INTERNATIONAL STANDARD

ISO 19005-1

First edition
2005-10-01

Corrected version
2005-12-01

Reviewed and confirmed in 2015

Document management — Electronic document file format for long-term preservation —

Part 1: Use of PDF 1.4 (PDF/A-1)

Gestion de documents — Format de fichier des documents électroniques pour une conservation à long terme —

Partie 1: Utilisation du PDF 1.4 (PDF/A-1)

Reference number
ISO 19005-1:2005(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

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 19005-1 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Application issues*, in collaboration with Technical Committees ISO/TC 130, *Graphic technology*, ISO/TC 42, *Photography* and ISO/TC 46, *Information and documentation*, Subcommittee SC 11, *Archives/records management*, in a Joint Working Group.

ISO 19005 consists of the following parts, under the general title *Document management — Electronic document file format for long-term preservation*:

— *Part 1: Use of PDF 1.4 (PDF/A-1)*

In this corrected version of ISO 19005-1:2005 paragraph 5 of the Foreword (above) has been augmented in order fo mention the collaboration of ISO Technical Committees ISO/TC 130, ISO/TC 42 and ISO/TC 46 in the preparation of this part of ISO 19005.

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, microform, or other hard copy format. Businesses, governments, libraries, archives and other institutions and individuals around the world use PDF to represent considerable bodies of important information. Much of this information must be kept for substantial lengths of time; some must be kept permanently. These PDF files must remain useable and accessible across multiple generations of technology. The future use of, and access to, these objects depends upon maintaining their visual appearance as well as their higher-order properties, such as the logical organization of pages, sections, and paragraphs, machine recoverable text stream in natural reading order, and a variety of administrative, preservation and descriptive metadata.

Adobe Systems Incorporated makes the PDF specification publicly available. However, the inclusive, feature-rich nature of the format requires that additional constraints be placed on its use to make it suitable for the long-term preservation of electronic documents.

The primary purpose of this part of ISO 19005 is to define a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.

A secondary purpose of this part of ISO 19005 is to provide a framework for recording the context and history of electronic documents in metadata within conforming files.

Another purpose of this part of ISO 19005 is to define a framework for representing the logical structure and other semantic information of electronic documents within conforming files.

These goals are accomplished by identifying the set of PDF components that may be used, and restrictions on the form of their use, within conforming PDF/A files.

By itself, PDF/A does not necessarily ensure that the visual appearance of the content accurately reflects any original source material used to create the conforming file; e.g. 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 part of ISO 19005. In addition, it is important for those organizations to implement policies and practices regarding the inspection of conforming files for correct visual appearance.

This part of ISO 19005 should be used as one component of an organization's electronic archival environment for long-term retention of documents. Successful implementation of this part of ISO 19005 for archival purposes depends upon:

— the retention requirements of an organization's archival environment, records management policies and procedures as specified in ISO 15489-1, [9];

— any additional requirements and conditions necessary to ensure the persistence of electronic documents and their characteristics over time, including, but not limited to, those defined by:

  — ISO 14721;

  — ISO/TR 15801, [10];

  — ISO/TR 18492, [12];

  — ISO 18509-1, [13];

  — ISO 18509-2, [14];

— quality assurance processes necessary to verify conformance with applicable requirements and conditions; e.g. an inspection regime to verify the quality and integrity of converted source data.

This part of ISO 19005 should lead to the development of various applications that read, render, write and validate conforming files. Different applications will incorporate various capabilities to prepare, interpret and process conforming files based on needs as perceived by the suppliers of those applications. However, it is important to note that a conforming application must be able to read and process appropriately all files complying with a specified conformance level.

This document has been created as Part 1 of ISO 19005 to allow the creation of future parts, which can provide compatibility with future versions of the underlying PDF specification without rendering this document or applications based on PDF Version 1.4 obsolete.

The following terms, referring to this specification or parts thereof, are recommended when referring to this specification when the full ISO name is not being used:

— “PDF/A” – a synonym for the ISO 19005 family of standards;

— “PDF/A-1” – a synonym for ISO 19005-1;

— “PDF/A-1a” – a synonym for ISO 19005-1 Level A conformance;

— “PDF/A-1b” – a synonym for ISO 19005-1 Level B conformance.

This part of ISO 19005, in conjunction with *PDF Reference* and *XMP Specification*, January 2004, provides sufficient information to interpret any conforming PDF/A file. *PDF Reference* contains a statement from Adobe Systems Incorporated concerning its intellectual property and its willingness to allow perpetual, royalty-free, non-exclusive use of that property in order to promote the use of PDF. Adobe has provided ISO with a similar statement relating to *XMP Specification*. In general, anyone may use *PDF Reference* and *XMP Specification* to create applications that read, write or otherwise process PDF/A files.

Patent claims regarding applications that read, render, write or otherwise process PDF/A files are outside the scope of this part of ISO 19005.

NPES and AIIM (accredited standards developing organizations) maintain an ongoing series of application notes for guiding developers and users of this part of ISO 19005. These application notes are available at <http://www.npes.org/standards/toolspdfa.html> and <http://www.aiim.org/pdfa/app-notes>. Both NPES and AIIM will also retain copies of the specific non-ISO normative references of this part of ISO 19005 which are publicly available electronic documents.

Document management — Electronic document file format for long-term preservation —

Part 1: Use of PDF 1.4 (PDF/A-1)

1 Scope

This part of ISO 19005 specifies how to use the Portable Document Format (PDF) 1.4 for long-term preservation of electronic documents. It is applicable to documents containing combinations of character, raster and vector data.

This part of ISO 19005 does not apply to:

— specific processes for converting paper or electronic documents to the PDF/A 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/IEC 646, *Information technology — ISO 7-bit coded character set for information interchange*

NOTE 1 The character encoding defined in ISO/IEC 646 is equivalent to ANSI X3.4 (ASCII) [1] and ECMA-6 [2].

ISO/IEC 9541-1, *Information technology — Font information interchange — Part 1: Architecture*

ISO/IEC 10646-1, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

NOTE 2 The character code values defined in ISO/IEC 10646-1 are equivalent to those of Unicode [22].

ISO 14721, *Space data and information transfer systems — Open archival information system — Reference model*

ISO 15930-4, *Graphic technology — Prepress digital data exchange using PDF — Part 4: Complete exchange of CMYK and spot colour printing data using PDF 1.4 (PDF/X-1a)*

Date and Time Formats, W3C Note, 15 September 1997. Available from Internet <http://www.w3.org/TR/NOTE-datetime>

Errata for PDF Reference, third edition, 18 June 2003. Available from Internet <http://partners.adobe.com/asn/acrobat/docs/PDF14errata.txt>

Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation, 4 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-xml-20040204>

ICC.1:1998-09, *File Format for Color Profiles*, International Color Consortium. Available from Internet <http://www.color.org/ICC-1_1998-09.PDF>

ICC.1A:1999-04, *Addendum 2 to Spec. ICC.1:1998-09*, International Color Consortium. Available from Internet <http://www.color.org/ICC-1A_1999-04.PDF>

PDF Reference: Adobe Portable Document Format, Version 1.4, Adobe Systems Incorporated – 3rd ed. (ISBN 0-201-75839-3). Available from Internet <http://partners.adobe.com/asn/acrobat/docs/File_Format_Specifications/PDFReference.pdf>

RDF/XML Syntax Specification (Revised), W3C Recommendation, 10 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>

Tags for the Identification of Languages, RFC 1766, March 1995. Available from Internet <http://www.ietf.org/rfc/rfc1766.txt>

XMP Specification, January 2004, Adobe Systems Incorporated. Available from Internet <http://partners.adobe.com/asn/tech/xmp/pdf/xmpspecification.pdf>

NOTE 3 AIIM and NPES (accredited standards developing organizations) maintain copies of the non-ISO references that are publicly available electronic documents.

3 Terms and definitions

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

3.1
conformance level
identified set of restrictions and requirements to which files and readers must comply

[ISO 15930-4]

3.2
cross reference table
PDF data structure that contains the byte offset of the start of indirect objects within the file

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

[ISO 15930-4]

3.4
electronic document
electronic representation of a page-oriented aggregation of text and graphic data, and metadata useful to identify, understand and render that data, that can be reproduced on paper or optical microform without significant loss of its information content

3.5
end-of-file marker
five character sequence **%%EOF** marking the end of a PDF file

3.6
end-of-line marker
EOL marker
one or two character sequence marking the end of a line of text, consisting of a **CARRIAGE RETURN** character (0Dh) or a **LINE FEED** character (0Ah) or a **CARRIAGE RETURN** followed immediately by a **LINE FEED**

3.7
font
identified collection of graphics that may be glyphs or other graphic elements

[ISO 15930-4]

3.8
glyph
recognizable abstract graphic symbol that is independent of any specific design

[ISO/IEC 9541-1]

3.9
ICC profile
colour profile conforming to the ICC specification and its addendum

[ICC.1:1998-09] and [ICC.1A:1999-04]

3.10
interactive reader
reader that requires or allows human interaction 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.

