

**DEPARTEMENT  
BILDUNG, KULTUR UND SPORT**

Abteilung Kultur

Bibliothek und Archiv Aargau

21. Januar 2019

**MERKBLATT 5**

**Migration in archivtaugliche Formate**

---

**1. Ziel dieses Merkblatts**

Dieses Merkblatt unterstützt Sie dabei, die Migration von elektronischen Dokumenten in archivtaugliche Formate gemäss geltender Richtlinie Aktenführung und Archivierung umzusetzen.

Unter 2. wird der Prozess der Migration skizziert. Unter 3. werden die zulässigen archivtauglichen Formate aufgeführt und erläutert. Unter 4. wird schliesslich die Rolle des Staatsarchivs und des technischen Ansprechpartners der Organisationseinheit (= OE) definiert.

**2. Prozess der Migration**

Migration ist der Prozess, um elektronische Dokumente von einem Dateiformat in ein anderes zu konvertieren. Die für die Ablieferung ans Staatsarchiv verwendeten Formate der Dokumente müssen archivtauglich und somit für die Langzeitarchivierung geeignet sein.

Nachfolgend werden die wichtigsten Schritte bei der Migration aufgeführt und beschrieben:

**2.1 Zeitpunkt definieren**

Die Migration in archivtaugliche Formate muss von der OE rechtzeitig geplant und spätestens vor der Ablieferung ans Staatsarchiv durchgeführt werden.

Aus den folgenden Gründen wird empfohlen, die Migration in archivtaugliche Formate so früh wie möglich im Data Lifecycle zu integrieren:

- a) Bei langen Aufbewahrungsfristen kann bereits im System des Aktenbildners die Migration in ein archivtaugliches Dateiformat erforderlich sein, um die Lesbarkeit während der Aufbewahrungsdauer zu garantieren.
- b) Eventuell sind Informationen in der Benutzerumgebung des Aktenbildners vorhanden, die verloren gehen können, wenn die Umwandlung in archivtaugliche Dateiformate erst unmittelbar vor der Ablieferung erfolgt.
- c) Bei jeder Migration können unerwartete Fehler auftreten, die eher von Personen bemerkt werden, die noch unmittelbar mit Entstehung und Inhalt der Unterlagen vertraut sind.

**2.2 Analyse**

Dateiformate lassen sich in folgende Formatkategorien einteilen: Text, PDF und Office-Dokumente, Tabellenkalkulationen, Datenbanken, Bild, Audio und Video. Für jede dieser Kategorien gibt es bestimmte archivtaugliche Dateiformate.

Unter 3. werden nur die gängigsten Dateiformate behandelt. Fragen Sie im Zweifelsfall bei Ihrem technischen Ansprechpartner oder – zwingend im Fall der Übergabe an das Staatsarchiv – bei diesem nach.

### **2.3 Konvertierung durchführen**

Für die Umwandlung von herkömmlichen Dateiformaten in archivtaugliche oder für die Erstellung von archivtauglichen Dateiformaten können unterschiedliche Programme verwendet werden. Die Qualität dieser Programme respektive der Konvertierungsergebnisse können variieren. Darüber hinaus ist es in gewissen Fällen sinnvoll, externe Dienstleister heranzuziehen.

### **3. Dateiformate**

Das Staatsarchiv macht in Hinsicht der Übernahme von elektronischen Daten zur Langzeitarchivierung Vorgaben, die auf nationalen und internationalen Standards basieren.

Insbesondere folgende:

- Schweizerisches Bundesarchiv (BAR): Standards für die Archivierung digitaler Unterlagen.
- Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST): Katalog archivischer Dateiformate, Kriterienkatalog.
- ISO-Normen.

Das Staatsarchiv zieht eine Beschränkung der zulässigen Dateiformate für die digitale Langzeitarchivierung vor. Ziel ist, die Dateiformate langfristig haltbar zu machen und eine effiziente Migration zu garantieren.

