

ARCHIVTAUGLICHE DATEIFORMATE

STANDARDS FÜR DIE ARCHIVIERUNG DIGITALER UNTERLAGEN

Verantwortliche Stelle: Schweizerisches Bundesarchiv
Ressort Innovation und Erhaltung

Stand: Juli 2007

1 EINLEITUNG

Unterlagen, welche mit Hilfe von Informations- und Kommunikationstechnologien empfangen oder erstellt wurden, werden im Folgenden als „digitale Unterlagen“ und ihre Aufbewahrung auf unbestimmte Zeit im Bundesarchiv als „digitale Archivierung“ bezeichnet.

Die anbietepflichtige Stelle sorgt dafür, dass die Unterlagen so aufbereitet sind, dass sie ohne zusätzlichen Aufwand im Hinblick auf ihre Archivwürdigkeit bewertet und gegebenenfalls archiviert werden können.¹

Die für die Ablieferung verwendeten Formate und Materialien müssen archivtauglich sein und den in Anhang 2 aufgeführten Merkmalen entsprechen. Ist dies nicht der Fall, lässt die abliefernde Stelle die Informationen auf geeignetes Trägermaterial und in geeignete Formate kopieren sowie in geeignete Behältnisse überführen.²

Das Bundesarchiv setzt die *Dateiformate* fest, die als archivtauglich gelten und somit den strengen Anforderungen genügen, die eine langfristige Verstehbarkeit garantieren.

Das Bundesarchiv favorisiert die Einschränkung der Anzahl Formate gegenüber der Vielfalt. Eine knappe, überschaubare und vorsichtig ausgewählte Menge an Formaten stellt die langfristige Verstehbarkeit in deutlich höherem Masse sicher als eine schwer kontrollierbare und aufwendig zu pflegende grosse Menge an Formaten, die oft von den gerade aktuellen Versionen der Computer-Applikationen abhängig ist.

Dieses Dokument gibt eine Übersicht der zur Zeit akzeptierten Formate.

¹ VBGA, Art. 5

² Weisungen über die Anbietepflicht und die Ablieferung von Unterlagen an das BAR vom 28. September 1999 (<http://www.bar.admin.ch/dokumentation/00437/index.html?lang=de>)

2 ARCHIVIERUNGSFORMATE

Zur Zeit existieren in Bundesarchiv folgende Standards für die Archivierung digitaler Unterlagen:

Anwendungsbereich	archivtaugliche Formate	Bemerkungen
Text (unstrukturiert)	"Nur Text" ("plain text")	UTF-8 UTF-16 ISO-8859-1 ISO 8859-15 US-ASCII
„Office“ Dokumente	PDF/A	entspricht PDF 1.4 mit Einschränkungen
Tabellen	CSV	comma separated values
Relationale Datenbanken	SIARD RDB DATA	
Rasterbilder	TIFF	
Audio	WAVE	

Allgemeiner Hinweis: Die Liste der akzeptierten Formate ist klein. Geringe Mengen an Dateien müssen in eines dieser Formate umgewandelt werden. Oft sind jedoch grössere Mengen an Dateien bereits in anderen Formaten vorhanden, z.B. PDF, JPEG oder MP3. Das Bundesarchiv bietet in diesen Fällen Beratung zum weiteren Vorgehen an.

Im folgenden werden die Formate kurz erläutert. Die ausführlichen Bundesarchiv-Standards mit Hinweisen und Einschränkungen sind im nächsten Kapitel zu finden.

Text

Geeignet für: einfache, unstrukturierte Textdaten

Eine unstrukturierte Textdatei ist zur Darstellung reinen Textinhaltes geeignet, der mit minimalen Strukturierungsmöglichkeiten auskommt (Zeilen) und ansonsten keine weitere Struktur- bzw. Darstellungsinformation benötigt, d.h., er enthält keine eingebetteten oder sichtbaren Anweisungen zur Darstellung (Fettdruck, eingerückt, farbig, etc.) oder Informationen zur Struktur (Titel, Abschnitt, Unterabschnitt, Inhaltsverzeichnis, etc.).

Es hat sich gezeigt, dass gerade diese in ihrer Art sehr einfachen Dateien sich als am längsten haltbar und verstehbar erweisen.

Beispiele sind einfache E-Mail-Meldungen ("Nur-Text", ohne Anhänge) oder Kurzbeschreibungen (README.TXT).

Folgende Zeichensätze sind erlaubt:

- US-ASCII
- ISO 8859-1 und 8859-15 (Latin-1 und Latin-9)
- Unicode (UTF-8, UTF-16)

Dateien in anderen Zeichensätzen sollen in Unicode, vorzugsweise UTF-8, umgewandelt werden. Besteht Unklarheit bezüglich des Zeichensatzes des Originaltextdokuments, darf die Umwandlung nicht vorgenommen werden.

