Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Swiss Federal Archives SFA
Information Preservation Division IPD

STANDARDS FOR ARCHIVING DIGITAL DOCUMENTS

— *Archivable File Formats*

Body responsible:	Swiss Federal Archives Information Preservation Division
Date:	April 2020
Version	2020/04, replaces version 2018/04

Change log

Version	Description, comment
06/2007	First version July 2007
01/2014	Expanded to include MPEG-4 (video), full revision
04/2018	Expanded to include PDF/A 2, JPEG 2000, XML/XSD, SIARD 2.1, FFV1, TIFF+EWF.XML, INTERLIS, full revision
04/2020	Expanded to include PDF/A-2b

Table of Contents

1 Introduction 3
1.1 Summary of the most important changes 3
2 Areas of application 4
2.1 Text 4
2.2 Image 4
2.3 Audio 5
2.4 Video 5
2.5 Tables and databases 5
2.6 Geodata 5
2.7 Other 5
2.7.1 XML/XSD 5
2.7.2 Web archiving 5
2.7.3 Other areas of application 6
3 Archivable file formats 7
TEXT – unstructured, plain text data 7
PDF/A – Portable Document Format / Archive 9
CSV – text file with delimiters 11
SIARD – Software Independent Archiving of Relational Databases 13
TIFF – Tagged Image File Format 14
JPEG 2000 – Joint Photographic Experts Group 2000 15
WAVE – audio format from Microsoft 16
FFV1 – video codec 17
MPEG-4 – video format 18
XML/XSD – eXtensible Markup Language 19
TIFF+EWF.XML – Tagged Image File Format and Extended World File 20
INTERLIS 21

1 INTRODUCTION

In this standard, documents received or created using information and communication technologies are referred to as “digital documents” and storing them for an indefinite period in the Federal Archives is referred to as “digital archiving”.

Bodies that are required to submit records must ensure that their documents are ready to be appraised for archival value and, if appropriate, archived without additional effort.[1]

The formats and materials used for submission must be archivable and correspond to the features set out in sections 2 and 3. The Federal Archives define the *file formats* that are regarded as archivable and therefore meet the stringent requirements for guaranteeing long-term interpretability.

The Federal Archives favour concentration on a limited number of formats over variety. A small, manageable and carefully selected collection of formats is much more likely to ensure long-term interpretability than a large number of formats that are difficult to control and time-consuming to maintain and are often dependent on the current versions of computer applications.

This document provides an overview of the currently accepted formats. Details of formats that are currently candidates for inclusion in this list of archivable formats can be found on the Federal Archives website.

1.1 Summary of the most important changes

New formats:

- PDF/A-2b

[1] ArchO, Art. 5

2 AREAS OF APPLICATION

The following standards for archiving digital documents are currently in place at the Federal Archives:

Area of application	**Archivable formats**	**Remarks**
Text (unstructured)	plain text	UTF-8 UTF-16 ISO-8859-1 ISO 8859-15 US-ASCII
“Office” documents	PDF/A	Corresponds to PDF 1.4 (PDF/A-1) and PDF 1.7 (PDF/A-2) with limitations
Tables	CSV	Text file with delimiters, encoding as with text (unstructured)
Relational databases	SIARD	
Raster images	TIFF JPEG 2000	
Audio	WAVE	
Video	MPEG-4 FFV1	
Structured (text) data	XML	UTF-8, UTF-16
Geodata	TIFF + EWF.XML INTERLIS	

The following is a brief explanation of the formats. The detailed Federal Archives standards, with notes and limitations, can be found in the next section.

2.1 Text

In the majority of cases, it is the (text) content of text data that matters. The term “text content” covers not only conventional text documents but also presentations and websites. Such content is often enhanced using formatting and images. The PDF/A format is suitable for archiving text content of this kind.

In addition to these “Office” text documents there is unstructured text. This does not contain any embedded or visible display instructions (bold, indented, in colour, etc.) or information on structure (title, section, sub-section, table of contents, etc.). Experience has shown that very simple files of this kind can be preserved and interpreted for the longest time. Examples include simple e-mails (plain text, without attachments), logfiles and short descriptions (README.TXT). The “plain text” format is suitable for archiving this content.

2.2 Image

A distinction is made between raster images and vector images. Raster images have a defined size and are essentially a collection of individual pixels. Vector images save the image data in the form of vectors and are therefore freely scalable. Vector images are especially suitable for diagrams and the like, but are entirely unsuited to photographs. As the occurrence of vector data in the Federal Administration has so far been limited, there has not yet been any need to archive them in a corresponding format. If a body submitting records needs to archive vector data, it is therefore requested to contact the Federal Archives in order to agree on how to proceed.