## Tabellarische Übersicht der archivtauglichen Formate

Dateiformate		Archivtaugliche Formate	Kommentare
Textdaten	unstrukturiert	TXT	Zeichenkodierung: ASCII oder UTF-8
	strukturiert	CSV	
	Markup	XML/XSD	
PDF, Office-Dokumente		PDF/A-2u	bereits vorhandene PDF/A-1-Dokumente sollen nicht migriert werden; PDF/A-2a und PDF/A-2b-Dokumente werden ebenfalls akzeptiert
Tabellenkalkulationen	einfach	CSV	Zeichenkodierung: ASCII oder UTF-8
	komplex	keine Migration	ursprüngliches Dateiformat belassen
	neu erstellte	OOXML	
Datenbanken		SIARD	
Bilddaten	Raster	TIFF, JPEG2000	bereits vorhandene JPEG-Dateien sollen nicht migriert werden
	Vektor	SVG, PDF/A-2u	
	Retrodigitalisierung (Textdokumente)	PDF/A-2u	bereits vorhandene PDF/A-1-Dokumente sollen nicht migriert werden
Audio-Daten		WAV	bereits vorhandene MP3-Dateien sollen nicht migriert werden
Video-Daten		nicht bestimmt	das jeweils zu verwendende Dateiformat (Containerformat und Codex) ist mit dem StAAG zu vereinbaren

### 3.1 Textdaten

#### Unstrukturierte Textdaten (TXT)

1. Unstrukturierte Textdaten bestehen nur aus reinem Text (ohne jegliche Formatierung). Sie werden oft mit der Dateiendung `.txt` versehen, wie z. B. *Einladung.txt*. Reine Text-Dateien erfüllen die meisten archivischen Anforderungen maximal.
2. Andere Dateiendungen sind ebenfalls möglich. Wenn Sie nicht sicher sind, ob die Datei Textdaten enthält, empfiehlt es sich, die Datei mit einem Texteditor (z. B. Notepad unter Windows) zu öffnen, um zu kontrollieren, ob es sich um eine Textdatei handelt. Im Zweifelsfall fragen Sie Ihren technischen Ansprechpartner.

3. Beispiele von unstrukturierten Textdaten sind Log-Dateien, kurze Beschreibungen, Memos oder unformatierte E-Mails.
4. Unstrukturierte Textdaten liegen bereits in einem archivtauglichen Dateiformat (vgl. unten Erläuterungen zur Zeichenkodierung) vor.

### Strukturierte Textdaten (CSV)

1. Strukturierte Textdaten bestehen ebenso nur aus reinem Text und sind tabellarisch strukturiert. Jede Zeile stellt einen Datensatz dar, die mehrere Felder aufweist, welche meistens durch ein Komma getrennt sind. Andere Trennzeichen (z. B. Tabulator) werden auch oft benutzt und sind genauso zulässig.
2. Das Trennzeichen darf innerhalb einer Datei nicht geändert werden.
3. Wird das Komma als Trennzeichen verwendet und kommt es innerhalb eines Feldes vor, muss das betroffene Feld durch Anführungszeichen eingefasst werden, z. B.: "abc,d",efg,hijkl
4. Anführungszeichen, die nicht als Trennzeichen verwendet werden, müssen durch ein zusätzliches Anführungszeichen markiert werden, z. B.: Monitor 24", 2kg → "Monitor 24"", 2kg"
5. Die erste Zeile einer CSV-Datei sollte die Namen der Felder in der Tabelle darstellen (Header) und muss wie alle anderen Zeilen formatiert sein, z. B.:

```
jahr,menge,geprüft
2015,34,ja
2012,66,nein
```

6. Wenn der erwähnte Header fehlt, muss die Struktur der Tabelle in der CSV-Datei extern (durch ein anderes Dokument) dokumentiert sein, da die Daten oft nicht ohne zusätzliche Metadaten interpretierbar sind. So kann z. B. die Zahlenreihe 34, 45, 81, 28, 56 sowohl als eine Liste von Alters- als auch eine Liste von Gewichtsangaben interpretiert werden.
7. CSV-Dateien müssen die Dateierdung .csv haben.
8. Strukturierte Textdaten liegen bereits in einem archivtauglichen Dateiformat vor und müssen nicht migriert werden (vgl. unten Erläuterungen zur Zeichenkodierung).