3.11
Level A conformance
conformance level encompassing all requirements of this part of ISO 19005

3.12
Level B conformance
conformance level encompassing the requirements of this part of ISO 19005 regarding the visual appearance of electronic documents, but not their structural or semantic properties

3.13
long-term
period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository, which may extend into the indefinite future

[ISO 14721]

3.14
PDF
Portable Document Format
file format defined in *PDF Reference* and its *Errata*

[ISO 15930-4]

3.15
reader
software application that is able to read and process files appropriately

[ISO 15930-4]

3.16
space character
text string character used to represent orthographic white space in the operands of text-showing operators

NOTE Commonly used space characters include **HORIZONTAL TABULATION** (U+0009), **LINE FEED** (U+000A), **VERTICAL TABULATION** (U+000B), **FORM FEED** (U+000C), **CARRIAGE RETURN** (U+000D), **SPACE** (U+0020), **NO-BREAK SPACE** (U+00A0), **EN SPACE** (U+2002), **EM SPACE** (U+2003), **FIGURE SPACE** (U+2007), **PUNCTUATION SPACE** (U+2008), **THIN SPACE** (U+2009), **HAIR SPACE** (U+200A), **ZERO WIDTH SPACE** (U+200B), and **IDEOGRAPHIC SPACE** (U+3000).

3.17
white-space character
NULL (00h), **HORIZONTAL TABULATION** (09h), **LINE FEED** (0Ah), **FORM FEED** (0Ch), **CARRIAGE RETURN** (0Dh) or **SPACE** (20h) character

3.18
writer
software application that is able to write files

[ISO 15930-4]

3.19
XMP packet
structured wrapper for serialized XMP metadata that can be embedded in a wide variety of file formats

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.

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

Token characters used to delimit objects and describe the structure of PDF files, as defined in *PDF Reference* 3.1, 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 **CARRIAGE RETURN** (0Dh).

Text string characters in content streams, as defined by *PDF Reference* 3.8.1, may be identified by their ISO/IEC 10646-1 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 **EN SPACE** (U+2002).

For the purposes of this part of ISO 19005, references to the “*PDF Reference*” are to *PDF Reference: Adobe Portable Document Format*, version 1.4, 3rd ed., as amended by *Errata for PDF Reference*, 3rd ed.

5 Conformance levels

5.1 General

This part of ISO 19005 defines a file format for representing electronic documents known as “PDF/A-1.” Conforming PDF/A-1 files shall adhere to all requirements of *PDF Reference* as modified by this part of ISO 19005. A conforming file may include any valid *PDF Reference* feature that is not explicitly forbidden by this part of ISO 19005. Features described in PDF specifications prior to Version 1.4 which are not explicitly described in *PDF Reference* should not be used. Neither the version number in the header of a PDF file nor

the value of the **Version** key in the document catalog dictionary shall be used in determining whether a file is in accordance with this part of ISO 19005.

NOTE 1 A conforming file is not obligated to use any PDF feature other than those explicitly required by *PDF Reference* or this part of ISO 19005.

NOTE 2 The proper mechanism by which a file can presumptively identify itself as being a PDF/A-1 file of a given conformance level is described in 6.7.11.

5.2 Level A conformance

Level A conforming files shall adhere to all of the requirements of this part of ISO 19005. A file meeting this conformance level is said to be a “conforming PDF/A -1a file.”

5.3 Level B conformance

In recognition of the varying preservation needs of the diverse user communities making use of PDF files, this part of ISO 19005 defines a Level B conformance level. Level B conforming files shall adhere to all of the requirements of this part of ISO 19005 except those of 6.3.8 and 6.8. A file meeting this conformance level is said to be a “conforming PDF/A-1b file.”

NOTE The Level B conformance requirements are intended to be those minimally necessary to ensure that the rendered visual appearance of a conforming file is preservable over the long term. However, Level B conforming files might not have sufficiently rich internal information to allow for the preservation of the document's logical structure and content text stream in natural reading order, which is provided by Level A conformance. The requirements for Level A conformance place greater responsibilities on writers of conforming files and those preparing such files, but these requirements allow for a higher level of document preservation service and confidence over time. Additionally, Level A conformance facilitates the accessibility of conforming files for physically impaired users.

5.4 Conforming readers

A conforming reader shall comply with all requirements regarding reader functional behaviour specified in this part of ISO 19005. The requirements of this part of ISO 19005 with respect to reader behaviour are stated in terms of general functional requirements applicable to all conforming readers. This part of ISO 19005 does not prescribe any specific technical design, user interface or implementation details of conforming readers.

The rendering of conforming files shall be performed as defined in *PDF Reference* subject to the further requirements specified by this part of ISO 19005. Features described in PDF specifications prior to Version 1.4 that are not explicitly described in *PDF Reference* may be ignored by conforming readers.

Conforming readers shall read and process appropriately all PDF/A-1 files complying with a specified conformance level. Level A conforming readers shall read and process appropriately all Level A and B conforming files. Level B conforming readers shall read and process appropriately all Level B conforming files.

6 Technical requirements

6.1 File structure

6.1.1 General

6.1.2 to 6.1.13 address overall file format issues and the base elements that form the general structure of a conforming file.

6.1.2 File header

The **%** character of the file header shall occur at byte offset 0 of the file.

The file header line shall be immediately followed by a comment consisting of a **%** character followed by at least four characters, each of whose encoded byte values shall have a decimal value greater than 127.

NOTE The presence of encoded character byte values greater than decimal 127 near the beginning of a file is used by various software tools and protocols to classify the file as containing 8-bit binary data that should be preserved during processing.

6.1.3 File trailer

The file trailer dictionary shall contain the **ID** keyword. The keyword **Encrypt** shall not be used in the trailer dictionary. No data shall follow the last end-of-file marker except a single optional end-of-line marker.

The file trailer referred to is either the last trailer dictionary in a PDF file, as described in *PDF Reference* 3.4.4 and 3.4.5, or the first page trailer in a linearized PDF file, as described in *PDF Reference* F.2. In a linearized file the **ID** keyword shall be present in both the first page trailer and the last trailer dictionaries and the value of both instances of the keyword shall be identical.

NOTE The explicit prohibition of the **Encrypt** keyword has the implicit effect of disallowing encryption and password-protected access permissions.

6.1.4 Cross reference table

In a cross reference subsection header the starting object number and the range shall be separated by a single **SPACE** character (20h).

The **xref** keyword and the cross reference subsection header shall be separated by a single EOL marker.

Any object whose offset is not referenced in the cross reference table shall be exempt from all requirements of this part of ISO 19005.

6.1.5 Document information dictionary

A document information dictionary may be defined in a conforming file. If defined, its elements shall be consistent with analogous XMP metadata properties as specified in 6.7.3.

6.1.6 String objects

Hexadecimal strings shall contain an even number of non-white-space characters, each in the range **0** to **9**, **A** to **F** or **a** to **f**.

6.1.7 Stream objects

The **stream** keyword shall be followed either by a **CARRIAGE RETURN** (0Dh) and **LINE FEED** (0Ah) character sequence or by a single **LINE FEED** character. The **endstream** keyword shall be preceded by an EOL marker.

The value of the **Length** key specified in the stream dictionary shall match the number of bytes in the file following the **LINE FEED** character after the **stream** keyword and preceding the EOL marker before the **endstream** keyword.

NOTE 1 These requirements remove potential ambiguity regarding the ending of stream content.

A stream object dictionary shall not contain the **F**, **FFilter**, or **FDecodeParams** keys.

NOTE 2 These keys are used to point to document content external to the file. The explicit prohibition of these keys has the implicit effect of disallowing external content that can create external dependencies and complicate preservation efforts.

6.1.8 Indirect objects

The object number and generation number shall be separated by a single white-space character. The generation number and **obj** keyword shall be separated by a single white-space character.

The object number and **endobj** keyword shall each be preceded by an EOL marker. The **obj** and **endobj** keywords shall each be followed by an EOL marker.

6.1.9 Linearized PDF

Linearization shall be permitted but any linearization information supplied within a file should be ignored by conforming readers.