Anmerkung: Im Gegensatz zu "einfachen" Textdateien enthalten strukturierte Textdateien nebst dem eigentlichen Inhalt Markierungen, die den Inhalt strukturieren bzw. Anweisungen zur Darstellung (Formatierung) enthalten. Die Markierungen selbst sind Textteile aus demselben Zeichensatz, z.B. Tags in XML, Steuercodes in PDF oder das Trennzeichen (häufig ein Komma) bei CSV. Ohne die Zusatzinformation über die Struktur und deren Bedeutung sind derartige Dateien nicht interpretierbar und dürfen nicht als reine Textdatei archiviert werden. Ebenfalls nicht geeignet ist das txt-Format für sehr umfangreiche Dokumente.

PDF/A

Geeignet für: „Office“ Dokumente

Das Portable Document Format (PDF) ist ein offenes (d.h. publiziertes), proprietäres Format zur Beschreibung von Druckseiten der Firma Adobe Systems Inc. Es eignet sich für die Speicherung von Dokumenten, deren Inhalt auf Druckseiten adäquat wiedergegeben werden kann. (Zum Beispiel Word-Dokumente, Excel-Dokumente, ... Für die Wiedergabe von HTML-Dateien eignet es sich hingegen oft nicht, da diese wegen Interaktionen, Scripting und nicht-barrierenfreien Designtricks oft auf Druckseiten nicht adäquat wiedergegeben werden können.)

Das von der ISO (Internationale Standardisierungsorganisation) standardisierte, offene, nicht-proprietäre Format PDF/A (PDF/Archive) basiert auf dem von der Herstellerfirma Adobe Systems Inc. publizierten, proprietären Format PDF 1.4 ergänzt durch folgende Einschränkungen, welche für die Archivtauglichkeit unverzichtbar sind:

- keine Referenzierung externer Daten (d.h. insbesondere: alle im Dokument verwendeten Schriftsätze müssen im Dokument eingebettet sein)
- keine Verschlüsselung und kein Passwortschutz
- kein ausführbarer (Programm-)Code
- keine Multimedia-Inhalte
- einige wenige Metadaten

Der im Jahr 2006 von der ISO in Kraft gesetzte Standard heisst genauer PDF/A-1. Die ISO sieht voraus, dass sie in Zukunft weitere Standardteile (z.B. über empfohlene Benutzung) verabschieden wird.

Von diesem Standard gibt es zwei Konformitätsgrade PDF/A-1a und PDF/A-1b. Jedes Dokument, welches dem strengeren Standard PDF/A-1a genügt, genügt auch dem weniger strengen Standard PDF/A-1b.

Das Bundesarchiv akzeptiert Dokumente, welche zum Standard PDF/A-1b konform sind, als PDF/A-konform. Es empfiehlt aber, wo immer möglich, den Standard PDF/A-1a einzuhalten. Dieser garantiert über die exakte Seitenwiedergabe hinaus auch eine höhere Barrierenfreiheit, zu welcher der Bund aufgrund interner Weisungen verpflichtet ist.

Generell kann jedes Dokument, das gedruckt werden kann, und dessen Druckversion den Inhalt des Dokuments adäquat wiedergibt, für die Archivierung in PDF/A umgewandelt werden.

CSV (Comma Separated Values)

Geeignet für: tabellenartig organisierte Daten

CSV-Daten sind strukturierte Text-Daten. Jede Zeile weist Felder (Kolonnen) auf, die durch ein Komma oder eine anderes ausgewähltes Zeichen getrennt sind. Damit lassen sich Tabellen sehr einfach archivieren.

Applikationen wie Excel oder OpenOffice erlauben das Speichern und Einlesen von CSV-Daten.

Sind die Tabellen umfangreich (über 50'000 Zeilen oder 200 Spalten) oder sind mehrere miteinander verknüpfte Tabellen vorhanden, ist das SIARD-Format als Alternative zu erwägen.

SIARD RDB-DATA

Geeignet für: Relationale Datenbanken

Der SIARD Prozess besteht aus einer Suite von Tools, um Daten aus relationalen Datenbanken zu archivieren.

Das SIARD A0 Process Tool extrahiert aus Relationalen Datenbanksystemen (RDBS) Strukturinformationen und Inhalte und wandelt sie in ein archivierbares Format um. Der IT-Leistungserbringer ist der typische Anwender dieses Tools.

Mit dem SIARD A1 Process Tool beschreibt die abliefernde Stelle die Datenbank, Tabellen und Kolonnen genauer.

Die SIARD Tools sind Java-basierte Applikationen und werden vom Bundesarchiv den abliefernden Stellen zur Verfügung gestellt.

TIFF (Tagged Image File Format)

Geeignet für: Bilddaten (Rasterbilddaten)

Das TIFF-Format speichert Bilder als Rastergrafik (auch Pixelgrafik oder Bitmap genannt), eine tabellenartige Anordnung von Pixel (Bildelemente), denen Farbwerte zugeordnet sind.

Das TIFF-Format ist flexibel und erweiterbar. Deshalb sind zur Sicherung der langfristigen Verstehbarkeit gewisse Einschränkungen erforderlich. Die wichtigsten sind hier kurz erwähnt (s.a. Anhang):

- keine Komprimierung (Ausnahme: Schwarz-weiss Bilder)
- proprietäre Erweiterungen werden vom BAR prinzipiell ignoriert. Dennoch müssen sie sich an den Standard halten

Digitale Bilder in anderen Formaten (z.B. JPEG, GIF) müssen in TIFF umgewandelt werden.

