INTERNATIONAL STANDARD

ISO
19005-2

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

Part 2:
Use of ISO 32000-1 (PDF/A-2)

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

Partie 2: Utilisation de l'ISO 32000-1 (PDF/A-2)

© ISO 2011

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 19005-2 was prepared by Technical Committee ISO/TC 171, *Document management applications*, Subcommittee SC 2, *Application issues* in cooperation with 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)*

⸺ *Part 2: Use of ISO 32000-1 (PDF/A-2)*

The following parts are under preparation:

⸺ *Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)*

Introduction

PDF is a digital format for representing page-based documents. PDF files can 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 needs to be kept for substantial lengths of time; some needs to be kept permanently. These PDF files need to remain useable and accessible across multiple generations of technology. However, the inclusive, feature-rich nature of the format requires that constraints be placed on its use to make it suitable for the long-term preservation of electronic documents. 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.

ISO 19005 has been created as a multi-part document, of which this is Part 2. This allows future parts to be created without rendering ISO 19005, or applications based on it, obsolete.

The primary purpose 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 static visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.

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

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

These goals are accomplished by identifying the set of PDF components that can 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 might need to impose additional requirements, such as the best practices in Annex C, 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.

PDF/A does not directly address the topic of authenticity, either for the underlying content to be visually represented or for the PDF/A file itself. Such authenticity is generally considered to be important for legal, regulatory and governance purposes and is beyond the scope of this International Standard.

This part of ISO 19005 is 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 following:

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

— 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 in ISO 14721, ISO/TR 15801, and ISO/TR 18492;

— the 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 is intended to 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 needs to be able to read and process appropriately all files complying with a specified conformance level.

This part of ISO 19005 extends the capabilities of ISO 19005-1. It is based on PDF version 1.7 (as defined in ISO 32000-1) rather than PDF version 1.4 (which is used as the basis of ISO 19005-1). These added capabilities are made possible through compliance with ISO 32000-1 and include

— improvements to tagged PDF (for enhanced accessibility),

— Compressed Object and XRef streams (for smaller file sizes),

— PDF/A-compliant file attachments, portable collections and PDF packages,

— transparency, and

— JPEG 2000 compression.

This part of ISO 19005 (in conjunction with its normative references) provides sufficient information to interpret any conforming PDF/A-2 file.

NPES and AIIM (accredited standards developing organizations) maintain an ongoing series of application notes for guiding developers and users of ISO 19005. These application notes are available at <http://www.npes.org/standards/toolspdfa.html> and <http://www.aiim.org/documents/Standards/PDF-A/ISO19005AppNotes.pdf>. Both NPES and AIIM 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 2:
Use of ISO 32000-1 (PDF/A-2)

1 Scope

This part of ISO 19005 specifies the use of the Portable Document Format (PDF) 1.7, as formalized in ISO 32000-1, for preserving the static visual representation of page-based electronic documents over time.

This part of ISO 19005 is not applicable 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, or

— 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*[1]

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*[2]

ISO 15076-1, *Image technology colour management — Architecture, profile format and data structure — Part 1: Based on ICC.1:2010*

ISO/IEC 15444-2:2004, *Information technology — JPEG 2000 image coding system: Extensions*

ISO 15930-7:2010, *Graphic technology — Prepress digital data exchange using PDF — Part 7: Complete exchange of printing data (PDF/X-4) and partial exchange of printing data with external profile reference (PDF/X-4p) using PDF 1.6*

ISO 19005-1, *Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)*

---

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

2) The character code values defined in ISO/IEC 10646 are equivalent to those of Unicode.

© ISO 2011 – All rights reserved

ISO 24517-1, *Document management — Engineering document format using PDF — Part 1: Use of PDF 1.6 (PDF/E-1)*

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

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

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

ICC.1:2001-12, *File Format for Color Profiles (Version 4.0.0)*, International Color Consortium. Available from <http://www.color.org/>

ICC.1:2003-09, *File Format for Color Profiles (Version 4.1.0)*, International Color Consortium. Available from <http://www.color.org/>

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

RFC 2315, *PKCS#7: Cryptographic Message Syntax Version 1.5*. Available from http://www.rfc-editor.org

RFC 3280, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*. Available from http://www.rfc-editor.org

Adobe Glyph List, 20 September 2002, Adobe Systems Incorporated. Available from <http://partners.adobe.com/public/developer/en/opentype/glyphlist.txt>

Adobe Supplement to ISO 32000-1, BaseVersion 1.7, ExtensionLevel 5, Adobe Systems Incorporated. Available from <http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/adobe_supplement_iso32000_1.pdf>

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 are required to comply

3.2
electronic document
electronic representation of a page-oriented aggregation of text, images and graphic data, and metadata useful to identifying and understanding that data, that can be reproduced on paper or other substrates, as well as rendered electronically on display devices, without significant loss of its information content

3.3
end-of-file marker
five-character sequence (**%%EOF**) marking the end of a PDF file

3.4
EOL marker
end-of-line marker
one- or two-character sequence marking the end of a line, consisting of a **CARRIAGE RETURN** character (0Dh) or a **LINE FEED** character (0Ah) or a **CARRIAGE RETURN** followed immediately by a **LINE FEED**

3.5
extension schema
conforming XMP schema that is not defined in the XMP Specification, nor in ISO 19005-1 or ISO 19005-2

3.6
font
identified collection of graphics that may be glyphs or other graphic elements

[ISO 32000-1]

3.7
font program
software program written in a special-purpose language, such as the *Type 1*, *TrueType*, or *OpenType* font format, that is understood by a specialized font interpreter

[ISO 32000-1]

3.8
interactive reader
reader that requires or allows human interaction with the content and other objects contained in the document during the software's processing phase

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

3.9
Level A conformance
conformance level encompassing all requirements of this part of ISO 19005

3.10
Level B conformance
conformance level encompassing the requirements of this part of ISO 19005 regarding the visual appearance of electronic documents but not those regarding their structural and semantic properties nor the requirement that all text have Unicode equivalents

3.11
Level U conformance
conformance level encompassing the requirements of this part of ISO 19005 regarding the visual appearance of electronic documents, together with the requirement that all text in the document have Unicode equivalents

3.12
long term
period of time long enough for there to be concern about the impacts on the information being held in a repository of changing technologies, including support for new media and data formats, and of a changing user community, and which may extend into the indefinite future

3.13
PDF
Portable Document Format
file format defined in ISO 32000-1:2008

3.14
reader
software application that is able to read and process PDF/A files

3.15
writer
software application that is able to write PDF/A files

3.16
XMP packet
structured wrapper for serialized XMP metadata that can be embedded in PDF as well as other 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. Some names can also be used as values, depending on the context, and so the styling of the content will be context specific.

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

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

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

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

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

The following terms, referring to ISO 19005, or parts thereof, are recommended when the full ISO name is not being used:

— "PDF/A" – a synonym for the ISO 19005 series 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.

— "PDF/A-2" – a synonym for ISO 19005-2;

— "PDF/A-2a" – a synonym for ISO 19005-2 Level A conformance;

— "PDF/A-2b" – a synonym for ISO 19005-2 Level B conformance;

— "PDF/A-2u" – a synonym for ISO 19005-2 Level U conformance.

5 Conformance levels

5.1 General

This part of ISO 19005 defines a file format for representing electronic documents known as "PDF/A-2". Conforming PDF/A-2 files shall adhere to all requirements of ISO 32000-1 as modified by this part of ISO 19005. A conforming file may include any valid ISO 32000-1 feature that is not explicitly forbidden by this part of ISO 19005. Features described in PDF specifications prior to Version 1.7 which are not explicitly described in ISO 32000-1 should not be used.

NOTE 1   A conforming file is not obligated to use any PDF feature other than those explicitly required by ISO 32000-1 or this part of ISO 19005.

As described in 6.1.2, the version number of a file may be any value from 1.0 to 1.7, and the value shall not be used in determining whether a file is in conformance with this part of ISO 19005.

NOTE 2   The proper mechanism by which a file can presumptively identify itself as being a PDF/A-2 file of a given conformance level is described in 6.6.4.

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-2a 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 6.2.11.7 and 6.7. A file meeting this conformance level is said to be a "conforming PDF/A-2b file".

NOTE 1   The Level B conformance requirements are intended to be the minimum 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.