6.1.10 Filters

The **LZWDecode** filter shall not be permitted.

NOTE The use of the LZW compression algorithm has been subject to intellectual property constraints.

6.1.11 Embedded files

A file specification dictionary, as defined in *PDF* 3.10.2, shall not contain the **EF** key. A file's name dictionary, as defined in *PDF Reference* 3.6.3, shall not contain the **EmbeddedFiles** key.

NOTE These keys are used to encapsulate files containing arbitrary content within a PDF file. The explicit prohibition of these keys has the implicit effect of disallowing embedded files that can create external dependencies and complicate preservation efforts.

6.1.12 Implementation limits

A conforming file shall not violate any of the architectural limits specified in *PDF Reference* Table C.1.

NOTE By complying with these limits, a conforming file is compatible with the widest possible range of readers.

6.1.13 Optional content

The document catalog dictionary shall not contain a key with the name **OCProperties**.

NOTE The explicit prohibition of the **OCProperties** key, which is allowed in *PDF 1.5* [19], has the implicit effect of disallowing optional content that generates alternative renderings of a document.

6.2 Graphics

6.2.1 General

6.2.2 to 6.2.10 describe restrictions placed on both conforming files and readers. They are intended to address graphical rendering issues that do not involve fonts and interactive elements.

6.2.2 Output intent

A conforming file may specify the colour characteristics of the device on which it is intended to be rendered by using a PDF/A-1 OutputIntent. A PDF/A-1 OutputIntent is an OutputIntent dictionary, as defined by *PDF Reference* 9.10.4, that is included in the file's **OutputIntents** array and has *GTS_PDFA1* as the value of its **S** key and a valid ICC profile stream as the value its **DestOutputProfile** key.

If a file's **OutputIntents** array contains more than one entry, then all entries that contain a **DestOutputProfile** key shall have as the value of that key the same indirect object, which shall be a valid ICC profile stream.

6.2.3 Colour spaces

6.2.3.1 General

All colours shall be specified in a device-independent manner, either directly by the use of a device-independent colour space, or indirectly by the use of an **OutputIntent**. A conforming file may use any colour space specified in *PDF Reference*, except as restricted in 6.2.3.2 to 6.2.3.4.

NOTE Specifying colour in the device-independent manner described within 6.2.3 enables predictable colour rendering based on a colorimetric definition and without reliance on assumptions or information external to the conforming file. It also provides a mechanism whereby a colorimetric definition can be associated with device-dependent colour data.

6.2.3.2 ICCBased colour spaces

All **ICCBased** colour spaces shall be embedded as ICC profile streams as described in *PDF Reference* 4.5.

A conforming reader shall render **ICCBased** colour spaces as specified by the ICC specification, and shall not use the **Alternate** colour space specified in an ICC profile stream dictionary.

6.2.3.3 Uncalibrated colour spaces

A conforming file may use either the **DeviceRGB** or **DeviceCMYK** colour space but shall not use both. If an uncalibrated colour space is used in a file then that file shall contain a PDF/A-1 OutputIntent, as defined in 6.2.2. **DeviceRGB** may be used only if the file has a PDF/A-1 OutputIntent that uses an **RGB** colour space. **DeviceCMYK** may be used only if the file has a PDF/A-1 OutputIntent that uses a **CMYK** colour space.

When rendering a **DeviceGray** colour specification in a file whose **OutputIntent** is an **RGB** profile, a conforming reader shall convert the **DeviceGray** colour specification to **RGB** by the method described in *PDF Reference* 6.2.1.

When rendering a **DeviceGray** colour specification in a file whose **OutputIntent** is a **CMYK** profile, a conforming reader shall convert the **DeviceGray** colour specification to **DeviceCMYK** by the method described in *PDF Reference* 6.2.2.

When rendering colours specified in a device-dependent colour space a conforming reader shall use the file's PDF/A-1 OutputIntent dictionary, as defined in 6.2.2, as the source colour space.

6.2.3.4 Separation and DeviceN colour spaces

A conforming reader shall obey the following rules when rendering colour spaces based on **DeviceN** or **Separation** colour spaces.

— If the named colourants in the colour space are all from the list **Cyan**, **Magenta**, **Yellow**, **Black**, the file has an **OutputIntent**, and that **OutputIntent** is a **CMYK** profile, then the colourants shall be treated as components of the colour space specified by the PDF/A-1 OutputIntent dictionary, as defined in 6.2.2, and the alternate colour space shall not be used.

— If the output device does not support the **Separation** colour space or **DeviceN** colourants, the **Alternate** colour space shall be used.

The **Alternate** colour space of a **Separation** or **DeviceN** colour space shall obey all restrictions on colour spaces specified in 6.2.3.2 and 6.2.3.3.

6.2.4 Images

An Image dictionary shall not contain the **Alternates** key or the **OPI** key.

If an Image dictionary contains the **Interpolate** key, its value shall be *false*.

Use of the **Intent** key shall conform to the rules given in 6.2.9.

6.2.5 Form XObjects

A form XObject dictionary shall not contain any of the following:

— the **OPI** key;

— the **Subtype2** key with a value of *PS*;

— the **PS** key.

NOTE In earlier versions of PDF the **Subtype2** key with a value of *PS* and the **PS** key were used to define arbitrary executable PostScript code streams, which have the potential to interfere with reliable and predictable rendering.

6.2.6 Reference XObjects

A conforming file shall not contain any reference XObjects.

NOTE Reference XObjects refer to arbitrary document content in external PDF files, creating external dependencies that complicate preservation efforts.

6.2.7 PostScript XObjects

A conforming file shall not contain any PostScript XObjects.

NOTE PostScript XObjects contain arbitrary executable PostScript code streams that have the potential to interfere with reliable and predictable rendering.

6.2.8 Extended graphics state

An ExtGState dictionary shall not contain the **TR** key. An ExtGState dictionary shall not contain the **TR2** key with a value other than *Default*. A conforming reader may ignore any instance of the **HT** key in an ExtGState dictionary.

Use of the **RI** key shall conform to the rules of 6.2.9.

6.2.9 Rendering intents

Where a rendering intent is specified, its value shall be one of the four values defined in *PDF Reference* *RelativeColorimetric*, *AbsoluteColorimetric*, *Perceptual* or *Saturation*.

NOTE The default rendering intent is *RelativeColorimetric*.

6.2.10 Content streams

A content stream shall not contain any operators not defined in *PDF Reference* even if such operators are bracketed by the **BX**/**EX** compatibility operators.

Use of the **ri** operator shall conform to the rules of 6.2.9.

NOTE 1 Content streams are used for page descriptions, e.g. the **Contents** stream of a page object or the stream of a form XObject, as well as for the appearance stream of annotations, including form fields or **Widget** annotations.

NOTE 2 In earlier versions of the PDF format a PostScript operator **PS** was defined. As this operator is not defined in *PDF Reference* its use is implicitly prohibited by 6.2.10.

6.3 Fonts

6.3.1 General

The intent of the requirements in 6.3.2 to 6.3.8 is to ensure that future rendering of the textual content of a conforming file matches, on a glyph by glyph basis, the static appearance of the file as originally created and to allow the recovery of semantic properties for each character of the textual content.

6.3.2 Font types

All fonts used in a conforming file shall conform to the font specifications defined in *PDF Reference* 5.5.

For the purposes of this part of ISO 19005, multiple master fonts shall be considered a special case of Type 1 fonts; any requirement explicitly stated with regard to Type 1 fonts shall be implicitly required with regard to multiple master fonts.

NOTE It is the responsibility of the writer to ensure the conformance of all fonts. This part of ISO 19005 does not prescribe the manner in which font conformance is determined.

6.3.3 Composite fonts

6.3.3.1 General

For any given composite (Type 0) font referenced within a conforming file, the **CIDSystemInfo** entries of its CIDFont and CMap dictionaries shall be compatible, as described in *PDF Reference* 5.6.2; in other words, the **Registry** and **Ordering** strings of the CIDSystemInfo dictionaries for that font shall be identical, unless the value of the CMap dictionary **UserCMap** key is *Identity-H* or *Identity-V*.

6.3.3.2 CIDFonts

For all Type 2 **CIDFonts**, the CIDFont dictionary shall contain a **CIDToGIDMap** entry that shall be a stream mapping from CIDs to glyph indices or the name *Identity*, as described in *PDF Reference* Table 5.13.