The Federal Archives offer two accepted formats for raster images: TIFF and JPEG2000. TIFF is the older and more widely used format. JPEG2000 is a more recent format that addresses some of the weaknesses of TIFF. For example, it permits the use of high-quality lossy compression. The Federal Archives recommend its use only in limited cases, but in some instances it may make very good sense. It also enables much better compression of images with little actual content, such as architects' plans that mainly consist of a few lines on a large, empty background.

2.3 Audio

In order to ensure high-quality preservation of audio data, the SFA currently accept only the WAVE format.

2.4 Video

Video archiving can quickly generate vast amounts of data. Consideration should therefore first be given to which video content needs to be preserved in high quality.

The FFV1 format is available for videos that are to be archived in high quality. It was developed specially for archiving. For optimum results, the SFA recommend generating the FFV1 file direct from the original (raw) video data.

For larger quantities of videos where image quality is secondary or that are already in MPEG-4 format, the latter can be used or retained. However, MPEG-4 is always a compromise between quality and preservability, and is explicitly recommended for use by the SFA only subject to those conditions.

2.5 Tables and databases

Where tables and databases are concerned, the complexity of the data model may dictate the choice of format; equally, technical considerations may weigh in favour of or against the use of a particular format.

For small numbers of simple tables without relations, the CSV format should be used, taking account of the encoding.

As soon as a large number of related tables need to be archived, the SIARD format is recommended.

Since large and complex tables and databases are never actually self-documenting, additional documentation is normally required, which also has to be archived.

2.6 Geodata

As with images, in geodata a distinction is made between raster data and vector data. The TIFF+EWF.XML format is used to archive geodata that are in the form of raster data. It consists of a TIFF image file and an accompanying EWF.XML file (Extended World File). The TIFF complies with the TIFF image standard while the EWF.XML file corresponds to an XML file in accordance with the EWF Schema Definition. The INTERLIS format is used to archive geodata that are in the form of vector data. Other archivable formats that are used to archive geodata include XML, SIARD, CSV and PDF.

Additional documentation on geodata archiving can be obtained from the Federal Archives.

2.7 Other

2.7.1 XML/XSD

The XML format can be used for structured (text) data. The most frequent case in which data are archived in XML is specialist applications that already hold certain data stored as XML or offer the option to export that data as XML.

2.7.2 Web archiving

Before web archiving, it is necessary to establish which web content is of archival value, so that this specific content can be preserved. For Federal Administration websites, this is mostly defined by the content (text) and not the form (layout/design). For this reason, the PDF/A format can often be used for archiving. Websites containing a large

amount of logic should be treated in the same way as specialist applications, so that for example SIARD and XML may be appropriate formats.

2.7.3 Other areas of application

If a submitting body has other areas of application that are not covered in this document, it should contact the Federal Archives with a view to finding the best possible solution for archiving the documents concerned.

3 ARCHIVABLE FILE FORMATS

TEXT – unstructured, plain text data

FORMAT

MANDATORY IDENTIFIER	TEXT – Text Plain
PRONOM PUIDs	x-fmt/16, x-fmt/21, x-fmt/22, x-fmt/62, x-fmt/111, x-fmt/282, x-fmt/283
FILE EXTENSION	Recommended file extension: .txt[2]
TYPE	File format and data format

BINDING SPECIFICATIONS

Permitted character encodings for plain text files:

- ISO Latin-1 (ISO 8859-1) and ISO Latin-9 (ISO 8859-15)
- Unicode 5.0 Universal Character Set (UCS) (ISO 10646:2003)
- US-ASCII (ANSI X3.4-1986) and US-ASCII safe characters

ISO Latin-1 (ISO 8859-1) and ISO Latin-9 (ISO 8859-15)

The ISO 8859 standard is a group of 15 character encodings for various alphabets.

Source: International Organization for Standardization, ISO/IEC 8859-1 "Information technology -- 8-bit single-byte coded graphic character sets -- Part 1: Latin alphabet No. 1"

Unicode

Unicode is an international standard in which a consistent digital code is defined for every meaningful character or text element of all known writing and character systems.

Unicode contains a 1:1 match for characters in the most important ISO character sets (e.g. those of the ISO 8859 standard series). This means that the same result occurs when converting from ISO to Unicode and back again. Today, most web browsers normally display these character sets perfectly, without the user noticing, using a Unicode-encoded font.

The Unicode character set is also referred to as the Universal Character Set (UCS), a term that comes from ISO standard 10646.

