INTERNATIONAL STANDARD

ISO
14289-1

First edition
2012-07-15

Corrected version
2012-08-01

Document management applications — Electronic document file format enhancement for accessibility — Part 1: Use of ISO 32000-1 (PDF/UA-1)

Applications de la gestion de documents — Amélioration de format du fichier du document électronique pour l'accessibilité — Partie 1: Utilisation de l'ISO 32000-1 (PDF/UA-1)

iTeh STANDARD PREVIEW
(standards.iteh.ai)

COPYRIGHT PROTECTED DOCUMENT

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO 14289-1 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Application issues*.

ISO 14289 consists of the following parts, under the general title *Document management applications — Electronic document file format enhancement for accessibility*:

— *Part 1: Use of ISO 32000-1 (PDF/UA-1)*

This corrected version of ISO 14289-1:2012 incorporates the following corrections:

— in 7.3 (line after second bullet), bold type has been applied to the word "Caption";

— in 7.3 (line before Note 1), "Table 333" has been replaced by "Table 323" and a reference to subclause 14.7.2 has been added;

— in 7.4.1 (second line), a reference to subclause 14.7.2 has been added;

— in 7.4.3 (first paragraph), the extra parenthesis after the word "system" has been removed;

— in 7.4.4 (first line), bold type has been applied to the letter "H";

— in 7.6 (second bullet), **LI** has been replaced by **L**;

— in 7.17 (last paragraph), a reference to 7.7.2 has been added;

— in 7.18.2 (first line), 7.17.1 has been replaced by 7.18.1;

— in 8.4 (third bullet), the word "the" has been added before "value";

— in 8.8 (note), a reference to 12.2 has been added.

Introduction

PDF is a digital format for representing documents. PDF files may be created natively in PDF form, converted from other electronic formats, or digitized from paper. Businesses, governments, libraries, archives and other institutions and individuals around the world use PDF to represent considerable bodies of important information. These PDF files should be made accessible to users with disabilities.

The accessibility of a document is dependent on the inclusion of a variety of semantic information in a document such as (but not limited to) machine-recoverable text presented in a declared language, logical structure of content, and organization of that content in pages, sections and paragraphs. Semantic information can also contain a variety of descriptive metadata, such as alternative text for images. The primary purpose of ISO 14289 (known as PDF/UA) is to define how to represent electronic documents in the PDF format in a manner that allows the file to be accessible. This is accomplished by identifying the set of PDF components that may be used and restrictions on the form of their use. PDF/UA is intended as a companion standard, to be used in conjunction with ISO 32000, ISO 19005, ISO 15930 and other standards as may apply for the purpose of achieving accessibility or as mandated by this International Standard. In order for PDF/UA conforming files to be truly accessible, requirements on conforming readers and conforming assistive technology are also stipulated herein. By itself, PDF/UA does not necessarily ensure that the visual appearance of the content accurately reflects any original source material used to create the conforming file. For example, the process used to create a conforming file might substitute fonts, reflow text, downsample images, or use lossy compression. Organizations that need to ensure that a conforming file is an accurate representation of original source material may need to impose additional requirements on the processes that generate the conforming file beyond those imposed by this International Standard. In addition, it is important for those organizations to implement policies and practices regarding the inspection of conforming files for correct placement of accessibility information.

AIIM (an accredited standards developing organization) maintains an ongoing series of application notes for guiding developers and users of this International Standard. These application notes will be displayed, when available, at http://www.aiim.org/resources/standards/committees/PDFUA/App-Notes. AIIM will also retain copies of the specific non-ISO normative references in this International Standard that are publicly available electronic documents.

iTeh STANDARD PREVIEW

(standards.iteh.ai)

1 Scope

This International Standard specifies how to use ISO 32000-1 (PDF/UA-1) to produce accessible electronic documents.

This International Standard is not applicable to:

— specific processes for converting paper or electronic documents to the PDF/UA format;

— specific technical design, user interface, implementation, or operational details of rendering;

— specific physical methods of storing these documents, such as media and storage conditions;

— required computer hardware and/or operating systems.

2 Normative references

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

ISO 32000-1:2008, *Document management — Portable document format — Part 1: PDF 1.7*

Web Content Accessibility Guidelines (WCAG) 2.0, W3C Recommendation, 11 December 2008. Available from the Internet <http://www.w3.org/TR/WCAG20>.

3 Terms and definitions

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

3.1
assistive technology
AT
software and/or hardware used by a person with a disability that provides alternative controls and/or renditions to facilitate their access to and usage of available functionality and information

3.2
alternate description
description of a structure element and its children

[SOURCE: ISO 32000-1:2008, 14.9.3]

NOTE   ISO 32000-1 uses both "alternate text" and "alternate description" to refer to the same item.

