INTERNATIONAL STANDARD

ISO 23504-1

First edition
2020-07

Corrected version
2020-09

Document management applications — Raster image transport and storage —

Part 1:
Use of ISO 32000 (PDF/R-1)

Applications de gestion de documents — Transport et stockage des images tramées —

Partie 1: Utilisation de l'ISO 32000 (PDF/R-1)

© ISO 2020

iTeh STANDARD PREVIEW
(standards.iteh.ai)

COPYRIGHT PROTECTED DOCUMENT

Contents

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 23504-1:2020
https://standards.iteh.ai/catalog/standards/sist/a728d3ea-25c7-46d1-a834-
23e01a8e18c9/iso-23504-1-2020

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.

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 ISO documents 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 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 of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Document file formats, EDMS systems and authenticity of information*.

This corrected version of ISO 23504-1:2020 incorporates the following corrections:

— Angled brackets inserted around 'total height' in the numerator of the second formula in A.4;

— ']' added to the line before '/Whitepoint' in A.8.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document describes PDF/R (Raster), a strict subset of the PDF file format, for storing, transporting and exchanging multi-page raster-image documents, especially scanned documents and photographs. PDF/R provides the portability of PDF while offering the core functionality of TIFF. Bitonal, grayscale and RGB images are supported. Compression options include JPEG, lossless CCITT Group 4 Fax and uncompressed.

This document describes the restrictions that differentiate a PDF/R file from a standard PDF file. Additionally, it specifies (see Clause 5) that a comment is used to identify files claiming to be PDF/R files. There is no intention herein to claim any intellectual property that is not present in the existing PDF standard, nor claim any IP that is covered therein.

PDF/R is intended to be a standard format for storing, transporting and exchanging scanned documents. As a subset of PDF, it takes advantage of the widespread support for viewing, printing and processing PDF files. As a narrowly restricted subset of PDF, it is much simpler to generate and interpret, allowing it to replace the TIFF and JPEG file formats for capture and delivery of scanner output.

PDF/R imposes many restrictions on PDF content and layout, for the following benefits:

— files can be read and written without a full PDF parser or generator;

— files can be created efficiently from raster images;

— files can be generated using a fixed-size raster data buffer;

— images can be located and read efficiently with comparatively simple code;

— PDF/R files can be quickly and easily identified as such by software;

— PDF/R supports effective and readily available compression algorithms.

PDF/R has important advantages over the full PDF format for storing scanned documents:

— the raster image data can be recovered;

— a complex rendering engine is not required;

— it provides a precise, well-defined target, simplifying engineering design and testing.

PDF/R retains optional PDF security features useful for protecting content:

— encryption is allowed for implementations that need to protect document content at rest.

PDF/R retains optional PDF digital signature features useful for authenticating content:

— one or more digital signatures may be used for implementations that require verification of the document origin, authenticity, date or time of creation, and so on.

PDF/R has important advantages over TIFF and JPEG for storing scanned documents:

— compared to TIFF, it has far fewer and simpler variants;

— compared to TIFF, compression is simpler and better standardized and supported;

— compared to TIFF, PDF files can be natively viewed and printed on more platforms;

— unlike JPEG, it is natively multi-page and handles bitonal images.

PDF/R was created by collaboration between the TWAIN Working Group, which originated the PDF/R concept, and the PDF Association, which provided PDF technology expertise and perspective as well as means of communicating with the PDF software industry to ensure a diverse range of relevant viewpoints was represented.

ISO 23504-1:2020
https://standards.iteh.ai/catalog/standards/sist/a728d3ea-25c7-46d1-a834-
e3e0fa8e8bdc/iso-23504-1-2020

Document management applications — Raster image transport and storage —

Part 1:
Use of ISO 32000 (PDF/R-1)

1 Scope

This document defines a subset of ISO 32000 suitable for storage, transport and exchange of multi-page raster-image documents, including but not limited to scanned documents. Bitonal, grayscale and RGB images are supported. Compression options for image data streams include JPEG, CCITT Group 4 Fax and uncompressed.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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*

ISO 32000-2[1]):2020, *Document management — Portable document format — Part 2: PDF 2.0*

3 Terms and definitions

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

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
page image
image of one side of a *physical page* (3.2)

3.2
physical page
physical media object with two sides

3.3
unencrypted PDF/R file
file conforming to this PDF/R specification that does not contain an **Encrypt** dictionary in the trailer dictionary

3.4
encrypted PDF/R file
file conforming to this PDF/R specification that does contain an **Encrypt** dictionary in the trailer dictionary

---

1) Under preparation. Stage at the time of publication: ISO DIS 32000-2.