### Markup-Textdaten (XML, XSD)

1. Markup-Textdaten werden zur Beschreibung strukturierter Informationen und Spezifikation der Interpretation von Daten eingesetzt.
2. XSD-Dateien sind Dateien (selber in XML-Format) mit XML-Schemadefinitionen, d. h. Dateien, die das Vokabular und die Strukturen in den XML-Dateien definieren. Sie sollen mit den von ihnen definierten XML-Dateien zusammen geliefert werden.
3. Fehlt eine definierende XSD-Datei, müssen das verwendete Vokabular und die Datenstruktur in den XML-Dateien separat dokumentiert werden, um ihre spätere Nutzung zu sichern.
4. XML- und XSD-Dateien müssen wohlgeformt sein, d. h. mit den Vorgaben der Spezifikation von XML konform sein. Die Wohlgeformtheit kann mit einem XML-Validierungstool geprüft werden.
5. Markup-Textdaten müssen nicht migriert werden (vgl. unten Erläuterungen zur Zeichenkodierung).

### Zeichenkodierung

1. Alle Textdaten (unstrukturiert, strukturiert und Markup) müssen entweder in ASCII oder in Unicode (in der Variante UTF-8, ohne BOM) kodiert sein.
2. Ist die Datei in einem anderen Zeichensatz kodiert, muss die Textdatei vor der Ablieferung in UTF-8 (ohne BOM) umgewandelt werden.
3. Um eine fehlerfrei Umwandlung durchzuführen, ist unabdingbar, die ursprüngliche Zeichenkodierung korrekt festzustellen. Wenn Sie nicht sicher sind, in welcher Zeichenkodierung der Text kodiert ist, lassen Sie sich von Ihrem technischen Ansprechpartner beraten.

### **3.2 PDF und Office-Dokumente (DOC, DOCX, RTF, ODT, PPT, PPTX, ODP, PUB...)**

1. Unter Office-Dokumenten werden alle Dokumente verstanden, die von den gängigen Office-Anwendungen produziert werden. PDF-Dokumente, die noch in keinem archivtauglichen Format vorliegen, zählen auch dazu.
2. Dokumente dieses Typs sind Textdokumente (nicht zu verwechseln mit den unter 3.1. erläuterten Textdaten), Präsentationen, Publikationen und PDF-Dateien. Tabellen werden separat unter 3.3. behandelt.
3. Textdokumente (wie z. B. von Microsoft Word oder OpenOffice Writer) haben typischerweise die Dateierendungen `.doc`, `.docx`, `.odt` oder `.rtf`.
4. Präsentationen (wie z. B. von Microsoft PowerPoint oder OpenOffice Impress) haben typischerweise die Dateierendungen `.ppt`, `.pptx` oder `.odp`.
5. Publikationen (wie z. B. von Microsoft Publisher) haben typischerweise die Endung `.pub`.
6. PDF-Dokumente haben die Endung `.pdf`. Es gibt mehrere Normen und Standards. Archivtauglich sind jedoch nur Dokumente im PDF/A-Standard. Das Staatsarchiv setzt auf die Verwendung von PDF/A-2u für neue Konvertierungen.
7. Alle hier erwähnten Office-Dokumente müssen vor der Ablieferung in PDF/A-2u (d. h. im Konformitätslevel U) konvertiert werden. Dies gilt auch für PDF-Dateien, die noch nicht im PDF/A-Format vorliegen.
8. Existierende PDF/A-1 und andere PDF/A-2 Dokumente werden weiterhin akzeptiert und müssen nicht migriert werden.

