

Internet Engineering Task Force (IETF)
Request for Comments: 8118
Obsoletes: 3778
Category: Informational
ISSN: 2070-1721

M. Hardy
L. Masinter
D. Markovic
Adobe Systems Incorporated
D. Johnson
PDF Association
M. Bailey
Global Graphics
March 2017

The application/pdf Media Type

Abstract

The Portable Document Format (PDF) is an ISO standard (ISO 32000-1:2008) defining a final-form document representation language in use for document exchange, including on the Internet, since 1993. This document provides an overview of the PDF format and updates the media type registration of "application/pdf". It obsoletes RFC 3778.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc8118>.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. History	3
3. Fragment Identifiers	3
4. Subset Standards	5
5. PDF Versions	6
6. PDF Implementations	7
7. Security Considerations	7
8. IANA Considerations	8
9. References	9
9.1. Normative References	9
9.2. Informative References	9
Appendix A. Changes since RFC 3778	11
Authors' Addresses	12

1. Introduction

This document is intended to provide updated information on the registration of the MIME Media Type "application/pdf" for documents in the PDF (Portable Document Format) syntax. It obsoletes [RFC3778].

PDF was originally envisioned as a way to reliably communicate and view printed information electronically across a wide variety of machine configurations, operating systems, and communication networks.

PDF is used to represent "final form" formatted documents. PDF pages may include text, images, graphics, and multimedia content such as video and audio. PDF is also capable of containing auxiliary structures, including annotations, bookmarks, file attachments, hyperlinks, logical structures, and metadata. These features are

useful for navigation and building collections of related documents, as well as for reviewing and commenting on documents. A rich JavaScript model has been defined for interacting with PDF documents.

The imaging model for PDF was originally based on the PostScript [PS] page description language, used to render complex text, images, and graphics in a device-independent and resolution-independent manner.

PDF supports encryption and digital signatures. The encryption capability is combined with access control information to facilitate management of the functionality available to the recipient. PDF supports the inclusion of document and object-level metadata through the eXtensible Metadata Platform [XMP].

2. History

PDF is used widely in the Internet community. The first version of PDF, 1.0, was published in 1993 by Adobe Systems Incorporated. Since then, PDF has grown to be a widely used format for capturing and exchanging formatted documents electronically across the Web, via email and virtually every other document-exchange mechanism. In 2008, PDF 1.7 was adopted as an ISO standard (ISO 32000-1:2008 [ISOPDF]) using the ISO "Fast-Track" process. That specification is technically identical to Adobe Portable Document Format version 1.7 [AdobePDF].

The ISO TC-171 committee developed a "refresh" of PDF, known as ISO 32000-2; the version is PDF 2.0 [ISOPDF2].

In addition to ISO 32000-1:2008 and ISO 32000-2, several subset standards have been defined to address specific use cases and standardized by the ISO. These standards include PDF for Archival (PDF/A) [ISOPDFA], PDF for Engineering (PDF/E) [ISOPDFE], PDF for Universal Accessibility (PDF/UA) [ISOPDFUA], PDF for Variable Data and Transactional Printing (PDF/VT) [ISOPDFVT], and PDF for Prepress Digital Data Exchange (PDF/X) [ISOPDFX]. The subset standards are fully compliant PDF files capable of being displayed in a general PDF viewer.

3. Fragment Identifiers

Fragment identifiers appear at the end of a URI and provide a way to reference an anchor to subordinate content within the target of the URI, or additional parameters to the process of opening the identified content. The syntax and semantics of fragment identifiers are referenced in the media type definition.

The specification of fragment identifiers for PDF appeared originally in [RFC3778] and is now included in ISO 32000-2 [ISOPDF2]. This section is a summary of that material. Any disagreements between [ISOPDF2] and this document should be resolved in favor of the ISO 32000-2 definition.

A fragment identifier for PDF has one or more parameters, separated by the ampersand (&) or pound (#) character. Each parameter consists of the parameter name, "=" (equal), and the parameter value; lists of values are comma-separated, and parameter value strings may be URI-encoded [RFC3986]. Parameters are processed left to right.

Coordinate values (such as <left>, <right>, and <width>) are expressed in the default user space coordinate system of the document: 1/72 of an inch measured down and to the right from the upper left corner of the (current) page ([ISOPDF2] 8.3.2.3 "User Space").