4   Notation

PDF operators, PDF keywords, the names of keys in PDF dictionaries, and other predefined names are written in bold font; operands of PDF operators or values of dictionary keys are written in italic 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 *Sig* value for the **FT** 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 uppercase in bold 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[2] character name written in uppercase in bold font followed by a parenthetic four-digit hexadecimal character code value with the prefix "U+".

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

5   Version identification

A PDF file conforming to the PDF/R specification is identified by one comment line near the end of the file, immediately before the last occurrence of the line in the file containing the **startxref** key. The comment shall be:

    %PDF-raster-x.y

where

"x" (the digit before the decimal point)   is the major version number

"y" (the digit after the decimal point)   is the minor version number

The PDF/R version number for PDF files conforming to this document shall be 1.0. New major versions may be incompatible with previous versions; new minor versions are expected to not break existing readers.

This comment line marks the file as intended to conform to this specification.

EXAMPLE

```
trailer
<<
/Info 58 0 R
/Size 59
/Root 1 0 R
/ID
[   <D7916DF85B0EE1998036EA145A1CE7B4>
<D7916DF85B0EE1998036EA145A1CE7B4>
]
>>
%PDF-raster-1.0
startxref
177317
%%EOF
```

---

2)   Under preparation. Stage at the time of publication: ISO/IEC DIS 10646.

6 Conformity requirements

6.1 General

A conforming PDF/R file shall conform to all requirements listed in 6.2, "PDF subset" to 6.8, "Encryption".

6.2 PDF subset

6.2.1 General

Conformity of unencrypted and encrypted PDF/R files only differs regarding the use of encryption. Encrypted PDF/R files make use of encryption features introduced in ISO 32000-2, and not available in ISO 32000-1. The definition of, and the requirements for, any other feature allowed in a PDF/R file do not differ between ISO 32000-1 and ISO 32000-2. For the sake of simplicity, all requirements for PDF/R files, with the exception of those for the use of encryption, are specified on the background of ISO 32000-1.

6.2.2 Unencrypted PDF/R files

A PDF/R-conforming file that is not encrypted shall adhere to all the requirements of ISO 32000-1 as modified by this document.

The header shall be one of the following:

— "%PDF-1.4";

— "%PDF-1.5";

— "%PDF-1.6";

— "%PDF-1.7".

NOTE   If the contents of the file are inconsistent with the version number in the header processing results will be implementation dependent.

No filters other than the following shall be used in an unencrypted PDF/R file:

— *FlateDecode*;

— *CCITTFaxDecode* (only for bitonal images);

— *DCTDecode* (only for 8-bit grayscale or RGB images).

6.2.3 Encrypted PDF/R files

A PDF/R-conforming file that is encrypted shall adhere to all requirements of ISO 32000-1, as modified by this document, with the following exceptions:

— the header shall be "%PDF-2.0";

— the file shall adhere to all requirements of ISO 32000-2:2020, 7.6, "Encryption", as modified by 6.8, "Encryption", in this document.

Only the following filters shall be allowed in an encrypted PDF/R file:

— *FlateDecode*;

— *CCITTFaxDecode* (only for bitonal images);

— *DCTDecode* (only for 8 bit grayscale or RGB images);

— *Crypt*.

6.2.4 Unencrypted and encrypted PDF/R files

All indirect references shall have a generation number equal to zero.

All objects referred to be indirect references shall be listed.

NOTE 1   This precludes indirect object references to a non-existent object as described in ISO 32000-1:2008, 7.3.9, "Null Object".

Stream dictionaries shall not contain a **Type** key with a value of *ObjStm*.

NOTE 2   This precludes the use of object streams described in ISO 32000-1:2008, 7.5.7, "Object streams".