WAVE

Geeignet für: Audiodaten

WAVE-Dateien sind Container für digitale Audiodaten, die vorwiegend PCM (pulse code modulation) kodierte Daten enthalten. Das Bundesarchiv akzeptiert PCM-kodierte Audiodaten und empfiehlt, diese in hoher Qualität bereitzustellen:

Im Anhang sind Hinweise zu Szenarien mit konkreten Vorschlägen.

3 STANDARDS

3.1 Standard für Textdaten (unstrukturiert)

TEXT – unstrukturierte "Nur-Text"-Daten

0000	FORMAT-REFERENZ BUNDESARCHIV
	CH-BAR-ARCHIVE-FORMATSPEC_TEXT-001.01
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV
	1.0 vom 2006-10-01
1000	VERBINDLICHER BEZEICHNER
	TEXT – Text Plain
1010	DATEI-ENDUNG
	<p>Empfohlene Datei-Endung:</p> <p>.txt</p> <p>Textdateien weisen manchmal andere Endungen auf (z.B., .log, .dat, .lst). Insbesondere ist dies bei strukturierten Textdateien der Fall. Es besteht kein Zwang, unstrukturierte Textdateien ggf. in .txt umzubenennen, insbesondere dann nicht, wenn in der Endung Bedeutung steckt und einen Hinweis auf deren ursprünglichen Verwendungszweck ergibt.</p>
1020	VERBINDLICHE SPEZIFIKATIONEN
	<p>Erlaubte Zeichensatzkodierungen für "Nur-Text"-Dateien:</p> <ul style="list-style-type: none"> • ISO Latin-1 (ISO 8859-1) und ISO Latin-9 (ISO 8859-15) • Unicode 5.0 Universal Character Set (UCS) (ISO 10646:2003) • US-ASCII (ANSI X3.4-1986) bzw. US-ASCII "safe characters"
1030	VERBINDLICHE SPEZIFIKATIONEN
	<p>ISO Latin-1 (ISO 8859-1) und ISO Latin-9 (ISO 8859-15)</p> <p>Der ISO 8859 Standard ist eine Gruppe von 15 Zeichensatzkodierungen für verschiedene Alphabete</p> <p>Bezugsquelle: International Organization for Standardization, ISO/IEC 8859-1 "Information technology -- 8-bit single-byte coded graphic character sets -- Part 1: Latin alphabet No. 1" (http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=28245&ICS1=35&ICS2=40&ICS3=)</p> <p>Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein</p>
1040	VERBINDLICHE SPEZIFIKATIONEN
	<p>Unicode</p> <p>Unicode ist ein internationaler Standard, in dem langfristig für jedes sinntragende Zeichen bzw. Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt wird.</p> <p>In Unicode finden Zeichen der wichtigsten ISO-Zeichensätze wie die ISO-Normen der Serie 8859 eine 1:1-Entsprechung (das bedeutet, dass bei einer Konvertierung von ISO zu Unicode und zurück das gleiche Ergebnis herauskommt). Heute erledigen die meisten Webbrowser die Darstellung dieser Zeichensätze mit einer Unicode-kodierten Schrift in der Regel perfekt und vom Benutzer unbemerkt.</p> <p>ISO 10646 ist die von ISO verwendete praktisch bedeutungsgleiche Bezeichnung des Unicode-Zeichensatzes; er wird dort als Universal Character Set (UCS) bezeichnet.</p> <p>Die folgenden Kodierungen von Unicode sind erlaubt:</p> <ul style="list-style-type: none"> • UTF-8 (eine 8-bit Kodierung mit variabler Länge, bietet eine maximale Kompatibilität mit US-ASCII) • UTF-16 (eine 16-bit Kodierung mit variabler Länge) <p>UCS-2, UTF-7 gelten als veraltet und sollten nicht mehr verwendet werden.</p>

Bezugsquellen: ISO 10646:2003 unter [Publically Available Standards](#)
(http://isotc.iso.org/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm)

Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein

1050 VERBINDLICHE SPEZIFIKATIONEN

US-ASCII (ANSI X3.4-1986) bzw. "safe characters"

Grundsätzlich ist der US-ASCII Zeichensatz gemäss ANSI X3.4-1986 Standard bzw. ISO/IEC 646-US oder ISO/IEC 646:1991-IRV (international reference version) erlaubt. Alle anderen Bezeichnungen "ASCII" sind nicht archivtauglich.

Aufgrund "nationaler Varianten" (z.B. in den ISO/IEC 646 Standards) sind einige Zeichen des US-ASCII Zeichensatzes @ [\] { | } unsicher und können im internationalen Austausch von Textdaten falsch übertragen bzw. falsch interpretiert werden.