6.3.3.3 CMaps

All CMaps used within a conforming file, except *Identity-H* and *Identity-V*, shall be embedded in that file as described in *PDF Reference* 5.6.4. For those CMaps that are embedded, the integer value of the **WMode** entry in the CMap dictionary shall be identical to the **WMode** value in the embedded CMap stream.

6.3.4 Embedded font programs

The font programs for all fonts used within a conforming file shall be embedded within that file, as defined in *PDF Reference* 5.8, except when the fonts are used exclusively with text rendering mode 3. A font is considered to be used if any of its glyphs are referenced in any of the following contexts:

— the **Contents** stream of a page object;

— the stream of a Form XObject;

— the appearance stream of an annotation, including form fields;

— the content stream of a Type 3 font glyph;

— the stream of a tiling pattern.

Only fonts that are legally embeddable in a file for unlimited, universal rendering shall be used.

All conforming readers shall use the embedded fonts, rather than other locally resident, substituted or simulated fonts, for rendering.

NOTE 1 As discussed in *PDF Reference* 5.2.5, text rendering mode 3 specifies that glyphs are not stroked, filled or used as a clipping boundary. A font referenced for use solely in this mode is therefore not rendered and is thus exempt from the embedding requirement.

NOTE 2 There is no exemption from the requirements of 6.3.4 for the 14 standard Type 1 fonts. Type 3 fonts are exempt from the requirements of 6.3.4 because the manner in which Type 3 fonts are defined ensures that they are always embedded within PDF files, although the mechanism used to embed them differs from that of *PDF Reference* 5.8.

NOTE 3 The requirements for font program metadata are described in 6.7.10.

NOTE 4 As stated in 6.3.5, font subsets are acceptable as long as the embedded font programs provide glyph definitions for all characters referenced within the file. Embedding the font programs allows any conforming reader to reproduce correctly all glyphs in the manner in which they were originally published without reference to possibly ephemeral external resources.

NOTE 5 This part of ISO 19005 precludes the embedding of fonts whose legality depends upon special agreement with the font copyright holder. Such an allowance places unacceptable burdens on an archive to verify the existence, validity and longevity of such claims.

6.3.5 Font subsets

As stated in 6.3.4, embedded font programs shall define all font glyphs referenced for rendering with a conforming file. Type 0 **CIDFont** and Type 1 and TrueType font subsets, as described in *PDF Reference* 5.5.3, may be used as long as the embedded font programs define all appropriate glyphs.

For all Type 1 font subsets referenced within a conforming file, the font descriptor dictionary shall include a **CharSet** string listing the character names defined in the font subset, as described in *PDF Reference* Table 5.18.

For all **CIDFont** subsets referenced within a conforming file, the font descriptor dictionary shall include a **CIDSet** stream identifying which CIDs are present in the embedded **CIDFont** file, as described in *PDF Reference* Table 5.20.

NOTE The use of font subsets allows a potentially substantial reduction in the size of conforming files.

6.3.6 Font metrics

For every font embedded in a conforming file, the glyph width information stored in the **Widths** entry of the font dictionary and in the embedded font program shall be consistent.

NOTE This requirement is necessary to ensure predictable font rendering, regardless of whether a given reader uses the metrics in the **Widths** entry or those in the font program.

6.3.7 Character encodings

All non-symbolic TrueType fonts shall specify *MacRomanEncoding* or *WinAnsiEncoding* as the value of the **Encoding** entry in the font dictionary. All symbolic TrueType fonts shall not specify an **Encoding** entry in the font dictionary, and their font programs' “cmap” tables shall contain exactly one encoding.

NOTE This requirement makes normative the suggested guidelines described in *PDF Reference* 5.5.5.

6.3.8 Unicode character maps

6.3.8 is applicable only for files meeting Level A conformance. For Level B conformance the requirements of 6.3.8 can be ignored.

The font dictionary shall include a **ToUnicode** entry whose value is a CMap stream object that maps character codes to Unicode values [22], as described in *PDF Reference* 5.9, unless the font meets any of the following three conditions:

— fonts that use the predefined encodings **MacRomanEncoding**, **MacExpertEncoding** or **WinAnsiEncoding**, or that use the predefined **Identity-H** or **Identity-V** CMaps;

— Type 1 fonts whose character names are taken from the Adobe standard Latin character set or the set of named characters in the Symbol font, as defined in *PDF Reference* Appendix D;

— Type 0 fonts whose descendant **CIDFont** uses the **Adobe-GB1**, **Adobe-CNS1**, **Adobe-Japan1** or **Adobe-Korea1** character collections.

NOTE Unicode mapping allows the retrieval of semantic properties about every character referenced in the file.

6.4 Transparency

If an **SMask** key appears in an **ExtGState** or **XObject** dictionary, its value shall be *None*.

A **Group** object with an **S** key with a value of *Transparency* shall not be included in a form **XObject**.

The following keys, if present in an **ExtGState** object, shall have the values shown:

— **BM** *Normal* or *Compatible*

— **CA** *1.0*

— **ca** *1.0*

NOTE These provisions prohibit the use of transparency within a conforming file. The visual effect of partially transparent graphics can be achieved using techniques other than the use of the *PDF Reference* transparency keys, including pre-rendered data or flattened vector objects. The use of such techniques does not prevent a file from being PDF/A-1 conformant.

6.5 Annotations

6.5.1 General

In addition to the rendering behaviour defined by *PDF Reference* as modified by this part of ISO 19005, conforming interactive readers shall provide a mechanism to display the values of the **Contents** key of annotation dictionaries.

NOTE This part of ISO 19005 does not prescribe the specific behaviour or technical implementation details that interactive readers may use to implement this functional requirement.

6.5.2 Annotation types

Annotation types not defined in *PDF Reference* shall not be permitted. Additionally, the **FileAttachment**, **Sound** and **Movie** types shall not be permitted.

NOTE Support for multimedia content is outside the scope of this part of ISO 19005.

6.5.3 Annotation dictionaries

An annotation dictionary shall not contain the **CA** key with a value other than *1.0*.

An annotation dictionary shall contain the **F** key. The **F** key's **Print** flag bit shall be set to *1* and its **Hidden**, **Invisible** and **NoView** flag bits shall be set to *0*.

Text annotations should set the **NoZoom** and **NoRotate** flag bits of the **F** key to *1*.

NOTE 1 The restrictions on annotation flags prevent the use of annotations that are hidden or that are viewable but not printable. The **NoZoom** and **NoRotate** flags are permitted, which allows the use of annotation types that have the same behaviour as the commonly-used text annotation type. By definition, text annotations exhibit the **NoZoom** and **NoRotate** behaviour even if the flags are not set, as described in *PDF Reference* 8.4.5; explicitly setting these flags removes any potential ambiguity between the annotation dictionary settings and reader behaviour.

An annotation dictionary shall not contain the **C** array or the **IC** array unless the colour space of the **DestOutputProfile** in the PDF/A-1 OutputIntent dictionary, defined in 6.2.2, is *RGB*.

NOTE 2 These provisions ensure that the device colour spaces used in annotations by mechanisms other than an appearance stream are indirectly defined by means of the PDF/A-1 OutputIntent.

If an annotation dictionary contains the **AP** key, the appearance dictionary that it defines as its value shall contain only the **N** key, whose value shall be a stream defining the appearance of the annotation.

NOTE 3 All of the provisions of 6.5.3 apply to all annotation types, including the **Widget** type used for form fields.

6.6 Actions

6.6.1 General

The **Launch**, **Sound**, **Movie**, **ResetForm**, **ImportData** and **JavaScript** actions shall not be permitted. Additionally, the deprecated **set-state** and **no-op** actions shall not be permitted. Named actions other than **NextPage**, **PrevPage**, **FirstPage**, and **LastPage** shall not be permitted. In response to each of the four allowed named actions, conforming interactive readers shall perform the appropriate action described in *PDF Reference* Table 8.45.

Interactive form fields shall not perform actions of any type.

NOTE 1 Support for multimedia content is outside the scope of this part of ISO 19005. The **ResetForm** action changes the rendered appearance of a form. The **ImportData** action imports form data from an external file. **JavaScript** actions permit an arbitrary executable code that has the potential to interfere with reliable and predictable rendering.

NOTE 2 Additional requirements for interactive form fields are specified in 6.9.

6.6.2 Trigger events

A Widget annotation dictionary or Field dictionary shall not include an **AA** entry for an additional-actions dictionary. The document catalog dictionary shall not include an **AA** entry for an additional-actions dictionary.

NOTE These additional-actions dictionaries define arbitrary JavaScript actions. The explicit prohibition of the **AA** entry has the implicit effect of disallowing JavaScript actions that can create external dependencies and complicate preservation efforts.

