Network Working Group
Internet-Draft
Obsoletes: 3778 (if approved)
Intended status: Informational
Expires: January 22, 2015

M. Hardy
L. Masinter
Adobe
D. Johnson
PDF Association

July 21, 2014

The application/pdf Media Type
draft-hardy-pdf-mime-00

Abstract

PDF, the 'Portable Document Format', 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 replaces RFC 3778.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 22, 2015.

Copyright Notice

Copyright (c) 2014 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. Accessibility for PDF 5
6. PDF Implementations 5
7. Security Considerations 5
8. IANA Considerations 6
9. References 7
Authors' Addresses 8

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. Additionally, this document provides a brief history of the PDF format, describes several of the key capabilities of the format and addresses some security concerns.

PDF is used widely in the Internet community. The first version of PDF, 1.0, was published in 1993 by Adobe Systems [REF needed]. 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.

PDF represents "final form" formatted documents with a fixed layout and appearance. PDF pages may include text, images, graphics and multimedia content such as video and audio. PDF is also capable of containing higher level structures including annotations, bookmarks, file attachments, hyperlinks, logical structure and metadata. A rich JavaScript model has been defined for interacting with PDF documents.

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 metadata through XMP [XMP] metadata as well as directly via PDF structures.

In addition to the ISO 32000-1:2008 PDF standard, several ISO PDF subset standards have been defined to address specific use cases.

These standards include PDF for Archival (PDF/A), PDF for Engineering (PDF/E), PDF for Universal Accessibility (PDF/UA), PDF for Variable Data and Transactional Printing (PDF/VT) and PDF for Prepress Digital Data Exchange (PDF/X). The subset standards are fully compliant PDF files capable of being displayed in a general PDF viewer.

PDF usage is widespread enough for 'application/pdf' to be used in other IETF specifications. RFC 2346 [RFC2346] describes how to better structure PDF files for international exchange of documents where different paper sizes are used; HTTP byte range retrieval is illustrated using application/pdf (RFC 2616 [RFC2616], Section 19.2); RFC 3297 [RFC3297] illustrates how PDF can be sent to a recipient in a way that identifies the user's ability to accept the PDF using content negotiation.

2. History

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

PDF relies on the same fundamental imaging model as the PostScript [PS] page description language to render complex text, images, and graphics in a device and resolution-independent manner, bringing this feature to the screen as well as the printer. However, unlike PostScript, PDF enforces page independence, ensuring that any page in a document can render without having to render previous pages. Additionally, PDF reduces the complexity of processing content to improve performance for interactive viewing. In addition to the rendering capabilities, PDF also includes objects, such as hypertext links and annotations, that are not part of the page itself, but are useful for navigation, building collections of related documents and for reviewing and commenting on documents.

The application/pdf media type was first registered in 1993 by Paul Lindner for use by the gopher protocol and was subsequently updated in 1994 by Steve Zilles.

3. Fragment Identifiers

A set of fragment identifiers [RFC2396] and their handling are defined in Adobe Technical Note 5428 [PDFOpen]. This section summarizes that material.

A fragment identifier consists of one or more PDF-open parameters in a single URL, separated by the ampersand (&) or pound (#) character. Each parameter implies an action to be performed and the value to be

used for that action. Actions are processed and executed from left to right as they appear in the character string that makes up the fragment identifier.

The PDF-open parameters allow the specification of a particular page or named destination to open. Named destinations are similar to the "anchors" used in HTML or the IDs used in XML. Once the target is specified, the view of the page in which it occurs can be specified, either by specifying the position of a viewing rectangle and its scale or size coordinates or by specifying a view relative to the viewing window in which the chosen page is to be presented.

The list of PDF-open parameters and the action they imply is:

namedest=<name>
Open to a specified named destination (which includes a view).

page=<pagenum>
Open the specified (physical) page.

zoom=<scale>,<left>,<top>
Set the <scale> and scrolling factors. <left>, and <top> are measured from the top left corner of the page, independent of the size of the page. The pair <left> and <top> are optional but both must appear if present.

view=<keyword>,<position>
Set the view to show some specified portion of the page or its bounding box; keywords are defined by Table 8.2 of the PDF Reference, version 1.5 (NEEDS UPDATING TO ISO REF). The <position> value is required for some of the keywords and not allowed for others.

viewrect=<left>,<top>,<wd>,<ht>
As with the zoom parameter, set the scale and scrolling factors, but using an explicit width and height instead of a scale percentage.

highlight=<lt>,<rt>,<top>,<btm>
Highlight a rectangle on the chosen page where <lt>, <rt>, <top>, and <btm> are the coordinates of the sides of the rectangle measured from the top left corner of the page.

