SLUB
Wir führen Wissen.

Langzeitarchivfähige Dateiformate

SLUB Dresden

Version 2.0.1, 2021-10-01

Inhaltsverzeichnis

Überblick . . . 1
Einleitung . . . 1
Risikoanalyse . . . 2
- Vorbetrachtung . . . 3
- Kriterien zur Beurteilung . . . 3

Risikobetrachtung ausgewählter Dateiformate des SLUBArchivs . . . 8
- PDF/A & PDF/UA (Portable Documents Format) . . . 8
- GIF (Graphics Interchange Format) . . . 8
- HTML, HTM (HyperText Markup Language) . . . 8
- SVG (Scalable Vector Graphics) . . . 8
- TIF (Tagged Image File Format) . . . 9
- WARC (WebARChive) . . . 9
- XML (eXtensible Markup Language) . . . 9
- DOCX, XLSX, PPTX (OfficeOpen XML) . . . 9
- ODF (OASIS Open Document Format for Office Applications) . . . 10
- PNG (Portable Network Graphics) . . . 10
- Matroska Multimediacontainer . . . 10

Release Note 2.0.1 vom 2021-10-01

- Neustrukturierung
- kleinere Darstellungsprobleme gefixt

Überblick

Dieses Dokument ist ein Teil der Übernahmespezifikation für das SLUBArchiv. Zur Übernahmespezifikation gehören die folgenden Dokumente [1]:

- In der **Übernahmevereinbarung** zwischen SLUB und Dienstnehmer sind die Daten, Ansprechpartner und organisatorischen Randbedingungen beschrieben. Dies schliesst die zu verwendenden Handreichungen für Dateiformate bzw. Objektgruppen mit ein.
- In der **Langzeitarchivfähige Dateiformate** sind die Formate aufgeführt, die die SLUB als potentiell archivfähig bewertet und für die die Funktionalitäten der Formaterkennung, Formatvalidierung und Metadatenextraktion in einem ausreichenden Maß durch das SLUBArchiv gewährleistet werden könnten. Die genaue Festlegung erfolgt spezifisch für jeden Workflow und jede Objektgruppe auf Basis der ermittelten signifikanten Eigenschaften.
- Die **SIP Spezifikation für automatischen Ingest SLUBArchiv** beschreibt den Aufbau der Ablieferungspakete (englisch: Submission Information Package, SIP) mit denen der Dienstnehmer die zu archivierenden Dokumente für das SLUBArchiv bereitstellt.
- Die **DIP Spezifikation für automatischen Access SLUBArchiv** beschreibt den Aufbau eines Auslieferungspaketes (englisch: Dissemination Information Package, DIP), welches für die automatische Weiterverarbeitung zielgruppengerechter Ausspielungen (Access) von im SLUBArchiv archivierter digitaler Datenobjekte (IE) geeignet ist
- Die **Workflow Spezifikation für automatisierte Interaktionen mit dem SLUBArchiv** beschreibt den Prozess der Übergabe zu archivierender Dokumente in das SLUBArchiv (Ingest / AIP Update), das Fehlerprotokoll und den Zugriff auf die archivierten digitaler Objekte (Access).
- Das Dokument **Spezifikation Rechteauszeichnung SLUBArchiv** beschreibt, wie rechtliche Informationen zu einem Datenobjekt kodiert und abgelegt werden müssen.
- Das Dokument **Webservice SLUBArchiv** beschreibt Funktionen, die Dienstnehmer nutzen können, um Informationen über ihre Daten im SLUBArchiv über einen Webservice abzufragen.
- Vom SLUBArchiv verwendete Begriffe sind im **Glossar SLUBArchiv** definiert.

Einleitung

Digitale Langzeitarchivierung stellt besondere Anforderungen an die verwendeten Dateiformate. Digitale Master müssen in offen spezifizierten Dateiformaten vorliegen, die eine leichte Konvertierbarkeit erlauben, um dem Formatwandel über die Jahre Rechnung zu tragen. Gleichzeitig müssen auch über mehrere Konvertierungen die signifikanten Eigenschaften erhalten bleiben, die eine zukünftige Nutzbarkeit erlauben.