6.6.3 Hypertext links

Conforming interactive readers may choose to make hyperlinks non-actionable, but in addition to the rendering behaviour defined by *PDF Reference*, as modified by this part of ISO 19005, they shall provide a mechanism to display the **F** and **D** keys of a **GoToR** action dictionary, the **URI** key of a **URI** action dictionary, and the **F** key of a **SubmitForm** action dictionary.

NOTE Since hyperlinks transfer the thread of execution outside the control of an interactive reader, this subclause allows an interactive reader to choose to make them not actionable. For purposes of archival disclosure of the complete information content of conforming files it is important for interactive readers to provide some mechanism to expose the destination of all hyperlinks. However, this part of ISO 19005 does not prescribe any specific behaviour or the technical implementation details that interactive readers might use to meet the functional requirement of this subclause.

6.7 Metadata

6.7.1 General

6.7.2 to 6.7.11 specify requirements for metadata within conforming files. Metadata is essential for effective management of a file throughout its life cycle. A file depends on metadata for identification and description, as well as for describing appropriate technical and administrative matters. As a result, writers of conforming files may have to comply with various domain-specific metadata requirements defined external to this part of ISO 19005. This part of ISO 19005 outlines a structured, consistent framework that supports a broad variety of metadata requirements.

6.7.2 Properties

The document catalog dictionary of a conforming file shall contain the **Metadata** key. The metadata stream that forms the value of that key shall conform to *XMP Specification*. All metadata properties embedded in a file shall be in XMP form except for document information dictionary entries that have no XMP analogues, as defined in 6.7.3. Properties specified in XMP form shall use either the predefined schemas defined in *XMP Specification* 4, or extension schemas that comply with *XMP Specification* 4, and 6.7.8. Metadata object stream dictionaries shall not contain the **Filter** key.

NOTE 1 The explicit prohibition of the **Filter** key has the implicit effect of preserving the contents of XMP metadata streams as plain text that is visible to non-PDF aware tools.

NOTE 2 An extension schema is any XMP schema that is not defined in *XMP Specification*.

6.7.3 Document information dictionary

A document information dictionary may appear within a conforming file. If it does appear, then all of its entries that have analogous properties in predefined XMP schemas, as defined by Table 1, shall also be embedded in the file in XMP form with equivalent values. Any document information dictionary entry not listed in Table 1 shall not be embedded using a predefined XMP schema property.

NOTE 1 Since a document information dictionary is allowed within a conforming file, it is possible for a single file to be both PDF/A-1 (ISO 19005-1) and PDF/X (ISO 15930-4 and ISO 15930-6 [11]) conformant.

Table 1 — Crosswalk between document information dictionary and XMP properties

Document information dictionary		XMP	
Entry	PDF type	Property	XMP type
Title	*text string*	**dc:title**	*Text*
Author	*text string*	**dc:creator**	*seq Text*
Subject	*text string*	**dc:subject**	*Text*
Keywords	*text string*	**pdf:Keywords**	*Text*
Creator	*text string*	**xmp:CreatorTool**	*Text*
Producer	*text string*	**pdf:Producer**	*Text*
CreationDate	*date*	**xmp:CreateDate**	*Date*
ModDate	*date*	**xmp:ModifyDate**	*Date*

NOTE The XML namespace URI for the **dc** prefix is <http://purl.org/dc/elements/1.1/>; the namespace URI for the **pdf** prefix is <http://ns.adobe.com/pdf/1.3/>; and the namespace URI for the **xmp** prefix is <http://ns.adobe.com/xap/1.0/>.

The value of the document information dictionary entries and their analogous XMP properties shall be equivalent. For properties that map from the PDF *text string* type to the XMP *Text* type, value equivalence shall be on a character-by-character basis, independent of encoding, comparing the numeric ISO/IEC 10646-1 code points for the characters.

NOTE 2 The explicit requirement for equivalence between the values of document information dictionary entries and their analogous XMP properties has the implicit effect of providing unambiguous interpretation of that property's value.

If the **dc:creator** property is present in XMP metadata then it shall be represented by an ordered *Text* array of length one whose single entry shall consist of one or more names. Equivalence between **Author** and **dc:creator**, shall be on a character-by-character basis, independent of encoding, comparing the numeric ISO/IEC 10646-1 code points for the characters.

EXAMPLE 1 The document information dictionary entry:

```
/Author (Peter, Paul and Mary)
```

is equivalent to the XMP property:

```
<dc:creator>
  <rdf:Seq>
    <rdf.:li>Peter, Paul, and Mary</rdf:li>
  </rdf:Seq>
</dc:creator>
```

Date properties are formatted as a variable-length sequence of temporal components ranging in granularity: year, month, day, hour, minute, second. For properties that map between the PDF *date* type, defined by *PDF Reference* 3.8.2, and the XMP *Date* type, defined by *Date and Time Format*s, value equivalence shall be on a component-by-component basis, relative to Coordinated Universal Time (UTC), i.e., correcting for local time zone offset.

EXAMPLE 2 The document information dictionary entries:

```
/CreationDate (D:20040402)
/ModDate (D:20040408091132-05'00')
```

are equivalent to the XMP properties:

```
<xmp:CreateDate>2004-04-02</xmp:CreateDate>
<xmp:ModifyDate>2004-04-08T14:11:32Z</xmp:ModifyDate>
```

6.7.4 Normalization

All XMP schemas should define the normalization rules that are applicable for their properties. For all metadata properties defined in schemas that do provide normalization rules, the property values shall be entered, saved and retained in the normalized fashion defined by those schemas to facilitate interchange and support consistent interpretation of metadata by conforming readers.

6.7.5 XMP header

The **bytes** and the **encoding** attributes shall not be used in the header of an XMP packet.

NOTE Both the **bytes** and **encoding** attributes are deprecated in *XMP Specification*.

6.7.6 File identifiers

A conforming file should have one or more metadata properties to characterize, categorize and otherwise identify the file. This part of ISO 19005 does not mandate any specific identification scheme. Identifiers may be externally based, such as an International Standard Book Number (ISBN)[4] or a Digital Object Identifier (DOI), or internally based, such as a Globally Unique Identifier/Universally Unique Identifier (GUID/UUID) or another designation assigned during workflow operations. Identifiers may be included through use of the

xmp:Identifier property; use of the **xmpMM:DocumentID**, **xmpMM:VersionID** and **xmpMM:RenditionClass** properties; or use of properties from an extension schema. Any identification system may be used so long as the properties comply with XMP requirements and this part of ISO 19005.

If a conforming file is changed in any way, even if only by the addition of an **xmpMM:History** entry as described in 6.7.7, then the changing identifier part of the file trailer dictionary **ID** key should be modified as described in *PDF Reference* 9.3.

NOTE The XML namespace URI for the **xmp** prefix is <http://ns.adobe.com/xap/1.0/>; the namespace URI for the **xmpMM** prefix is <http://ns.adobe.com/xap/1.0/mm/>.

6.7.7 File provenance information

In order to describe all high-level user actions taken to create, transform or otherwise instantiate a conforming file, each of those actions should be recorded in the **xmpMM:History** property. For each action that is recorded:

— the **action**, **parameters** and **when** fields shall be specified;

— the **softwareAgent** field should be specified;

— the **instanceID** field shall not be specified.

NOTE 1 The XML namespace URI for the prefix **xmpMM** is <http://ns.adobe.com/xap/1.0/mm/>.

NOTE 2 Applications with specific auditing requirements may need to record additional types of action or additional details about actions beyond those defined by predefined XMP schemas. Examples of additional types of action include those that change the appearance of the document, such as downsampling or font substitution. Examples of additional details include the identity of the human agent that instigated or performed the action or the environment in which the action occurred.

In cases where original sources such as paper, microform or electronic files are transformed into conforming files, **xmpMM:History** should describe all high-level processing (e.g. transformed from PDF 1.4 to PDF/A-1); alterations to file content or functionality (e.g. embedded JavaScript and audio objects were not retained); handling of pre-existing metadata (e.g. all document information dictionary values converted to XMP); and any other significant aspects of the transformation process.

For all conforming files, whether created natively or by conversion from sources such as paper, microform, or other electronic formats, **xmpMM:History** should describe all subsequent high-level workflow processes (e.g. descriptions of activities and handoffs); citations to policies governing file handling (e.g. titles of official directives under which files are collected, processed, and used); names and versions of software tools; any other matters that are needed to indicate the context of the file's creation and use.