NOTE 2   A Level B conforming file can include features from 6.2.11.7 and 6.7 but still be identified as Level B.

5.4   Level U 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 U conformance level. Level U conforming files shall adhere to all of the requirements of this part of ISO 19005, except those of 6.7. A file meeting this conformance level is said to be a "conforming PDF/A-2u file".

NOTE 1   The Level U conformance requirements are intended to be those necessary to ensure that not only is the rendered visual appearance of a conforming file preservable over the long term, but that any text contained in the document can be reliably extracted as a series of Unicode codepoints. However, Level U 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.

NOTE 2   A Level U conforming file can include features from 6.7 but still be identified as Level U.

NOTE 3   Level U is new to this part of ISO 19005 and therefore does not have an equivalent in ISO 19005-1.

5.5   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 and other processing of conforming files shall be performed as defined in ISO 32000-1, subject to the additional restrictions specified by this part of ISO 19005. Features described in PDF specifications that are not explicitly described in ISO 32000-1 shall be ignored by conforming readers.

Conforming PDF/A-2 readers shall read and process appropriately all PDF/A-2 files. In addition, conforming PDF/A-2 readers shall read and process appropriately all PDF/A-1 files as defined by ISO 19005-1.

6 Technical requirements

6.1 File structure

6.1.1 General

Overall file format issues and the base elements that form the general structure of a conforming file are addressed in 6.1.2 to 6.1.12.

Any data contained in a conforming file that is not described in ISO 32000-1 or in this part of ISO 19005 should be ignored by a conforming reader and shall not be used to render content on a page.

6.1.2 File header

The file header shall begin at byte zero and shall consist of "%PDF-1.n" followed by a single EOL marker, where 'n' is a single digit number between 0 (30h) and 7 (37h).

The aforementioned EOL marker shall be immediately followed by a **%** (25h) character followed by at least four bytes, each of whose encoded byte values shall have a decimal value greater than 127.

NOTE   The presence of encoded 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 needs to be preserved during processing.

6.1.3 File trailer

The file trailer dictionary shall contain the **ID** keyword whose value shall be File Identifiers as defined in ISO 32000-1:2008, 14.4.

NOTE 1   No data can follow the last end-of-file marker except a single optional end-of-line marker as described in ISO 32000-1:2008, 7.5.5.

The keyword **Encrypt** shall not be present in the trailer dictionary.

NOTE 2   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

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

Any indirect object whose offset is not referenced in any cross-reference table, nor in any cross-reference stream, shall be exempt from all requirements of this part of ISO 19005 and may be ignored by a conforming reader. If a conforming reader does not ignore such indirect objects, they shall never influence the way content is rendered.

6.1.5 Document information dictionary

A document information dictionary may be present in a conforming file and a PDF/A-2 compliant reader shall ignore it.

NOTE   Metadata can be included in a document through the use of XMP metadata streams as specified in 6.6.3.

6.1.6 String objects

The number of hexadecimal digits in a hexadecimal string shall always be even.

NOTE   This avoids the need for the provision in ISO 32000-1 about the absence of the final hexadecimal digit.

6.1.7   Stream objects

6.1.7.1   General

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

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

NOTE 1   These keys are used to point to data 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.

NOTE 2   Since an inline image dictionary is not a stream object, this provision allows the presence of the **F** key in an inline image dictionary as the abbreviation for **Filter**.

6.1.7.2   Filters

All standard stream filters listed in ISO 32000-1:2008, 7.4, Table 6 may be used, with the exception of *LZWDecode*. In addition, the *Crypt* filter shall not be used unless the value of the **Name** key in the decode parameters dictionary is *Identity*. Filters that are not listed in ISO 32000-1:2008, 7.4, Table 6 shall not be used.

NOTE   The *Crypt* filter is used to apply encryption and access control to the file.

6.1.8   Name objects

Font names, names of colourants in Separation and DeviceN colour spaces, and structure type names, after expansion of character sequences escaped with a NUMBER SIGN (23h), if any, shall be valid UTF-8 character sequences.

These requirements make normative the recommendations set out in ISO 32000-1:2008, 7.3.5.

All other name objects should adhere to these same restrictions.

6.1.9   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.

6.1.10   Inline image dictionaries

The value of the **F** key in the Inline Image dictionary shall not be *LZW, LZWDecode, Crypt*, a value not listed in ISO 32000-1:2008, Table 6, or an array containing any such value.

6.1.11   Linearized PDF

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

NOTE   As defined in ISO 32000-1:2008, Annex F, a PDF is not linearized if the value of the **L** key in the linearization dictionary does not match the actual length of the PDF file. This implies that an incremental update to a linearized PDF will render it non-linearized.

6.1.12 Permissions

No keys other than **UR3** and **DocMDP** shall be present in a permissions dictionary (ISO 32000-1:2008, 12.8.4, Table 258). If **DocMDP** is present, then the Signature References dictionary (ISO 32000-1:2008, 12.8.1, Table 253) shall not contain the keys **DigestLocation**, **DigestMethod**, and **DigestValue**.

NOTE   These restrictions are present to ensure that functionality such as obsolete versions of the "User Rights" dictionary do not appear in a document conforming to this part of ISO 19005.

6.1.13 Implementation limits

A conforming file shall not contain any integer greater than 2147483647.

A conforming file shall not contain any integer less than -2147483648.

A conforming file shall not contain any real number outside the range of $\pm 3.403 \times 10^{38}$.

A conforming file shall not contain any real number closer to zero than $\pm 1.175 \times 10^{-38}$.

A conforming file shall not contain any string longer than 32767 bytes.

A conforming file shall not contain any name longer than 127 bytes.

A conforming file shall not contain more than 8388607 indirect objects.

A conforming file shall not nest q/Q pairs by more than 28 nesting levels.

A conforming file shall not contain a DeviceN colour space with more than 32 colourants.

A conforming file shall not contain a CID value greater than 65535.

NOTE 1   These values are derived from ISO 32000-1:2008, Table C.1.

The size of any of the page boundaries described in ISO 32000-1:2008, 14.11.2 shall not be less than 3 units in either direction, nor shall it be greater than 14 400 units in either direction.

This requirement makes normative a recommendation from ISO 32000-1:2008, C.2.

NOTE 2   By complying with these limits, a conforming file is compatible with the widest possible range of readers.

6.2   Graphics

6.2.1   General

Restrictions that shall be placed on both conforming files and readers with respect to the graphical elements described in ISO 32000-1:2008, 7.8 are described in 6.2.2 to 6.2.11. A conforming reader shall render these graphical elements onto their respective PDF pages according to the rendering requirements of ISO 32000-1 as modified by this part of ISO 19005.

A conforming interactive reader may put additional user interface elements around, above or below the graphical elements of the page. These user interface elements may be a presentation of other PDF objects (such as bookmarks or page thumbnails) or they may represent non-PDF objects. In all cases, the user interface elements and their contents shall not be required to conform to the requirements of 6.2.2 to 6.2.11.

6.2.2   Content streams

Content streams shall not contain any operators not defined in ISO 32000-1 even if such operators are bracketed by the **BX**/**EX** compatibility operators.

NOTE 1   In earlier versions of PDF, a PostScript operator **PS** was defined. As this operator is not defined in ISO 32000-1, its use is implicitly prohibited by this clause.

NOTE 2   Content streams, as defined in ISO 32000-1:2008, 7.8.2, can be used for page descriptions, Form XObjects, Type 1 Patterns, and Type 3 fonts, as well as for the appearances of annotations.

The use of the rendering intent operator ('**ri**') shall conform to the requirements of 6.2.6.

The use of the flatness operator ('**i**') shall conform to the requirements of 6.2.7.

A content stream that references other objects, such as images and fonts that are necessary to fully render or process the stream, shall have an explicitly associated **Resources** dictionary as described in ISO 32000-1:2008, 7.8.3. Any named resource present in the resources dictionary, but whose name is not referenced from the associated content stream, is not used for rendering and therefore shall be exempt from all requirements of this part of ISO 19005.

6.2.3   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 OutputIntent. A PDF/A OutputIntent shall be identified as an **OutputIntent** dictionary, as defined by ISO 32000-1:2008, 14.11.5, that is included in the file's **OutputIntents** array. It shall have *GTS_PDFA1* as the value of its **S** key and a valid ICC profile stream as the value of its **DestOutputProfile** key.