The following parameters identify subordinate content of a PDF file but also may be used to set the document view to make the (start of) the identified content visible:

page=<pageNum>

Identifies a specified (physical) page; the first page in the document has a pageNum value of 1.

nameddest=<name>

Identifies a named destination ([ISOPDF2] 12.3.2.4 "Named destinations").

structelem=<structID>

A byte string with URI encoding; identifies the structure element with the ID key within a StructElem dictionary of the document.

comment=<commentID>

The value of an annotation name, which is defined by the NM key in the corresponding annotation dictionary of the selected page ([ISOPDF2] 12.5.2 "Annotation dictionaries").

ef=<name>

Identifies the embedded file where the parameter string <name> matches a file specification dictionary in the EmbeddedFiles name tree. If the "ef" parameter is not at the end of the fragment identifier, then the rest of the fragment identifier (after the ampersand or hash delimiter) is applied to the embedded file according to its own media type. This allows identification of content within the embedded file (which itself might be a PDF file).

NOTE: When attempting to open a PDF file that is not from a trusted source, the processor may choose to prompt the user or even prevent the file from being opened.

These parameters operate on the view of the PDF document when it is opened:

zoom=<scale>,<left>,<top>

<scale> is the percentage to which the document should be zoomed, where a value of 100 corresponds to a zoom of 100%. <left> and <top> are optional, but both must be specified if either is included.

view=<keyword>,<position>

The arguments correspond to those found in [ISOPDF2] 12.3.2.2 "Explicit destinations". <keyword> is one of the keywords defined in [ISOPDF2] "Table 149: Destination syntax" with appropriate position values.

viewrect=<left>,<top>,<width>,<height>

Set the view rectangle.

highlight=<left>,<right>,<top>,<bottom>

Highlight the specified rectangle.

search=<wordList>

Open the document and search for one or more words, selecting the first matching word in the document. <wordList> is a string enclosed in quotation marks, where individual words are separated by the space character (or %20).

fdf=<URI>

This parameter imports data into PDF form fields. The URI is either a relative or absolute URI to a Forms Data Format (FDF) or XML FDF (XFDF) file. The fdf parameter should be specified as the last parameter to a given URI.

4. Subset Standards

Several subsets of PDF have been published as distinct ISO standards:

- o PDF/X [ISOPDFX], initially released in 2001 as PDF/X-1a, specifies how to use PDF for graphics exchange, with the aim to facilitate correct and predictable printing by print service providers. The standard has gone through multiple revisions over the years and has several published parts, the most recently released being

part 8, specifying different levels of conformance: PDF/X-1a:2001, PDF/X-3:2002, PDF/X-1a:2003, PDF/X-3:2003, PDF/X-4, PDF/X-4p, PDF/X-5g, PDF/X-5pg, and PDF/X-5n.

- o PDF/A [ISOPDFA], initially released in 2005, specifies how to use PDF for long-term preservation (archiving) of electronic documents. It prohibits PDF features that are not well suited to long-term archiving of documents, including JavaScript or executable file launches. Its requirements for PDF/A viewers include color management guidelines and support for embedded fonts. There are three parts of this standard and a total of eight conformance levels: PDF/A-1a, PDF/A-1b, PDF/A-2a, PDF/A-2b, PDF/A-2u, PDF/A-3a, PDF/A-3b, and PDF/A-3u.
- o PDF/E, initially released in 2008 as PDF/E-1 [ISOPDFE], specifies how to use PDF in engineering workflows, such as manufacturing, construction, and geospatial analysis. Future revisions of PDF/E are supposed to include support for 3D PDF workflows.
- o PDF/VT, initially released in 2010, specifies how to use PDF in variable and transactional printing. It is based on PDF/X and places additional restrictions on PDF content elements and supporting metadata. It specifies three conformance levels: PDF/VT-1, PDF/VT-2, and PDF/VT-2s [ISOPDFVT].
- o PDF/UA [ISOPDFUA], initially released in 2012 as PDF/UA-1, specifies how to create accessible electronic documents. It requires the use of ISO 32000's Tagged PDF feature and adds many requirements regarding semantic correctness in applying logical structures to content in PDF documents.

