INTERNATIONAL STANDARD

ISO/IEC 21320-1

First edition
2015-10-15

Information technology — Document Container File —

Part 1:
Core

Technologies de l'information — Fichier conteneur de document —

Partie 1: Données de base

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information.

The committee responsible for this document is ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

ISO/IEC 21320 consists of the following parts, under the general title *Information technology — Document Container File*:

— *Part 1: Core*

Introduction

It is often useful to combine multiple digital resources into a single digital resource to make them easier to store and process. The combined digital resource may also use data compression to minimize the space needed for storage.

This part of ISO/IEC 21320 specifies a format for this purpose. The format is a compatible profile of that defined by the Zip Application Note of PKWARE® Inc.

The use of Zip-based archive files for digital documents is increasingly common. Within ISO/IEC JTC 1, there are the examples of ISO/IEC 26300 series and ISO/IEC 29500. Outside of the International Standardization system, there are similar uses such as those of the EPUB Open Container Format (from the IDPF) and Widget Packaging and XML Configuration (from the W3C).

The technology defined by the Zip Application Note has been in wide use in ICT industries for over twenty years, and the specification has been freely available for much of that time. However, it has never been formally standardized and this lack of standardization presents several challenges to standards (including ISO/IEC Standards) that wish to reference it, including the following:

— stability of reference: what is the correct reference to give for the Zip Application Note and how can it be ensured that this reference remains available?;

— intellectual property rights: what, if any, patents are necessary to implement this technology, and is there a subset that may be freely implemented?;

— cultural and linguistic adaptability: is the Zip Application Note sufficient by itself, or is additional expository material needed to define best practices for global use, e.g. the use of IRIs for file names?;

— interoperability within domain: is there a technology subset that will provide greater interoperability within the domain of Document Container Files than permitting all features of the Zip Application Note?.

Information technology — Document Container File —

Part 1:
Core

1   Scope

This part of ISO/IEC 21320 specifies the core requirements for

— document container files, and

— implementations that produce and/or consume document container files.

This part of ISO/IEC 21320 normatively references the Zip File Format Specification version 6.3.3 of PKWARE® Inc. Document container files are conforming Zip files as specified by that document.

2   Normative references

The following documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

NOTE   Each of the following documents has a unique identifier that is used to cite the document in the text. The unique identifier consists of the part of the reference up to the first comma.

ISO/IEC 10646:2014, *Information technology — Universal Coded Character Set (UCS)*

Appnote,[1] APPNOTE.TX — ZIP File Format Specification, PKWARE® Inc., September 2012

IETF RFC 1951,[2] DEFLATE Compressed Data Format Specification version 1.3, May 1996

3   Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
byte
sequence of 8 bits

3.2
digital resource
sequence of *bytes* (3.1)

3.3
document container file
digital resource that conforms to this part of ISO/IEC 21320

3.4
implementation
processor that operates on *document container files* (3.3)

---

1)   Available at http://www.pkware.com/documents/APPNOTE/APPNOTE-6.3.3.TXT

2)   Available at http://www.ietf.org/rfc/rfc1951.txt

4 Requirements

A digital resource is conformant to this part of ISO/IEC 21320 when it is in accord with the provisions of this part of ISO/IEC 21320. Such a digital resource is a document container file.

An implementation is conformant to this part of ISO/IEC 21320 when it processes conformant digital resources in accord with the provisions of this part of ISO/IEC 21320.

5 Profile of Appnote

5.1 General

The provisions of this part of ISO/IEC 21320 are the provisions of Appnote as modified by the annotations provided in Table 1.

The provisions of Appnote shall be determined by interpreting Appnote in its own terms, and according to its own context.

EXAMPLE   Appnote's Clause 3 has its own "Notations" which differ from those used by ISO and IEC. Appnote's own notations apply for the purposes of its interpretation. So, for instance, "must" in the Appnote is to be read as equivalent to "shall" in ISO and IEC Standards.

5.2 Annotations to Appnote

Table 1 is a two-column table which provides annotations to Appnote. The left-hand column of Table 1 contains section numbers of Appnote, and text in parentheses, which together identify a portion of Appnote. The right-hand column of Table 1 gives an annotation to the sections so identified. Each annotation applies to the corresponding identified section of Appnote.

NOTE   The annotations to Appnote can have an effect beyond the section of the Appnote to which they explicitly apply. For example, if a specific technical feature is prohibited by an annotation, no other mention of that feature can be interpreted as re-admitting that feature to this part of ISO/IEC 21320.