NOTE 1   PDF/A requires that an OutputIntent be present when uncalibrated colour spaces are used (see 6.2.4.3 for more details). It has this requirement in order to ensure reliable rendering of colour through the indirect use of the OutputIntent profile provided.

NOTE 2   The value for GTS_PDFA1 was maintained for this part of ISO 19005 to enable greater compatibility with ISO 19005-1.

In addition, the **DestOutputProfileRef** key, as defined in ISO 15930-7:2010, Annex A, shall not be present in any PDF/X OutputIntent.

NOTE 3   Disallowing the **DestOutputProfileRef** key maintains the intent of this part of ISO 19005 of ensuring self-contained documents with no external references. However, it does mean that a single PDF is unable to be compliant with both PDF/A-2 and PDF/X-4p.

If a file's **OutputIntents** array contains more than one entry, as might be the case where a file is compliant with this part of ISO 19005 and at the same time with PDF/X-4 or PDF/E-1, 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.

The profile stream that is the value of the **DestOutputProfile** key shall either be an output profile (Device Class = "prtr") or a monitor profile (Device Class = "mntr"). The profiles shall have a colour space of either "GRAY", "RGB", or "CMYK". If present in the **DestOutputProfile** stream object, the **Alternate** key shall be ignored by a PDF/A-2 conforming reader.

6.2.4   Colour spaces

6.2.4.1   General

All colours shall be specified in a device-independent manner, either directly by the use of device-independent colour spaces, or indirectly by the means of the **DestOutputProfile** in the PDF/A **OutputIntent**. A conforming file may use any colour space specified in ISO 32000-1, except as restricted in 6.2.4.2 to 6.2.4.5.

NOTE   Specifying colour in a device independent manner as described within 6.2.4 enables predictable colour rendering based on a colorimetric definition and without reliance on heuristic assumptions or on information external to the conforming file. It also provides a mechanism whereby a colorimetric definition can be associated with device-dependent colour data.

6.2.4.2   ICCBased colour spaces

The profile that forms the stream of an **ICCBased** colour space shall conform to ICC.1:1998-09, ICC.1:2001-12, ICC.1:2003-09 or ISO 15076-1.

NOTE 1   ISO 32000-1 allows the use of all versions of ICC profiles up to ICC.1:2003-09. For practical reasons based on the behaviour of profile creation software, this part of ISO 19005 also allows the use of ISO 15076-1, which is believed to be technically identical in all respects relevant to its use here, other than the value of the profile version number.

A conforming reader shall render **ICCBased** colour spaces as specified by ISO 32000-1 and the ICC specification, and shall not use the **Alternate** colour space specified in an ICC profile stream dictionary.

Overprint mode (as set by the **OPM** value in an **ExtGState** dictionary) shall not be one (1) when an **ICCBased** CMYK colour space is used and when overprinting for stroke or fill or both is set to true.

NOTE 2   This prohibition avoids unpredictable overprinting behaviour when overprint mode is 1 if implicit colour conversion is applied as described in ISO 32000-1:2008, 8.6.7.

6.2.4.3   Uncalibrated/Device colour spaces

DeviceRGB shall only be used if a device independent **DefaultRGB** colour space has been set when the **DeviceRGB** colour space is used, or if the file has a PDF/A **OutputIntent** that contains an RGB destination profile.

DeviceCMYK shall only be used if a device independent **DefaultCMYK** colour space has been set or if a DeviceN-based **DefaultCMYK** colour space has been set when the **DeviceCMYK** colour space is used or the file has a PDF/A **OutputIntent** that contains a **CMYK** destination profile.

DeviceGray shall only be used if a device independent **DefaultGray** colour space has been set when the **DeviceGray** colour space is used, or if a PDF/A **OutputIntent** is present.

NOTE 1   As described in ISO 32000-1:2008, 8.6.5.6, colours that are specified in a device colour space (**DeviceGray**, **DeviceRGB**, or **DeviceCMYK**) are device dependent. By setting default colour spaces, a conforming writer can request that such colours be systematically transformed (remapped) into device independent CIE-based colour spaces.

NOTE 2   A DeviceN-based DefaultCMYK colour space is subject to all provisions in 6.2.4.4, thus making it device independent.

When rendering colours specified in **DeviceRGB** or **DeviceCMYK,** and no matching device independent default colour space has been set, a conforming reader shall use the profile in the file's PDF/A **OutputIntent** dictionary as the source colour space.

When rendering colours specified in **DeviceGray** and no device independent **DefaultGray** colour space has been set, a conforming reader shall render the **DeviceGray** colour as follows:

⸺ If the PDF/A **OutputIntent** contains a 'GRAY' destination profile, that profile shall be used as the source colour space when rendering the colour.

⎯ If the PDF/A **OutputIntent** contains an 'RGB' destination profile, then the conforming reader shall convert the **DeviceGray** colour to RGB by the method described in ISO 32000-1:2008, 10.3.2, and shall use the RGB destination profile as the source colour space when rendering the colour.

⎯ If the PDF/A **OutputIntent** contains a 'CMYK' destination profile, then the conforming reader shall convert the **DeviceGray** colour to CMYK by the method described in ISO 32000-1:2008, 10.3.3, and shall use the CMYK destination profile as the source colour space when rendering the colour.

6.2.4.4   Separation and DeviceN colour spaces

If the named colourants in the colour space are all from the list **Cyan**, **Magenta**, **Yellow**, and **Black**, and if the file has a PDF/A **OutputIntent**, and if that **OutputIntent** contains a **CMYK** destination profile, then a conforming reader shall treat the colourants as components of the colour space specified by the destination profile in the PDF/A **OutputIntent** dictionary, as defined in 6.2.3.

NOTE 1   All other aspects of rendering Separation colour spaces are described in ISO 32000-1:2008, 8.6.6.4.

NOTE 2   All other aspects of rendering DeviceN and NChannel colour spaces are described in ISO 32000-1:2008, 8.6.6.5.

The alternate space of a **Separation** or **DeviceN** colour space shall obey all restrictions on colour spaces specified in 6.2.4.2 and 6.2.4.3.

For any spot colour used in a **DeviceN** or **NChannel** colour space, an entry in the **Colorants** dictionary shall be present. Any **Separation** colour space which appears in a **Colorant** dictionary shall obey the same restrictions as any other **Separation** colour space.

NOTE 3   Although the **Colorant** key is defined in ISO 32000-1 as optional in a **DeviceN** colour space attributes dictionary, this requirement makes it mandatory for this part of ISO 19005.

All **Separation** arrays within a single PDF/A-2 file (including those in **Colorants** dictionaries) that have the same *name* shall have the same *tintTransform* and *alternateSpace.* In evaluating equivalence, the PDF objects shall be compared, rather than the computational result of the use of those PDF objects. Compression and whether or not an object is direct or indirect shall be ignored.

NOTE 4   A PDF/A-2 writer might need to synchronize multiple *alternateSpace* and *tintTransform* entries when creating a PDF/A-2 file.

The **Separation** arrays in the **Colorants** dictionary of **DeviceN** and **NChannel** colour spaces should be consistent with the *tintTransform* and *alternateSpace* of the **DeviceN** or **NChannel** colour space itself.

6.2.4.5   Indexed and Pattern colour spaces

Indexed and **Pattern** colour spaces are indirect methods of specifying colour. All the requirements of 6.2.4 apply to the underlying colour spaces of **Indexed** and **Pattern** colour spaces.

6.2.5   Extended graphics state

An **ExtGState** dictionary shall not contain the **TR** or **HTP** keys. 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.

NOTE 1   The **HTP** key was present in early versions of PDF but was removed by PDF 1.3.

The **TransferFunction** key in a halftone dictionary shall be used only as required by ISO 32000-1.

All halftones in a conforming PDF/A-2 file shall have the value **1** or **5** for the **HalftoneType** key.

NOTE 2   This prohibits the use of threshold screens that will produce different appearances at different resolutions.

Halftones in a conforming PDF/A-2 file shall not contain a **HalftoneName** key.

The use of the **RI** key shall conform to the requirements of 6.2.6.

The use of the **FL** key shall conform to the requirements of 6.2.7.

Conforming readers shall ignore the **BG, BG2**, **UCR** and **UCR2** functions when rendering the PDF.

