



Weisung betreffend Dateiformate für die Langzeitarchivierung

Vom 1. April 2024

1. Rechtsgrundlagen und Geltungsbereich

Die vorliegende Weisung wird gestützt auf § 5 Abs. 4 Gesetz über das Archivwesen (Archivgesetz, SG 153.600) erlassen. Sie ist verbindlich für die öffentlichen Organe gemäss § 3 Abs. 5 Archivgesetz.

2. Einleitung

Die Langzeitarchivierung digitaler Unterlagen (Records) basiert im Staatsarchiv auf dem Ansatz der Dateiformatmigration. Um diesen Ansatz in der Praxis umzusetzen, ist es erforderlich, die Vielfalt der Dateiformate, die in operativen Systemen verwendet werden, bei der Ablieferung an das Staatsarchiv auf eine kleine Menge von archivtauglichen Dateiformaten zu reduzieren. Diese Weisung definiert diejenigen Dateiformate, die das Staatsarchiv als archivtauglich beurteilt.

Basis dieser Vorgaben ist der Katalog der archivischen Dateiformate (KaD) der «Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST)». Dort finden sich auch weitere Informationen zu den einzelnen Dateiformaten.

Die Vorgaben beziehen sich auf von einem öffentlichen Organ ursprünglich digital generierte Records oder solche, die durch ersetzendes Scannen generiert wurden, sowie Records, die das öffentliche Organ im Zuge des Geschäftsgangs von Dritten in digitaler Form erhalten hat.

3. Formatkonvertierung

Die Formatkonvertierung muss kontrolliert und protokolliert erfolgen, damit die Authentizität und Integrität der Records erhalten bleibt. Da bei der Konvertierung von einem Dateiformat zu einem anderen wesentliche Eigenschaften der Records verloren gehen können, ist auf nicht notwendige Konvertierungen zu verzichten.

Die Konvertierung soll aus Gründen der Nachvollziehbarkeit nach Möglichkeit im Quellsystem erfolgen; möglichst rasch nach Ende der Erstellung und Bearbeitung der Records resp. Abschluss des Dossiers und spätestens bei der Generierung des SIP, dem Ablieferungspaket an das Staatsarchiv. Dies setzt allerdings voraus, dass das Quellsystem über die erforderlichen Konvertierungswerkzeuge verfügt. In allen anderen Fällen muss die Konvertierung vor der Generierung des SIP mit dem Staatsarchiv geregelt werden.

Die Einführung von Konvertierungswerkzeugen in Quellsystemen oder als alleinstehende Werkzeuge ist mit dem Staatsarchiv vor der Beschaffung und Implementierung seitens des öffentlichen Organs abzustimmen.

4. Archivtaugliche Dateiformate

Einleitung

Nachstehend werden die vom Staatsarchiv als archivtauglich definierte Dateiformate aufgelistet und knapp beschrieben. Weitergehende Hinweise sind im Katalog archivischer Dateiformate der KOST enthalten.

Wenn im Quellsystem des öffentlichen Organs Dateiformate enthalten sind, die in der nachstehenden Auflistung nicht enthalten sind, muss das öffentliche Organ vor einer Konvertierung in andere Dateiformate oder der Ablieferung zur Archivierung mit dem Staatsarchiv Kontakt aufnehmen.

Die Dateiformate werden in folgende Formatkategorien gegliedert:

- Text
- Unformatierter Text
- Tabellenkalkulation
- E-Mail
- Bild, Rasterdaten
- Bild, Vektordaten
- Audio
- Video
- Datenbanken
- Strukturierte Daten
- Webseiten (Hypertext)
- Geodaten
- CAD
- Statistik

Unterschieden wird pro Formatkategorie zwischen erwünschten und akzeptierten Formaten. Grundsätzlich sind nur die erwünschten Formate bei der Ablieferung zur Archivierung zu verwenden. In Ausnahmefällen dürfen auch Records in akzeptierten Formaten abgeliefert werden. Ob ein Ausnahmefall vorliegt, ist mit dem Staatsarchiv vor der Ablieferung seitens des öffentlichen Organs zu klären.