Deswegen ist es wichtig, vor der permanenten Archivierung eine Menge an Dateiformaten festzulegen, die für prinzipiell langzeitarchivfähig gehalten werden und deren Einspeisung in das

Langzeitarchiv erlaubt werden darf.

Langzeitarchivfähige Dateiformate müssen folgenden Kriterien genügen:

- offen standardisiert (notfalls offen spezifiziert); kein proprietäres Format
- weit verbreitet
- geringe Komplexität
- ohne Zugriffsschutzmechanismen wie Kopierschutz, Verschlüsselung, DRM
- selbstdokumentierend
- robust
- keine Abhängigkeiten zu anderen Dateiformaten
- lizenzfrei
- validierbar

Die offene Standardisierung von Dateiformaten erlaubt es im Fall der Formatobsoleszenz, Lese- und Konvertierungsprogramme neu zu erstellen. Weitverbreitete Dateiformate bringen es mit sich, dass das Formatwissen über diese stärker verbreitet und für die Nachwelt eher dokumentiert ist. Hinzu kommt, dass die Verfügbarkeit von geeigneter Software mit hoher Wahrscheinlichkeit länger gegeben ist.

Die geringe Komplexität eines Dateiformates geht oft mit einer besseren Verständlichkeit und einer weniger fehleranfälligen Implementierung einher.

Unter selbstdokumentierenden Dateiformaten versteht man jene, die über eine bestimmte Signatur (magic byte, meist am Dateianfang) eine einfache Identifikation des Dateiformates erlauben. Dateiformate sind dann robust, wenn sich einzelne Bitfehler auf den Inhalt der Datei nur gering oder gar nicht auswirken. Aus diesem Grund wird oft von komprimierten Datenformaten abgeraten. Wenn, wie im Falle des Videodatenformates Matroska/FFV1, verlustfreie Datenkompression mit dem gezielten Hinzufügen von Redundanz, zB. durch CRC, kombiniert wird, kann die Forderung nach Robustheit erfüllt werden.

Sinnvoll ist es, wenn Dateiformate vollständig spezifiziert sind und nicht auf andere Dokumente angewiesen sind. Dies erhöht die Chance Lese- und Konvertierungsprogramme nur anhand der vorhandenen Spezifikation zu erstellen. Lizenzbehaftete, proprietäre Spezifikationen erschweren den Erhalt, da eine Sicherung der Spezifikationsdokumente notwendig ist und Lizenzgeber über die Zeit nicht mehr verfügbar sein können. Dateiformate sollten aus dem selben Grund auch keine Zugriffsschutzmechanismen enthalten, da der Zugriff auf den Inhalt über die Zeit verloren gehen kann (oder gar nicht möglich ist).

Sinnvoll ist ebenso, dass für die Dateiformate in der digitalen Langzeitarchivierung geeignete Validatoren oder zumindest Referenzimplementierungen zur Verfügung stehen, die eine Beurteilung ermöglichen, inwieweit Dateien von Programmen interpretiert werden können.

Proprietäre Datenformate halten diesen Anforderungen nicht stand.

Risikoanalyse

Vorbetrachtung

Basierend auf der Risikodefinition "Risiko ist Eintrittswahrscheinlichkeit mal Schadenshöhe" ist eine Risiko-Bewertung der im Archiv befindlichen Dateiformate vorzunehmen. Arbeitsgrundlage ist der Entwurf der Digital Preservation Risk Matrix der US National Archives (NARA) https://github.com/usnationalarchives/digital-preservation/tree/master/Digital_Preservation_Risk_Matrix.

Das Risiko besteht für das Archiv notwendige Formatmigrationen dann, wenn bei genutzten Dateiformaten Obsoleszenz droht. Die Schadenshöhe ist dabei an den Aufwand der Migration, bzw. Totalverlust der Dateien bei nicht stattfindender Migration gekoppelt. Die Eintrittswahrscheinlichkeit wird durch mehrere Parameter, wie Grad der Verbreitung, Komplexität des Dateiformates, Quelloffene Implementierung, offene Spezifikationen und Abkündigung von Softwareprodukten etc. bestimmt.