All of these subset standards use the "application/pdf" media type. The subset standards are generally not exclusive, so it is possible to construct a PDF file that conforms to, for example, both PDF/A-2b and PDF/X-4 subset standards.

PDF documents claiming conformance to one or more of the subset standards use XMP metadata to identify levels of conformance. PDF processors should examine document metadata streams for such subset standards identifiers and, if appropriate, label documents as such when presenting them to the user.

5. PDF Versions

The PDF format has gone through several revisions, primarily for the addition of features. PDF features have generally been added in a way that older viewers "fail gracefully", because they can just ignore features they do not recognize. Even so, the older the PDF

version produced, the more legacy viewers will support that version, but the fewer features will be enabled. The "application/pdf" media type is used for all versions. See [ISOPDF2] Annex I, "PDF Versions and Compatibility".

6. PDF Implementations

PDF files are experienced through a reader or viewer of PDF files. For most of the common platforms in use (iOS, OS X, Windows, Android, ChromeOS, Kindle) and for most browsers (Edge, Safari, Chrome, Firefox), PDF viewing is built in. In addition, there are many PDF viewers available for download and installation. The PDF specification was published and freely available since the format was introduced in 1993, so hundreds of companies and organizations make tools for PDF creation, viewing, and manipulation.

7. Security Considerations

PDF is certainly a complex media type as per Section 4.6 of [RFC6838], which sets requirements for security analysis of media type registrations. [RFC3778] (which this document obsoletes) contained a detailed analysis of some of the security issues for PDF implementations known at the time. While the analysis isn't necessarily wrong, the threat analysis is much too limited, and the mitigations are somewhat out of date. There is now extensive literature on security threats involving PDF implementations and how to avoid them, consistent with broad implementation over decades. We are not registering a new media type but rather are making a primarily administrative update. With those caveats:

The PDF file format allows several constructs that may compromise security if handled inadequately by PDF processors. For example:

- o PDF may contain scripts to customize the displaying and processing of PDF files. These scripts are expressed in a version of JavaScript and are intended for execution by the PDF processor.
- o A PDF file may refer to other PDF files for portions of content. PDF processors may be expected to find and use these external files when processing the document.
- o PDF may act as a container for various files embedded in it (for example, as attached files). PDF processors may offer functionality to open and display such files or store them on the system, such as with the "ef" open action. The PDF specification places no restrictions on types of files that may be embedded, so

PDF processors should be extremely careful to prevent unwanted execution of attached executables or decompression of attached archives that may store dangerous files in the host file system.

- o PDF files may contain links to content on the Internet. PDF processors may offer functionality to show such content upon following the link.
- o The fragment identifier syntax (Section 3) contains directives for opening ("ef") or including ("fdf") additional material.

PDF interpreters executing any scripts or programs related to these constructs must be extremely careful to ensure that untrusted software is executed in a protected environment.

In addition, the PDF processor itself, as well as its plugins, scripts, etc., may be a source of insecurity, by either obvious or subtle means.

8. IANA Considerations

This document updates the registration of "application/pdf", a media type registration previously defined in [RFC3778], using the registration template defined in [RFC6838]:

Type name: application

Subtype name: pdf

Required parameters: none

Optional parameter: none

Encoding considerations: binary

Security considerations: See Section 7 of this document.

Interoperability considerations: See Section 5 of this document.

Published specification: ISO 32000-2 (PDF 2.0) [ISOPDF2] is the most recent.

Applications that use this media type: See Section 6 of this document.

Fragment identifier considerations: See Section 3 of this document.

Additional information:

Deprecated alias names for this type: none

Magic number(s): All PDF files start with the characters "%PDF-" followed by the PDF version number, e.g., "%PDF-1.7" or "%PDF-2.0". These characters are in US-ASCII encoding.

File extension(s): .pdf

Macintosh file type code(s): "PDF "

Person & email address to contact for further information:

Duff Johnson <duff@duff-johnson.com>, Peter Wyatt
<Peter.wyatt@cisra.canon.com.au>, ISO 32000 Project Leaders.

Intended usage: COMMON

Restrictions on usage: none

Author: Authors of this document

Change controller: ISO; in particular, ISO 32000 is by ISO TC 171/SC 02/WG 08, "PDF specification". Duff Johnson <duff@duff-johnson.com> and Peter Wyatt <Peter.wyatt@cisra.canon.com.au> are current ISO 32000 Project Leaders.

