PDF

Portable Document Format (**PDF**), standardized as **ISO 32000**, is a file format developed by Adobe in 1992 to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems.[2][3] Based on the PostScript language, each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991.[4] PDF was standardized as ISO 32000 in 2008.[5] The last edition as ISO 32000-2:2020 was published in December 2020.[6]

PDF files may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and form-fields, layers, rich media (including video content), three-dimensional objects using U3D or PRC, and various other data formats. The PDF specification also provides for encryption and digital signatures, file attachments, and metadata to enable workflows requiring these features.

Portable Document Format

Adobe PDF icon

Filename extension	`.pdf`

Internet media type	`application/pdf`,[1]
	`application/x-pdf`
	`application/x-bzpdf`
	`application/x-gzpdf`

Type code	PDF [1] (including a single trailing space)
Uniform Type Identifier (UTI)	com.adobe.pdf
Magic number	%PDF
Developed by	Adobe Inc. (1991–2008) ISO (2008–)
Initial release	June 15, 1993
Latest release	2.0
Extended to	PDF/A, PDF/E, PDF/UA, PDF/VT, PDF/X
Standard	ISO 32000-2
Open format?	Yes
Website	iso.org/standard/75839.html (https://iso.org/standard/75839.html)

History

The development of PDF began in 1991 when John Warnock wrote a paper for a project then code-named Camelot, in which he proposed the creation of a simplified version of PostScript called Interchange PostScript (IPS).[7] Unlike traditional PostScript, which was tightly focused on rendering print jobs to output devices, IPS would be optimized for displaying pages to any screen and any platform.[7]

Adobe Systems made the PDF specification available free of charge in 1993. In the early years PDF was popular mainly in desktop publishing workflows, and competed with several other formats, including DjVu, Envoy, Common Ground Digital Paper, Farallon Replica and even Adobe's own PostScript format.

PDF was a proprietary format controlled by Adobe until it was released as an open standard on July 1, 2008, and published by the International Organization for Standardization as ISO 32000-1:2008,[8][9] at which time control of the specification passed to an ISO Committee of volunteer industry experts. In 2008, Adobe published a Public Patent License to ISO 32000-1 granting royalty-free rights for all patents owned by Adobe necessary to make, use, sell, and distribute PDF-compliant implementations.[10]

PDF 1.7, the sixth edition of the PDF specification that became ISO 32000-1, includes some proprietary technologies defined only by Adobe, such as Adobe XML Forms Architecture (XFA) and JavaScript extension for Acrobat, which are referenced by ISO 32000-1 as normative and indispensable for the full implementation of the ISO 32000-1 specification.[11] These proprietary technologies are not standardized, and their specification is published only on Adobe's website.[12][13][14] Many of them are not supported by popular third-party implementations of PDF.

ISO published version 2.0 of PDF, ISO 32000-2 in 2017, available for purchase, replacing the free specification provided by Adobe.[15] In December 2020, the second edition of PDF 2.0, ISO 32000-2:2020, was published, with clarifications, corrections, and critical updates to normative references[16] (ISO 32000-2 does not include any proprietary technologies as normative references).[17] In April 2023 the PDF Association made ISO 32000-2 available for download free of charge.[15]

Technical details

A PDF file is often a combination of vector graphics, text, and bitmap graphics. The basic types of content in a PDF are:

- Typeset text stored as content streams (i.e., not encoded in plain text);
- Vector graphics for illustrations and designs that consist of shapes and lines;
- Raster graphics for photographs and other types of images; and
- Other multimedia objects.

In later PDF revisions, a PDF document can also support links (inside document or web page), forms, JavaScript (initially available as a plugin for Acrobat 3.0), or any other types of embedded contents that can be handled using plug-ins.

PDF combines three technologies:

- An equivalent subset of the PostScript page description programming language but in declarative form, for generating the layout and graphics.
- A font-embedding/replacement system to allow fonts to travel with the documents.
- A structured storage system to bundle these elements and any associated content into a single file, with data compression where appropriate.

PostScript language

PostScript is a page description language run in an interpreter to generate an image.[7] It can handle graphics and has standard features of programming languages such as branching and looping.[7] PDF is a subset of PostScript, simplified to remove such control flow features, while graphics commands remain.[7]