Text

Diese Formatkategorie umfasst Records mit Inhalten in Textform mit Formatierungen, die Tabellen, Bilder und grafische Darstellungen beinhalten können. Das klassische Verwaltungsschriftgut fällt überwiegend in diese Kategorie. Auch Formate für Präsentationen werden hierunter subsumiert. Office Anwendungen wie Microsoft Office mit Word und Powerpoint oder Libreoffice generieren entsprechende Records.

Erwünscht:

- PDF/A-2u

Akzeptiert:

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

Hinweise:

- Bei den PDF/A-1- und PDF/A-2-Ausprägungen gelten folgende Einschränkungen:
 - JBIG-Komprimierung darf nicht verwendet werden.
 - Rasterbilder sind als JPEG2000 einzubetten.
- Die akzeptierten PDF/A-1 und PDF/A-2-Formatausprägungen dürfen verwendet werden, sofern Daten in diesen Formaten bereits vorliegen oder PDF/A-2u nicht generiert werden kann.
- PDF/A-3 ist nicht archivtauglich und darf nicht verwendet werden.
- Dateien im PDF-Format sind nach PDF/A-2u zu konvertieren.

Unformatierter Text

Diese Formatkategorie umfasst Unterlagen mit Text, der nicht oder nur minimal formatiert ist. Dies können einfache Textdokumente wie Readme-Dateien sein.

Erwünscht:

- TXT

Akzeptiert:

[Keine weiteren Dateiformate definiert.]

Hinweise:

- TXT soll insbesondere verwendet werden, wenn Unterlagen bereits in diesem Format vorliegen.
- Bei TXT ist im Regelfall als Zeichencodierung UTF-8 einzusetzen.

Tabellenkalkulation

Die Formatkategorie umfasst die Darstellung numerischer und alphanumerischer Daten in Tabellenform und deren Verarbeitung mithilfe von Funktionen. Hierunter fallen Exceltabellen sowie weitere Tabellenwerke, etwa aus Buchhaltung und Controlling oder Statistiken. Die Tabellenstruktur kann um Grafiken ergänzt sein.

Erwünscht:

- XLSX (ab Microsoft Office 2013)
- ODS (ab ODF Version 1.2)

Akzeptiert:

- CSV

Hinweise:

- XLS: ist in XSLX zu überführen - mit Microsoft Office 2013 oder neuer.
- CSV soll nur verwendet werden, falls ausschliesslich Tabellenwerte erhalten bleiben sollen. Beim Einsatz von CSV sind folgende Parameter zu dokumentieren: Trennzeichen zwischen Spalten, Header, Quotation, Zeilenumbrüche, Zeichensatz.
- Bei CSV ist nach Möglichkeit als Zeichencodierung UTF-8 einzusetzen.

E-Mail

In diese Kategorie fallen Records, die von E-Mailanwendungen generiert werden. Sie können weitere Records als Beilagen (Attachments) beinhalten.

Erwünscht:

- PDF/A-2u

Akzeptiert:

- PDF/A-2a
- PDF/A-2b
- PDF/A-1a
- PDF/A-1b
- EML

Hinweis:

- EML darf verwendet werden, wenn Dateien in diesem Format bereits vorliegen.
- Anhänge sind in erwünschte Dateiformate zu konvertieren, sofern es sich nicht bereits um erwünschte oder akzeptierte Dateiformate handelt.

Bild, Rasterdaten

Records dieser Formatkategorie umfassen visuelle Darstellungen und Grafiken, die als Rastergrafik gespeichert werden.

Erwünscht:

- TIFF - in der Ausprägung der KOST
- JPEG2000 (J2F) lossless

Akzeptiert:

- JPEG2000 lossy oder visually lossless
- JPEG
- PNG
- DNG
- PDF/A-2u

Hinweise:

- Wenn Records in einem der akzeptierten Formate vorliegen, ist bei der Ablieferung an das Staatsarchiv keine Konvertierung vorzunehmen.
- DNG: nur akzeptiert, wenn RAW-Formate erhalten werden sollen; dies ist mit dem Staatsarchiv vorgängig abzustimmen.
- PDF/A-2u: ausschliesslich zu verwenden für Records, die beim ersetzenden Scannen entstanden sind beim Öffentlichen Organ.