3.3
dictionary
associative table containing key-value pairs, specifying the name and value of an attribute for objects, which is generally used to collect and tie together the attributes of a complex object

[SOURCE: ISO 32000-1:2008, 4.17]

3.4
interactive reader
reader that requires or allows human interaction with the content and other objects contained in the document during the software's processing phase

NOTE   A file viewing tool is an example of an interactive reader; a raster image processor is an example of a reader that is not interactive.

[SOURCE: ISO 19005-2:2011, definition 3.8]

3.5
Portable Document Format
PDF
file format defined in ISO 32000-1:2008

3.6
reader
software application that is able to read and process PDF files appropriately

[SOURCE: ISO 32000-1:2008, 2.2]

3.7
replacement text
text that is an exact replacement for the structure element and its children in cases where an AT device needs to read the element inline with the surrounding content

[SOURCE: ISO 32000-1:2008, 14.9.4]

3.8
writer
software application that is able to write PDF files

[SOURCE: ISO 32000-1:2008, 2.3]

4   Notation

PDF operators, PDF keywords, the names of keys in PDF dictionaries, and other predefined names are written in bold sans serif font; operands of PDF operators or values of dictionary keys are written in italic sans serif font. Some names can also be used as values, depending on the context, and so the styling of the content will be context specific.

EXAMPLE 1   The *Default* value for the **TR2** key.

Token characters used to delimit objects and describe the structure of PDF files, as defined in ISO 32000-1:2008, 7.2.1, may be identified by their ISO/IEC 646 character name written in upper case in bold sans serif font followed by a parenthetic two digit hexadecimal character value with the suffix "h".

EXAMPLE 2   **CARRIAGE RETURN** (0Dh).

Text string characters, as defined in ISO 32000-1:2008, 7.9.2, may be identified by their ISO/IEC 10646 character name written in uppercase in bold sans serif font followed by a parenthetic four digit hexadecimal character code value with the prefix "U+".

EXAMPLE 3   **EN SPACE** (U+2002).

5   Version identification

The PDF/UA version and conformance of a file shall be specified using the PDF/UA Identification extension schema defined in Table 1. All extension schemas referenced from any metadata stream in a conforming file shall have their descriptions embedded within the referencing metadata stream or the metadata stream that is the value of the **Metadata** key in the **Catalog**.

The Identification schema namespace URI is http://www.aiim.org/pdfua/ns/id/. The required schema namespace prefix is **pdfuaid**.

Table 1 — PDF/UA Identification schema

Property	Value type	Category	Description
pdfuaid:part	*Open Choice of Integer*	Internal	Required PDF/UA version identifier
pdfuaid:amd	*Open Choice of Text*	Internal	Optional PDF/UA amendment identifier
pdfuaid:corr	*Open Choice of Text*	Internal	Optional PDF/UA corrigenda identifier

The value of **pdfuaid:part** shall be the part number of this International Standard to which the file conforms. If the file conforms to a version of this International Standard that is defined by an amendment to a part, then the value of **pdfuaid:amd** shall be the amendment number and year, separated by a colon.

The values of the **pdfuaid:part, pdfuaid:amd** and **pdfuaid:corr** properties do not by themselves determine conformance with any part of this International Standard.

6 Conformance requirements

6.1 General

ISO 14289 defines a file format for representing electronic documents known as "PDF/UA" and also defines how conforming readers and conforming assistive technology (AT) process such documents.

The version number of a file may be any value from 1.0 to 1.7, and the value shall not be used in determining whether a file is in conformance with this part of ISO 14289.

NOTE   The proper mechanism by which a file can presumptively identify itself as being a PDF/UA file of a given conformance level is described in Clause 5.

6.2 Conforming files

Conforming files shall adhere to all requirements of ISO 32000-1:2008 as modified by this part of ISO 14289. A conforming file may include any valid ISO 32000-1:2008 feature that is not explicitly forbidden by this part of ISO 14289. Features described in PDF specifications prior to ISO 32000-1:2008 which are not explicitly defined in ISO 32000-1 should not be used.

NOTE   A conforming file is not obligated to use any PDF feature other than those explicitly required by ISO 32000-1 or this part of ISO 14289.

A conforming file shall contain PDF/UA version identification as defined in Clause 5.

A conforming file shall adhere to all file format provisions in Clause 7.

6.3 Conforming reader

A conforming PDF/UA reader shall adhere to conforming reader provisions in ISO 32000-1:2008 as well as to all reader requirements of this part of ISO 14289.

NOTE   A conforming reader can also be integrated with conforming assistive technology.

6.4 Conforming assistive technology

A conforming AT shall adhere to all provisions in Clause 9.

7 File format requirements

7.1 General

All real content shall be tagged as defined in ISO 32000-1:2008, 14.8. Artifacts (ISO 32000-1:2008, 14.8.2.2.2) shall not be tagged in the structure tree.

Content shall be marked in the structure tree with semantically appropriate tags in a logical reading order. Content that does not render and does not have a rendered equivalent shall be marked as an artifact.

NOTE 1   Because they are not considered real content, artifacts are not present in the structure tree according to ISO 32000-1:2008, 14.8.2.2. Real content consists of graphics objects in page content with semantic significance. Artifacts are graphics objects outside of the real content.

