Network Working Group	M. Hardy
Internet-Draft	L. Masinter
Obsoletes: 3778 (if approved)	D. Markovic
Intended status: Informational	Adobe Systems Incorporated
Expires: March 8, 2017	D. Johnson
	PDF Association
	M. Bailey
	Global Graphics
	September 4, 2016

The application/pdf Media Type

draft-hardy-pdf-mime-04

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

Copyright Notice

Table of Contents

1. Introduction
2. History
3. Fragment Identifiers
4. Subset Standards
5. PDF Versions
6. PDF Implementations
7. Security Considerations
8. IANA Considerations
9. References
    9.1. Normative References
    9.2. Informative References

Appendix A. Changes since RFC 3778

Authors' Addresses

1. Introduction

This document is intended to provide updated information on the registration of the MIME Media Type application/pdf for documents defined in the PDF [ISOPDF], "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 structure and metadata. These features are useful for navigation, building collections of related documents and for reviewing and commenting on documents. A rich JavaScript model has been defined for interacting with PDF documents.

PDF used the imaging model of the PostScript [PS] page description language to render complex text, images, and graphics in a device 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 e-mail and virtually every other document exchange mechanism. In 2008, PDF 1.7 was published as an ISO standard [ISOPDF], ISO 32000-1:2008. It was adopted using ISO Fast-Track process and is technically identical to Adobe Portable Document Format version 1.7 [AdobePDF] referenced by [RFC3778].

The ISO TC-171 committee is presently working on a refresh of PDF, known as ISO 32000-2, with a version of PDF 2.0, expected to be published in 2017.

In addition to ISO 32000-1:2008 and 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

A set of fragment identifiers [RFC3986] and their handling are defined in ISO 32000-2 [ISOPDF2]. This section summarizes that material; any disagreements between that document and this should be resolved in favor of the ISO definition, once that has been approved.

A fragment identifier is comprised of one or more parameters separated by the AMPERSAND (&) character. Each parameter implies an action to be performed on the document and provides values to be used for that action; the values for a parameter are introduced by an EQUAL SIGN (=) and separated by a COMMA (,); values which are strings appear in the fragment identifier using URI's percent-hex escaping -- spaces, reserved and non-ASCII strings are included by %nn encoding the UTF-8 of each character. Actions shall be processed and executed from left to right as they appear in the character string that makes up the fragment identifier.

The parameters listed in this section operate on the document at the point it is opened; for this reason they are sometimes referred to as PDF open parameters. The fragment identifier should be processed immediately after document-specified open parameters have been processed.

The table below lists the PDF open parameters relevant to PDF. All coordinate values (left, right, top, and bottom) are expressed in the default user space coordinate system (1/72 of an inch measured down and to the right from the upper-left corner); see [ISOPDF] Section 8.3.2.3 "User Space".

PDF Open Parameters

Parameter Name	**Arguments**	**Description**
nameddest	*name*	Open the document to the specified named destination. The argument provided is a string which shall correspond to the name of a destination in the target document.
page	*pageNum*	Open the document to the specified page number. The argument shall be a positive integer number. The first page in the document has a pageNum value of 1.
zoom	*scale scale,left,top*	Open the document with the specified zoom level and optional offset. The scale argument shall be either an integer or floating point value representing the percentage to which the document should be zoomed, where a value of 100 would correspond to a zoom of 100%. The left and top arguments are optional, but shall both be specified if either is included. The left and top arguments shall be integer or floating point values representing the offset from the left and top of the page in a coordinate system where 0,0 represents the top left corner of the page.
view	*keyword,position*	Open the document with the specified destination set as the view. The arguments shall correspond to those found in [ISOPDF2] 12.3.2.2, "Explicit destinations". The keyword shall correspond to one of the keywords defined in [ISOPDF2] Table 149, "Destination syntax" with appropriate position values.
viewrect	*left,top,width,height*	Open the document with the specified window view rectangle. The left and top arguments shall be integer or floating point values representing the offset from the left and top of the page in a coordinate system where 0,0 represents the top left corner of the page. The width and height arguments shall be integer or floating point values representing the width and height of the view.
highlight	*left,right,top,bottom*	Open the document with the specified rectangle highlighted. Each argument shall be an integer or floating point value representing the rectangle measured from the top left corner of the page.