Table 1 — Annotations to Appnote

Appnote section	Annotation
4.3.3 (general Format of a Zip file)	Resources conforming to this part of ISO/IEC 21320 shall not span multiple volumes, and shall not be split into user-defined segment sizes.
4.3.6 (overall Zip file format)	Resources conforming to this part of ISO/IEC 21320 shall not contain either "encryption header" or "archive decryption header" constructs.
4.3.8 (file data)	Files contained in archives shall not be encrypted using the mechanisms described in Appnote.
4.3.9.6 (data descriptor)	The "central directory encryption method" mentioned by this subclause of Appnote shall not be used.

NOTE 1 Although the above provision for Appnote,4.4.3 permits values for "version needed to extract" which imply support for features of value 1.1 ("File is a volume label"), 2.0 ("File is a folder [...]" etc.), 2.1 ("File is compressed using Deflate64™"), 2.5 ("File is compressed using PKWARE DCL Implode") and 2.7 ("File is a patch data set"), none of these features are permitted by this part of ISO/IEC 21320.

NOTE 2 To promote global use, implementers are strongly encouraged to support bit 11 of the "general purpose bit flag" (Appnote, 4.4.4) being set.

NOTE 3 For file names, any Unicode character is allowed. However, different file systems impose different restrictions on permissible characters, and RFC 3987 introduces its own restrictions. As a result, it might not be possible to create a file name or IRI from a zip item name.

NOTE 4 Some file systems perform Unicode normalization or case normalization of file names while others do not. Thus, multiple different zip item names in a single Zip archive may become identical after normalization.

<div align="center">**Table 1** *(continued)*</div>

Appnote section	Annotation
4.3.10 (archive decryption header)	The "archive decryption header" feature specified by this subclause of Appnote shall not be used.
4.3.13 (digital signature)	The "digital signature" feature specified by this subclause of Appnote shall not be used.
4.4.1.5 (explanation of fields)	Archives shall not be split or spanned.
4.4.3 (version needed to extract)	Resources conforming to this part of ISO/IEC 21320 shall not contain a "version needed to extract" value greater than 45. The ZIP64™ extension feature version 1 may be used. The ZIP64™ extension feature version 2 shall not be used.
4.4.4 (general purpose bit flag)	Bit 0, bits 4 to 10 and 12 to 15, shall not be set. Bit 11 should be set. If the value of any byte in the file name or file comment is greater than 0x7F, bit 11 shall be set. If bit 11 is set, the file names and comment fields for the document container file shall be Unicode strings encoded using the UTF-8 encoding scheme as specified in ISO/IEC 10646:2014, Annex D.
4.4.5 (compression method)	The compression method shall be either 0 ("stored") or 8 ("deflated").
4.7 (manifest files)	This clause of Appnote shall be disregarded.
5.5 (deflating)	Files compressed using this method shall use the mechanisms specified in IETF RFC 1951.
6.0 (traditional PKWARE encryption)	This clause of Appnote shall be disregarded.
7.0 (strong encryption specification)	This clause of Appnote shall be disregarded.
8.0 (splitting and spanning Zip files)	This clause of Appnote shall be disregarded.

NOTE 1   Although the above provision for Appnote,4.4.3 permits values for "version needed to extract" which imply support for features of value 1.1 ("File is a volume label"), 2.0 ("File is a folder [...]" etc.), 2.1 ("File is compressed using Deflate64™"), 2.5 ("File is compressed using PKWARE DCL Implode") and 2.7 ("File is a patch data set"), none of these features are permitted by this part of ISO/IEC 21320.

NOTE 2   To promote global use, implementers are strongly encouraged to support bit 11 of the "general purpose bit flag" (Appnote, 4.4.4) being set.

NOTE 3   For file names, any Unicode character is allowed. However, different file systems impose different restrictions on permissible characters, and RFC 3987 introduces its own restrictions. As a result, it might not be possible to create a file name or IRI from a zip item name.

NOTE 4   Some file systems perform Unicode normalization or case normalization of file names while others do not. Thus, multiple different zip item names in a single Zip archive may become identical after normalization.

Annex A
(informative)

Informal summary

While this part of ISO/IEC 21320 formally states its provisions in the normative portions of the text, informally, the principal technical characteristics of this part of ISO/IEC 21320, as against Appnote, can be summarized as follows:

— files stored in archives may only be stored uncompressed, or using the deflate mechanism;

— encryption features of Appnote are prohibited;

— digital signature features of Appnote are prohibited;

— patched data features of Appnote are prohibited;

— archives may not span multiple volumes or be segmented.