The following Unicode encodings are permitted:

UTF-8 (an 8-bit encoding of variable length, offers maximum compatibility with US-ASCII)

UTF-16 (a 16-bit encoding of variable length)

UCS-2, UTF-7 are regarded as outdated and should no longer be used.

Sources: ISO 10646:2003 at Publicly Available Standards

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

In principle, the US-ASCII character set as defined in the ANSI X3.4-1986 standard and ISO/IEC 646-US or ISO/IEC 646:1991-IRV (international reference version) is permitted. All other "ASCII" designations are not archivable.

Owing to "national variants" (e.g. in the ISO/IEC 646 standards), some characters of the US-ASCII character set (e.g. @ [\ ] { | }) are unsafe and may be transferred or interpreted incorrectly when exchanging text data internationally.

It is advisable to use only safe characters. In addition to the alphabetical characters ("A" to "Z" and "a" to "z"), numbers ("0" to "9") and the space (" "), only the following characters should be regarded as safe:

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

Sources:

Wikipedia, the free encyclopedia: ASCII

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

OWNERS

[2] Text files – especially structured text files – sometimes have other endings (e.g. .log, .dat, .lst). It is not compulsory to rename them as .txt, especially if the ending contains a meaning and an indication of their original use.

	ISO/IEC Standards: International Organization for Standardization ANSI Standards: American National Standards Institute
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**
	Text data (simple texts, unstructured) An unstructured text file is suitable for displaying pure text content with minimal structuring (lines) that does not require any other structural or display information, i.e. does not contain embedded or visible display instructions (bold, indented, colour, etc.) or information on its structure (title, section, sub-section, table of contents, etc.). Examples include simple e-mails (plain text, without attachments), logfiles and short descriptions (README.TXT). If the text data are structured, other standards such as CSV, PDF/A or XML should be used.
	BINDING RESTRICTIONS ON USE
	Unstructured, plain text data must not contain any control characters except line breaks (LF, CR) and page breaks (FF) as well as the tab character (TAB).
	The null character (NUL) must not be used**.**
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**
	Files that are in character sets other than those listed above should be converted into Unicode, preferably UTF-8. In case of doubt regarding the character set of the original text document, no conversion may be carried out.
	NOTES AND COMMENTS
	Where substantial quantities of text files with unknown character encoding are to be archived, the Federal Archives should be consulted in advance.

PDF/A – Portable Document Format / Archive

FORMAT	**MANDATORY IDENTIFIER**	PDF/A – PDF/Archive
	PRONOM PUIDs	fmt/95 (PDF/A-1a), fmt/354 (PDF/A-1b), fmt/476 (PDF/A-2a), fmt/477 (PDF/A-2b), fmt/478 (PDF/A-2u)
	FILE EXTENSION	Recommended file extensions: .pdfa, .pdf
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	ISO PDF/A-1 ISO 19005-1:2005 Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1) This ISO standard is a limitation of PDF 1.4.	
	ISO PDF/A-1 Corrigendum ISO 19005-1:2005/Cor.2:2011 Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1); TECHNICAL CORRIGENDUM 2 Source: International Organization for Standardization	
	ISO PDF/A-2 ISO 19005-2:2011 Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2) This ISO standard is a limitation of PDF 1.7. Source: International Organization for Standardization	
	PDF 1.4 PDF Reference third edition, Adobe Portable Document Format, Version 1.4, Addison Wesley, 2001, ISBN 0-201-75839-3 PDF 1.4 Specification of Version 1.4 of the PDF (Portable Document Format) page description language from Adobe Systems Inc. at http://www.adobe.com/devnet/pdf/pdf_reference.html.	
	PDF 1.7 ISO 32000-1:2008 Document management – Portable document format – Part 1 : PDF 1.7	
	XMP XMP Specification: XMP Adding Intelligence to Media, Adobe Systems Inc., 2004 XMP Specification of the Extensible Metadata Platform. See: http://www.adobe.com/devnet/xmp.html ISO standard: https://www.iso.org/standard/57421.html	
	OWNERS	
	ISO/IEC Standards: International Organization for Standardization Adobe Systems Inc. Standards: Adobe Systems Inc.	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	Printable "Office" documents. A file is appropriate for archiving in PDF/A format if a printed version adequately reproduces its content, i.e. the information content of the PDF version corresponds to that of the printed version. With PDF/A-2, embedded files are also permitted, provided they are compatible with PDF/A-1 or PDF/A-2. In such cases, they can only be visually reproduced correctly on a computer screen, as these characteristics cannot be printed. Since characters and words are stored as characters rather than pixels in the PDF format, the PDF/A format is preferable to the TIFF format for displaying pages wherever it contains textual character information and is not a simple page wrapper for an image file.	
	BINDING RESTRICTIONS ON USE	