Kriterien zur Beurteilung

Bewertung und Gewichtung der Kriterien

Die Bewertung der Kriterien erfolgt zwecks einfacherem Handling nach einem Punktesystem. Die Punkte werden pro Kriterium addiert.

Die Bewertung erfolgt aus Sicht des vom SLUBArchiv gewählten Dateiformat-Profils. Bei Zweifeln oder bei zusammengesetzten Dateiformaten, wie zB. Matroska-FFV1, wird bei Widersprüchen in den Antworten die Punktezahl *Sonstige* bzw. die pessimistische Variante gewählt.

Die pessimistische Bewertung eines Kriteriums ist dem Umstand geschuldet, dass es sinniger ist ein Risiko zu über- statt zu unterschätzen.

Es können gleiche Fragen in unterschiedlichen Kriterien vorkommen. Der Hintergrund ist der, dass die zugrundeliegende Eigenschaft Auswirkung auf beide Kriterien haben kann.

Um die Nachvollziehbarkeit der Beurteilung zu gewährleisten, sollte eine kurze Begründung hinterlegt werden.

Die Bewertung eines Dateiformates anhand der vorliegenden Kriterien dient zur Bestandsaufnahme der möglichen Risiken, die durch die Verwendung von bestimmten Dateiformaten im Archiv bestehen.

Sie ersetzt nicht eine kritische Reflektion über die Ergebnisse der Bewertung und ist nicht hinreichend für die Auswahl eines archivtauglichen Dateiformates.

Der Fall, dass ein oder mehrere Kriterien mit 0 bewertet werden, ist daher nicht zwingend ein Ausschlusskriterium für die Ungeeignetheit eines Dateiformates. Genausowenig gilt der umgedrehte Fall.

> Die Bewertung ist nur ein Indikator, der eine objektivierte Auseinandersetzung mit den möglichen Risiken ermöglicht und damit Basis für ein Risikomanagment.

Offenlegungsgrad (offen)

Dieses Kriterium ermittelt, inwieweit ein Dateiformat offen dokumentiert ist. Je niedriger die Punktzahl, desto höher das Risiko, dass ein Dateiformat nicht mehr migrierbar ist, weil keine Konvertierungsprogramme entwickelt werden können.

- Ist das Dateiformat offen spezifiziert, d.h. das Recht die Spezifikation zu lesen, anzuwenden und basierend darauf Werkzeuge zu entwickeln ist nicht eingeschränkt?? (Ja=2 / Nein=0 / Sonstige=0)
- Gibt es aktuell verfügbare und lauffähige Validierungswerkzeuge? (Ja=2 / Nein=0 /Sonstige=1)
- Sind diese Validierungswerkzeuge Opensource? (Ja=1 / Nein=0 / Sonstige=0)
- Ist das Dateiformat durch eine internationale Standardisierungsorganisation geprüft und als Standard veröffentlicht? (Ja=2 / Nein=0 / Sonstige=1)
- Ist die Spezifikation kostenfrei zugänglich? (Ja=1 / Nein=0 / Sonstige=0)
- Ist die Spezifikation komplett verfügbar? (Ja=2 / Nein=0 / Sonstige=1)

Adoptionsgrad (verbreitet)

Dieses Kriterium definiert inwieweit ein Dateiformat verbreitet ist. Je niedriger die Punktzahl, desto höher das Risiko, dass ein Dateiformat nicht mehr migrierbar ist, weil niemand das Dateiformat kennt und es nicht einmal mehr anzeigbar ist.