Conforming readers shall respect the **OP**, **op** and **OPM** entries in **ExtGState** dictionaries as described in ISO 32000-1:2008, 8.6.7 when rendering the PDF. When rendering to a device that does not natively support all colourants to be rendered, a conforming reader shall simulate the overprinting of the colourants as if they had been rendered to a device that did natively support them.

NOTE 3   Having a conforming reader respect these entries, both when viewing on screen and when printing, ensures a consistent rendering between these two types of output.

6.2.6   Rendering intents

Where a rendering intent is specified, its value shall be one of the four values defined in ISO 32000-1:2008, Table 70: *RelativeColorimetric*, *AbsoluteColorimetric*, *Perceptual* or *Saturation*.

NOTE   The default rendering intent is *RelativeColorimetric*.

6.2.7   Flatness

Conforming readers shall ignore the actual value of the **FL** entry of the Extended Graphic State or the operand value for the **i** operator. Instead, when a conforming reader renders content, it shall choose a value suitable to enable efficient rendering without introducing visible artefacts.

6.2.8   Images

6.2.8.1   General

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*. For an inline image, the **I** key shall have a value of *false*.

Use of the **Intent** key shall conform to the requirements given in 6.2.6.

6.2.8.2   Thumbnail images

A conforming reader shall never substitute the rendering of a page by making use of thumbnail images, regardless of whether such thumbnail images are retrieved from a **Thumbs** entry in a page dictionary, an **xmp:Thumbnails** entry in the document, a page level XMP metadata stream or from any other data in the file.

6.2.8.3   JPEG2000

JPEG2000 compression shall be used as specified in ISO 32000-1:2008. Only the JPX baseline set of features, as restricted or extended by ISO 32000-1:2008 and this subclause, shall be used.

NOTE 1   The JPX baseline set of features is defined in ISO/IEC 15444-2:2004, M.9.2.

The number of colour channels in the JPEG2000 data shall be 1, 3 or 4.

If the number of colour space specifications in the JPEG2000 data is greater than 1, there shall be exactly one colour space specification that has the value *0x01* in the **APPROX** field. If the specified colour space specification uses an ICC profile, then that profile shall conform to the requirements of ISO 32000-1:2008, 8.6.5.5.

NOTE 2 The value *0x01* in the **APPROX** field identifies the colour space with the best colour fidelity available.

The value of the **METH** entry in its 'colr' box shall be *0x01, 0x02* or *0x03*. A conforming reader shall use only that colour space and shall ignore all other colour space specifications.

JPEG2000 enumerated colour space 19 (*CIEJab*) shall not be used.

JPEG2000 enumerated colour space 12 (*CMYK*), which is part of JPX but not JPX baseline, may be used.

Where the JPEG2000 image effectively uses *DeviceGray, DeviceRGB* or *DeviceCMYK,* whether through the **ColorSpace** entry in the Image XObject or in the absence thereof through the colour space definition in the JPEG2000 data, the provisions of 6.2.4.3 shall apply.

NOTE 3 s-YCC and es-YCC — the two YCC flavours allowed in baseline JPX — are just alternative representations of sRGB and esRGB. Details can be found in ISO 15444-2.

NOTE 4 ISO 32000-1 states that a **ColorSpace** entry in an Image XObject containing JPEG2000-compressed data overrides any colour space defined within the JPEG2000 data stream itself. It further requires that the number of colour channels in the JPEG2000 data has to match the number of components in the colour space defined in the **ColorSpace** entry of the Image XObject; the PDF producer has to ensure that the samples are consistent with the colour space used.

The bit-depth of the JPEG2000 data shall have a value in the range 1 to 38. All colour channels in the JPEG2000 data shall have the same bit-depth.

Images compressed using the JPEG2000 compression method shall be created and read as described in ISO/IEC 15444-2:2004.

NOTE 5 This section is based on ISO 15930-7:2010, 6.27 (PDF/X-4). Having it included in this part of ISO 19005 ensures the use of a well-defined subset of the complete ISO 15444-2 specification that is consistent with other International Standards. In addition, this section provides for a subset of JPEG2000 that is also aligned with ISO 24517-1 (PDF/E-1).

6.2.9 XObjects

6.2.9.1 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.

6.2.9.2 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.

6.2.9.3   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.

6.2.10 Transparency

PDF transparency (as described in ISO 32000-1:2008, Clause 11) may be used in a PDF/A-2 file. The method that a conforming reader should use to determine whether a given page contains any graphical elements whose associated graphic state contains transparency or which are otherwise involved in a transparency operation is defined in Annex A.

A conforming reader shall use the document's PDF/A **OutputIntent** as the default blending colour space (ISO 32000-1:2008, 11.3.4). If the document does not contain a PDF/A **OutputIntent**, then all **Page** objects that contain transparency shall include the **Group** key, and the attribute dictionary that forms the value of that **Group** key shall include a **CS** entry whose value shall be used as the default blending colour space.

NOTE   This requirement ensures that there is always an explicitly defined transparency blending space specified for any content which has associated transparency.

The value for any **CS** key in any transparency group's attribute dictionary shall conform to the restrictions on colour spaces set out in 6.2.4.

Only blend modes that are specified in ISO 32000-1:2008 shall be used for the value of the **BM** key in an extended graphic state dictionary. A PDF/A-2 compliant reader shall process these blend modes as described in ISO 32000-1:2008, 11.3.5, and as amended by the Adobe Supplement to ISO 32000-1, BaseVersion 1.7, ExtensionLevel 5, Section 3.

6.2.11 Fonts

6.2.11.1   General

The intent of the requirements given in 6.2.11.2 to 6.2.11.8 is to ensure that the 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, when possible, to allow the recovery of semantic properties for each character of the textual content. Unless a requirement specifically states that it shall only apply to text that would be rendered by a conforming reader, they shall apply to any font including those used exclusively with text rendering mode 3.

NOTE   A font referenced solely in text rendering mode 3 (ISO 32000-1:2008, 9.3.6) is not rendered and is thus exempt from the requirements that impact the visual representation of the glyphs of a font.

6.2.11.2   Font types

All fonts and font programs used in a conforming file, regardless of rendering mode usage, shall conform to the provisions in ISO 32000-1:2008, 9.6 and 9.7, as well as to the font specifications referenced by these provisions.

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

Multiple master fonts shall be considered a special case of Type 1 fonts; any requirement stated with regard to Type 1 fonts shall also be required with regard to multiple master fonts.

6.2.11.3   Composite fonts

6.2.11.3.1   General

For any given composite (Type 0) font within a conforming file, the **CIDSystemInfo** entry in its CIDFont dictionary and its Encoding dictionary shall have the following relationship:

— If the **Encoding** key in the Type 0 font dictionary is *Identity-H* or *Identity-V*, any values of **Registry**, **Ordering**, and **Supplement** may be used in the **CIDSystemInfo** entry of the CIDFont.

— Otherwise, the corresponding **Registry** and **Ordering** strings in both CIDSystemInfo dictionaries shall be identical, and the value of the **Supplement** key in the CIDSystemInfo dictionary of the CIDFont shall be greater than or equal to the **Supplement** key in the CIDSystemInfo dictionary of the CMap.

NOTE   The requirement for the **Supplement** key ensures that the font includes glyphs for all CIDs which can be referenced by the CMap.

6.2.11.3.2   CIDFonts

ISO 32000-1:2008, 9.7.4, Table 117 requires that all embedded Type 2 **CIDFonts** in 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 ISO 32000-1:2008, 9.7.4, Table 117.

6.2.11.3.3   CMaps

All CMaps used within a PDF/A-2 file, except those listed in ISO 32000-1:2008, 9.7.5.2, Table 118, shall be embedded in that file as described in ISO 32000-1:2008, 9.7.5. 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.

A **CMap** shall not reference any other **CMap** except those listed in ISO 32000-1:2008, 9.7.5.2, Table 118.

6.2.11.4   Embedding

6.2.11.4.1   General

The font programs for all fonts used for rendering within a conforming file shall be embedded within that file, as defined in ISO 32000-1:2008, 9.9. A font is considered to be used if at least one of its glyphs is referenced from a content stream (6.2.2).

NOTE 1   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 external resources.

NOTE 2   As discussed in ISO 32000-1:2008, 9.3.6, 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.