Es wird empfohlen, nur Zeichencodes zu verwenden, die als sicher gelten, die sogenannten "safe characters". Neben den alphabetischen Zeichen ("A" bis "Z" und "a" bis "z") den Ziffern ("0" bis "9"), dem Leerschlag (" "), sind nur die folgenden Zeichen als sicher zu betrachten:

`! " % & ' () * + , - . / : ; < = > ?`

Bezugsquellen:

[ASCII](#) - From Wikipedia, the free encyclopedia (<http://en.wikipedia.org/wiki/ASCII>)

[ISO/IEC 646:1991](#) Information technology -- ISO 7-bit coded character set for information interchange, IRV international reference version

(<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=4777&ICS1=35&ICS2=40&ICS3=>)

Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein

1100 EIGENTÜMER

ISO/IEC Standards: [International Organization for Standardization](#) (<http://www.iso.org>)

ANSI Standards: [American National Standards Institute](#) (<http://www.ansi.org>)

2000 TYP

Dateiformat und Datenformat

2010 ANWENDUNGSGEBIETE

Textdaten (einfache Texte, unstrukturiert)

Eine unstrukturierte Textdatei ist zur Darstellung reinen Textinhaltes geeignet, der mit minimalen Strukturierungsmöglichkeiten auskommt (Zeilen) und ansonsten keine weitere Struktur- bzw. Darstellungsinformation benötigt, d.h., er enthält keine eingebetteten oder sichtbare Anweisungen zur Darstellung (Fettdruck, eingerückt, farbig, etc.) oder Informationen zur Struktur (Titel, Abschnitt, Unterabschnitt, Inhaltsverzeichnis, etc.).

Beispiele sind einfache E-Mail-Meldungen ("Nur-Text", ohne Anhänge) oder Kurzbeschreibungen (LIESMICH.TXT, README.TXT).

Sind die Textdaten strukturiert, sind andere Standards zu verwenden:

- für tabellenartige Struktur: CSV
- für Office-Dokumente: PDF/A

5000 VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG

Unstrukturierte "Nur-Text"-Daten dürfen keine Steuerzeichen enthalten ausser der Zeilen- (LF, CR) und Seitenumbrüche (FF) sowie dem Tabulatorzeichen (TAB).

5010 VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG

Das Null-Zeichen (NUL) darf nicht vorkommen.

7000 EMPFEHLUNGEN DES BUNDESARCHIVS ZUR HERSTELLUNG DIESES ARCHIVFORMATS

Dateien in anderen Zeichensätzen als die unter (1020) aufgeführten sollen in Unicode, vorzugsweise UTF-8, umgewandelt werden. Besteht Unklarheit bezüglich des Zeichensatzes des Originaltextdokuments, darf die Umwandlung nicht vorgenommen werden.

9900 BEMERKUNGEN UND HINWEISE

Liegen grössere Mengen zu archivierender Text-Dateien mit unbekannter Zeichenkodierung, ist vorgängig das Bundesarchiv zu konsultieren.

3.2 Standard für druckbare Dokumente (strukturiert)

PDF/A – Portable Document Format / Archive	
0000	FORMAT-REFERENZ BUNDESARCHIV
	CH-BAR-ARCHIVE-FORMATSPEC_PDF/A-001.01
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV
	1.0 vom 2007-07-02
1000	VERBINDLICHER BEZEICHNER
	PDF/A – PDF/Archive
1010	DATEI-ENDUNG
	<p>Empfohlene Datei-Endung:</p> <p>.pdfa, .pdf</p> <p>PDF-Dokumente gemäss PDF 1.4 mit Endung .pdf genügen oft den Einschränkungen von PDF/A. Entscheidend für die Konformität zu PDF/A ist die Validierung des Formats, nicht die Datei-Endung.</p>
1020	VERBINDLICHE SPEZIFIKATIONEN
	<p>ISO PDF/A-1</p> <p>ISO 19005-1:2005 Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)</p> <p>Dieser ISO-Standard ist eine Einschränkung von PDF 1.4. Die ISO plant, ihm weitere Standards zur Benutzung etc. folgen zu lassen.</p> <p>Bezugsquelle: International Organization for Standardization, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=38920&ICS1=37&ICS2=100&ICS3=99</p> <p>Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein</p>
1030	VERBINDLICHE SPEZIFIKATIONEN
	<p>ISO PDF/A-1 Korrigendum</p> <p>ISO 19005-1:2005/Cor.1:2007 Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1); TECHNICAL CORRIGENDUM 1</p> <p>Korrigendum zu PDF/A-1</p> <p>Bezugsquelle: International Organization for Standardization, http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=45613&ICS1=37&ICS2=100&ICS3=99</p> <p>Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein</p>
1040	VERBINDLICHE SPEZIFIKATIONEN
	<p>PDF 1.4</p> <p>PDF Reference third edition, Adobe Portable Document Format, Version 1.4, Addison Wesley, 2001, ISBN 0-201-75839-3</p> <p>PDF 1.4 Spezifikation der Version 1.4 der Seitenbeschreibungssprache PDF (Portable Document Format) von Adobe Systems Inc. auch unter http://www.adobe.com/devnet/pdf/pdf_reference.html.</p>
1050	VERBINDLICHE SPEZIFIKATIONEN
	<p>XMP</p> <p>XMP Specification: XMP Adding Intelligence to Media, Adobe Systems Inc., 2004</p> <p>XMP Spezifikation der „Extensible Metadata Platform“. Siehe: http://partners.adobe.com/public/developer/xmp/sdk/index.html</p>
1100	EIGENTÜMER