- Wird das Dateiformat durch Behörden und öffentliche Einrichtungen häufig verwendet? (Ja=1 / Nein=0 / Sonstige=0)
- Wird das Dateiformat auch außerhalb von Behörden und öffentlichen Einrichtungen häufig verwendet? (Ja=2 / Nein=0 / Sonstige=0)
- Wird das Dateiformat aktiv betreut? (durch eine Community=3 / einzelne Person=1 / einzelne öffentliche Einrichtung=2 / gar nicht=0)
- Sind für das Dateiformat verschiedene Anzeigeprogramme verfügbar? (Ja=2 / Nein=0 / Sonstige=0)
- Hat die Archiv- und Bibliothekscommunity dieses Dateiformat als Archivfähiges Datenformat allgemein anerkannt? (Ja=2 / Nein=0 / Sonstige=0)

Transparenzgrad (transparent)

Dieses Kriterium definiert die einfache Durchführbarkeit von Analysen. Je niedriger die Punktezahl, desto höher ist das Risiko Dateien nur mit unverhältnismäßig hohem Aufwand analysieren zu können.

- Ist das Dateiformat menschenlesbar und mittels Texteditor anzeig- und bearbeitbar? (Ja=1 / Nein=0 / Sonstige=0)
- Ist die verfügbare Spezifikation des Dateiformates detailliert genug, so dass man das

Dateiformat mittels Hexeditor in Grundzügen analysieren kann? (Ja=2 / Nein=0 / Sonstige=0)

- Verwendet das Dateiformat Standardzeichen oder andere bekannte Kodierungsmethoden, zB. UTF8 für Zeichen bzw. Zahlen nach IEEE? (Ja=1 / Nein=0 / Sonstige=0)
- Ist Software, die das Dateiformat erzeugen kann für wenig Geld oder gar kostenlos erhältlich? (Ja=2 / Nein= 0 /Sonstige=1)
- Ist Software, die das Dateiformat erzeugen kann für aktuelle Computerplattformen verfügbar? (Ja=1 / Nein=0 / Sonstige=0)
- Ist das Dateiformat komprimiert bzw. benutzt es Datenkompression? (Ja=0 & Nein=2 / Sonstige=1)
- Erlaubt das Dateiformat nutzerdefinierte Datenkompression, die zu Veränderungen der essentiellen Charakteristiken, zB. Kompressionsartefakte, der Datei führen? (Ja=0 / Nein=2 / Sonstige=0)

Grad der Selbstdokumentation (selbstdokumentiert)

Dieses Kriterium definiert, inwieweit das Dateifomat selbstdokumentiert ist. Je niedriger die Punktzahl, desto höher das Risiko, dass man ohne externe Metadaten eine Datei nicht zuordnen und analysieren kann.

- Enthält das Format eine Selbstbeschreibung durch deskriptive Metadaten? (Ja=2 / Nein=0 / Sonstige= 0)
- Enthält das Format eine Selbstbeschreibung durch technische Metadaten? (Ja=2 / Nein=0 / Sonstige=0)
- Enthält das Format eine Selbstbeschreibung durch administrative Metadaten? (Ja=1 / Nein=0 / Sonstige=0)
- Liegen diese Metadaten in einem anerkannten, internationalen Metadatenstandard vor? (Ja=2 / Nein=0 / Sonstige=0)
- Sind die Formatmetadaten robust genug für ein akkurate Dateianalyse? (Ja=2 / Nein=0 / Sonstige=1)
- Erlaubt das Format den Nachweis der Authentizität durch eingebettete elektronische Signaturen ohne die Nutzbarkeit einzuschränken? (Ja=2 / Nein=0 / Sonstige=1)

Grad der Abhängigkeit von externer Hard- und Software (technologieunabhängig)

Dieses Kriterium definiert die Unabhängigkeit von spezifischer Soft- und Hardware und ist ein Indiz für proprietäre Formate. Je niedriger die Punktezahl, desto höher das Risiko, dass ein Erhalt der Nutzbarkeit durch Emulation und Formatmigration wegen fehlender Abspielumgebung nicht mehr durchführbar sind.

