Technical overview [edit]

File structure [edit]

A PDF file is a 7-bit ASCII file, except for certain elements that may have binary content. A PDF file starts with a header containing the magic number and the version of the format such as `%PDF-1.7`. The format is a subset of a COS ("Carousel" Object Structure) format.[15] A COS tree file consists primarily of *objects*, of which there are eight types:[16]

- Boolean values, representing *true* or *false*
- Numbers
- Strings, enclosed within parentheses (`(...)`), may contain 8-bit characters.
- Names, starting with a forward slash (`/`)
- Arrays, ordered collections of objects enclosed within square brackets (`[...]`)
- Dictionaries, collections of objects indexed by Names enclosed within double pointy brackets (`<<...>>`)
- Streams, usually containing large amounts of data, which can be compressed and binary
- The null object

Furthermore, there may be comments, introduced with the percent sign (`%`). Comments may contain 8-bit characters.

Objects may be either *direct* (embedded in another object) or *indirect*. Indirect objects are numbered with an *object number* and a *generation number* and defined between the `obj` and `endobj` keywords. An index table, also called the cross-reference table and marked with the `xref` keyword, follows the main body and gives the byte offset of each indirect object from the start of the file.[17] This design allows for efficient random access to the objects

CLOSE ×

Portable Document Format (**PDF**) (redundantly: PDF format) i
e format developed by Adobe in the 1990s to present documents
ding text formatting and images, in a manner independent of
ication software, hardware, and operating systems.[2][3] Based o
PostScript language, each PDF file encapsulates a complete
cription of a fixed-layout flat document, including the text, fonts,
or graphics, raster images and other information needed to displ
DF was standardized as ISO 32000 in 2008, and no longer
ires any royalties for its implementation.[4]

ay, PDF files may contain a variety of content besides flat text an
hics including logical structuring elements, interactive elements

Contents [hide]

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 [edit]

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

PDF

From Wikipedia, the free encyclopedia

For other uses, see PDF (disambiguation).

Contents [hide]

The **Portable Document Format** (**PDF**) (redundantly: PDF format) 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.[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 was standardized as ISO 32000 in 2008, and no longer requires any royalties for its implementation.[4]

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.[*citation needed*] 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` [note 1]
Internet media type	`application/pdf` ,[1] `application/x-pdf` `application/x-bzpdf` `application/x-gzpdf`
Type code	'PDF '[1] (including a single space)
Uniform Type Identifier (UTI)	com.adobe.pdf
Magic number	`%PDF`
Developed by	ISO Originally Adobe
Initial release	15 June 1993; 26 years ago
Latest release	2.0
Extended to	PDF/A, PDF/E, PDF/UA, PDF/VT, PDF/X
Standard	ISO 32000-2
Open format?	Yes
Website	www.iso.org/standard/63534.html