	ISO/IEC Standards: International Organization for Standardization (http://www.iso.org) Adobe Systems Inc. Standards: Adobe Systems Inc. (http://www.adobe.com)
2000	TYP
	Dateiformat und Datenformat
2010	ANWENDUNGSGEBIETE
	<p>Druckbare „Office“ Dokumente.</p> <p>Eine Datei wird dann mit Vorteil im PDF/A-Format archiviert, wenn ihre Druckversion ihren Inhalt adäquat wiedergibt. Der Informationsgehalt der PDF-Version entspricht dem der Druckversion.</p> <p>Da Zeichen und Wörter im PDF-Format als Zeichen und nicht als Bildpunkte gespeichert sind, ist das PDF/A-Format dem TIFF-Format als Seitendarstellung vorzuziehen, wo immer es textuelle Zeicheninformation enthält und nicht ein einfacher Seiten-Wrapper für eine Bilddatei ist.</p>
5000	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
	PDF/A-Daten müssen mindestens als PDF/A-1b validiert sein.
7000	EMPFEHLUNGEN DES BUNDESARCHIVS ZUR HERSTELLUNG DIESES ARCHIVFORMATS
	Für die Verwandlung von Word- und Excel-Dateien in PDF/A sind die entsprechenden Plug-Ins von Adobe zu verwenden. Der IT-Betrieb der Bundesverwaltung ist dazu anzuhalten, diese Plug-Ins mit der Standard-Einstellung für die Erzeugung von PDF/A zur Verfügung zu stellen.
9900	BEMERKUNGEN UND HINWEISE
	Dokumente im PDF/A-Format sind schwer zu verändern. (Auch in diesem Punkt gleichen sie einer ausgedruckten Version.) Die Konversion ins PDF/A-Format ist dann vorzunehmen, wenn das Dokument nicht mehr geändert werden muss. Die Konversion ist möglichst früh vom Autor des Dokuments vorzunehmen (und nicht erst zehn Jahre später durch das Bundesarchiv), weil einzig der Autor beurteilen kann, ob die Druckversion das Dokument adäquat wiedergibt.
	<i>Ende der Archivformat-Spezifikation CH-BAR-ARCHIVE-FORMATSPEC_PDFA-001.01</i>

3.3 Standard für CSV (comma separated values)

CSV – comma separated values

0000	FORMAT-REFERENZ BUNDESARCHIV
CH-BAR-ARCHIVE-FORMATSPEC_CSV-001.01	
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV
2.0.1 vom 2006-12-04	
1000	VERBINDLICHER BEZEICHNER
CSV – Comma Separated Values	
1010	DATEI-ENDUNG
Verbindliche Datei-Endung: .csv	
1020	VERBINDLICHE SPEZIFIKATIONEN
RFC 4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files	
Bezugsquellen: RFC4180 - Shafranovich, Y., "Common Format and MIME Type for Comma-Separated Values (CSV) Files", RFC 4180, October 2005. http://www.ietf.org/rfc/rfc4180.txt	
1030	VERBINDLICHE SPEZIFIKATIONEN
Erlaubte Zeichensatzkodierungen für CSV-Dateien: <ul style="list-style-type: none">• ISO Latin-1 (ISO 8859-1) und ISO Latin-9 (ISO 8859-15)• Unicode 5.0 Universal Character Set (UCS) (ISO 10646:2003)• US-ASCII "safe characters" (ANSI X3.4-1986) Genaue Hinweise zu den genannten Zeichensatzkodierungen sind im Standard für "Nur-Text"-Daten aufgeführt.	
1040	VERBINDLICHE SPEZIFIKATIONEN
Obwohl verschiedene Spezifikationen und Implementationen für CSV existieren, gibt es keinen formalen Standard, der alle existierenden Varianten von CSV-Dateien handhaben kann. Das Memo RFC 4180 beschreibt das Format, welches von den meisten Implementationen verstanden wird. Dieses ist die Basis des BAR-Standards.	
Die Zeilen der Tabelle werden im folgenden "Records" und die Kolonnen "Felder" genannt. In einer CSV-Datei sind die Felder durch ein Trennzeichen (meist ein Komma) getrennt.	
1. Jeder Record ist grundsätzlich auf einer Zeile (Ausnahme: s. Punkt 6), die durch einen Zeilenumbruch terminiert wird (ASCII LF oder ASCII CRLF oder ASCII CR). Z.B.:	
aaa,bbb,ccc CRLF	
xxx,yyy,zzz CRLF	
2. Der letzte Record im File kann aber muss nicht durch einen Zeilenumbruch abgeschlossen sein:	
aaa,bbb,ccc CRLF	
xxx,yyy,zzz	
3. Empfehlenswert ist die die erste Zeile als Header (Kopfzeile) darzustellen, mit gleichem Format wie die übrigen Zeilen. Im Header stehen die Namen der Felder der Tabelle.	
Das CSV-Format beinhaltet keine Information über die Präsenz eines Headers. Diese muss extern bereitgestellt werden. (z.B. mit dem optionalen "header" Parameter des MIME Typs). Beispiel:	