- Benötigt das Dateiformat eine ganz spezielle Hardwareumgebung (zB. Grafikkarten, Soundkarten, Speicheranforderungen, Bluerayplayer, Spielekonsole), um damit zu interagieren, es nutzen und migrieren zu können? (Ja=0 / Nein=3 / Sonstige=1)
- Benötigt das Dateiformat proprietäre Software, um es anschauen, nutzen und migrieren zu

können? (Ja=0 / Nein=3 / Sonstige=1)

- Benötigt das Dateiformat spezielle Plugins oder Scripte, um es anschauen, nutzen und migrieren zu können? (Ja=0 / Nein=2 / Sonstige=0)
- Benötigt das Dateiformat ein bestimmtes Betriebssystem, um es anschauen, nutzen und migrieren zu können? (Ja=0 / Nein=1 / Sonstige=0)

Grad der Lizenz- und Patentabhängigkeit (rechtefrei)

Dieses Kriterium definiert das Risiko der Behinderungen von Eigenentwicklungen für die Formatmigration. je niedriger die Punktezahl, desto höher das Risiko, dass ein Dateiformat nicht mehr migrierbar ist, weil keine Konvertierungsprogramme entwickelt werden können.

- Ist das Format Gegenstand von Lizenz- oder Patentansprüchen, die die Entwicklung von Open-Source-Tools zum Öffnen und Verwalten der Dateien behindern könnten? (Ja=0 / Nein=3 / Sonstige=1)
- Sind mit dem Format Gebühren verbunden, die sich aus den Lizenz- oder Patentansprüchen ergeben? (Ja=0 / Nein=2 / Sonstige=1)
- Unterliegt die Formatspezifikation Open-Source-Lizenzbedingungen? (Ja=3 / Nein=0 / Sonstige=1)

Grad der technischen Schutzmaßnahmen (drmfrei)

Dieses Kriterium definiert die Gefährdung der Nutzbarkeit durch nicht mehr entschlüsselbare oder verwendbare Inhalte. Je niedriger die Punktezahl, desto höher das Risiko dass eine Datei nicht mehr nutzbar ist.

- Verfügt das Format über die Möglichkeit, die gesamte oder einen Teil der resultierenden Datei zu verschlüsseln? (Ja=0 / Nein=2 / Sonstige=0)
- Erfordert das Format die Verwendung von Verschlüsselung? (Ja=0 / Nein=2 / Sonstige=1)
- Können technische Schutzmaßnahmen (z. B. Digital Rights Management) angewendet werden? (Ja=0 / Nein=2 / Sonstige=0)
- Erlaubt das Format eingebettete Informationen zur Durchsetzbarkeit von Rechtsansprüchen, die die zukünftige Nutzbarkeit gefährden könnten, wie z. B. Wasserzeichen, zeitlich befristete Lizenzschlüssel? (Ja=0 / Nein=2 / Sonstige=0)

Grad der Alterung (aktuell)

Dieses Kriterium ist sowohl ein Indiz für mögliche Formatobsoleszenz, als auch für die Stabilität von Dateiformaten. Je niedriger die Punktezahl, desto höher das Risiko, dass eine Datei nicht gelesen werden kann, weil die Dateiformatspezifikation veraltet oder zu frisch ist.

- Wann wurde die Formatspezifikation erstmals erstellt? (>30Jahre=0 / >20Jahre=1 / >10 Jahre=2 | <10 Jahre =1)
- Wann wurde die Formatspezifikation letztmalig aktualisiert? (>30 Jahre=0 / >20 Jahre=1 / >10Jahre=2 | <10 Jahre =3)

- Ist die Formatspezifikation unter einem üblichen persistenten Identifier bzw. persistenter URL auffindbar? (Ja=2 / Nein=0 / Sonstiges=0)

Komplexitätsgrad (einfach)

Dieses Kriterium definiert, wie umfangreich Formatwissen aufgebaut werden muss und wieviel Aufwand für die Fehleranalyse und Entwicklung von Migrationslösungen aufgebracht werden muss. Je niedriger die Punktezahl, desto höher das Risiko, dass Formatwissen nur mit unverhältnismäßig hohem Aufwand aufgebaut und Dateien nicht hinreichend analysisiert werden können.