NOTE 2   This content also applies whether or not the content is included in one or more optional content groups.

Non-standard structure types are permitted. However, they shall be mapped to the nearest functionally equivalent standard type, as defined in ISO 32000-1:2008, 14.8.4, in the role map dictionary of the structure tree root. This mapping may be indirect; within the role map a non-standard type can map directly to another non-standard type, but eventually the mapping shall terminate at a standard type.

NOTE 3   Non-standard structure types are types other than those defined in ISO 32000.

Standard tags defined in ISO 32000-1:2008, 14.8.4, shall not be remapped.

Flickering, blinking, or flashing shall not be used (WCAG 2.0, Guideline 2.3).

Information shall not be conveyed by contrast, colour, format or layout, or by combinations thereof, unless the content is tagged to reflect all intended meaning.

NOTE 4   WCAG 2.0, Guideline 1.4 explains issues regarding contrast, colour and other formatting for accessibility.

All information conveyed with sound should also be available without sound.

The **Metadata** stream in the document's catalog dictionary shall contain a dc:title entry, where dc is the recommended prefix for the Dublin Core metadata schema as defined in the XMP specification, which clearly identifies the document. A document information dictionary may be present in a conforming file and an ISO 14289-1 conforming reader shall ignore it.

The **ViewerPreferences** dictionary of the document's **Catalog** dictionary shall be present and shall contain at least the key **DisplayDocTitle** with a value of *true*, as described in ISO 32000-1:2008, 12.2, Table 150.

Documents consisting of raster-based images may be processed to generate machine-readable content. In such cases, errors resulting from the content-generation process shall be corrected and the content shall be tagged according to Clause 7.

NOTE 5   Methods used for correcting such errors are dependent on the tools used for processing documents consisting of raster-based images.

Files claiming conformance with this International Standard shall have a **Suspects** value of *false* (ISO 32000-1:2008, Table 321).

7.2 Text

Content shall be tagged in logical reading order. The most semantically appropriate tag shall be used for each logical element in the document content.

Character codes shall map to Unicode as described in ISO 32000-1:2008, 14.8.2.4.2. Characters not included in the Unicode specification may use the Unicode private use area or declare another character encoding.

Natural language shall be declared as discussed in ISO 32000-1:2008, 14.9.2 and/or as described in ISO 32000-1:2008, 7.9.2. Changes in natural language shall be declared. Changes in natural language inside

text strings (e.g. inside alternate descriptions) shall be declared using a language identifier as described in ISO 32000-1:2008, 14.9.2.2.

Stretchable characters such as parentheses or brackets (often drawn by combining several individual glyphs to form the appearance of a single glyph) shall be tagged using **ActualText**, as specified in ISO 32000-1:2008, 14.9.4.

7.3   Graphics

Graphics objects, other than text objects, shall be tagged with a **Figure** tag as described in ISO 32000-1:2008, 14.8.4.5, Table 340. If any of the following exceptions are true, then the graphic shall be tagged as an artifact:

— the graphic does not represent meaningful content, or

— the graphic appears as a background to a link annotation, in which case, the alternative text on the link shall describe both the graphic and the link.

A caption accompanying a figure shall be tagged with a **Caption** tag.

Figure tags shall include an alternative representation or replacement text that represents the contents marked with the Figure tag as noted in ISO 32000-1:2008, 14.7.2, Table 323.

NOTE 1   See also WCAG 2.0, Guideline 1.1.

If text represented in a graphic is not text in a natural language that is meant to be read by a human reader, alternative text describing the nature or purpose of the graphic shall be provided.

NOTE 2   Text that is a type sample or a sample of the writing system used by a language are examples of text that is not in a natural language.

Graphics that possess semantic value only in combination with other graphics shall be tagged with a single **Figure** tag for each group.

When a more accessible representation exists, the more accessible representation should be used instead of a graphic.

NOTE 3   An example where more accessible representations exist is text.

7.4   Headings

7.4.1   General

A conforming document shall use heading tags for all headings as detailed in ISO 32000-1:2008, 14.8.4.3.2. The **T** key in ISO 32000-1:2008, 14.7.2, Table 323 should be used to denote document sections.

NOTE   'Chapter 1' or 'Bibliography' are examples of document section headings

7.4.2   Numbered headings

For documents that are not strongly structured, as described in ISO 32000-1:2008, 14.8.4.3.5, heading tags shall be used as follows.

— If any heading tags are used, **H1** shall be the first.

— A document may use more than one instance of any specific tag level. For example, a tag level may be repeated if document content requires it.

NOTE 1   **H1 H2 H3 H3** is a valid sequence if the content has one top-level heading, one second-level heading, and two consecutive third-level headings.

— If document semantics require a descending sequence of headers, such a sequence shall proceed in strict numerical order and shall not skip an intervening heading level. **H1 H2 H3** is permissible, while **H1 H3** is not.