Parameter Name	Arguments	Description
structelem	*structID*	Open to the page on which the first content item, hierarchically contained within the structure element identified by the structure ID, is located. If no content is contained within the hierarchy of the structure element or the ID does not match a structure element, the page number shall be treated as the first page within the document. The structID shall be a byte string with URI encoding that will be matched to the ID key within a StructElem dictionary.
comment	*commentID*	Open the document with the specified comment selected. The commentID shall be the value of an annotation name, which is defined by the NM key in the corresponding annotation dictionary (see 12.5.2 "Annotation dictionaries", Table 167). If the comment parameter is combined with another parameter that defines a specific page to be displayed, then the comment parameter shall appear after that in the URI. Note: The NM key is unique to a specific page, but is not guaranteed to be unique to a document. Unless the page on which the comment resides has been selected prior to the comment parameter, the comment will not be selected.
search	*wordList*	Open the document and search for one or more words, selecting the first matching word in the document. The wordList argument defines the search words and shall be a string enclosed within quotation marks comprised of individual words separated by space characters. Note that the space characters must be encoded.
fdf	*URI*	Open the document and then import the data from the specified FDF or XFDF file (see [ISOPDF] Section 12.7.8). The URI shall be either a relative or absolute URI to an FDF or XFDF file. The fdf parameter should be specified as the last parameter to a given URI. Note: The fdf parameter is recommended to be the last parameter so that the document can open directly to the appropriate view.
ef	*name*	Open the embedded file contained within the EmbeddedFiles name tree identified by the name. The name argument shall be a byte string used to match a file specification dictionary in the EmbeddedFiles name tree.

4. Subset Standards

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

- PDF/X, initially released in 2001 as PDF/X-1a [ISOPDFX], specifies how to use PDF for graphics exchange, with the aim to fascilitate 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-5, PDF/X-5g, PDF/X-5pg and PDF/X-5n.
- PDF/A, initially released in 2005, specifies how to use PDF for long-term preservation (archiving) of electronic documents. It prohibits PDF features which 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.
- 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.
- PDF/VT, initially released in 2010, specifies how to use PDF in variable and transactional printing. It is based on PDF/X, and adds adidtional restrictions on PDF content elements and supporting metadata. It specifies three conformance levels: PDF/VT-1, PDF/VT-2 and PDF/VT-2s [ISOPDFVT].

- PDF/UA, initially released in 2012 as PDF/UA-1 [ISOPDFUA], specifies how to create accessible electonic documents. It requires 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 application/pdf media type. The subset standards are generally not exclusive, so it is possible to construct a PDF file which 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 apropriate, label documents as such when presenting them to the user.

5. PDF Versions

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. See [ISOPDF] 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 install. 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

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

- 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.
- PDF file may refer to other PDF files for portions of content. PDF processors are expected to find these external files and load them in order to display the document.
- 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. THe PDF specification places no restrictions on types of files which may be embedded, so PDF processors should be extremely careful to prevent unwanted execution of attached executables or decompression of attached archives which may store dangerous files in the host file system.
- PDF files may contain links to content on the internet. PDF processors may offer functionality to show such content upon following the link.

PDF interpreters executing any scripts or programs related to these constructs must be extremely careful to insure 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 as 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 32000-1:2008 (PDF 1.7) [ISOPDF]. ISO 32000-2 (PDF 2.0) [ISOPDF2] is currently under development.

Applications which 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". 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.

Also available free from Adobe.

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

Currently under development - publication expected in 2017. This becomes a Normative Reference on approval.

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, "Extensible metadata platform (XMP) specification -- Part 1: Data model, serialization and core properties", ISO 16684-1, 2012.

Not available for free, but there are a number of descriptive resources, e.g.,

[PS] Adobe Systems Incorporated, "PostScript Language Reference, third edition", 1999.

[AdobePDF] Adobe Systems Incorporated, "PDF Reference, sixth edition", 2006.

[RFC6838] Freed, N., Klensin, J. and T. Hansen, "Media Type Specifications and Registration Procedures", BCP 13, RFC 6838, DOI 10.17487/RFC6838, January 2013.

[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.

[RFC3778] Taft, E., Pravetz, J., Zilles, S. and L. Masinter, "The application/pdf Media Type", RFC 3778, DOI 10.17487/RFC3778, May 2004.

Appendix A. Changes since RFC 3778

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

- 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 updated.
- The overview of PDF capabilitiies, the history of PDF, and the descriptions of PDF subsets were updated to reflect more recent relevant history.
- The section on Fragment identifiers was updated to closely reflect the material which has been added to ISO-32000-2.
- The status of popular PDF implementations was updated.
- The Security Considerations were updated to match the current understanding of PDF vulnerabilities.
- The registration template was updated to match RFC 6838.

Authors' Addresses

Matthew Hardy
Adobe Systems Incorporated
345 Park Ave
San Jose, CA 95110
USA
EMail: mahardy@adobe.com

Larry Masinter
Adobe Systems Incorporated
345 Park Ave
San Jose, CA 95110

USA
EMail: masinter@adobe.com
URI: http://larry.masinter.net

Dejan Markovic
Adobe Systems Incorporated
345 Park Ave
San Jose, CA 95110
USA
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
UK
EMail: martin.bailey@globalgraphics.com
URI: http://www.globalgraphics.com