
**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)



This document is a preview generated by EMS



COPYRIGHT PROTECTED DOCUMENT

© ISO 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
4 Notation	4
5 Conformance levels	4
5.1 General	4
5.2 Level A conformance	5
5.3 Level B conformance	5
5.4 Level U conformance	5
5.5 Conforming readers	5
6 Technical requirements	6
6.1 File structure	6
6.2 Graphics	8
6.3 Annotations	17
6.4 Interactive forms	18
6.5 Action	19
6.6 Metadata	20
6.7 Logical structure	25
6.8 Embedded files	27
6.9 Optional content	27
6.10 Use of alternate presentations and transitions	28
6.11 Document requirements	28
Annex A (normative) Method for determining transparency on a page	29
Annex B (normative) Requirements for digital signatures in PDF/A	31
Annex C (informative) Best practices for PDF/A	32
Annex D (informative) Incorporation of XFA datasets into a PDF/A-2 conforming file	34
Bibliography	35

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;

ISO 19005-2:2011(E)

- 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 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**