME-Typen für Video
- IANA: Liste aller MIME-Typen für Video
- Angela Kailus: Handreichung für ein FAIRES Management kulturwissenschaftlicher Forschungsdaten. 10. Empfehlungen für Dateiformate
- Delft Digitalisierung ethnologischer Filmbestand: Zielformate: Archivmaster und Derivative Copy
- Deutsche Nationalbibliothek (DNB): Leitfaden für die digitale Langzeitarchivierung audiovisueller Medien
- Österreichische Mediathek: Digitalisierung: FFV1 an der Österreichischen Mediathek
- Sustainability of Digital Formats: Planning for Library of Congress Collections: FF Video Codec 1, Version 0, 1 and 3

3D

> Die Digitalisate in Bild, Text, Ton, Video und **3D** brauchen unterschiedliche Dateiformate für je Archivkopie, Nutzungskopie und Ansichtskopie.

Hochwertige und hochauflösende Daten können in den Formaten *.OBJ oder *.DAE gespeichert werden.

3D-Dateiformate

Zweck	Erwünschtes Dateiformat
Archivkopie	• glTF/GLB • X3D • STL • OBJ • DAE • PLY • WRL • DICOM • IFC
Nutzungskopie	• glTF/GLB • X3D • STL • OBJ • DAE • PLY • WRL • DICOM • IFC
Ansichtskopie	• glTF/GLB *(Viewer vorhanden)* • X3D *(derzeit noch kein Viewer)*

Gängige Formate für Anwendungen der **Virtual Reality** und **Augmented Reality** sind *.USDZ und *.glTF.

Für den **3D-Druck** ist *.STL und für die Webvisualisierung sind *.X3D und *.GLB weitverbreitet.

Die Liste der 3D-Formate ist nicht vollständig, da 3D-Modelle für unterschiedliche Einsatzzwecke erstellt werden und mit diesen entsprechende optimierte Formate einhergehen. Ein wichtiges Kriterium für die angeführten 3D-Formate ist der Zugang zu 3D-Inhalten in Industriestandardformaten, die von den Nutzern angezeigt werden können, ohne dass sie zusätzliche Software installieren müssen. Das Einbetten von 3D-Inhalten in PDF-Dokumenten wird nicht empfohlen.

Literaturhinweise:

- Kulturpool Forum: 3D-Digitalisierung (Anmeldung nötig)
- IANA: Liste aller MIME-Typen für 3D (*model*)
- Angela Kailus: Handreichung für ein FAIRES Management kulturwissenschaftlicher Forschungsdaten. 10. Empfehlungen für Dateiformate
- CARARE: 3D und Virtuality Training Hub: 3D and Virtual Reality
- Europeana Pro: 3D Content in Europeana
- Europeana Pro: Introducing the Europeana Common Culture 3D pilot (11.12.2020, aktualisiert 6.11.2023)
- European Commission (2022): Basic principles and tips for 3D digitisation of cultural heritage
- Fileformat.com: Erfahren Sie mehr über 3D-Dateiformate und APIs, die 3D-Dateien öffnen und erstellen können (Original Englisch)
- Web3D Consortium: Open Standards for Real-Time 3D Communication

Weitere Ressourcen zu Datenformaten

Weitere allgemeine Ressourcen zu Dateiformaten:

- ARCHE (A Resource Centre for Humanities Related Research in Austria): Filenames, Formats & Metadata
- Brandon University: Best practices for file formats
  National Archives: Appendix A: Tables of File Formats
- CARARE Metadata Schema: The Schema
- Cornell University Library: Recommended File Formats
- CLARIN: Introduction: standards in CLARIN
- Digitale Editionen an der ÖNB: Technische Dokumentation
- digiS: Parameter bei der 2D Digitalisierung – eine Orientierungshilfe
- Europeana Pro: Digital Transformation Task Force
- EFG Data Provider Handbook (2021)
- Europeana: Media Policy
- Europeana Knowledge Base: Example records - content & metadata tiers
- Federal Agencies Digital Guidelines Initiative: Guidelines: Technical Guidelines for Digitizing Cultural Heritage Materials
- Library of Congress: Recommended Formats Statement
- Missouri S&T: Digitization Best Practices
- Sustainability of Digital Formats: Planning for Library of Congress Collections: Format Descriptions
- Sustainability of Digital Formats: Planning for Library of Congress Collections: Formats, Evaluation Factors, and Relationships
- UNESCO: Managing low-cost digitization projects in least developed countries and small island developing states: a manual
- W3C: Best Practices/Publishing Statistical Data In Linked Data Format