	PDF/A-1 data must as a minimum be validated as PDF/A-1b. PDF/A-2 data must as a minimum be validated as PDF/A-2b.
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**
	Many PDF/A converters are available. In some cases they differ greatly in their orientation (e.g. focus on automated conversion or manual conversion, source formats supported, system requirements, etc.), so it is impossible to give a general recommendation. KOST compiles a periodic study of the available PDF/A converters.
	NOTES AND COMMENTS
	Documents in PDF/A format are difficult to change (in this respect, they are also like a printed version). Conversion to PDF/A format should be undertaken when the document no longer needs to be changed. It should be done as early as possible by the author, because the author is the only person who can judge whether the printed version reproduces the document adequately.
	The following is a list of the main changes in PDF/A-2 compared with PDF/A-1: • JPEG 2000 compression: In PDF/A-2, JPEG 2000 compression is permitted, allowing for better image quality compared with the file size. • Transparency: Transparency is permitted in PDF/A-2. It is prohibited in PDF/A-1, which can lead to problems when converting. PDF/A-2 thus also allows the use of watermarks. • Embedded files: PDF/A-2 allows embedding of files, provided they are compatible with the PDF/A-1 or PDF/A-2 standard. • File size: PDF/A-2 files may be more than 10 GB in size, larger than PDF/A-1. • "Optional content": PDF/A-2 has the option of displaying or hiding layers, which can change the way the document is displayed. This is often used in Computer Aided Design (CAD).

CSV – text file with delimiters

FORMAT

MANDATORY IDENTIFIER	CSV – text file with delimiters
PRONOM PUIDs	fmt/18
FILE EXTENSION	Mandatory file extension: .csv
TYPE	File format and data format

BINDING SPECIFICATIONS

RFC 4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files

Sources: 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

Permitted character encodings for CSV files:

- ISO Latin-1 (ISO 8859-1) and ISO Latin-9 (ISO 8859-15)
- Unicode 5.0 Universal Character Set (UCS) (ISO 10646:2003)
- US-ASCII safe characters (ANSI X3.4-1986)

Precise guidance on the character encodings can be found in the standard for plain text data.

Although there are various specifications and implementations for CSV, there is no formal standard that can handle all existing variants of CSV files. Memo RFC 4180 describes the format that is understood by most implementations. It forms the basis for the SFA standard.

In what follows, the lines in the table are referred to as "records" and the columns as "fields". In a CSV file the fields are separated by a delimiter (mostly a comma).

1. Each record is normally on one line (exception, see point 6) which is ended by a line break (ASCII LF or ASCII CRLF or ASCII CR), e.g.:

   aaa,bbb,ccc CRLF

   xxx,yyy,zzz CRLF

2. The last record can (but does not have to) end with a line break:

   aaa,bbb,ccc CRLF

   xxx,yyy,zzz

3. It is advisable to display the first line as a header in the same format as the remaining lines. The header contains the names of the fields in the table.
   The CSV format does not contain any information about the presence of a header. This must be supplied externally. (e.g. using the optional "header" parameter of the MIME type).
   Example:

   Fieldname1,Fieldname2,Fieldname3 CRLF

   aaa,bbb,ccc CRLF

   xxx,yyy,zzz CRLF

4. Within each header and each record are a number of fields that are separated by commas. Each line must contain the same number of fields. Spaces are significant and must not be ignored. The last field must not be followed by a comma:

   aaa,bbb,cc cc,ddd

5. Each field can be (but does not have to be) enclosed within double quotes. If fields are not enclosed within double quotes, no double quotes may appear within the fields:

   "aaa",bbb,"ccc" CRLF

   xxx,yyy,zzz CRLF

6. Fields that contain line breaks (CRLF), double quotes or commas must be enclosed within double quotes:

   "aaa","b CRLF

   bb","ccc" CRLF

   xxx,yyy,"z,zz" CRLF

   Where possible, this special case should be avoided, as many programs (e.g. Excel, Access) interpret it incorrectly.

7. If a double quote appears within a field, it must be marked by a preceding double quote. The field must also be enclosed within double quotes:

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

Permitted SFA extension to RFC 4180

8. A delimiter other than the comma may be used. The rules above apply, subject to the necessary changes. The following delimiters are permitted:

Character	ASCII code
,	0x2C
;	0x3B
|	0x7C
#	0x23

The Federal Archives must normally be contacted in advance in the case of deviations that go beyond rules 1.-8.