In cases where XMP metadata properties have been changed or deleted as a file moves through its life cycle, **xmpMM:History** should describe those changes by including entries whose **parameters** fields specify the name of the properties and their previous values. This recommendation applies to all metadata properties except the **xmpMM:History** itself. If a metadata property has been deleted, the **action** field of its entry in **xmpMM:History** shall be *pdfa:deleted*.

6.7.8 Extension schemas

All extension schemas used in a conforming file shall have their descriptions embedded within that file in the metadata stream defined by 6.7.2. These descriptions shall be specified using the PDF/A extension schema description schema defined in this clause.

NOTE 1 An extension schema is any XMP schema that is not defined in *XMP Specification*.

The extension schema description schema defined in Table 2 uses the namespace URI <http://www.aiim.org/pdfa/ns/schema>. The required schema namespace prefix is **pdfaSchema**.

NOTE 2 According to the W3C XML Namespace recommendation [18], namespace URI's are for identification purposes only and are not required to be actionable links. None of the namespace URI's defined for XMP extension schemas in this part of ISO 19005 is an actionable link. Attempting to dereference or follow any of these links will not result in a valid web page.

Table 2 — PDF/A extension schema description schema

Property	Value type	Category	Description
pdfaSchema:schema	*Text*	External	Optional description of schema
pdfaSchema:namespaceURI	*URI*	External	Schema namespace URI
pdfaSchema:prefix	*Text*	External	Preferred schema namespace prefix
pdfaSchema:property	*seq Property*	Internal	Description of schema properties
pdfaSchema:valueType	*seq ValueType*	Internal	Description of schema-specific value types

The **Property** type defined in Table 3 is an XMP structure containing the description of a schema property. The field namespace URI is <http://www.aiim.org/pdfa/ns/property>. The required field namespace prefix is **pdfaProperty**.

Table 3 — PDF/A property type schema

Field name	Value type	Description
pdfaProperty:name	*Text*	Property name
pdfaProperty:valueType	*Open Choice of Text*	Value type of the property, drawn from *XMP Specification* 4, or an embedded PDF/A value type extension schema
pdfaProperty:category	*Closed Choice of Text*	Property category: *internal* or *external*
pdfaProperty:description	*Text*	Description of the property

The preferred values for **pdfaProperty:valueType** should be the non-deprecated property value types defined in *XMP Specification* 2004, 4. Array types shall be preceded by their container type: *alt*, *bag* or *seq*, separated from the base type by a single white-space character.

The **ValueType** type defined in Table 4 is an XMP structure containing the description of all property value types used by embedded extension schemas that are not defined in *XMP Specification* 4. The field namespace URI is <http://www.aiim.org/pdfa/ns/type>. The required field namespace prefix is **pdfaType**.

Table 4 — PDF/A value type schema

Field name	Value type	Description
pdfaType:type	*Text*	Property value type name
pdfaType:namespaceURI	*URI*	Property value type field namespace URI
pdfaType:prefix	*Text*	Preferred value type field namespace prefix
pdfaType:description	*Text*	Description of the property value type
pdfaType:field	*seq Field*	Optional description of the structured fields

The **Field** type defined in Table 5 is an XMP structure containing the description of a property value type field. The field namespace URI is <http://www.aiim.org/pdfa/ns/field>. The required field namespace prefix is **pdfaField**.

Table 5 — PDF/A field schema

Field name	Value type	Description
pdfaField:name	*Text*	Field name
pdfaField:valueType	*Open Choice of Text*	Field value type, drawn from *XMP Specification* 2004, 4, or an embedded PDF/A value type extension schema
pdfaField:description	*Text*	Field description

6.7.9 Validation

All content of all XMP packets shall be well-formed as defined by *Extensible Markup Language (XML) 1.0 (Third Edition)*, 2.1, and *RDF/XML Syntax Specification (Revised)*, 7. If possible, at the time a writer creates or resaves a conforming file all of the content of that file's XMP packets should be validated.

6.7.10 Font metadata

For all embedded Type 0, Type 1, or TrueType font programs, the embedded font file stream dictionary should include a **Metadata** entry whose value is an XMP metadata stream. The following XMP metadata elements should be supplied: **xmp:Title**, giving the value of the **FontName** key from the font's font descriptor dictionary; **xmpRights:Copyright**, giving the copyright statement; **xmpRights:Marked**, with the Boolean value *true*; **xmpRights:Owner**, giving the legal owner of the font; **xmpRights:UsageTerms**, giving a statement of the licensing terms under which the font is being used. Additional XMP metadata may be included at the discretion of the file writer.

NOTE 1 Font rights information is helpful in order to preserve the identity and scope of the intellectual property rights of the font copyright holder. While many fonts embed statements of copyright and licensing terms within the font itself, this is not a uniform practice. Therefore it is advantageous to require the explicit representation of rights statements in the conforming file. Even though this may be redundant, it obviates the necessity for some future system to have the ability to parse through the particular internal structure of font programs.

NOTE 2 The XML namespace URI for the **xmp** prefix is <http://ns.adobe.com/xap/1.0/>; the namespace URI for the **xmpRights** prefix is <http://ns.adobe.com/xap/1.0/rights/>.

6.7.11 Version and conformance level identification

The PDF/A version and conformance level of a file shall be specified using the PDF/A Identification extension schema defined in this subclause.

The Identification schema defined in Table 6 uses the namespace URI <http://www.aiim.org/pdfa/ns/id>. The required schema namespace prefix is **pdfaid**.

Table 6 — PDF/A identification schema

Property	Value type	Category	Description
pdfaid:part	*Open Choice of Integer*	Internal	PDF/A version identifier
pdfaid:amd	*Open Choice of Text*	Internal	Optional PDF/A amendment identifier
pdfaid:conformance	*Closed Choice of Text*	Internal	PDF/A conformance level: *A* or *B*

The value of **pdfaid:part** shall be the part number of ISO 19005 to which the file conforms. If the file conforms to a version of ISO 19005 that is defined by an amendment to a part, then the value of **pdfaid:amd** shall be the amendment number and year, separated by a colon.

A Level A conforming file shall specify the value of **pdfaid:conformance** as *A*. A Level B conforming file shall specify the value of **pdfaid:conformance** as *B*.

The values of the **pdfaid:part**, **pdfaid:amd**, and **pdfaid:conformance** properties do not by themselves determine conformance with a part of ISO 19005. The actual determination of conformance shall be performed as specified in Clause 5.

6.8 Logical structure

6.8.1 General

Subclause 6.8 is applicable only for files meeting Level A conformance. For Level B conformance the requirements of 6.8 can be ignored.

The intent of the requirements in 6.8.2 to 6.8.8 is to ensure the recovery of the textual content of a conforming file as a sequence of words defined in the natural reading order of the language in which they are written. Similarly, it ensures that the individual characters of each word are recoverable in their natural reading order. Furthermore, these requirements allow the recovery of higher-level semantic information concerning the logical structure of the document.

PDF/A-1 writers should not add structural or semantic information that is not explicitly or implicitly present in the source material solely for the purpose of achieving conformance. Examples of such information are structure hierarchy, natural language specification, alternative descriptions, non-textual annotations, replacement text and expansions of abbreviations and acronyms.

NOTE It is inadvisable for writers to generate structural or semantic information using automated processes without appropriate verification.

6.8.2 Tagged PDF

6.8.2.1 General

A Level A conforming file shall meet all of the requirements set forth for Tagged PDF in *PDF Reference* 9.7.

NOTE Tagged PDF defines conventions for explicitly declaring and describing the logical structural aspects of document content.

6.8.2.2 Mark information dictionary

The document catalog dictionary shall include a **MarkInfo** dictionary whose sole entry, **Marked**, shall have a value of *true*.

NOTE This setting indicates that the file conforms to the Tagged PDF conventions.

6.8.3 Artifacts

6.8.3.1 General

Pagination features such as running heads or page numbers, cosmetic layout features such as footnote rules or background screens, and production aids such as cut marks and colour bars should be specified as pagination, layout, and page artifacts, respectively, as described in *PDF Reference* 9.7.2.

6.8.3.2 Word breaks

For languages and script systems that normally use space characters to indicate word breaks, the following additional restriction shall apply:

Within show strings, word breaks shall be explicitly indicated by the presence of one or more space characters between all of the individual words in the show string. If a word ends at a show string boundary, one or more space characters shall be inserted at the end of the show string. Note that a single word may span two or more show strings; word breaks are indicated only by the explicit presence of one or more space characters, not by the boundaries of a show string. For the purposes of indicating word breaks, a sequence of two or more consecutive space characters is semantically equivalent to a single spacing character.