All specified actions are executed in order; later actions will override the effects of previous actions; for this reason, page actions should appear before zoom actions. Commands are not case sensitive (except for the value of a named destination).

4. Subset Standards

TODO: Describe the subset standards, their history and include references to the ISO documents.

5. Accessibility for PDF

TODO: Describe the Accessibility capabilities of PDF.

6. PDF Implementations

There are a number of widely available, independently implemented, interoperable implementations of PDF for a wide variety of platforms and systems. Since the PDF specification was published and freely available since the format was introduced in 1993, hundreds of companies and organizations, including web-browser developers, make PDF creation, viewing, and manipulation tools for many years prior to ISO standardization of PDF.

TODO: Update the above list to ensure relevance to update market conditions...

7. Security Considerations

TODO: Clean up of this section is still required...

An "application/pdf" resource contains information to be parsed and processed by the recipient's PDF system. Because PDF is both a representation of formatted documents and a container system for the resources need to reproduce or view said documents, it is possible that a PDF file has embedded resources not described in the PDF Reference.

Although it is not a defined feature of PDF, a PDF processor could extract these resources and store them on the recipients system. Furthermore, a PDF processor may accept and execute "plug-in" modules accessible to the recipient. These may also access material in the PDF file or on the recipients system. Therefore, care in establishing the source, security, and reliability of such plug-ins is recommended. Message-sending software should not make use of arbitrary plug-ins without prior agreement on their presence at the intended recipients. Message-receiving and -displaying software should make sure that any non-standard plug-ins are secure and do not present a security threat.

PDF may contain "scripts" to customize the displaying and processing of PDF files. These scripts are expressed in a version of JavaScript. They are intended for execution by the PDF processor.

User agents executing such scripts or programs must be extremely careful to insure that untrusted software is executed in a protected environment.

In general, any information stored outside of the direct control of the user -- including referenced application software or plug-ins and embedded files, scripts or other material not covered in the PDF Reference -- can be a source of insecurity, by either obvious or subtle means. For example, a script can modify the content of a document prior to its being displayed. Thus, the security of any PDF document may be dependent on the resources referenced by that document.

8. IANA Considerations

This document updates the registration of 'application/pdf', a media type registration as defined in Multipurpose Internet Mail Extensions MIME) Part Four: Registration Procedures [RFC2048]:

MIME media type name: application

MIME subtype name: pdf

Required parameters: none

Optional parameter: none

Encoding considerations: PDF files frequently contain binary data, and thus must be encoded in non-binary contexts.

Security considerations: See Section 7 of this document.

Interoperability considerations: See Section 6 of this document.

Published specification: ISO 32000-1:2008 (PDF 1.7) [ISOPDF].

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

Additional information:

Magic number(s): All PDF files start with the characters '%PDF-' using 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 "

For further information: Duff Johnson <duff.johnson@pdfa.org>, Cherie Ekholm <cheriee@microsoft.com>, ISO 32000 Project Leaders

Intended usage: COMMON

Author/Change controller: Duff Johnson <duff.johnson@pdfa.org>, Cherie Ekholm <cheriee@microsoft.com>, ISO 32000 Project Leaders

9. References

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

Also available free from Adobe Systems.

[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., [1]

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

Available at: [2]

[PDFOpen] Adobe Systems Incorporated, "PDF Open Parameters", Technical Note 5428, May 2003.

Available at: [3]

[RFC2048] Freed, N., Klensin, J., and J. Postel, "Multipurpose Internet Mail Extensions (MIME) Part Four: Registration Procedures", BCP 13, RFC 2048, November 1996.

[RFC2346] Palme, J., "Making Postscript and PDF International", RFC 2346, May 1998.

[RFC2396] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifiers (URI): Generic Syntax", RFC 2396, August 1998.

[RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.

[RFC3297] Klyne, G., Iwazaki, R., and D. Crocker, "Content Negotiation for Messaging Services based on Email", RFC 3297, July 2002.

Authors' Addresses

Matthew Hardy
Adobe
345 Park Ave
San Jose, CA 95110
USA

Email: mahardy@adobe.com

Larry Masinter
Adobe
345 Park Ave
San Jose, CA 95110
USA

Email: masinter@adobe.com
URI: http://larry.masinter.net

Duff Johnson
PDF Association
Neue Kantstrasse 14
Berlin 14057
Germany

Email: duff.johnson@pdfa.org
URI: http://www.pdfa.org