Feldname1, Feldname2, Feldname3 CRLF

aaa,bbb,ccc CRLF

xxx,yyy,zzz CRLF

4. Innerhalb des Headers und jedes Records stehen mehrere Felder, die durch Kommas voneinander getrennt sind. Jede Zeile sollte die gleiche Anzahl Felder beinhalten. Leerzeichen sind von Bedeutung und dürfen nicht ignoriert werden. Das letzte Feld darf nicht von einem Komma gefolgt sein:

aaa,bbb,cc cc,ddd

5. Jedes Feld kann aber muss nicht durch Anführungszeichen (double quotes) eingefasst sein. Falls Felder nicht mit Anführungszeichen eingefasst sind, dürfen innerhalb der Felder keine Anführungszeichen auftreten:

"aaa",bbb,"ccc" CRLF

xxx,yyy,zzz CRLF

6. Felder, die Zeilenumbrüche (CRLF), Anführungszeichen oder Kommas enthalten, müssen in Anführungszeichen eingefasst sein:

"aaa", "b CRLF

bb", "ccc" CRLF

xxx,yyy, "z, zz" CRLF

Nach Möglichkeit ist dieser Spezialfall zu vermeiden, weil er von vielen Programmen (z.B. Excel, Access) falsch interpretiert wird.

7. Falls ein Anführungszeichen innerhalb eines Feldes auftritt, muss es durch ein vorangehendes markiert sein. Das Feld muss zudem mit Anführungszeichen eingefasst sein:

"aaa", "b" "bb", "ccc" CRLF

1050 VERBINDLICHE SPEZIFIKATIONEN

Zulässige BAR-Erweiterung zum RFC 4180

8. Ein anderes Trennzeichen (Delimiter) als Komma kann gewählt werden. Es gelten analog die obigen Regeln (1011). Folgende Trennzeichen sind gestattet:

<u>Zeichen</u>	<u>ASCII-Code</u>
;	0x3B
	0x7C
#	0x23

Grundsätzlich muss bei Abweichungen, die über Regeln 1.-8. hinausgehen, das Bundesarchiv vorgängig kontaktiert werden.

1100 EIGENTÜMER

RFC Standards: The Internet Engineering Task Force <http://www.ietf.org>

2000 TYP

Dateiformat und Datenformat

2010 ANWENDUNGSGEBIETE

Tabellenartig organisierte Daten.

Excel-Tabellen, wenn der Inhalt der Tabelle von Bedeutung ist (im Gegensatz zur Darstellung).

Einzelne Tabellen aus *kleinen* Datenbanken (MS Access, MySQL, etc.). Falls die Datenbanken umfangreich sind, mehrere Tabellen enthalten, die miteinander in Beziehung stehen, ist SIARD anzuwenden (s. Standard für relationale Datenbanken).

5000 VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG

Zeilen einer CSV-Datei müssen immer die gleiche Anzahl Felder enthalten. CSV-Dateien, die von dieser Regel abweichen, sind nicht erlaubt.

9900 BEMERKUNGEN UND HINWEISE

Besteht bei Excel- bzw. kleineren Datenbanken Unklarheit, ob CSV oder SIARD geeigneter ist, erteilt das Bundesarchiv gerne Auskunft.