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

NOTE 3   This part of ISO 19005 precludes the embedding of font programs whose legality depends upon special agreement with the copyright holder. Such an allowance places unacceptable burdens on an archive to verify the existence, validity and longevity of such claims.

Embedded fonts shall define all glyphs referenced for rendering within the conforming file.

NOTE 4   As stated in 6.2.11.4.2, subsets of a font are acceptable as long as the embedded font provides glyph definitions for all characters referenced within the file.

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

NOTE 5   There is no exemption from the requirements of 6.2.11.4 for the 14 standard Type 1 fonts.

6.2.11.4.2   Subset embedding

ISO 32000-1:2008, 9.6 permits the embedding of subsets of font programs.

NOTE 1   The use of subsets of a font and its associated font program allows a potentially substantial reduction in the size of conforming files.

If the **FontDescriptor** dictionary of an embedded Type 1 font contains a **CharSet** string, then it shall list the character names of all glyphs present in the font program, regardless of whether a glyph in the font is referenced or used by the PDF or not.

NOTE 2   The above requirement makes normative the statements in ISO 32000-1:2008, 9.8.

If the **FontDescriptor** dictionary of an embedded CID font contains a **CIDSet** stream, then it shall identify all CIDs which are present in the font program, regardless of whether a CID in the font is referenced or used by the PDF or not.

NOTE 3   The above requirement makes normative the statements in ISO 32000-1:2008, 9.8.3.1, Table 124.

6.2.11.5   Font metrics

For every font embedded in a conforming file and used for rendering, the glyph width information in the font dictionary and in the embedded font program shall be consistent. For ISO 19005, consistent is defined to be a difference of no more than 1/1000 unit.

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

6.2.11.6   Character encodings

For all non-symbolic TrueType fonts used for rendering, the embedded TrueType font program shall contain one or several non-symbolic **cmap** entries such that all necessary glyph lookups can be carried out.

All non-symbolic TrueType fonts shall have either *MacRomanEncoding* or *WinAnsiEncoding* as the value for the **Encoding** key in the Font dictionary or as the value for the **BaseEncoding** key in the dictionary which is the value of the **Encoding** key in the Font dictionary.

In addition, no non-symbolic TrueType font shall define a **Differences** array unless all of the glyph names in the **Differences** array are listed in the Adobe Glyph List and the embedded font program contains at least the Microsoft Unicode (3,1 – Platform ID=3, Encoding ID=1) encoding in the "cmap" table.

Symbolic TrueType fonts shall not contain an **Encoding** entry in the font dictionary, and the "cmap" table in the embedded font program shall either contain exactly one encoding or it shall contain, at least, the Microsoft Symbol (3,0 – Platform ID=3, Encoding ID=0) encoding.

In all cases, for TrueType fonts that are to be rendered, character codes shall be able to be mapped to glyphs according to ISO 32000-1, 9.6.6.4 without the use of a non-standard mapping chosen by the conforming reader.

6.2.11.7   Unicode character maps

6.2.11.7.1   General

6.2.11.7 is applicable only for files meeting Level A or Level U conformance. For Level B conformance, the requirements of 6.2.11.7 may be ignored by a conforming writer.

6.2.11.7.2   Level A and Level U conformance

The font dictionary of all fonts, regardless of their rendering mode usage, shall include a **ToUnicode** entry whose value is a CMap stream object that maps character codes for at least all referenced glyphs to Unicode values, as described in ISO 32000-1:2008, 9.10.3, unless the font falls under at least one of the following four categories:

— fonts that use the predefined encodings **MacRomanEncoding**, **MacExpertEncoding** or **WinAnsiEncoding**;

— Type 1 and Type 3 fonts where the glyph names of the glyphs referenced are all contained in the Adobe Glyph List or the set of named characters in the Symbol font, as defined in ISO 32000-1:2008, Annex D;

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

— non-symbolic TrueType fonts.

NOTE 1   Unicode mapping allows the retrieval of semantic properties about every character referenced in the file.

The Unicode values specified in the **ToUnicode** CMap shall all be greater than zero (0), but not equal to either U+FEFF or U+FFFE.

NOTE 2   This requirement ensures that the values in the ToUnicode CMap will be useful values and not simply placeholders.

6.2.11.7.3   Level A conformance

For Level A conformance only, for any character, regardless of its rendering mode, that is mapped to a code or codes in the Unicode Private Use Area (PUA), an **ActualText** entry as described in ISO 32000-1:2008,14.9.4 shall be present for this character or a sequence of characters of which such a character is a part.

6.2.11.8   Use of .notdef glyph

A PDF/A-2 compliant document shall not contain a reference to the .**notdef** glyph from any of the text showing operators, regardless of text rendering mode, in any content stream.

NOTE   Since the .**notdef** glyph does not have any semantic value, this requirement is made to avoid any ambiguity which might result from its use.

6.3   Annotations

6.3.1   Annotation types

Annotation types not defined in ISO 32000-1 shall not be permitted. Additionally, the **3D**, **Sound, Screen** and **Movie** types shall not be permitted.

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

6.3.2   Annotation dictionaries

Except for annotation dictionaries whose **Subtype** value is *Popup*, all annotation dictionaries shall contain the **F** key. If present, the **F** key's **Print** flag bit shall be set to *1* and its **Hidden**, **Invisible, ToggleNoView,** 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   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 ISO 32000-1:2008*,* 12.5.3; explicitly setting these flags removes any potential ambiguity between the annotation dictionary settings and reader behaviour.

6.3.3   Annotation appearances

Every annotation (including those whose **Subtype** value is *Widget*, as used for form fields), except for the two cases listed below, shall have at least one appearance dictionary:

— annotations where the value of the **Rect** key consists of an array where value 1 is equal to value 3 and value 2 is equal to value 4;

— annotations whose **Subtype** value is *Popup* or *Link*.

A conforming reader shall render the appearance dictionary without regard to any other keys and values in the annotation dictionary and shall ignore the values of the **C, IC, Border, BS, BE, CA, H, DA, Q, DS, LE, LL, LLE,** and **Sy** keys.

NOTE 1   Requiring an appearance dictionary for each annotation ensures the reliable rendering of the annotations.

For all annotation dictionaries containing an **AP** key, the appearance dictionary that it defines as its value shall contain only the **N** key. If an annotation dictionary's **Subtype** key has a value of *Widget* and its **FT** key has a value of *Btn*, the value of the **N** key shall be an appearance subdictionary, otherwise the value of the **N** key shall be an appearance stream.

NOTE 2   In accordance with the requirements of ISO 32000-1:2008, 12.7.4.2.3 and 12.7.4.2.4, a Button form field needs to have multiple appearance states, each one associated with the specific values that the button can take.

6.3.4   Display of annotation contents

In addition to the rendering behaviour defined by ISO 32000-1 and as modified by this part of ISO 19005, conforming interactive readers shall provide a mechanism to display the values of the **Contents** key of all annotation dictionaries, including those whose **Subtype** value is *Widget*, except for Widgets of type **Sig** (Digital Signature).

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

6.4   Interactive forms

6.4.1   General

The intent of the requirements of 6.4 is to ensure that there is no ambiguity about the rendering of form fields.

The value of a form field shall not be used by a conforming reader when rendering the field. Instead, the conforming reader shall follow the requirements of 6.3.3 and render the appearance dictionary.

NOTE 1   The value of a field, except that of a pushbutton field, is the value of the **V** key in the field dictionary. A pushbutton field has no value.

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*.

NOTE 2   Annotations of type Widget are also subject to all provisions of 6.3.2 and 6.3.3.

6.4.2   XFA forms

The document's interactive form dictionary that forms the value of the **AcroForm** key in the document's **Catalog** of a PDF/A-2 file, if present, shall not contain the **XFA** key. In addition, a document's **Catalog** shall not contain the **NeedsRendering** key.

NOTE 1   This prohibits the use of XML-based XFA forms.

NOTE 2   In order to enable the preservation of the data from an XFA form in a PDF/A compliant document, provisions for moving that data from the XFA key to another part of the PDF format are described in Annex D.

6.4.3   Digital signatures

As permitted by ISO 32000-1:2008, 12.8.1, a PDF/A-2 conforming file may contain document, certifying or user rights signatures. Such signatures shall be specified in the PDF through the use of *signature fields* in accordance with ISO 32000-1:2008, 12.7.4.5. All annotations associated with signature fields shall meet the requirements of 6.3.2 and 6.3.3 of this part of ISO 19005.