Bild, Vektordaten

Records dieser Formatkategorie umfassen visuelle Darstellungen und Grafiken, die als Vektorgrafik gespeichert werden. Im Gegensatz zu Rastergrafiken können Vektorgrafiken ohne Qualitätseinbussen beliebig skaliert werden. Vektorgrafiken werden oft eingesetzt für Logos, Infografiken und technische Dokumentationen, die in ganz unterschiedlichen Grössen verwendet werden – etwa die Verwendung eines Logos auf Visitenkarten und auf Plakaten.

Erwünscht:

[Keine Dateiformate definiert.]

Akzeptiert

- SVG

Hinweise:

- In dieser Formatkategorie hat sich noch kein Dateiformat etabliert, das vom Staatsarchiv uneingeschränkt empfohlen wird und das daher als «erwünscht» eingestuft ist. Es gibt jedoch mit SVG ein Dateiformat, das akzeptiert wird.

Audio

Records dieser Formatkategorie enthalten Sprachaufzeichnungen oder Musik, die als akustische Information wiedergegeben werden.

Erwünscht:

- WAVE PCM

Akzeptiert:

- MP3
- FLAC
- ALAC

Hinweise:

- Wenn Records in einem der akzeptierten Formate vorliegen, ist bei der Ablieferung an das Staatsarchiv keine Konvertierung vorzunehmen.

Video

Records dieser Formatkategorie enthalten audiovisuelle Aufzeichnungen mit Bewegtbildern.

Erwünscht:

- FFV1 mit Matroska als Container

Akzeptiert:

- Digital Video DV
- MPEG-2
- MJPEG2000 (lossless)
- MJPEG2000 (DCP)
- MPEG-4
- Uncompressed Video

Hinweise:

- Bei Videodaten gibt es kein per se optimales Dateiformat für die Langzeitarchivierung. In Abhängigkeit von vorliegenden Daten, den Nutzungsformen und den Nutzergruppen kann die Entscheidung zur Übernahme situativ unterschiedlich ausfallen.
- Es ist zu unterscheiden zwischen dem Video-Codec und dem Containerformat.
- MJPEG200, MPEG-2 und MPEG-4 definieren sowohl einen Codec als auch ein Containerformat.
- Die Container AVI und MOV werden oft für uncompressed Video verwendet. Wenn Unterlagen in diesen Containern vorliegen, muss das Staatsarchiv vor einer Konvertierung konsultiert werden.

- Wenn Unterlagen in einem der akzeptierten Formate vorliegen, ist bei der Ablieferung an das Staatsarchiv keine Konvertierung vorzunehmen, da gerade bei Unterlagen dieser Formatkategorie jede Konvertierung zu Artefakten führen kann.
- Wenn andere als die hier genannten Video-Codec und Containerformate vorliegen, muss das Staatsarchiv vor einer Konvertierung konsultiert werden.

Datenbanken

Bei Datenbanken handelt es sich um strukturierte Daten, die durch spezifische Datenbank-anwendungen verwaltet werden. Auch komplexe Tabellenkalkulationsinhalte können als Datenbank eingestuft werden.

Erwünscht:

- SIARD

Akzeptiert:

- CSV mit Metadaten

Hinweis:

- Datenbanken werden nur als Datenbanken archiviert, wenn dies aufgrund der Bewertung des Staatsarchivs entschieden wird.
- SIARD kann nur relationale Datenbanken behandeln
- Sowohl bei SIARD als auch CSV ist die Dokumentation der Datenbank ein wesentlicher Bestandteil der Archivierung. Inhalt und Umfang der Dokumentation sind mit dem Staatsarchiv zu vereinbaren.

Strukturierte Daten

Strukturierte Daten sind Daten, welche in Form von Zahlen oder Zeichenfolgen in Tabellen (Reihen und Spalten) festgehalten werden. Die Anzahl der Spalten, der Datentyp einer Spalte, sowie Beziehungen zu anderen Tabellen sind definiert.