6.8.3.3 Structure hierarchy

The logical structure of the conforming file shall be described by a structure hierarchy rooted in the **StructTreeRoot** entry of the document catalog dictionary, as described in *PDF Reference* 9.6.

Each structure element dictionary in the structure hierarchy shall have a **Type** entry with the name value of **StructElem**.

Writers of conforming files should attempt to capture a document's logical structure hierarchy to the finest granularity possible, making use of the standard structure types for grouping elements, block-level structure elements, paragraph-like elements, list elements, table elements, inline-level structure elements, link elements and illustration elements, as defined in *PDF Reference* 9.7.4, to the fullest extent possible.

NOTE The explicit description of a document's logical structure will prove valuable to future efforts to recover the document's full semantic value for the purposes of rendering or migration to other data formats.

6.8.3.4 Structure types

The definition of block-level structuring elements should follow the strongly structured paradigm as described in *PDF Reference* 9.7.4.

All non-standard structure types shall be mapped to the nearest functionally equivalent standard type, as defined in *PDF Reference* 9.7.4, 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 must terminate at a standard type.

6.8.4 Natural language specification

The default natural language for all text in a file should be specified by the **Lang** entry in the document catalog dictionary.

All textual content within a file which differs from the default language should be indicated by use of a **Lang** property attached to a marked-content sequence, or by a **Lang** entry in a structure element dictionary, as described in *PDF Reference* 9.8.1.

If the **Lang** entry is present in the document catalog dictionary or in a structure element dictionary or property list, its value shall be a language identifier as defined by RFC 1766*, Tags for the Identification of Languages*, as described in *PDF Reference* 9.8.1.

All text strings encoded in Unicode whose language is not the default natural language for the file or not the natural language defined by the innermost enclosing structure element or marked-content sequence should indicate their language using the internal escape sequence described in *PDF Reference* 3.8.1.

NOTE The distinction between words foreign to a language and foreign words incorporated by common usage into a language is problematic. The intent of these requirements is to allow for future unambiguous semantic interpretation of textual content.

6.8.5 Alternate descriptions

All structure elements whose content does not have a natural predetermined textual analogue, e.g. images, formulae, etc., should supply an alternate text description using the **Alt** entry in the structure element dictionary, as described in *PDF Reference* 9.8.2.

NOTE Alternate descriptions provide textual descriptions that aid in the proper interpretation of otherwise opaque non-textual content.

6.8.6 Non-textual annotations

For annotation types that do not display text, the **Contents** key of an annotation dictionary should be specified with an alternative description of the annotation's contents in human-readable form.

6.8.7 Replacement text

All textual structure elements that are represented in a non-standard manner, e.g., custom characters or inline graphics, should supply replacement text using the **ActualText** entry in the structure element dictionary, as described in *PDF Reference* 9.8.3.

NOTE Replacement text provides textual equivalents that aid in the proper interpretation of otherwise opaque, unusual representations of textual components.

6.8.8 Expansions of abbreviations and acronyms

All instances of abbreviations and acronyms in textual content should be placed in a marked-content sequence with a **Span** tag whose **E** property provides a textual expansion of the abbreviation or acronym, as described in *PDF Reference* 9.8.4.

NOTE Abbreviation and acronym expansion provides textual equivalents that aid in the proper interpretation of otherwise opaque nomenclature.

6.9 Interactive Forms

The intent of the requirements of this subclause is to ensure that there is no ambiguity about the rendering of form fields.

A conforming reader shall not use form fields to change the rendered representation of the page or the content of the file at any time. A Widget annotation dictionary or Field dictionary shall not contain the **A** or **AA** keys.

The **NeedAppearances** flag of the interactive form dictionary shall either not be present or shall be *false*.

Every form field shall have an appearance dictionary associated with the field's data. A conforming reader shall render the field according to the appearance dictionary without regard to the form data.

NOTE Requiring an appearance dictionary ensures the reliable rendering of the form.

Annex A
(informative)

PDF/A-1 conformance summary

A.1 General

The information in A.1 to A.3 is provided as a convenience to gain a quick general overview of the scope of the variance of the PDF/A-1 requirements from those of *PDF Reference*. However, it is not intended as a comprehensive list of all PDF/A-1 requirements. The normative statement of all PDF/A-1 requirements is found in Clauses 2 to 6. In the event of any discrepancy between the information presented in this informative annex and the normative text, the normative text is always considered to present the definitive statement of the requirements.

Tables A.1 and A.2 list the PDF 1.4 operators, objects and keys within objects for which the requirements of this part of ISO 19005 vary from *PDF Reference* for the purposes of PDF/A-1 conformance. These tables indicate the status of the operator, object or key, and the normative clause where that status is defined. The following status values are used.

— *Required* The operator, object or key is required in conforming files.

— *Prohibited* The operator, object or key is prohibited from conforming files.

— *Restricted* The operator, object or key may appear in conforming files, but only subject to specific constraints on its use, contents or value.

— *Recommended* The operator, object or key should appear in conforming files.

— *Ignored* The operator, object or key may appear in conforming files but is ignored by conforming readers.

If a reference to a PDF dictionary object is included in tables, but keys within that object are not explicitly listed, then all keys within that object and its descendants, if any, inherit their status from the object that is shown in the table. An object is a descendant from another object, called its ancestor, if any of the following conditions are true:

— the object is the value of a key in the ancestor object;

— the ancestor is an array and the object is an element of that array;

— the object is a descendant of a descendant of the ancestor object.

A.2 Operators

All operators defined in *PDF Reference* for use in **Contents** streams may be included in a conforming file, subject to the conditions shown in Table A.1.

Table A.1 — Operator status

Operator	Status	Subclause
CS	*Restricted*	6.2.3
cs	*Restricted*	6.2.3
K	*Restricted*	6.2.3
k	*Restricted*	6.2.3
RG	*Restricted*	6.2.3
rg	*Restricted*	6.2.3
ri	*Restricted*	6.2.9
Operators not defined in *PDF Reference*	*Prohibited*	6.2.10

A.3 Objects and keys

All objects and keys defined in *PDF Reference* may be included in a conforming file, subject to the conditions shown in Table A.2. Some of the requirements for keys are relative to a specific key/value pair. In such cases the relevant value is shown following the key.

Table A.2 — Object and key status

Object	Key (and value)	Status	Subclause
AcroForm	**NeedAppearances**	*Restricted*	6.9
Action	**N *NOP***	*Prohibited*	6.6.1
	S *Named*	*Restricted*	6.6.1
	S *ImportData* **S *JavaScript*** **S *Launch*** **S *Movie*** **S *ResetForm*** **S *SetState*** **S *Sound***	*Prohibited*	6.6.1
Annot	**AA**	*Prohibited*	6.6.2
	CA	*Restricted*	6.5.3
	Contents	*Recommended* (for Level A conformance of non-textual annotations)	6.8.6
	Subtype *FileAttachment* **Subtype** *Sound* **Subtype** *Movie*	*Prohibited*	6.5.2
Artifact property list dictionary		*Recommended* (for Level A conformance)	6.8.3

Table A.2 (*continued*)

Object	Key (and value)	Status	Subclause
Catalog	**AA**	*Prohibited*	6.6.2
	Lang	*Recommended* (for Level A conformance)	6.8.4
	Metadata	*Required*	6.7.2
	Names	*Restricted*	6.1.11
	OCProperties	*Prohibited*	6.1.13
	OutputIntents	*Restricted*	6.2.2
	StructTreeRoot	*Recommended* (for Level A conformance)	6.8.3.3
CMap	**CIDSystemInfo**	*Restricted*	6.3.3.1
	WMode	*Restricted*	6.3.3.3
ExtGState	**BM**	*Restricted*	6.4
	CA	*Restricted*	6.4
	ca	*Restricted*	6.4
	HT	*Ignored*	6.2.8
	SMask	*Restricted*	6.4
	TR	*Prohibited*	6.2.8
	TR2	*Restricted*	6.2.8
Field dictionary	**AA**	*Prohibited*	6.6.2
Filespec	**EF**	*Prohibited*	6.1.11
Filters	**LZWDecode**	*Prohibited*	6.1.10
Font	**FontDescriptor**	*Required* (unless Type3)	6.3.4
	ToUnicode	*Required* [a] (for Level A conformance)	6.3.8
	Type	*Restricted*	6.3.2
	Widths	*Ignored*	6.3.6
Font **(Subtype** *CIDFontType0* or *CIDFontType2***)**	**CIDSystemInfo**	*Restricted*	6.3.3.1
Font **(Subtype** *CIDFontType2***)**	**CIDtoGIDMap**	*Restricted*	6.3.3.2
Font **(Subtype** *TrueType***)**	**Encoding**	*Prohibited* (if symbolic font) *Restricted* (if non-symbolic)	6.3.7
Font file stream	**Metadata**	*Recommended*	6.7.10
FontDescriptor	**CharSet**	*Required*	6.3.5
	CIDSet	*Required* (if CIDFont)	6.3.5
	FontFile or **FontFile2** or **FontFile3**	*Required*	6.3.5