OWNERS

RFC standards: The Internet Engineering Task Force http://www.ietf.org

RULES AND RESTRICTIONS

AREAS OF APPLICATION

Data organised in the form of tables.

Excel tables if the content of the table is important (as opposed to how it is displayed).

Individual tables from small databases (MS Access, MySQL, etc.). If the databases are large or contain a number of related tables, SIARD should be used (see standard for relational databases).

BINDING RESTRICTIONS ON USE

Lines in a CSV file must always contain the same number of fields. CSV files that deviate from this rule are not permitted.

RECOMMENDATIONS AND NOTES

FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT

-

NOTES AND COMMENTS

If it is unclear whether CSV or SIARD is more suitable for Excel or smaller databases, the SFA will be happy to offer guidance.

SIARD – Software Independent Archiving of Relational Databases

FORMAT	**MANDATORY IDENTIFIER**	SIARD format
	PRONOM PUIDs	fmt/161 (SIARD 1.0), fmt/XXX[3] (SIARD 2.1)
	FILE EXTENSION	Mandatory file extension: .siard
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	eCH-0165: SIARD Format Specification: Version 2.1 Source: https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html or https://kost-ceco.ch/cms/index.php?siard_de	
	eCH-0165: SIARD Format Specification: Version 1.0 Source: https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0165&documentVersion=1.0	
	OWNERS	
	©2005-2018 Swiss Federal Archives	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	Relational databases	
	BINDING RESTRICTIONS ON USE	
	If the tables contain columns of the BLOB (binary large object) type, files stored within them must also comply with the archivable formats prescribed by the SFA.	
	Version 2.0 is not accepted by the Federal Archives.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	The use of SIARD Suite is recommended for archiving relational databases. Source: Swiss Federal Archives. https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html	
	The csv2siard tool allows CSV files to be converted into SIARD files. Source: Coordination centre for the long-term archiving of electronic documents (KOST). http://kost-ceco.ch/cms/index.php?csv2siard_de	
	NOTES AND COMMENTS	
	A SIARD file is not normally sufficiently self-documenting to guarantee interpretability of the archived data. Documentation on the source system and data model should therefore be archived in addition to the SIARD file. In each case, a decision must be taken on what has to be documented. The documentation may include (examples): • Graphic data model (entity relationship diagram, ERD) • Data descriptions / code lists • System specification / system description • User handbook / regulations on use / training materials / screenshots	

[3] There is not yet defined a PRONOM PUID for SIARD format 2.1. The correct PUID will be added as soon as possible.

TIFF – Tagged Image File Format

FORMAT	**MANDATORY IDENTIFIER**	TIFF – Tagged Image File Format
	PRONOM PUIDs	fmt/353
	FILE EXTENSION	Recommended file extensions: .tif, .tiff
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	Revision 6.0 of 1992-06-03 and revision 6.0.1 of 1995-10-15. Source: http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf	
	OWNERS	
	©1986-1988, 1992 Adobe Systems Inc., U.S.A.	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	Raster images (black/white, grey scales, colour)	
	BINDING RESTRICTIONS ON USE	
	TIFF files must be validated as TIFF 6.	
	PROPRIETARY EXTENSIONS: Format extensions by software manufacturers are only permitted if they adhere strictly to the relevant requirements of the standard. In case of doubt, the submitting body should obtain written confirmations from the software manufacturer. In particular, proprietary extensions ("private fields and values") must be implemented above the tag number or constant number 32767. NB: Even if they are implemented in accordance with the standard, these extensions are in principle ignored by the Federal Archives. In particular, the "TIFF Enhancements for Adobe Photoshop®" are also ignored, which chiefly means that the image source tag 37724 written by Adobe™ Photoshop® is ignored. Use of the two other "Advanced TIFF" options of Adobe™ Photoshop® – ZIP/zlib and JPEG compressions – is excluded. Also expressly ignored are the "TIFF Enhancements for Adobe™ PageMaker® 6.0" and the "Kodak™ TIFF Extensions". Adobe™ Photoshop "TIFF Enhancements for Adobe™ Photoshop®": http://partners.adobe.com/public/developer/en/tiff/TIFFphotoshop.pdf TIFF Enhancements for Adobe™ PageMaker® 6.0: http://partners.adobe.com/public/developer/en/tiff/TIFFPM6.pdf Kodak™ TIFF Extensions: Contact the Eastman Kodak Company direct. For other software, see the manufacturer's documentation	
	MULTIPAGE TIFF: Integration of multiple pages within a single TIFF file (multipage TIFF, more than one IFD) is not permitted. If the same image is to be archived with different resolutions or bit depths, different files in different folders are to be created, so that the various versions can easily be ordered separately. If a document comprising more than one page is to be archived, it should be stored in PDF/A format.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	-	
	NOTES AND COMMENTS	
	TIFF uses 4-byte file offsets, which means the maximum amount of image data in a TIFF file is limited to 4 GB (compressed). Where TIFF files are in the "GeoTIFF" format (http://trac.osgeo.org/geotiff/), they should be converted into the archivable TIFF+EWF.XML format.	