Erwünscht:

- XML
- JSON

Akzeptiert:

- CSV

Hinweis:

- Aufgrund der unterschiedlichen Handhabung der Maskierung von Trennzeichen und Zeilenumbruch bei verschiedenen CSV-Versionen wird CSV lediglich akzeptiert. Auch

können mit CSV anders als mit XML und JSON komplexe Datenstrukturen nur schwer abgebildet werden.

- Falls Daten in (semi-)strukturierten Textdateien aus Datenbanken stammen, ist die Vorgabe zu Datenbankformaten zu konsultieren.

Webseiten (Hypertext)

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.

Erwünscht:

- WARC

Akzeptiert:

- HTML

Hinweise:

- Das akzeptierte Format HTML kann verwendet werden für sehr kleine Webseiten
- Das Staatsarchiv verfügt über die Infrastruktur zur Sicherung und Archivierung von Webseiten. Daher sind Webseiten, die archiviert werden sollen, im Rahmen des Anbietersprozesses als solche anzumelden. Die Sicherung der Webseite und die Generierung des WARC-Files erfolgen anschliessend direkt durch das Staatsarchiv.
- Ist die Ablieferung von Webseiten resp. Hypertextdaten geplant, ist in jedem Fall vorgängig der Aussonderung mit dem Staatsarchiv Kontakt aufzunehmen.

Geodaten

Bei Geodaten handelt es sich um Unterlagen aus Geoinformationssystemen bzw. Geografischen Informationssystemen (GIS-Systeme). Sie beinhalten raumbezogene Informationen.

Erwünscht:

- Vektordaten: INTERLIS 2

Akzeptiert:

- Vektordaten: INTERLIS 1
- Bild- und Grafik-Rasterdaten: Rasterdaten mit ESRI World-File
- Produkte und Karten: Vektorgrafik in PDF

Hinweise:

- Ist die Ablieferung von Geodaten geplant, ist in jedem Fall vorgängig mit dem Staatsarchiv Kontakt aufzunehmen.

CAD

CAD (Computer-Aided Design) bezeichnet das computerunterstützte Konstruieren von unterschiedlichen Produkten wie Autos, Flugzeugen oder Bauwerken. 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.

Erwünscht:

[Keine Dateiformate definiert.]

Akzeptiert:

- DWG
- IFC
- Vektorgrafik in PDF

Hinweise:

- IFC: verwendet bei integrierter Produktion
- Vektorgrafik in PDF: zu verwenden bei Konstruktions- und Architekturplänen.
- Ist die Ablieferung von CAD-Daten geplant, ist in jedem Fall vorgängig der Aussonderung mit dem Staatsarchiv Kontakt aufzunehmen.
- In dieser Formatkategorie hat sich noch kein Dateiformat etabliert, das vom Staatsarchiv uneingeschränkt empfohlen wird und das daher als «erwünscht» eingestuft ist. Es gibt jedoch mehrere Dateiformate, die akzeptiert werden.

Statistik

Bei Statistikdaten handelt es sich um strukturierte Daten, die in Statistikanwendungen generiert und verwaltet werden.

Erwünscht:

[Keine Dateiformate definiert.]

Akzeptiert:

- SPSS Austauschformat POR
- SAS transport file format
- SDMX
- CSV mit Metadaten

Hinweise:

- Ist die Ablieferung von Statistikdaten geplant, ist in jedem Fall vorgängig der Aussonderung mit dem Staatsarchiv Kontakt aufzunehmen.

- In dieser Formatkategorie hat sich noch kein Dateiformat etabliert, das vom Staatsarchiv uneingeschränkt empfohlen wird und das daher als «erwünscht» eingestuft ist. Es gibt jedoch mehrere Dateiformate, die akzeptiert werden.

Basel, 1. April 2024

A handwritten signature in black ink, appearing to read 'Esther Baur', written in a cursive style.

Esther Baur
Staatsarchivarin