

Kanton St.Gallen
Departement des Innern

Amt für Kultur
Staatsarchiv St.Gallen



Archivische Dateiformate

Version vom: Februar 2025
Ersetzt Version vom: September 2019

© Staatsarchiv St.Gallen

Inhaltsverzeichnis

Einleitung	3
Formatkategorie 'Text'	3
Formatkategorie 'Bild'	3
Formatkategorie 'Audio'	3
Formatkategorie 'Video'	4
Formatkategorie 'Tabellenkalkulation'	4
Formatkategorie 'Datenbanken'	4
Formatkategorie 'Hypertext' (Website)	4
Formatkategorie 'GIS'	4
Formatkategorie 'CAD/CAM'	4
Formatkategorie 'Strukturierte Daten'	5
Konvertierung	5

Einleitung

Nur eine kleine Auswahl der vielen verschiedenen Dateiformate ist für die Archivierung tauglich. Hinsichtlich der Übernahme von digitalen Daten zur Langzeitarchivierung macht das Staatsarchiv hiermit Vorgaben, die auf dem *Katalog archivischer Dateiformate* (KaD) der KOST¹ basieren, und bietet zusätzlich Empfehlungen bezüglich Konvertierung.² Konvertierungen sind nötig, falls Dateien nicht von Beginn an in archivischen Dateiformaten erstellt werden.

Dateiformate lassen sich in zehn Formatkategorien einteilen: Text, Bild, Audio, Video, Tabellenkalkulation, Datenbanken, Hypertext, GIS, CAD/CAM und Strukturierte Daten. Für jede dieser Kategorien gibt es bestimmte geeignete Dateiformate, welche aufgrund verschiedener Kriterien, nach momentanem Wissensstand, für die Archivierung tauglich sind. Die Kriterien sind Offenheit, Lizenzfreiheit, Verbreitung, Funktionalität, Implementierung, Speicherdichte und Verifizierbarkeit. Zusätzlich werden die bisherige Verwendung und die Perspektive des jeweiligen Formates in die Bewertung miteinbezogen.

Nicht für alle Formatkategorien gibt es eindeutige Vorgaben. Unter Umständen müssen der konkrete Anwendungsfall und die Aufbewahrungsdauer der Daten in den Formatentscheid miteinbezogen werden.

Formatkategorie 'Text'

Kurzbeschreibung

Unter Textdaten werden Daten zur Repräsentation von wenig strukturiertem Text verstanden, wie sie unter anderem von Textverarbeitungssoftware oder Texteditoren erzeugt werden (z.B. Office-Dateien, E-Mail etc.).

Vorgabe

Erwünscht: PDF/A-2u

Akzeptiert: PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2b

Formatkategorie 'Bild'

Kurzbeschreibung

Bei Bilddaten ist zwischen Raster- und Vektorgrafiken zu unterscheiden. Ein einziges ideales Archivformat für Bilddaten gibt es nicht.

Vorgabe

Rastergrafiken: TIFF, JPEG

Vektorgrafiken: Rücksprache mit Staatsarchiv nötig

Formatkategorie 'Audio'

Kurzbeschreibung

Audiodaten enthalten Informationen, die durch das menschliche Ohr wahrgenommen werden sollen.

Vorgabe

Erwünscht: WAV

Akzeptiert: MP3

¹ Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen. Version 7.0 vom Juni 2024. URL: https://kost-ceco.ch/cms/kad_rating_de.html [Stand: 05. Februar 2025].

² Umwandlung oder Erstellung von archivischen Dateiformaten.

Formatkategorie 'Video'

Kurzbeschreibung

Bei Videos ist je nach Entstehungszusammenhang, Archivauftrag und Bewertungsentscheid jeweils ein bestimmtes Format das in diesem Fall am besten geeignete.

Vorgabe

Erwünscht: MKV (FFV1)

Akzeptiert: MP4 (H.264 oder H.265)

Formatkategorie 'Tabellenkalkulation'

Kurzbeschreibung

Tabellenkalkulationssoftware stellt numerische und alphanumerische Daten dar und erlaubt insbesondere deren Verarbeitung mittels Funktionen.

Vorgabe

Erwünscht: ODF oder OOXML (z.B. XLSX)

Akzeptiert: CSV

Formatkategorie 'Datenbanken'

Kurzbeschreibung

Bei der Archivierung von Datenbanken, heute in der Regel von relationalen Datenbanken, ist nicht nur das Zielformat von Bedeutung, sondern auch die Methode der Archivierung.

Vorgabe

In sich abgeschlossene Datenbanken:

Erwünscht: SIARD

Akzeptiert: CSV (mit Metadaten)

Datenbanken als Teil eines umfassenderen Systems: Rücksprache mit Staatsarchiv nötig

Formatkategorie 'Hypertext' (Website)

Kurzbeschreibung

Hypertext bezeichnet eine Menge von Texten, die über Verknüpfungen (Links) miteinander verbunden sind. Die Gestaltung der Texte wie auch die Verknüpfungen erfolgen mit einer Auszeichnungssprache (markup language). Die bekannteste Hypertext-Auszeichnungssprache ist HTML.

Vorgabe

Erwünscht: WARC

Akzeptiert: PDF/A-2

Formatkategorie 'GIS'

Kurzbeschreibung

Geoinformationssysteme bzw. Geografische Informationssysteme (GIS-Systeme) sind Informationssysteme zur Erfassung, Bearbeitung, Organisation und Darstellung räumlicher Daten. Die für diese Aufgabe notwendigen Vektordaten und raumbezogenen Metadaten werden gemeinhin in Datenbanken verwaltet.

Vorgabe

Vektordaten: INTERLIS 1

Bild- und Grafik-Rasterdaten: Rasterdaten mit World-File (je nach zugrundeliegendem Bildformat z.B. JGW)

Produkte und Karten: Vektorgrafik in PDF

Formatkategorie 'CAD/CAM'

Kurzbeschreibung

CAD (Computer-Aided Design) bezeichnet das computerunterstützte Konstruieren von unterschiedlichen Produkten wie Autos, Flugzeuge, Bauwerke. Im engeren Sinne wird

darunter das Zeichnen von Konstruktions- und Architekturplänen mit Computerunterstützung verstanden. Das Produkt ist in der Regel eine 2D- oder 3D-Vektorgrafik mit Metadaten zu den einzelnen Vektorelementen

Vorgabe

CAD-Daten: DWG

Integrierte Produktion (CAx): IFC

Konstruktions- und Architekturpläne: Vektorgrafik in PDF

Formatkategorie 'Strukturierte Daten'

Kurzbeschreibung

Strukturierte Daten können sowohl in reinen, semi-strukturierten Textformaten, in Datenbanken oder in Tabellenkalkulationsformaten abgelegt werden.

Vorgabe

Erwünscht: XML, JSON

Akzeptiert: CSV

Konvertierung

Kurzbeschreibung

Für die Umwandlung von Dateien in archivische Dateiformate oder für die Erstellung solcher Daten können unterschiedliche Programme verwendet werden. Die Qualität der Konvertierungsergebnisse ist je nach Programm unterschiedlich.

Vorgabe

Rücksprache mit Staatsarchiv nötig