JPEG 2000 – Joint Photographic Experts Group 2000

FORMAT	**MANDATORY IDENTIFIER**	JPEG 2000
	PRONOM PUIDs	x-fmt/392
	FILE EXTENSION	Recommended file extensions: .jp2
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	ISO/IEC 15444 ISO/IEC 15444-1:2016 Information technology – JPEG 2000 image coding system – Part 1: Core coding system Source: International Organization for Standardization	
	Some elements of the standard are subject to patent rights in accordance with the Common Patent Policy for ITU-T/ITU-R/ISO/IEC: The patent holders have undertaken to keep the standard free of charge.	
	OWNERS	
	ISO/IEC Standards: International Organization for Standardization Joint Photographic Experts Group	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	JPEG 2000 is a standard issued by the Joint Photographic Experts Group for the lossless and lossy encoding of raster images in high quality. It is suitable for a wide range of applications, from digital (surveillance) cameras and scanners to high-resolution medical imaging systems. JPEG 2000 supports very large images.	
	BINDING RESTRICTIONS ON USE	
	The only permitted file format is JP2 (Part 1, Annex I, ISO/IEC 15444), which allows images to be stored individually. The JPEG 2000 codestream contained in the JP2 file must satisfy the restrictions of Profile-0 ("J2P0"). (The current standard ISO/IEC 15444-1:2016 comprises a number of profiles that were not in the previous standard ISO/IEC 15444-1:2004; the new profiles relate to digital cinema and are not relevant to archiving.)	
	The JPX file format (Part 2, Annex L and Annex M) is not permitted. It allows images to be stored with extended properties, which may not be reproduced correctly using conventional programs. The Compound Image File Format JPM (Part 6) is likewise not permitted. It allows for the storage of images assembled from a number of individual components. Reproduction of the assembled images using conventional programs cannot be guaranteed.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	Not all programs that allow images to be stored in the JPEG 2000 format produce .jp2 files that comply with the standard. The Jpylyzer program can be used for validation (http://jpylyzer.openpreservation.org).	
	NOTES AND COMMENTS	
	For individual images, lossless encoding is recommended. For a larger number of images, minimal lossy compression is recommended.	
	JPEG 2000 is a complex format, and allows for a correspondingly large number of configuration options. If large numbers of images are to be encoded using JPEG 2000, it is advisable to conduct tests in order to establish the most suitable configuration.	
	The Federal Archives are aware that the official reference software Open JPEG 2000 may produce inadequate results. It may therefore be advisable to examine professional alternatives.	

WAVE – audio format from Microsoft

FORMAT

MANDATORY IDENTIFIER	WAVE ALIAS IDENTIFIER: WAVEFORMAT, PCMWAVEFORMAT
PRONOM PUIDs	fmt/1, fmt/2, fmt/6, fmt/141
FILE EXTENSION	Mandatory file extension: .wav
TYPE	File format and data format

BINDING SPECIFICATIONS

There is no published standard for WAVE files. The WAVE format is an implementation of the Resource Interchange File Format (RIFF) from Microsoft Corporation. This has been released as a publication.

Sources: Multimedia Programming Interface and Data Specifications 1.0, published by IBM Corporation and Microsoft Corporation, August 1991. WAVEFORMAT (structure) from Microsoft.

OWNERS

RULES AND RESTRICTIONS

AREAS OF APPLICATION

Audio data

BINDING RESTRICTIONS ON USE

Only the default Microsoft Linear Pulse Code Modulation (LPCM) codec is permitted. Byte offset 20 (0x14) must contain the value 1 as 2-byte value (short) (in the little-endian byte order).

RECOMMENDATIONS AND NOTES

FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT

Creating from analogue sources

The following sampling rate and quantizations are recommended when digitising from analogue sources:

Sampling rate	Quantization
48 KHz	16 or 24 bits

Creating from digital sources

If the audio data are already in digital form, the existing sampling rate and quantization should be retained. For CD Audio and Audio DAT these are:

Original	Sampling rate	Quantization
CD Audio	44.1 KHz	16 bits
DAT	44.1 KHz / 48 KHz	16 bits

NOTES AND COMMENTS

The WAVE format is very closely related to the CD Audio (CDA) format. However, its documentation is extremely difficult to access. Widely used programs (Nero etc.) convert CDA <-> WAVE.