Table A.2 (*continued*)

Object	Key (and value)	Status	Subclause
Group	**S**	*Restricted*	6.4
MarkInfo	**Marked *true***	*Required* (for Level A conformance)	6.8.2.2
Page	**AA**	*Prohibited*	6.6.2
PDF/A output intent dictionary	**DestOutputProfile**	*Restricted*	6.2.2
	S	*Restricted*	6.2.2
Span dictionary	**E**	*Recommended* (for Level A conformance)	6.8.8
	Lang	*Recommended* (for Level A conformance of non-default language content)	6.8.4
Stream dictionary	**Alternate**	*Ignored*	6.2.3.2
	F	*Prohibited*	6.1.7
	FDecodeParams	*Prohibited*	6.1.7
	FFilter	*Prohibited*	6.1.7
	ICCBased	*Restricted*	6.2.3.2
Structure element dictionary	**ActualText**	*Recommended* (for Level A conformance of non-standard elements)	6.8.7
	Alt	*Recommended* (for Level A conformance of non-textual elements)	6.8.5
	Type *StructElem*	*Required* (for Level A conformance)	6.8.3.3
Trailer	**Encrypt**	*Prohibited*	6.1.3
Trailer	**ID**	*Required*	6.1.3
XObject	**Subtype *PS***	*Prohibited*	6.2.5
XObject (**Subtype** *Form*)	**Group**	*Restricted*	6.4
	OPI	*Prohibited*	6.2.5
	Ref	*Prohibited*	6.2.6
XObject (**Subtype** *Image*)	**Alternates**	*Prohibited*	6.2.4
	Intent	*Restricted*	6.2.4
	OPI	*Prohibited*	6.2.4
	SMask	*Restricted*	6.4

[a] There are three specific exemptions from this status defined in 6.3.8.

Annex B
(informative)

Best practices for PDF/A

B.1 Use of non-XMP metadata

Use of non-XMP metadata at the file level is strongly discouraged as there is no assurance that such metadata can be preserved in accordance with this specification. In cases where non-XMP metadata are present, the preference is to convert them to XMP, embed them in the file, and describe the conversion in the **xmpMM:History** property. The **xmpMM:History** property should also be used to indicate any non-XMP elements that have not been converted.

Failure to preserve metadata may cause problems in locating, interpreting, managing, and authenticating a file in the future, which may in turn diminish or cancel its archival value.

B.2 Natural language identifiers

Languages should be identified using ISO 639-1[3], ISO 3166-1[5] or IANA[17] registered identifiers. Private use identifiers should be used only if the language does not have a defined identifier within ISO 639-1, ISO 3166-1 or IANA registry. In the event that a language is truly unknown, the identifier *x-unknown* should be used.

NOTE The use of ISO 639-1[3], ISO 3166-1[5] and IANA-registered identifiers is defined in RFC 1766, *Tags for the Identification of Languages*, which PDF uses as the basis for its language identifiers. ISO 639-2[24] defines three-letter language identifiers that are not allowed under RFC 1766.

B.3 Recommendations for capturing or converting documents to PDF/A

For archival preservation purposes, this Best Practices statement provides recommendations for processes that capture or convert documents to PDF/A format to ensure that the resulting conforming files retain their quality and integrity as records. Archival institutions and other organizations with long-term preservation requirements should encourage the use of Level A conformance as described in 5.1 and the additional guidelines of this clause.

ISO 15489-1:2001[9], 7.1, specifies that “to support the continuing conduct of business, comply with the regulatory environment and provide necessary accountability, organizations should create and maintain authentic, reliable and useable records, and protect the integrity of those records for as long as required”.

The regulatory environment for submitting documents to an organization’s archival institution may include requirements, standards and policies for electronic documents that stipulate document quality rules such as minimum image resolution, compression restrictions, or prohibited processes that either alter or dispose of approved data. For archival preservation purposes, the quality and integrity of documents created according to these legal and regulatory requirements, applicable standards and organizational policy should be retained when they are captured or converted to PDF/A.

To meet this critical archival need, PDF/A capture or conversion processes should replicate the exact content and quality of the source document within the conforming file. The following are examples of software development guidelines that accomplish this.

— Writers of conforming files should not use lossy compression, subsampling, downsampling or any other process that either alters the content or degrades the quality of source data in the conforming file.

— Software should not substitute searchable text, based on optical character recognition, for the original scanned text within the bit-mapped image of documents that are scanned to conforming files from paper or converted to conforming files from image formats.

NOTE Optical character recognition processes may involve loss of data through imprecise interpretation of scanned characters.

Bibliography

[1] ANSI X3.4, *Information Systems — Coded Character Sets — 7-Bit American National Standard Code for Information Interchange (7-Bit ASCII)*

[2] ECMA-6, *7-Bit coded Character Set*, Available from Internet <http://www.ecma-international.org/publications/files/ecma-st/Ecma-006.pdf>

[3] ISO 639-1, *Codes for the representation of names of languages — Part 1: Alpha-2 code*

[4] ISO 2108, *Information and documentation — International standard book number (ISBN)*

[5] ISO 3166-1, *Codes for the representation of names of countries and their subdivisions — Part 1: Country codes*

[6] ISO/IEC 10646-1:2000/Amd 1:2002, *Amendment 1: Mathematical symbols and other characters*

[7] ISO/IEC 10646-2:2001, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 2: Supplementary Planes*

[8] ISO/IEC 14492:2001, *Information technology — Lossy/lossless coding of bi-level images*

NOTE This International Standard is equivalent to ITU-T Recommendation T.88, *Information technology — Coded representation of picture and audio information — Lossy/lossless coding of bi-level images*

[9] ISO 15489-1:2001, *Information and documentation — Records management — Part 1: General*

[10] ISO/TR 15801, *Electronic imaging — Information stored electronically — Recommendations for trustworthiness and reliability*

[11] ISO 15930-6, *Graphic technology — Prepress digital data exchange using PDF — Part 6: Complete exchange of printing data suitable for colour-managed workflows using PDF 1.4 (PDF/X-3)*

[12] ISO/TR 18492, *Long-term preservation of electronic document-based information*

[13] ISO 18509-1, *Electronic archival storage — Specifications relative to the design and operation of information processing systems in view of ensuring the storage and integrity on recordings stored in these systems — Part 1: Long term access strategy*

[14] ISO 18509-2, *Electronic archival storage — Specifications relative to the design and operation of information processing systems in view of ensuring the storage and integrity on recordings stored in these systems — Part 2: Technical specifications*

[15] ITU-T Recommendation T.4, *Standardization of Group 3 facsimile terminals for document transmission*

[16] ITU-T Recommendation T.563, *Terminal characteristics for Group 4 facsimile apparatus*

[17] *Language Tags*, IANA. Available from Internet <http://www.iana.org/assignments/language-tags>

[18] *Namespaces in XML 1.1*, W3C Recommendation, February 4, 2004. Available from Internet <http://www.w3.org/TR/2004/REC-xml-names11-20040204>

[19] *PDF Reference: Adobe Portable Document Format*, Version 1.5, Adobe Systems Incorporated — 4th edition. Available from Internet <http://partners.adobe.com/asn/acrobat/sdk/public/docs/PDFReference15_v6.pdf>

[20] RFC 1950, *ZLIB Compressed Data Format Specification version 3.3*, May 1996. Available from Internet <http://www.ietf.org/rfc/rfc1950.txt>

[21] RFC 1951, *DEFLATE Compressed Data Format Specification version 1.3*, May 1996. Available from Internet <http://www.ietf.org/rfc/rfc1951.txt>

[22] *The Unicode Standard*, Unicode Consortium. Available from Internet <http://www.unicode.org/versions/>

[23] *Unicode Standard Annex #15, Unicode Normalization Forms*, Unicode Consortium, 17 April 2003. Available from Internet <http://www.unicode.org/unicode/reports/tr15/>

[24] ISO 639-2, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

ICS 35.240.30; 37.100.99

Price based on 29 pages