- Verweist die Spezifikation explizit auf andere Dateiformat-Standards? (Ja=0 / Nein=2 / Sonstige=1)
- Erlaubt die Spezifikation Varianten des Dateiformats? (Ja=0 / Nein=3 / Sonstige=1)
- Besteht die Spezifikation aus mehreren Teilspezifikationen, die für sich veröffentlicht wurden und nicht alle aufeinander verweisen? (Ja=0 / Nein=2 / Sonstige=1)
- Verwendet das Datenformat für Daten und Metadaten ein eingeschränktes, kontrolliertes Vokabular (Ja=1 / Nein=0 / Sonstiges=1)
- Ist die Spezifikation umfangreich? (<100 Seiten=3 / < 1000 Seiten=2 / < 5000 Seiten=1 / >5000 Seiten=0)
- Erlaubt die Spezifikation mehrere Kodierungsvarianten von Daten? (Ja=0 / Nein=2 / Sonstige=1)
- Ist das Datenformat multiplex fähig, sprich: Mit anderen Datenformaten im Bitstrom mischbar, zB. weil es Datenpaket orientiert aufgebaut ist (wie zB. MP3)? (Ja=0 / Nein=2 / Sonstige=0)

Grad der Robustheit (robust)

Dieses Kriterium definiert wie gut ein Dateiformat mit fehlerhaften Übertragungen umgehen kann. Je niedriger die Punktezahl, desto höher das Risiko, dass Inhalte von beschädigten Dateien nicht wieder hergestellt werden können.

- Ist das Dateiformat menschenlesbar und mittels Texteditor anzeig- und bearbeitbar? (Ja=2 / Nein=0 / Sonstige=0)
- Ist das Dateiformat XML basiert? (Ja=2 / Nein=0 / Sonstige=0)
- Verwendet das Dateiformat einen eingeschränkten Bytewerte-Bereich für die Daten? (Ja=1 / Nein=0 / Sonstige=0)
- Verwendet das Dateiformat in Voreinstellung Kompression? (Ja=0 / Nein=2 / Sonstige=0)
- Verwendet das Dateiformat als zentralen Einstiegspunkt den Dateianfang das Dateiende oder ist es Datenblock-orientiert? (Dateianfang=2 / Datenblock=3 / Dateiende=0 / Unbekannt=1)
- Sind die Headerinformationen Prüfsummen gesichert? (Ja=2 / Nein=0 / Sonstige=0)
- Sind alle Daten außerhalb des Headers Prüfsummen gesichert? (Ja, alle=2 / Ja, nur Metadaten=1 / Ja, nur Nutzdaten=1 / Nein=0 / Sonstige=0)
- Sind 1-Bitfehler erkennbar? (Ja=2 / Nein=0 / Sonstige=1)
- Sind 1-Bitfehler korrigierbar (zB. bei CRC32)? (Ja=2 / Nein=0 / Sonstige=1)

- Sind n-Bitfehler erkennbar? (Ja=2 / Nein=0 / Sonstige=1)

Risikobetrachtung ausgewählter Dateiformate des SLUBArchivs

Dieses Dokument listet Dateiformate auf, die nach **aktuellem** Stand der Bearbeitung durch die SLUB grundsätzlich für die dauerhafte Aufnahme in das Langzeitarchiv der SLUB in Frage kommen.

> Die **exakte Festlegung**, welches Dateiformat im SLUBArchiv verwendet wird, erfolgt durch das SLUBArchiv **spezifisch** für jeden Workflow und jede Objektgruppe auf Basis der zu diesem Zweck ermittelten und dokumentierten **signifikanten Eigenschaften**.

Für bestimmte Dateiformate hat die SLUB Handreichungen herausgegeben, die die Kriterien für die Aufnahme in das SLUBArchiv genauer spezifizieren.