### **3.3 Tabellenkalkulationen (XLS, XLSX, ODS, OOXML...)**

1. Momentan liegt für Tabellen kein spezifisches archivtaugliches Format vor, deshalb wird grundsätzlich auf eine Migration verzichtet. Die nachfolgenden Bemerkungen sind aber dennoch zu beachten.
2. Einfache Tabellen (bei denen die visuelle Formatierung keine wesentliche Information darstellt oder Tabellen, die keine Graphiken oder Formeln enthalten) sollen in das CSV-Dateiformat migriert werden (vgl. unter 3.1 Strukturierte Textdaten). Besteht eine Tabellenkalkulation aus mehreren Tabellenblättern, muss jedes Tabellenblatt einzeln migriert werden.
3. Komplexe Tabellen (bei denen die visuelle Formatierung eine wesentliche Information darstellt oder Tabellen, die Graphiken oder Formeln enthalten) müssen nicht migriert werden und können im originalen Dateiformat abgeliefert werden.
4. Werden Tabellen neu gespeichert oder migriert, soll als Zielformat nur OOXML verwendet werden.

### **3.4 Datenbanken**

1. Datenbanken stellen komplexe Datenstrukturen dar, die oft Teil der fachspezifischen elektronischen Systeme der OE sind. Die wesentlichen Informationen befinden sich in diesem Fall meistens nicht in der Datenbank allein, sondern auch als Metadaten, die ausserhalb der Datenbank liegen und von dem elektronischen System verwaltet werden. Daher ist eine langfristige Archivierung nur sinnvoll, wenn diese das gesamte System einschliesst.
2. Das Staatsarchiv setzt deshalb auf eine Migration der Daten aus den Datenbanken und fachspezifischen elektronischen Systemen nach SIARD (E-Government-Standard eCH-0165: SIARD-Formatspezifikation).
3. Im Fall von komplexen elektronischen Systemen ist es im Hinblick auf eine optimale Archivierung das Beste, den jeweiligen Software-Hersteller und den zuständigen technischen Ansprechpartner heranzuziehen. Nur so kann gewährleistet werden, dass sowohl die eigentlichen Daten in der Datenbank als auch alle anderen extern vorhandenen Metadaten gesichert werden.

### **3.5 Bilddaten (TIFF, JPEG, PNG, BMP, SVG...)**

1. Raster-Bilddaten (Rastergrafik) sollen in das TIFF- (in der Version 6, ohne JPEG-Kompression und nicht Multipage) oder in das JPEG2000-Dateiformat konvertiert werden. Das Staatsarchiv zieht das JPEG2000-Dateiformat vor. Bereits vorhandene TIFF- oder JPEG-Dateien sollen nicht migriert werden.
2. Vektor-Bilddaten (Vektorgrafik) sollen in das SVG-Dateiformat migriert werden. Alternativ ist auch eine Migration in PDF/A-2u möglich. Damit gehen aber die Vektoreigenschaften verloren.
3. Bilddaten in JPEG-Format eignen sich nicht für die dauerhafte Archivierung. Allerdings ist es ein sehr verbreitetes Format und oft befinden sich zahlreiche Bilddaten in diesem Format in File-Ablagen. Da aber der Informationsverlust durch die Kompression bereits geschehen und eine Migration daher nicht mehr sinnvoll ist, werden vorhandene Bilddaten auch in diesem Format akzeptiert.
4. Bilddaten aus Papierakten oder sonstigen Textdokumenten (z. B. als Ergebnis einer Retrodigitalisierung) sollen in PDF/A-2u migriert werden, da hier nicht die Bildqualität, sondern die Reduktion von Speicherbelegung und die Benutzbarkeit im Vordergrund stehen.
5. Abhängig vom Inhaltstyp der Ablieferung (z. B. Digitalisierung von Fotobeständen oder Retrodigitalisierung älterer Textdokumente) kann es sinnvoll sein, im Voraus das Staatsarchiv zu kontaktieren, um spezifische Anforderungen der Ablieferung im Detail zu klären.