FFV1 – video codec

FORMAT	**MANDATORY IDENTIFIER**	FFV1 – FF Video Codec 1
	PRONOM PUIDs	fmt/569 (Matroska container[4])
	FILE EXTENSION	Mandatory file extensions: .mkv
	TYPE	Compression process (codec) for video
	BINDING SPECIFICATIONS	
	FFV1 Video Codec Specification Source: http://www.ffmpeg.org/~michael/ffv1.html	
	Draft IETF (Internet Engineering Task Force) Specification Source: https://tools.ietf.org/pdf/draft-niedermayer-cellar-ffv1-01.pdf	
	Matroska Specifications Source: https://www.matroska.org/technical/specs/index.html	
	OWNERS	
	Michael Niedermayer (FFmpeg project) The format is public domain and explicitly license-free.	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	Video data	
	BINDING RESTRICTIONS ON USE	
	The SFA only accept FFV1 files in version 3 (FFV1.3) of 2013. MKV (Matroska) must be used as the container. WAVE must be used as the codec for audio (see WAVE).	
	The "GOP size" (ffmpeg argument "-g") must be set to 1.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	For best results, FFV1 videos should be created directly from the original uncompressed video material where possible.	
	In case of doubt, the slicing (ffmpeg argument "-slices") should be kept smaller rather than larger, because a large slicing value can unnecessarily increase the file size without gaining additional encoding performance. To reduce the file size further, multi-pass encoding can be used (ffmpeg argument "-pass"). Ideally, the ffmpeg argument "-slicecrc 1" should be set, in order to store error detection information for each slice.	
	NOTES AND COMMENTS	
	FFV1 is a lossless intra-frame video codec (each individual image is compressed) that was developed specially for archiving. FFV1 is a pure image codec; sound is stored separately as WAVE in the Matroska container. FFV1 supports various image formats and colour depths without chroma subsampling, as well as a number of colour spaces (internally YCbCr for YUV data and JPEG 2000 RCT for RGB data).	

[4] The FFV1 data stream is packaged in a Matroska container.

MPEG-4 – video format

FORMAT	**MANDATORY IDENTIFIER**	MPEG-4
	PRONOM PUIDs	fmt/199
	FILE EXTENSION	Mandatory file extensions: .mp4 .mp4v
	TYPE	File format (container) and compression process (codec) for video and audio
	BINDING SPECIFICATIONS	
	ISO/IEC 14496-10 Coding of audio-visual objects -- Part 10: Advanced Video Coding ISO/IEC 14496-3 Coding of audio-visual objects -- Part 3: Audio ISO/IEC 14496-14 Coding of audio-visual objects -- Part 14: MP4 file format ISO/IEC 14496-17 Coding of audio-visual objects -- Part 17: Timed Text subtitle format Sources: International Organization for Standardization	
	OWNERS	
	ISO/IEC Standards: International Organization for Standardization	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	Video data with limited quality requirements	
	BINDING RESTRICTIONS ON USE	
	The following codecs are permitted: • Video: MPEG-4 part 10 (also referred to as MPEG-4 AVC or ITU H.264) • Audio: MPEG-4 AAC (Advanced Audio Coding), described in MPEG-4 part 3 The following container for the video and audio streams is permitted: • MP4, described in MPEG-4 part 14 Additionally, timed text subtitles are permitted where present.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	MPEG-4 is always a compromise between quality and preservability and is explicitly recommended by the SFA only if original material is available only as MPEG-4 or substantial amounts of video data are to be archived with a low quality requirement. In case of doubt, please contact the Federal Archives in advance.	
	Creating from analogue sources The sampler should create the native MPEG-4 format direct from the analogue sources. Under no circumstances may another, heavily compressing video format (such as WMC or VC1) be used as an intermediate stage.	
	Creating from digital sources If the video data are already in digital form, the existing coding should be retained where possible. a) the video codec is MPEG-4 If the file format is already MPEG-4 part 14, nothing more needs to be done. Other file formats must be repackaged in MPEG-4 part 14, taking care to ensure that no renewed coding takes place. b) the video codec is not MPEG-4 Normally, the video and audio streams must be transcoded into MPEG-4 AVC and MPEG-4 AAC. However, since this may lead to substantial losses, the exact procedure must be agreed with the Federal Archives in advance. When transcoding, the resolution of the video content must be retained; no resizing may be undertaken.	

