

Portable Document Format



The Portable Document Format (PDF) is a file format developed by Adobe in the 1990s to present documents, including text formatting and images, in a manner independent of application software, hardware, and operating systems. 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 was standardized as an open format, ISO 32000, in 2008, and no longer requires any royalties for its implementation.

Today, 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) and 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.

History and standardization

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 a variety of formats such as 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, 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 that are necessary to make, use, sell, and distribute PDF compliant implementations.

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. These proprietary technologies are not standardized and their specification is published only on Adobe's website. Many of them are also not supported by popular third-party implementations of PDF.

On July 28, 2017, ISO 32000-2:2017 (PDF 2.0) was published. ISO 32000-2 does not include any proprietary technologies as normative references.

Technical foundations

The PDF combines three technologies:

- A subset of the PostScript page description programming language, 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

PostScript is a page description language run in an interpreter to generate an image, a process requiring many resources. It can handle graphics and standard features of programming languages such as if and loop commands. PDF is largely based on PostScript but simplified to remove flow control features like these, while graphics commands such as `lineto` remain.

Often, the PostScript-like PDF code is generated from a source PostScript file. The graphics commands that are output by the PostScript code are collected and tokenized. Any files, graphics, or fonts to which the document refers also are collected. Then, everything is compressed to a single file. Therefore, the entire PostScript world (fonts, layout, measurements) remains intact.

As a document format, PDF has several advantages over PostScript:

- PDF contains tokenized and interpreted results of the PostScript source code, for direct correspondence between changes to items in the PDF page description and changes to the resulting page appearance.
- PDF (from version 1.4) supports graphic transparency; PostScript does not.
- PostScript is an interpreted programming language with an implicit global state, so instructions accompanying the description of one page can affect the appearance of any following page. Therefore, all preceding pages in a PostScript document must be processed to determine the correct appearance of a given page, whereas each page in a PDF document is unaffected by the others. As a result, PDF viewers allow the user to quickly jump to the final pages of a long document, whereas a PostScript viewer needs to process all pages sequentially before being able to display the destination page (unless the optional PostScript Document Structuring Conventions have been carefully complied with).

Content

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

- 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 image
- Multimedia objects in the document

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

PDF 1.6 supports interactive 3D documents embedded in the PDF - 3D drawings can be embedded using U3D or PRC and various other data formats.

Two PDF files that look similar on a computer screen may be of very different sizes. For example, a high resolution raster image takes more space than a low resolution one. Typically higher resolution is needed for printing documents than for displaying them on screen. Other things that may increase the size of a file is embedding full fonts, especially for Asiatic scripts, and storing text as graphics.

File formats and Adobe Acrobat versions

The PDF file format has changed several times, and continues to evolve, along with the release of new versions of Adobe Acrobat. There have been nine versions of PDF and the corresponding version of the software:

Date	PDF Version	Acrobat Version
1993	PDF 1.0	Acrobat 1
1994	PDF 1.1	Acrobat 2
1996	PDF 1.2	Acrobat 3
1999	PDF 1.3	Acrobat 4
2001	PDF 1.4	Acrobat 5
2003	PDF 1.5	Acrobat 6
2005	PDF 1.6	Acrobat 7
2006	PDF 1.7	Acrobat 8
2008	PDF 1.7	Acrobat 9
2009	PDF 1.7	Acrobat 9.1
2010	PDF 1.7	Acrobat X
2012	PDF 1.7	Acrobat XI
2017	PDF 2.0	Acrobat DC

Text is available under the Creative Commons Attribution Share Alike License.