### **3.6 Audio-Daten (AAC, AIFF, ALAC, FLAC, MP3, OGG, WMA...)**

1. Bisher haben sich noch keine Audio-Daten-Formate, die eine verlustfreie Kompression erlauben, als archivtauglich durchgesetzt.
2. Grundsätzlich sollen Audio-Daten im unkomprimierten WAV-Format (RIFF WAVE, mit Pulse-Code-Modulation (PCM) kodiert, d. h. nicht komprimiert) ans Staatsarchiv abgeliefert werden.
3. Das MP3-Format eignet sich nicht für die Langzeitarchivierung, da es die Audio-Daten mit Qualitätsverlust komprimiert. Da es sehr verbreitet ist, werden viele Audio-Dateien in diesem Format vorhanden sein. Eine Migration in das WAV-Format ist nicht sinnvoll, da der Verlust bereits stattgefunden hat. Daher werden vorhandene Audio-Daten im MP3-Format akzeptiert.
4. Audio-Daten in allen anderen Audio-Formaten müssen ins WAV-Format (RIFF WAVE) migriert werden.

### **3.7 Video-Daten (AVI, MOV, WMV, MPEG, M4V...)**

1. Verschiedene Gründe haben dazu geführt, dass sich aus den vielen heute verwendeten Formaten noch kein Video-Daten-Format für die Langzeitarchivierung durchgesetzt hat:
  - a) Video-Daten können je nach Format und Kompressionsgrad sehr viel Speicherplatz belegen. So kann z. B. eine Stunde Videomaterial mehr als 1TB gross sein.
  - b) Es ist nicht immer die bestmögliche Qualität erforderlich. So sind z. B. die Anforderungen an Werbevideos und an Videos aus einer Kunstaussstellung unterschiedlich.
2. Der Qualitätsverlust bei bereits verlustbehaftet komprimierten Video-Daten-Formaten kann nicht mehr rückgängig gemacht werden. Allenfalls stellt sich die Frage, ob eine stärkere Komprimierung im konkreten Fall sinnvoll ist.
3. Liegen mehrere Video-Datei-Formate derselben Video-Daten vor, soll das ursprüngliche Format übernommen werden, da es die wenigsten Transformationen erfahren hat. Dies ist vom Zusammenspiel zwischen Speicherbedarf und erforderlicher Video-Qualität abhängig; eventuell ist eine Migration in ein stärker komprimiertes Format sinnvoll.
4. Im Hinblick auf eine Migration von Video-Daten soll vorgängig das Staatsarchiv bezüglich archivtauglicher Formate kontaktiert werden.

### **3.8 Spezialfall: E-Mails**

1. Beilagen in E-Mails müssen ausgegliedert gespeichert und entsprechend den Anweisungen dieses Merkblatts in archivtaugliche Formate migriert werden.
2. E-Mails, welche Bilder, Überschriften, Tabellen, formatierte Listen oder formatierten Text enthalten, müssen in PDF/A-2u migriert werden.

### **4. Nutzungseinschränkungen**

Zur Gewährleistung der langfristigen Nachvollziehbarkeit und Nutzung müssen die elektronischen Unterlagen ans Staatsarchiv unverschlüsselt, ohne Nutzungseinschränkungen (inkl. DRM) und ohne Passwortschutz abgeliefert werden.

### **5. Rollen des Staatsarchivs und des technischen Ansprechpartners**

Das Staatsarchiv berät die OE bezüglich der archivtauglichen Formate und der Migration von elektronischen Text-, Bild-, Audio- und Videodaten. Für die Umsetzung ist die OE zusammen mit ihrem technischen Ansprechpartner verantwortlich. Dieser unterstützt Sie bei den folgenden Prozessschritten der Migration:

- a) Analyse der elektronischen Systeme in Bezug auf die erforderliche Funktionalität,
- b) Entscheidung über den notwendigen Einsatz von Drittapplikationen und deren Installation,
- c) Erstellung von Prozessen und Dokumentationen, die eine weitgehende Automatisierung der Konvertierung ermöglichen.

Andrea Voellmin  
Leiterin Bibliothek und Archiv Aargau