When generating signature appearances and any other PDF objects as part of the signing process, a conforming reader shall ensure that it does not invalidate compliance with this part of ISO 19005, specifically concerning any content added to the widget's appearance.

Additional requirements for the use of digital signatures in a PDF/A conforming file can be found in Annex B.

6.5   Action

6.5.1   General

The **Launch**, **Sound**, **Movie**, **ResetForm**, **ImportData, Hide, SetOCGState, Rendition, Trans, GoTo3DView** 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 ISO 32000-1:2008*,*12.6.4.11, Table 211.

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. **Hide** actions set an annotation's Hidden flag. **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 6.4.

6.5.2   Trigger events

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

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

6.5.3   Handling of GoToR, URI and SubmitForm actions

While permitted to be present in a conforming file, there are three types of actions for which a conforming interactive reader shall provide special treatment – the **GoToR**, the **URI** and the **SubmitForm** actions. The conforming interactive reader 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.

In addition, since the actual invocation of these three actions by conforming interactive readers involves the locating of and interacting with other files that might or might not be conforming, the reader may choose to not allow the actual invocation of these actions.

NOTE   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 such actions. However, this part of ISO 19005 does not prescribe any specific behaviour or the technical implementation details that interactive readers might use to meet these functional requirements.

6.6   Metadata

6.6.1   General

Requirements for metadata within conforming files are specified in 6.6.2 to 6.6.6. 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 possibly 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.

6.6.2   Metadata streams

6.6.2.1   General

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

In addition, all metadata streams present in the PDF shall conform to the XMP Specification.

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 the XMP Specification.

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

6.6.2.2   Namespaces and prefixes

According to the W3C XML Namespace recommendation[16], namespace prefixes are shortcuts to namespace URIs. No significance is given to the prefix itself, except where a specific prefix is identified as required, any prefix can be used. The prefixes in Table 1 should be used for all properties using the namespaces identified by the URIs listed in that table.

In addition, namespace URIs are for identification purposes only and are not required to be actionable links. None of the namespace URIs defined in this part of ISO 19005 is guaranteed to be an actionable link. Attempting to de-reference or follow any of these links might not result in a valid web page.

URI	Prefix
<http://purl.org/dc/elements/1.1/>	dc
<http://ns.adobe.com/pdf/1.3/>	pdf
<http://ns.adobe.com/xap/1.0/>	xmp
<http://ns.adobe.com/xap/1.0/mm/>	xmpMM

6.6.2.3 Schemas

6.6.2.3.1 General

All properties specified in XMP form shall use either the predefined schemas defined in the XMP Specification, ISO 19005-1 or this part of ISO 19005, or any extension schemas that comply with 6.6.2.3.2.

6.6.2.3.2 Extension schemas

All extension schemas referenced from any metadata stream in a conforming file shall have their descriptions embedded within the referencing metadata stream or the metadata stream that is the value of the **Metadata** key in the **Catalog**. Any schemas present in the metadata stream referenced from the **Catalog** shall be inherited by and apply to all metadata streams; however, all other schemas shall be considered only in the context of the stream in which it is embedded. Schemas present in metadata streams other than that of the **Catalog** may extend or replace some or all of a schema that was inherited from the **Catalog**'s stream.

NOTE   The reason for putting extension schemas in the document's metadata stream is to avoid duplication of common schemas across multiple metadata streams.

Extension schemas shall be specified using the PDF/A extension schema container schema defined in 6.6.2.3.3. All fields described in each of the tables in 6.6.2.3.3 shall be present in any extension schema container schema.

6.6.2.3.3 Extension schema container schemas

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

Table 2 — PDF/A extension schema container schema

Property	Value type	Category	Explanation
pdfaExtension:schemas	Bag Schema	Internal	Container for all embedded extension schema descriptions

The Schema type defined in Table 3 is an XMP structure containing the definition of an extension schema. The field namespace URI is <http://www.aiim.org/pdfa/ns/schema#>. The required field namespace prefix is **pdfaSchema**.

Table 3 — PDF/A Schema value type

Field name	Value type	Explanation
pdfaSchema:schema	*Text*	Description of schema
pdfaSchema:namespaceURI	*URI*	Schema namespace URI
pdfaSchema:prefix	*Text*	Preferred schema namespace prefix
pdfaSchema:property	*Seq Property*	List of schema properties
pdfaSchema:valueType	*Seq ValueType*	Description of schema-specific value types

The **Property** type defined in Table 4 is an XMP structure containing the definition of a schema property. The field namespace URI is <http://www.aiim.org/pdfa/ns/property#>. The required field namespace prefix is **pdfaProperty**.

Table 4 — PDF/A Property value type

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

The values for **pdfaProperty:valueType** shall either be value types defined in the XMP Specification or custom value types defined within the extension schema.

The **ValueType** type defined in Table 5 is an XMP structure containing the definition of all property value types used by embedded extension schemas that are not defined in the XMP Specification. The field namespace URI is <http://www.aiim.org/pdfa/ns/type#>. The required field namespace prefix is **pdfaType**.

Table 5 — PDF/A ValueType value type

Field name	Value type	Explanation
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*	Description of the structured fields

The **Field** type defined in Table 6 is an XMP structure containing the definition 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**.

Not for Resale

Table 6 — PDF/A Field value type

Field name	Value type	Explanation
pdfaField:name	*Text*	Field name
pdfaField:valueType	*Open Choice of Text*	Field value type drawn from the XMP Specification or an embedded PDF/A value type extension schema inside the same extension schema as where the field that makes use of this value type definition is defined
pdfaField:description	*Text*	Field description

6.6.3   Document information dictionary

A document information dictionary may appear within a conforming file. If it does appear, a compliant PDF/A-2 reader shall ignore it. A PDF/A-2 conforming writer should ensure that the values in the document information dictionary are consistent with the corresponding values in the document's metadata stream as listed in Table 7.

NOTE   Since a document information dictionary is allowed within a conforming file, it is possible for a single file to be conformant with multiple standards, including this part of ISO 19005, PDF/X (ISO 15930-1, ISO 15930-3, ISO 15930-4, ISO 15930-6, and ISO 15930-7) and PDF/E-1 (ISO 24517-1).

Table 7 — Crosswalk between document information dictionary and XMP properties

Document information dictionary		XMP	
Entry	PDF type	Property	XMP type
Title	*text string*	**dc:title["x-default"]**	*Text*
Author	*text string*	**dc:creator[0]**	*ProperName*
Subject	*text string*	**dc:description["x-default"]**	*Text*
Keywords	*text string*	**pdf:Keywords**	*Text*
Creator	*text string*	**xmp:CreatorTool**	*AgentName*
Producer	*text string*	**pdf:Producer**	*AgentName*
CreationDate	*date*	**xmp:CreateDate**	*Date*
ModDate	*date*	**xmp:ModifyDate**	*Date*

6.6.4   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 PDF/A Identification schema defined in Table 8 uses the namespace URI <http://www.aiim.org/pdfa/ns/id/>. The required schema namespace prefix is **pdfaid**.

Table 8 — PDF/A Identification schema

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

The value of **pdfaid:part** shall be the part number of ISO 19005 to which the file conforms. Files prepared in compliance with this part of ISO 19005 shall use a value of *2*.

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.

If the file conforms to a version of ISO 19005 that is defined by a corrigendum to a part, then the value of **pdfaid:corr** shall be the corrigendum 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*. A Level U conforming file shall specify the value of **pdfaid:conformance** as *U*.

The values of the **pdfaid:part**, **pdfaid:amd**, **pdfaid:corr** 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.

6.6.5   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) 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 the use of properties such as the **xmp:Identifier** property, the **xmpMM:InstanceID**, **xmpMM:DocumentID**, **xmpMM:VersionID** properties, or use of properties from an extension schema. Since any identification system may be used so long as the properties comply with XMP requirements and this part of ISO 19005, the previous list shall not be considered as exhaustive.

If an **xmpMM:History** entry, as described in 6.6.6, is added to a conforming file, then the changing identifier part of the file trailer dictionary **ID** key shall be modified according to 6.1.3.

6.6.6   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 inside the XMP metadata stream that is the value of the **Metadata** entry in the document catalog dictionary. For each action that is recorded:

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

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

⎯ the **instanceID** field should be specified.