9. References

9.1. Normative References

[ISOPDF] ISO, "Document management -- Portable document format -- Part 1: PDF 1.7", ISO 32000-1:2008, 2008.

[ISOPDF2] ISO, "Document management -- Portable document format -- Part 2: PDF 2.0", ISO 32000-2:2017, 2017.

9.2. Informative References

[ISOPDFX] ISO, "Graphic technology -- Prepress digital data exchange using PDF -- Part 8: Partial exchange of printing data using PDF 1.6 (PDF/X-5)", ISO 15930-8:2008, 2008.

[ISOPDFA] ISO, "Document management -- Electronic document file format for long-term preservation -- Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3)", ISO 19005-3:2012, 2012.

- [ISOPDFE] ISO, "Document management -- Engineering document format using PDF -- Part 1: Use of PDF 1.6 (PDF/E-1)", ISO 24517-1:2008, 2008.
- [ISOPDFVT] ISO, "Graphic technology -- Variable data exchange -- Part 2: Using PDF/X-4 and PDF/X-5 (PDF/VT-1 and PDF/VT-2)", ISO 16612-2:2010, 2010.
- [ISOPDFUA] ISO, "Document management applications -- Electronic document file format enhancement for accessibility -- Part 1: Use of ISO 32000-1 (PDF/UA-1)", ISO 14289-1:2014, 2014.
- [XMP] ISO, "Graphic technology -- Extensible metadata platform (XMP) specification -- Part 1: Data model, serialization and core properties", ISO 16684-1, 2012.
- [PS] Adobe Systems Incorporated, "PostScript Language Reference, third edition", 1999, <<https://www.adobe.com/products/postscript/pdfs/PLRM.pdf>>.
- [AdobePDF] Adobe Systems Incorporated, "PDF Reference, sixth edition", 2006, <http://www.adobe.com/content/dam/Adobe/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf>.
- [RFC6838] Freed, N., Klensin, J., and T. Hansen, "Media Type Specifications and Registration Procedures", BCP 13, RFC 6838, DOI 10.17487/RFC6838, January 2013, <<http://www.rfc-editor.org/info/rfc6838>>.
- [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, DOI 10.17487/RFC3986, January 2005, <<http://www.rfc-editor.org/info/rfc3986>>.
- [RFC3778] Taft, E., Pravetz, J., Zilles, S., and L. Masinter, "The application/pdf Media Type", RFC 3778, DOI 10.17487/RFC3778, May 2004, <<http://www.rfc-editor.org/info/rfc3778>>.

Appendix A. Changes since RFC 3778

This specification replaces RFC 3778, which previously defined the "application/pdf" Media Type. Differences include the following:

- o To reflect the transition from a proprietary specification by Adobe to an open ISO standard, the Change Controller has changed from Adobe to ISO, and references have been updated.
- o The overview of PDF capabilities, the history of PDF, and the descriptions of PDF subsets were updated to reflect more recent relevant history.
- o The section on fragment identifiers was updated to closely reflect the material that has been added to ISO-32000-2.
- o The status of popular PDF implementations was updated.
- o The Security Considerations section was updated to match the current understanding of PDF vulnerabilities.
- o The registration template was updated to match RFC 6838.

Authors' Addresses

Matthew Hardy
Adobe Systems Incorporated
345 Park Ave.
San Jose, CA 95110
United States of America

Email: mahardy@adobe.com

Larry Masinter
Adobe Systems Incorporated
345 Park Ave.
San Jose, CA 95110
United States of America

Email: masinter@adobe.com
URI: <http://LarryMasinter.net>

Dejan Markovic
Adobe Systems Incorporated
345 Park Ave.
San Jose, CA 95110
United States of America

Email: dmarkovi@adobe.com

Duff Johnson
PDF Association
Neue Kantstrasse 14
Berlin 14057
Germany

Email: duff.johnson@pdfa.org

Martin Bailey
Global Graphics
2030 Cambourne Business Park
Cambridge CB23 6DW
United Kingdom

Email: martin.bailey@globalgraphics.com
URI: <http://www.globalgraphics.com>