XML/XSD – eXtensible Markup Language

FORMAT	**MANDATORY IDENTIFIER**	XML - eXtensible Markup Language
	PRONOM PUIDs	fmt/101 (XML), x-fmt/280 (XSD)
	FILE ENDING	Mandatory file extensions: .xml, .xsd
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	Extensible Markup Language (XML) 1.0 (Fifth Edition), 2008 Source: W3 Consortium	
	Permitted character encodings: • Unicode 5.0 Universal Character Set (UCS) (ISO 10646:2003)	
	OWNERS	
	W3 Consortium XML is a freely accessible standard	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	XML is a simple, flexible text format. It is used to describe structured data and specify the structure or significance of data and is a popular format for exchanging structured data.	
	BINDING RESTRICTIONS ON USE	
	These rules apply only to XML texts with non-standardised vocabularies. For XML texts with standardised vocabularies, the corresponding specifications apply (e.g. SVG, RDF/XML). Important: The SFA currently accept XML files with standardised vocabularies (e.g. SVG, RDF/XML) only in exceptional cases, as they are not accepted archivable file formats at present.	
	The XML text must be well formed, i.e. it must comply with the rules of grammar, vocabulary and syntax for XML.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	Wherever possible, XML schema definitions (XSD) should be supplied with the XML files. Widely used and freely available XSDs, such as XHTML, are an exception to this rule. In case of doubt, the XSD should also be supplied.	
	NOTES AND COMMENTS	
	To enable later (re-)use of the XML file's content, the non-standard vocabulary used and the correct data structure must be documented unambiguously. This documentation may be formulated in everyday language and complemented with graphics.	
	An XML validator (e.g. https://validator.w3.org/) can be used to check whether an XML text is well formed.	

TIFF+EWF.XML – Tagged Image File Format and Extended World File

FORMAT	**MANDATORY IDENTIFIER**	TIFF+EWF.XML – Tagged Image File Format and Extended World File
	PRONOM PUIDs	fmt/353 (TIFF), fmt/101 (XML), x-fmt/280 (XSD)
	FILE EXTENSION	Recommended file extensions: .tif, .tiff Mandatory file extensions: .xml, .xsd
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	"Information sheet: Specification for an archivable geoformat for image and graphic raster data" Source: SFA	
	OWNERS	
	TIFF: ©1986-1988, 1992 Adobe Systems Inc., U.S.A. XML is a freely accessible standard TIFF+EWF.XML: SFA	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	TIFF+EWF.XML is a format jointly developed by the SFA and swisstopo for archiving image and graphic raster data, thematic raster data and height raster data with a geographical connection.	
	BINDING RESTRICTIONS ON USE	
	Specifications and rules are set out in the document "Information sheet: Specification for an archivable geoformat for image and graphic raster data". This can be obtained from the SFA.	
	The rules for the archivable TIFF and XML formats apply.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	The recommendations for the archivable TIFF and XML formats apply.	
	NOTES AND COMMENTS	
	The notes and comments for the archivable TIFF and XML formats apply.	

INTERLIS

FORMAT	**MANDATORY IDENTIFIER**	INTERLIS – The GeoLanguage
	PRONOM PUIDs	fmt/1014, fmt/1012, fmt/654 (INTERLIS 1, 2.2, 2.3 Model Files), fmt/1013, fmt/1011, fmt/653 (INTERLIS 1, 2.2, 2.3 Transfer Files)
	FILE EXTENSION	Recommended file extensions: .xtf .xml (INTERLIS 2.x Transfer Files) Mandatory file extensions: .ili (Model Files), .itf (INTERLIS 1 Transfer File)
	TYPE	File format and data format
	BINDING SPECIFICATIONS	
	INTERLIS 1 Reference Manual (SN 612 030) INTERLIS 2.2 Reference Manual (formerly SN 612 031) INTERLIS 2.3 Reference Manual (eCH-0031; SN 612 031)	
	OWNERS	
	The INTERLIS specification is maintained and updated by Coordination, Geo-Information and Services (COGIS).	
RULES AND RESTRICTIONS	**AREAS OF APPLICATION**	
	INTERLIS is used for thematic vector data and height vector data with a geographical connection and for geodata models.	
	BINDING RESTRICTIONS ON USE	
	The rules for the archivable XML format apply to XML-based formats.	
RECOMMENDATIONS AND NOTES	**FEDERAL ARCHIVES RECOMMENDATIONS ON CREATING THIS ARCHIVAL FORMAT**	
	The recommendations for the archivable XML format apply to XML-based formats.	
	NOTES AND COMMENTS	
	The notes and comments for the archivable XML format apply to XML-based formats.	