NOTE 1   Applications with specific auditing requirements might 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 ISO 32000-1 to PDF/A-2), alterations to file content or functionality (e.g. embedded JavaScript and audio objects 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, and 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.

NOTE 2   The "Tracking Document History" section of the Partners Guide to XMP for Dynamic Media describes a more comprehensive solution to document history tracking via XMP metadata that some conforming readers might choose to implement instead.

6.7   Logical structure

6.7.1   General

Subclause 6.7 is applicable only to files meeting Level A conformance. For Level B and Level U conformance, the requirements of this subclause may be ignored.

The intent of the requirements in 6.7.2 to 6.7.8 is to provide guidance in incorporating higher-level semantic information in PDF/A-2 conforming documents based on the recommendations in ISO 32000-1:2008, 14.7 and 14.8. Such information will help to ensure the recovery of the textual content of a conforming file in the natural reading order of the language in which they are written. In addition, the presence of the structure will enable richer accessibility of the PDF for those users with disabilities as detailed in ISO 32000-1:2008, 14.9.

NOTE 1   Examples of such information are structure hierarchy, natural language specification, alternative descriptions, non-textual annotations, replacement text and expansions of abbreviations and acronyms.

PDF/A-2 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.

NOTE 2   It is inadvisable for writers to generate structural or semantic information using automated processes without appropriate verification.

6.7.2   Tagged PDF

6.7.2.1   General

A Level A conforming file shall meet all of the requirements set forth for Tagged PDF in ISO 32000-1:2008, 14.8.

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

6.7.2.2   Mark information dictionary

The document catalog dictionary shall include a **MarkInfo** dictionary containing an entry, **Marked**, whose value shall be *true*.

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

6.7.3   Artefacts

6.7.3.1   Specification of artefacts

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 artefacts, respectively, as described in ISO 32000-1:2008, 14.8.2.2.1 and 14.8.2.2.2.

6.7.3.2 Word boundaries

For languages and script systems that normally use space characters (ISO 32000-1, 4.42) to indicate word boundaries, the following additional restriction shall apply:

Within show strings, word boundaries 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 unless immediately followed by a punctuation character.

NOTE A single word can span two or more show strings; word boundaries 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 boundaries, a sequence of two or more consecutive space characters is semantically equivalent to a single spacing character.

6.7.3.3 Structure hierarchy

The logical structure of the conforming file shall be described by a structure hierarchy rooted in the **StructTreeRoot** entry of the document's **Catalog** dictionary, as described in ISO 32000-1:2008, 14.7.

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 ISO 32000-1:2008, 14.8.4, 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.

6.7.3.4 Structure types

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

6.7.4 Natural language specification

The default natural language for all text in a file should be specified by the **Lang** entry in the document's **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 ISO 32000-1:2008, 14.9.2.

If the **Lang** entry is present in the document's **Catalog** dictionary or in a structure element dictionary or property list, its value shall be a language identifier as described in ISO 32000-1:2008, 14.9.2.

NOTE 1 Annex C of this part of ISO 19005 also gives some guidance for best practices in this area.

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 ISO 32000-1:2008, 7.9.2.

NOTE 2 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.

6.7.5   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 ISO 32000-1:2008, 14.9.3.

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

6.7.6   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.

6.7.7   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 ISO 32000-1:2008, 14.9.4.

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

6.7.8   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 ISO 32000-1:2008, 14.9.5.

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

6.8   Embedded files

A file specification dictionary, as defined in ISO 32000-1:2008, 7.11.3, may contain the **EF** key, provided that the embedded file is compliant with either ISO 19005-1 or this part of ISO 19005. The file specification dictionary for an embedded file shall contain the **F** and **UF** keys and should contain the **Desc** key.

A file's name dictionary, as defined in ISO 32000-1:2008, 7.7.4, may contain the **EmbeddedFiles** key, provided that all of the embedded files are compliant with either ISO 19005-1 or this part of ISO 19005.

NOTE   The prohibition of non-PDF/A compliant documents has the implicit effect of disallowing embedded files that can create external dependencies and complicate preservation efforts.

A conforming reader shall provide a mechanism to display the name strings from the value of the **EmbeddedFiles** key in the names dictionary of a conforming file. In addition, a conforming reader may also choose to display information from the associated embedded file stream dictionaries or their **Params** dictionary.

6.9   Optional content

Optional content may be used in PDF/A-2 files to allow multiple variants of a document to be supplied in a single file. Common use cases for this include multilingual documents, regional versioning or different object groupings on a CAD-type drawing.

A variant consists of one or more optional content groups (OCGs), which are associated through an optional content membership dictionary (OCMD) and an optional content configuration dictionary (OCCD). Each optional content configuration dictionary determines which OCGs are grouped together to form a single variant.

The document's **Catalog** may contain the **OCProperties** key. The presence of **OCProperties** indicates that the file contains variants, and the requirements of this section apply.

In the absence of explicit instructions to the contrary, a PDF/A-2 reader shall render the file in the default state set by the value of the **D** key in the **OCProperties** dictionary, as specified in "Determining the State of Optional Content Groups" (ISO 32000-1:2008, 8.11.4).

The **OCProperties** dictionary may also contain the **Configs** key. If a **Configs** key is present, then each element of the array that forms the value of the **Configs** key shall define a single variant.

Each optional content configuration dictionary that forms the value of the **D** key, or that is an element in the array that forms the value of the **Configs** key in the **OCProperties** dictionary, shall contain the **Name** key, which is the identifier of the variant, whose value shall be unique amongst all optional content configuration dictionaries within the PDF/A-2 file.

NOTE 1   It is recommended that all values for the **Name** key be selected in such a way as to allow unambiguous identification of the correct content that is to be printed or displayed.

If an optional content configuration dictionary contains the **Order** key, the array which is the value of this **Order** key shall contain references to all OCGs in the conforming file.

A conforming interactive reader shall provide a means to display the contents of the **Order** key from any OCCDs present in the conforming file that contain an **Order** key or that inherit the **Order** key from the default OCCD. In addition, if a conforming file contains OCCDs in addition to the default OCCD, then a conforming interactive reader shall provide a means to display the list of OCCDs from which a user can choose which one to view and print.

NOTE 2   The OCGs in an Order array can be structured using a hierarchy of arrays and not simply a flat list.

The **AS** key shall not appear in any optional content configuration dictionary.

NOTE 3   This prevents the automatic adjustment of states based on usage information (ISO 32000-1:2008, 8.11.4.5).

NOTE 4   The requirements of 6.2.11 apply for all fonts used in all optional content, even where a particular exchange will not result in some optional content being rendered.

A conforming reader shall not use the value of the **Intent** key.

6.10 Use of alternate presentations and transitions

There shall be no **AlternatePresentations** entry in the document's name dictionary. There shall be no **PresSteps** entry in any Page dictionary.

NOTE   These restrictions prohibit the use of the slide show alternate presentation, which can cause the on-screen presentation to differ from what is seen when printing the same file.

A PDF/A-2 conforming interactive reader shall ignore the **Trans** and **Dur** keys present in a **Page** dictionary.

6.11 Document requirements

The document catalog shall not contain the **Requirements** key.

NOTE   All PDF/A-2 conforming readers meet the requirements of this part of ISO 19005 and therefore conforming documents do not have specific requirements.

Annex A
(normative)

Method for determining transparency on a page

A.1 General

This section describes the method that a conforming reader shall use to determine whether a given page contains any graphical elements whose associated graphic state contains transparency or which are otherwise involved in a transparency operation.

A.2 Page content

For each graphical element to be rendered on the page, the graphic state shall be checked for any of the following conditions being present. If they are, then the element is considered involved in transparency and therefore the page contains transparency.

— **SMask** key is present and its value is of type dictionary.

— **ca** key is present and its value is less than 1.

— **CA** key is present and its value is less than 1.

— **BM** key is present and its value is not *Normal.*

If an element's graphic state sets a colourspace that is a Type 1 Pattern, then the **Pattern** resource shall be treated as a Form XObject and processed according to A.3.

In addition to the graphic state, certain types of graphical elements need additional processing. Any graphical element that represents a Form XObject shall be processed according to A.3. Graphical elements associated with Image XObjects shall be processed according to A.4. Text elements shall be processed according to A.5.