Ende der Archivformat-Spezifikation CH-BAR-ARCHIVE-FORMATSPEC_CSV-001.01

3.4 Standard für relationale Datenbanken

SIARD – Software Independent Archiving of Relational Databases

0000	FORMAT-REFERENZ BUNDESARCHIV
	CH-BAR-ARCHIVE-FORMATSPEC_SIARD-001.01
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV
	1.0 vom 2007-02-27
1000	VERBINDLICHER BEZEICHNER
	SIARD RDB DATA – Software Independent Archiving of Relational Databases
1010	DATEI-ENDUNG
	Die SIARD RDB Daten bestehen aus einer Menge von Dateien, die grösstenteils durch die eingesetzten Tools automatisch mit der richtigen Datei-Endung erzeugt werden. Einzig die Dateien der begleitenden Dokumentation (im Ordner "...") müssen sich an die Standards der jeweiligen Dateiformate halten.
1020	VERBINDLICHE SPEZIFIKATIONEN
	<p>Die Suite der SIARD Process Tools definiert de facto den SIARD Format-Standard. Der Einsatz dieser Tools garantiert die korrekte Herstellung der archivtauglichen Struktur für Daten aus einer relationalen Datenbank.</p> <p>User Manual for Process SIARD – A0 „Software Independent Archiving from Relational Databases“, 02.12.2004 User Manual for Process SIARD – A1 „Software Independent Archiving from Relational Databases“, 04.05.2005</p> <p>Bezugsquelle: Schweizerisches Bundesarchiv, 3003 Bern Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Ja</p>
1030	EIGENTÜMER
	©2005 Schweizerisches Bundesarchiv
2000	TYP
	Dateiformat und Datenformat
2010	ANWENDUNGSGEBIETE
	Relationale Datenbanken
5000	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
	Die Dateien der begleitenden Dokumentation einer mit SIARD archivierten relationalen Datenbank müssen in einem vom Bundesarchiv vorgegebenen Standard-Archivformat sein.
5010	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
	Enthalten die Tabellen Kolonnen vom Typ BLOB (binary large object), müssen gegebenenfalls darin abgelegte Dateien in einem vom Bundesarchiv vorgegebenen Standard-Archivformat sein.
5020	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
	Enthalten die Tabellen Kolonnen vom Typ CLOB (character large object), muss die (1) Zeichenkodierung bekannt sein und (2) einer vom Bundesarchiv genehmigten Zeichenkodierung entsprechen (s. Standard für Textdaten, 1020).
7000	EMPFEHLUNGEN DES BUNDESARCHIVS ZUR HERSTELLUNG DIESES ARCHIVFORMATS
	<p>Der Einsatz der SIARD Process Tools ist zwingend vorgeschrieben.</p> <p>Bezugsquelle: Schweizerisches Bundesarchiv.</p>
9900	BEMERKUNGEN UND HINWEISE
	Bei kleinen, einfachen Datenbanken kann anstelle des SIARD-Prozesses auch das CSV-Format eingesetzt werden. Dies gilt z.B. für einzelne wenige Tabellen geringen Umfangs (weniger als 50'000 Records), die

einfach zu verstehende Beziehungen aufweisen.

Besteht bei kleineren Datenbanken Unklarheit, ob SIARD oder CSV geeigneter ist, erteilt das Bundesarchiv weitere Auskünfte.

Ende der Archivformat-Spezifikation CH-BAR-ARCHIVE-FORMATSPEC_SIARD-001.01

3.5 Standard für Rasterbilder

TIFF – TAGGED IMAGE FILE FORMAT

0000	FORMAT-REFERENZ BUNDESARCHIV
CH-BAR-ARCHIVE-FORMATSPEC_IMAGE_001.01	
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV
1.0 vom 2002-01-01	
1000	VERBINDLICHER BEZEICHNER
TIFF – Tagged Image File Format	
1010	DATEI-ENDUNG
Verbindliche Datei-Endungen: .tif, .tiff	
1020	VERBINDLICHE SPEZIFIKATIONEN
Revision 6.0 vom 1992-06-03 und Revision 6.0.1 vom 1995-10-15 Bezugsquelle: http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein	
1030	EIGENTÜMER
©1986-1988, 1992 Adobe Systems Inc., U.S.A.	
2000	TYP
Dateiformat und Datenformat	
2010	ANWENDUNGSGEBIETE
Rasterbilder (Schwarz-Weiss, Graustufen, Farbe)	
5000	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
TIFF-Dateien müssen als TIFF 6 validieren.	
5010	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG
<p> PROPRIETÄRE ERWEITERUNGEN: Format-Erweiterungen durch Software-Hersteller sind nur zulässig, wenn sie sich strikt an die diesbezüglichen Anforderungen des Standards (1020) halten. Im Zweifelsfall sind durch die abliefernde Stelle schriftliche Bestätigungen des Software-Herstellers einzuholen. Insbesondere müssen proprietäre Erweiterungen („private fields and values“) gemäss (1020) oberhalb der Tag-Nummer resp. Konstanten-Nummern 32767 implementiert sein. Achtung: Auch wenn sie konform zum Standard implementiert sind, werden diese Erweiterungen vom Bundesarchiv prinzipiell ignoriert. </p> <p> Insbesondere werden auch die „TIFF Enhancements for Adobe Photoshop®“ ignoriert, was vor allem bedeutet, dass der von Adobe™ Photoshop® geschriebene Image-Source Tag 37724 ignoriert wird. Die Verwendung der zwei weiteren „Advanced TIFF“-Optionen von Adobe™ Photoshop® – ZIP/zlib- und JPEG-Kompressionen – ist ausgeschlossen. Ebenfalls ausdrücklich ignoriert werden die „TIFF Enhancements for Adobe™ PageMaker® 6.0“ und die „Kodak™ TIFF Extensions“. </p> <p> Allgemein: vgl. (1020) </p> <p> Adobe™ Photoshop „TIFF Enhancements für Adobe™ Photoshop®“: http://partners.adobe.com/asn/developer/pdfs/tn/TIFFphotoshop.pdf </p> <p> TIFF Enhancements for Adobe™ PageMaker® 6.0: http://partners.adobe.com/asn/developer/pdfs/tn/TIFFPM6.pdf </p> <p> Kodak™ TIFF Extensions: wenden Sie sich direkt an die Eastman Kodak Company (http://www.kodak.com/) </p> <p> Für andere Software: siehe Dokumentation des Herstellers </p>	

5030 VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG

MULTIPAGE-TIFF: Die Einbindung von mehreren Seiten in eine einzelne TIFF-Datei (Multipage-TIFF, mehrere IFDs) sind nicht erlaubt.