PDF/A & PDF/UA (Portable Documents Format)

- **archivfähige Formatversionen:** PDF/A-1 (ISO 19005-1:2005), PDF/A-2 (ISO 19005-2:2011)
- **PRONOM-Formate:** fmt/354, fmt/95
- **Randbedingungen:**
  - maßgebend ist die *Handreichung Publikationen und Pflichtexemplare* des SLUB Archiv

GIF (Graphics Interchange Format)

- **archivfähige Formatversionen:** GIF87a (nicht: GIF89a)
- **PRONOM-Formate:** fmt/3
- **Randbedingungen:**
  - Bilddaten ausschließlich unkomprimiert

HTML, HTM (HyperText Markup Language)

- **archivfähige Formatversionen:** HTML 3.2 (nur Bestandsdaten), HTML 4.01 strict, HTML 5
- **PRONOM-Formate:** fmt/98 (HTML 3.2), fmt/100 (HTML 4.01), fmt/471 (HTML 5)
- **Randbedingungen:**
  - nur statische Seiten ohne ausführbaren Code
  - Archivierung nur mit zugehörigem CSS (wenn vorhanden)
  - Archivierung vorzugsweise als WARC (fmt/289)

SVG (Scalable Vector Graphics)

- **archivfähige Formatversionen:** SVG 1.0, SVG 1.1, SVG 1.2 Tiny
- **PRONOM-Formate:** fmt/91, fmt/92, fmt/413
- **Randbedingungen:** (keine)

TIF (Tagged Image File Format)

- **archivfähige Formatversionen:** TIFF Rev. 6.0 Part 1 (Baseline TIFF), GeoTIFF, Exif
- **PRONOM-Formate:** fmt/10, fmt/155, x-fmt/387
- **Randbedingungen:**
  - maßgebend ist die *Handreichung TIFF* und ggf. die Ergänzung durch *Handreichung GeoTIFF* des SLUBArchiv

WARC (WebARChive)

- **archivfähige Formatversionen:** WARC (ISO 28500)
- **PRONOM-Formate:** fmt/289
- **ACHTUNG:** Die weitere Behandlung dieses Formates ist noch nicht abschließend geklärt (z. B.: Was passiert mit den eingeschlossenen und mglw. nicht archivfähigen Bilddateien?)

XML (eXtensible Markup Language)

- **archivfähige Formatversionen:** (alle) (u. a. Alto-XML, METS, MODS, DC)
- **PRONOM-Formate:** fmt/101
- **Randbedingungen:**
  - UTF-8-codiert
  - well-formed
  - validierbar gegen ein frei zugängliches standardisiertes Schema (RelaxNG, DTD, XML-Schema)

DOCX, XLSX, PPTX (OfficeOpen XML)

- **archivfähige Formatversionen:** docx 2013, xlsx 2013, pptx 2013
- **PRONOM-Formate:** ExL-Fmt-62, fmt/189
- **Randbedingungen:**
  - mit MS Office 2013 oder neuerer Version erstellt (ältere Formate sind nicht standardkonform)

ODF (OASIS Open Document Format for Office Applications)

- **archivfähige Formatversionen:** odt (OpenDocument-Text), ods (OpenDocument-Spreadsheet), odp (OpenDocument-Präsentation), odg (OpenDocument-Grafik)
- **PRONOM-Formate:** fmt/136, fmt/290, fmt/291, x-fmt/3, fmt/137, fmt/294, fmt/295, fmt/138, fmt/292, fmt/293, fmt/139, fmt/296, fmt/297
- **Randbedingungen:** (keine)

PNG (Portable Network Graphics)

- **archivfähige Formatversionen:** png
- **PRONOM-Formate:** fmt/11, fmt/12, fmt/13
- **Randbedingungen:** (keine)

Matroska Multimediacontainer

- **archivfähige Formatversionen:** mkv (FFV1 v3), mka (linear PCM)
- **PRONOM-Formate:** fmt/569
- **Randbedingungen:**
  - maßgebend ist die *Handreichung retrodigitalisierter Film und Video* des SLUB Archiv

[1] Die genannten Dokumente sind auf der Webseite des SLUBArchivs unter https://slubarchiv.slub-dresden.de/technische-standards-fuer-die-ablieferung-von-digitalen-dokumenten/ veröffentlicht und sind dort, technisch bedingt, spezifischer benannt.