6.3 Catalog dictionary

The **Catalog** dictionary shall contain the entries required by ISO 32000-1:2008, Table 28. It shall not contain any optional entries except zero, one or more of the following entries: **Version**, **ViewerPreferences**, **PageLayout**, **PageMode**, **AcroForm**, and **Metadata**.

6.4 Metadata

6.4.1 General

The **Catalog** dictionary of a conforming file may contain the **Metadata** key for which the value is a metadata stream as defined in ISO 32000-1:2008, 14.3.2.

Page dictionaries may contain the **Metadata** key for which the value is a metadata stream as defined in ISO 32000-1:2008, 14.3.2. This metadata stream, if present, shall contain entries with metadata specific to the page object.

6.4.2 Document level and page level metadata streams

The document level metadata stream and page level metadata streams may use properties defined in ISO 16684-1:2019 (XMP)[5] or custom properties. Where custom properties are used, namespaces shall be used in such a fashion that conflicts are avoided with other entries using the same property name. Each organization wishing to define and use its own custom properties shall define a suitable namespace based on a URL that is under the organization's control.

EXAMPLE 1   Examples for namespaces based on which custom properties can be defined:

— http://ns.twain.org/ns/pdfraster/v1/extra_metadata

— http://ns.twain.org/ns/pdfraster/v1/some_other_fields

— http://ns.some_company.com/ns/pdf_raster/version_1/company_specific_fields

EXAMPLE 2   Properties using the same name that are based on different namespaces:

```
<rdf:Description
   rdf:about=""
   xmlns:org_a="http://ns.org_a.com/pdfraster/1.0/"
   xmlns:org_b="http://ns.org_b.com/pdfraster/1.0/"
   <org_a:JobID>ABC-123</org_a:JobID>
   <org_b:JobID>987-654-321:tre-hgf-bvc</org_b:JobID>
</rdf:Description>
```

The TWAIN Working Group provides guidance regarding metadata properties for scanned images[1].

6.4.3   Document information dictionary

A document information dictionary may appear within a conforming file. It shall contain no entries other than **Creator**, **Producer**, **CreationDate**, and **ModDate**.

6.4.4   XMP Metadata

If an XMP metadata stream is present, each of the entries in the document information dictionary, shall be represented by the corresponding XMP property value. Table 1 indicates the mapping between document information dictionary and XMP properties.

Table 1 — Mapping document information dictionary to corresponding XMP properties

Document information dictionary		Document level metadata stream	
Entry	PDF type	Property	XMP type
Creator	text string	xmp:CreatorTool	AgentName
Producer	text string	pdf:Producer	AgentName
CreationDate	date	xmp:CreateDate	Date
ModDate	date	xmp:ModifyDate	Date

6.5   Page objects

6.5.1   General

Each page image is represented by a PDF page object. The page object is a dictionary that shall be constructed as mandated by ISO 32000-1:2008.

Each page object shall contain the entries required by ISO 32000-1:2008, Table 30, and shall contain one **Contents** entry, and shall not contain any optional entries except zero, one or more of the following entries: **Rotate**, **Metadata**, **Annots**, and **PZ**.

6.5.2   Page tree nodes

Page tree nodes shall not contain any entries other than those required by ISO 32000-1:2008, Table 29.

NOTE   This provision effectively prohibits the inheritance of such entries. This also applies to the **MediaBox** key. Thus, inheritance of the **MediaBox** key is not possible in a PDF/R file.

6.5.3   Media box

Each page object shall contain a **MediaBox** entry for which the value shall be of the form [0 0 w h], where w is the width of the page and h is the height.

NOTE 1   The **MediaBox** is defined in default user space coordinate units with a default value of 1/72 inch (see ISO 32000-1:2008, 8.3.2.3, "User space").

NOTE 2   The **MediaBox** reflects the size of the page and thus the page image represented on it prior to any rotation specified by the **Rotate** entry.

EXAMPLE   An ISO A4 sized page would have a **MediaBox** value of [0 0 595.27559 841.88976].

See Annex A.3, "(informative) Calculating the MediaBox" for a detailed example.