Annex B
(informative)

Filenames and interoperability

B.1 Requirement

The "Zip" format was originally developed as a general purpose archive format, it had to support various platforms with various character encoding before ISO/IEC 10646 standardization. Therefore, Appnote specifies few restrictions for filenames in the archive. For compatibility, this part of ISO/IEC 21320 does not require additional restrictions on filenames which are valid according to Appnote.

However, some digital document packages using UTF-8 filenames in Zip archive do define restrictions on filenames. In this Annex, these restrictions are summarized to help implementers avoid unexpected information loss during the document transcoding.

B.2 Normalizations

In the archive formats for many XML-based documents, the identifier of a file in the archive, or the internal reference from one file in the archive to another file in the same archive, is given with a path-rootless IRI (ISO/IEC 29500-2:2012 (OOXML OPC), ISO/IEC 26300 series (ODF Package), EPUB OCF, Widget Packaging and XML Configuration, etc.). Because the IRI comparator should not apply the normalization by itself, the IRI in the documents should be already normalized by NFC (See RFC 3987, 5.3.2.2 and 7.5, NFKC is recommended for safer normalization). OOXML OPC requests that the filename in the archive is already an IRI, thus the filename is expected to be normalized by NFC. EPUB OCF and Adobe UCF does not require that filenames are IRIs, but does require that two filenames in the same directory are different after the normalization and the case normalization.

There are some operating systems that apply NFD or its variant during the location of the file in their native filesystems, and many general purpose Zip utilities do not convert filenames by default. The extraction and repackage of the document package in such environment can break the package consistency, if the document package assumes the filenames in NFC.

B.3 Prohibited characters

The IRI specification prohibits some Unicode characters as raw characters in the resource names; some characters are reserved for special purpose (like U+0020 SPACE, U+0025 PERCENT SIGN, U+002F SOLIDUS, etc.) and PUA characters are prohibited. Some packages have stronger restrictions on the permitted character set for safer interchange, even if they are preserved by NFC and permitted by IRI. For example, Tag and Variation characters (U+E0000 - U+E0FFF) are prohibited by EPUB OCF, therefore, IVS (defined in ISO/IEC 10646:2014, 16.5) is unavailable in OCF, as raw sequence in the filenames. These known restrictions are summarized in Table B.1.

B.4 Table of characters

In Table B.1, the symbols in the five right-most columns signify the following:

— O: character usage is explicitly permitted;

— ×: character usage is prohibited;

— blank cell signifies character usage is not explicitly specified.

Table B.1 — Table of characters

Codepoint or range	Character or block name		JAR INDEX LIST	Widget Zip relative path	IRI root-less path OOXML OPC	EPUB OCF	Adobe UCF
U+0000 – U+001F	C0		U+0000, U+000A, U+000D are prohibited		×	×	×
U+0020	SPACE			O	×		
U+0022	QUOTATION MARK	"			×	×	×
U+0023	NUMBER SIGN	#			×		
U+0025	PERCENT SIGN	%		O	permitted for percent encoding		
U+002A	ASTERISK	*			×		×
U+002E	FULL STOP	.		O		except as the last character	except as the last character
U+002F	SOLIDUS	/			permitted as the segment separator	permitted as the segment separator	permitted as the segment separator
U+003A	COLON	:				×	×
U+003C	LESS THAN SIGN	<			×	×	×
U+003E	GREATER THAN SIGN	>			×	×	×
U+003F	QUESTION MARK	?			×	×	×
U+005B	LEFT SQUARE BRACKET	[		O	×		
U+005C	REVERSE SOLIDUS	\			×	×	×
U+005D	RIGHT SQUARE BRACKET	]		O	×		
U+005E	CIRCUMFLEX ACCENT	^			×		
U+007B	LEFT CURLY BRACKET	{			×		
U+007C	VERTICAL LINE	|			×		
U+007D	RIGHT CURLY BRACKET	}			×		
U+007F – U+009F	DEL, C1				×	×	
U+E000 – U+F8FF	PUA				×	×	
U+FDD0 – U+FDEF, U+FFF0 – U+FFFF				O	×	×	
U+E00000 – U+E0FFF	Tag and Variation Character			O		×	