Wenn dasselbe Bild in unterschiedlicher Auflösung oder Bittiefe archiviert werden soll, sind verschiedene Dateien in verschiedenen Ordnern herzustellen, damit die verschiedenen Versionen leicht separat bestellt werden können.

Wenn ein Dokument mit mehreren Seiten archiviert werden soll, ist es im PDF/A-Format abzuspeichern.

9900 BEMERKUNGEN UND HINWEISE

TIFF benutzt 4-Byte File-Offsets, weshalb die Bild-Datenmenge einer TIFF-Datei max. 4 GB (komprimiert) betragen kann.

Liegen grosse Mengen zu archivierender TIFF-Dateien im Format „GeoTIFF“ vor (<http://www.remotesensing.org/geotiff/geotiff.html>), so ist das Bundesarchiv zu konsultieren.

Ende der Archivformat-Spezifikation CH-BAR-ARCHIVE-FORMATSPEC_IMAGE-001.01

3.6 Standard für Audiodaten

WAVE – Audioformat von Microsoft					
0000	FORMAT-REFERENZ BUNDESARCHIV				
CH-BAR-ARCHIVE-FORMATSPEC_WAVE_001.01					
0010	VERSION DIESER FORMAT-REFERENZ BUNDESARCHIV				
1.0 vom 2007-02-27					
1000	VERBINDLICHER BEZEICHNER				
WAVE					
1005	ALIAS BEZEICHNER				
WAVEFORMAT PCMWAVEFORMAT					
1010	DATEI-ENDUNG				
Verbindliche Datei-Endung: .wav					
1020	VERBINDLICHE SPEZIFIKATIONEN				
<p>Es existiert kein publizierter Standard für WAVE Dateien. Das WAVE-Format ist eine Implementierung des Resource Interchange Formats (RIFF) der Microsoft Corporation. Dieses ist als Publikation freigegeben.</p> <p>Bezugsquelle: Multimedia Programming Interface and Data Specifications 1.0, herausgegeben durch IBM Corporation und Microsoft Corporation, August 1991 (http://www.kk.ij4u.or.jp/~kondo/wave/mpidata.txt). Auch erhältlich bei: http://www.tactilemedia.com/info/MCI_Control_Info.html.</p> <p>WAVEFORMAT (structure) from Microsoft. (http://msdn.microsoft.com/library/default.asp?url=/library/en-us/multimed/htm/_win32_waveformat_str.asp)</p> <p>Abgabe der Original-Spezifikation durch das Schweiz. Bundesarchiv: Nein</p>					
1030	EIGENTÜMER				
©1991 Microsoft Corporation					
2000	TYP				
Dateiformat und Datenformat					
2010	ANWENDUNGSGEBIETE				
Audiodaten					
5000	VERBINDLICHE EINSCHRÄNKUNGEN DER ANWENDUNG				
Erlaubt ist nur der default Microsoft Pulse Code Modulation (LPCM) Codec. Byte-Offset 20 (0x16) muss den Wert 1 als 2-Byte Wert (short) enthalten (in der little-endian byte order).					
7000	EMPFEHLUNGEN DES BUNDESARCHIVS ZUR HERSTELLUNG DIESES ARCHIVFORMATS				
<p>Herstellung aus <u>analogen</u> Quellen</p> <p>Folgende Abtastrate und Quantifizierungen sind bei Digitalisierungen aus analogen Quellen empfohlen:</p> <table border="1" style="margin-left: 40px;"> <thead> <tr> <th>Abtastrate</th> <th>Quantifizierung</th> </tr> </thead> <tbody> <tr> <td>48 KHz</td> <td>16 oder 24 bits</td> </tr> </tbody> </table>		Abtastrate	Quantifizierung	48 KHz	16 oder 24 bits
Abtastrate	Quantifizierung				
48 KHz	16 oder 24 bits				

7010 EMPFEHLUNGEN DES BUNDESARCHIVS ZUR HERSTELLUNG DIESES ARCHIVFORMATS

Herstellung aus digitalen Quellen

Liegen die Audiodaten bereits in digitaler Form vor, soll die bestehende Abtastrate und Quantifizierung beibehalten werden. Bei CD Audio und Audio DAT sind dies:

Original	Abtastrate	Quantifizierung
CD Audio	44.1 KHz	16 bits
DAT	44.1 KHz / 48 KHz	16 bits

9900 BEMERKUNGEN UND HINWEISE

Das WAVE-Format ist sehr nahe verwandt mit dem CD-Audio (CDA-)-Format. Dessen Dokumentation ist allerdings äusserst schwer zugänglich. Gängige Programme (Nero etc.) verwandeln CDA <-> WAVE.

Ende der Archivformat-Spezifikation CH-BAR-ARCHIVE-FORMATSPEC_WAVE-001.01