INTERNATIONAL STANDARD

ISO 19005-4

First edition
2020-11

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

Part 4: Use of ISO 32000-2 (PDF/A-4)

Reference number
ISO 19005-4:2020(E)

© ISO 2020



COPYRIGHT PROTECTED DOCUMENT

© ISO 2020

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **v**
Introduction ... **vi**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Notation ... **2**
5 Conformance ... **3**
5.1 Conforming Files ... 3
5.2 Conforming Processor ... 4
6 Technical requirements ... **4**
6.1 File structure ... 4
6.1.1 General ... 4
6.1.2 File header ... 4
6.1.3 File trailer ... 4
6.1.4 Cross reference table ... 4
6.1.5 String objects ... 5
6.1.6 Stream objects ... 5
6.1.7 Name objects ... 5
6.1.8 Indirect objects ... 5
6.1.9 Inline image dictionaries ... 5
6.1.10 Linearized PDF ... 6
6.1.11 Permissions ... 6
6.1.12 Document catalog dictionary ... 6
6.2 Graphics ... 6
6.2.1 General ... 6
6.2.2 Content streams ... 6
6.2.3 Output intent ... 7
6.2.4 Colour spaces ... 7
6.2.5 Extended graphics state ... 9
6.2.6 Flatness ... 10
6.2.7 Images ... 10
6.2.8 XObjects ... 11
6.2.9 Transparency ... 11
6.2.10 Fonts ... 11
6.3 Annotations ... 15
6.3.1 Annotation types ... 15
6.3.2 Annotation dictionaries ... 15
6.3.3 Annotation appearances ... 15
6.3.4 Display of annotation contents ... 16
6.4 Interactive forms ... 16
6.4.1 General ... 16
6.4.2 XFA forms ... 16
6.5 Digital signatures ... 16
6.5.1 General ... 16
6.5.2 Signature profiles ... 17
6.5.3 Proof of existence ... 17
6.5.4 Validation ... 17
6.6 Action ... 18
6.6.1 General ... 18
6.6.2 Handling of JavaScript actions ... 18
6.6.3 Trigger events ... 18
6.6.4 Handling of GoToR, GoToE, URI and SubmitForm actions ... 19

© ISO 2020 – All rights reserved

6.7 Metadata 19
6.7.1 General 19
6.7.2 Metadata streams 19
6.7.3 Version identification 20
6.7.4 File identifiers 20
6.7.5 File provenance information 21
6.8 Logical structure 21
6.9 Embedded files 21
6.10 Optional content 22
6.11 Use of alternate presentations and transitions 23
6.12 Document requirements 23
6.13 Print scaling 23
6.14 Geospatial 23
6.15 Measurement Properties 23
Annex A (normative) **Requirements for a PDF/A with Embedded Files (PDF/A-4f)** **24**
Annex B (normative) **Requirements for a PDF/A for Engineering (PDF/A-4e)** **25**
Bibliography **27**

 © ISO 2020 – All rights reserved

Foreword

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

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

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

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

A list of all parts in the ISO 19005 series can be found on the ISO website.

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

© ISO 2020 – All rights reserved

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.

This document is part of a series of documents, of which this is part 4. This allows future parts to be created without rendering this document or applications based on this document 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 visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files. Some parts of ISO 19005 also define a framework for representing the logical structure and other semantic information of electronic documents within conforming files. These goals are accomplished by identifying the set of PDF components that can be used, and restrictions on the form of their use, within conforming PDF/A files.

Parts 3 & 4 in this series enable PDF documents to serve as containers for other file formats, so that a single physical file can contain not only the visual representation but also other representations including the original authored version, richer semantic formats, and others. This standard does not address the long-term suitability of formats, that may be embedded, other than those compliant with any part of this document.

This document is based on PDF version 2.0 (as defined in ISO 32000-2:—[1)]) and as such provides recommendations in how to properly archive content that uses some of the newer features present there including page level output intents, associated files and improvements to tagged PDF.

This document also introduces some new directions in archiving non-static content that can be present in PDF documents, such as form fields and ECMAScript. It seeks to preserve more information in the file (by not requiring its removal during the archival process) and puts a greater burden on conforming viewers to ensure that such information does not alter the visual appearance of the file during consumption.

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 on the processes that generate the conforming file beyond those imposed by this document. In addition, it is important for those organizations to implement policies and practices regarding the inspection of conforming files for correct visual appearance.

This document is one component of an organization's electronic archival environment for long-term retention of documents. Successful implementation of this document for archival purposes depends upon:

— the retention requirements of an organization's archival environment, records management policies and procedures as specified in ISO 15489-1[2];

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

 © ISO 2020 – All rights reserved

— any additional conditions necessary to ensure the persistence of electronic documents and their characteristics over time, including, but not limited to, those defined by ISO 14721[1], and ISO/TR 18492[3];

— 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 document 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 document (in conjunction with its normative references) provides sufficient information to interpret any conforming PDF/A-4 file.

The PDF Association may maintain an ongoing series of application notes for guiding developers and users of this document. These application notes are available at https://www.pdfa.org/. The PDF Association will also retain copies of the specific non-ISO normative references of this document which are publicly available electronic documents.

© ISO 2020 – All rights reserved

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

Part 4: Use of ISO 32000-2 (PDF/A-4)

1 Scope

This document specifies the use of the Portable Document Format (PDF) 2.0, as formalized in ISO 32000-2:—, for preserving the static visual representation of page based electronic documents over time in addition to allowing any type of other content to be included as an embedded file or attachment.

This document does not apply to:

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

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

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

— required computer hardware and/or operating systems.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 14533-3, *Processes, data elements and documents in commerce, industry and administration — Long term signature profiles — Part 3: Long term signature profiles for PDF Advanced Electronic Signatures (PAdES)*

ISO 16684-2, *Graphic technology — Extensible metadata platform (XMP) — Part 2: Description of XMP schemas using RELAX NG*

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

ISO 19005-2, *Document management — Electronic document file format for long-term preservation — Part 2: Use of ISO 32000-1 (PDF/A-2)*

ISO 32000-2:—, *Document management — Portable document format — Part 2: PDF 2.0*

XMP Specification Part 2: Additional Properties – Adobe Systems – https://www.adobe.com/content/dam/acom/en/devnet/xmp/pdfs/XMPSDKReleasecc-2020/XMPSpecificationPart2.pdf

3 Terms and definitions

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

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

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

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

© ISO 2020 – All rights reserved

3.1
end-of-file marker
five-character sequence %%EOF marking the end of a PDF file

3.2
extension schema
conforming XMP schema that is not defined in the XMP Specification nor part 1 or part 2 of this document

3.3
font
identified collection of graphics that may be glyphs or other graphic elements

[SOURCE: ISO 32000-2:—]

3.4
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* (3.3) interpreter

[SOURCE: ISO 32000-2:—]

3.5
interactive processor
processor (3.8) that requires or allows human interaction with the content and other objects contained in the document during the software's processing phase

Note 1 to entry: A file viewing tool is an example of an interactive processor; a raster image processor is an example of a processor that is not interactive.

3.6
long-term
period of time long enough for there to be concern about the impacts of changing technologies

3.7
PDF
Portable Document Format
file format defined in ISO 32000-2:—

3.8
processor
software application that is able to read and process conforming files

3.9
writer
software application that is able to write conforming files

3.10
XMP packet
structured wrapper for serialized XMP metadata

[SOURCE: ISO 32000-2:—]

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 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.

© ISO 2020 – All rights reserved