Since Annotations require an appearance stream which is drawn by a conforming reader on top of the page content, it is possible that their presence might cause a page without any transparency to acquire some transparency. Therefore, all annotations objects in the page dictionary's **Annots** array shall have their appearance streams, if present, processed as a Form XObject, according to A.3.

A.3 Form XObjects

If the XObject's dictionary contains a **Group** key, and the value of the **Group** is *Transparency*, then the XObject is part of a transparency group and any page on which it is placed contains transparency.

In addition, the content stream of the Form XObject shall be processed according to A.2.

A.4 Image XObjects

If the XObject's dictionary contains an **SMask** key and its value is of type stream, then the image has transparency associated with it and any page on which it is placed contains transparency.

If the XObject's dictionary contains an **SMaskInData** key whose value is greater than zero (0), then the image has transparency associated with it and any page on which it is placed contains transparency.

A.5 Text objects

For any graphical element that represents text drawing, its text state shall be checked to determine the type of font being used for rendering. If the **Subtype** of the font dictionary is *Type3*, then each object in its CharProc array shall be processed as a Form XObject according to A.3.

Annex B
(normative)

Requirements for digital signatures in PDF/A

B.1 Signing

When computing the digest for the file, it shall be computed over the entire file, including the signature dictionary but excluding the PDF Signature itself. This range is then indicated by the **ByteRange** entry of the signature dictionary.

NOTE 1 This makes normative a recommendation in ISO 32000-1:2008, 12.8.1.

NOTE 2 By restricting the ByteRange entry in this way, it ensures that there are no bytes in the PDF that are not covered by the digest, other than the PDF Signature itself.

The PDF Signature (a DER-encoded PKCS#7 binary data object) shall be placed into the **Contents** entry of the signature dictionary. The PKCS#7 object shall conform to the PKCS#7 specification in RFC 2315. At minimum, it shall include the signer's X.509 signing certificate and there shall only be a single signer (e.g. a single "SignerInfo" structure) in the PDF Signature.

NOTE 3 Although ISO 32000-1 also allows the value of the **Contents** entry of the signature dictionary to be a DER-encoded PKCS#1 binary data object, that format is not recommended.

As recommended by ISO 32000-1:2008, 12.8.3.3.1, timestamping and revocation information should be included in order to improve the long-term non-repudiation properties of the signature. This revocation information, and as much of the complete chain of certificates as is available, shall be captured and validated before completing the creation of the PDF Signature. In addition, the revocation information shall be a signed attribute of the PDF Signature.

ISO 32000-1:2008 allows the inclusion of one or more RFC 5755 attribute certificates to be associated with the signer certificate. However, a conforming writer should not include them as they are not widely supported; use of this attribute will thus reduce interoperability.

NOTE 4 A conforming reader is not required to process any attribute certificates.

ISO 32000-1:2008 permits values for **Filter** and **SubFilter** other than those documented in ISO 32000-1:2008, 12.8.3.3.1. In order to validate signatures in PDF/A, the conforming reader shall be able to call the appropriate signature handler. In order to maximize interoperability, the two values for **SubFilter** listed in ISO 32000-1:2008, 12.8.3.3.1 (i.e., *adbe.pkcs7.detached* and *adbe.pkcs7.sha1*) shall be supported by a conforming reader.

B.2 Validation

A conforming reader shall perform the following steps when validating any signatures in a conforming file:

— Verify that the document digest matches that in the signature as specified in ISO 32000-1, 12.8.1.

— Validate the path of certificates used to verify the binding between the subject distinguished name and subject public key as specified in RFC 3280. The validity checks shall be carried out at the time indicated, either by time-stamp or some other trusted indication of the signing time. The revocation status shall be checked.

When validating the PDF Signature, a conforming reader may ignore any embedded revocation information in favour of alternative storage or referenced data as per its own policies.

Annex C
(informative)

Best practices for PDF/A

C.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. Preferably, the **xmpMM:History** property is also used to indicate any non-XMP elements that have not been converted.

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

C.2 Natural language identifiers

Natural language is declared as discussed in ISO 32000-1:2008, 14.9.2, and ISO 32000-1:2008, 7.9.2.

Language codes are identified using ISO 639-2 or ISO 3166-2 registered identifiers. Private use identifiers are used only if the language does not have a defined identifier within ISO 639-2 or ISO 3166-2. In the event that a language is truly unknown, the identifier *x-unknown* can be used.

NOTE   ISO 32000-1:2008 uses ISO 639-2 or ISO 3166-2 as the basis for its language identifiers.

Additionally, the following recommendations can be followed when possible:

— Documents not expressed in a natural language can declare the root language as *zxx*.

— Documents expressed in a language unknown to the author or creator can declare the root language as *und*.

— Documents with equal proportions of multiple languages can declare the root language as *mul* and use structure elements to group and tag each content block with the correct code for the language of the content.

— Changes in natural language are declared.

— Text direction is declared and changes in text direction are declared.

— When the meaning is ambiguous to the intended readership, abbreviations, acronyms, initialisms, and short forms are tagged with **Abbr** and their expansion should be given as per ISO 32000-1:2008, 14.9.5.

C.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 5.2 and the additional guidelines of this clause.

ISO 15489-1:2001, 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 can be retained when they are captured or converted to PDF/A.

To meet this critical archival need, PDF/A capture or conversion processes 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 are not using 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 is not substituting 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 can involve loss of data through imprecise interpretation of scanned characters.

C.4 Image processing

To avoid degrading an image compressed using lossy compression (such as JPEG), a conforming writer should either include the original compressed image in the PDF/A file without altering it or decompress the image and recompress it using a lossless compression algorithm.

Annex D
(informative)

Incorporation of XFA datasets into a PDF/A-2 conforming file

The *<xfa:datasets>* element of the XFA, comprising *<xfa:data>* elements, can be conserved in a PDF/A-2 conforming file by storing its data stream as a value in a names tree that itself is the value of the **XFAResources** key of the **Names** dictionary of the document catalog dictionary.

A conforming reader shall not use any value in the names tree associated with **XFAResources** for rendering.

NOTE   The **Names** dictionary is an area of PDF that represents a collection of document-level data which can be easily located by PDF processing tools thus allowing the data to be easily extracted into its original XML format for processing.

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 <http://www.ecma-international.org/publications/files/ecma-st/Ecma-006.pdf>

[3]   ISO 639-2:1998, *Codes for the representation of names of languages — Part 2: Alpha-3 code*

[4]   ISO 3166-2, *Codes for the representation of names of countries and their subdivisions — Part 2: Country subdivision code*

[5]   ISO 14721:2003, *Space data and information transfer systems — Open archival information system — Reference model*

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

[7]   ISO/TR 15801, *Document management — Information stored electronically — Recommendations for trustworthiness and reliability*

[8]   ISO 15930-1:2001, *Graphic technology — Prepress digital data exchange — Use of PDF — Part 1: Complete exchange using CMYK data (PDF/X-1 and PDF/X-1a)*

[9]   ISO 15930-3:2002, *Graphic technology — Prepress digital data exchange — Use of PDF — Part 3: Complete exchange suitable for colour-managed workflows (PDF/X-3)*

[10]   ISO 15930-4:2003, *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)*

[11]   ISO 15930-6:2003, *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]   *Partners Guide to XMP for Dynamic Media*, Adobe Systems Incorporated. Available from <http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/DynamicMediaXMPPartnerGuide.pdf>

[14]   *The Unicode Standard*, Version 4.0. The Unicode Consortium. Boston, MA: Addison-Wesley, 2003, ISBN 0-321-18578-1

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

[16]   *Namespaces in XML 1.0 (Second Edition)*, W3C Recommendation, 16 August 2006. Available from <http://www.w3.org/TR/2006/REC-xml-names-20060816/>

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

[18]   ICC.1:2004-10, *File Format for Color Profiles*, International Color Consortium. Available from <http://www.color.org/ICC-1_2004-10.PDF>

[19]   RFC 5755, *An Internet Attribute Certificate Profile for Authorization*. Available from http://www.rfc-editor.org[3)]

[20]   *XMP: Extensible Metadata Platform*, September 2005, Adobe Systems Incorporated. Available from <http://www.aiim.org/documents/standards/pdf/xmp_specification.pdf>

---

3)   Cancels and replaces RFC 3281.

ISO 19005-2:2011(E)

ICS 35.240.30; 37.100.99

Price based on 36 pages