

**Portable Document Format (PDF)** is a file format developed by Adobe Systems for representing documents in a manner that is independent of the original application software, hardware, and operating system used to create those documents. A PDF file can describe documents containing any combination of text, graphics, and images in a device independent and resolution independent format. These documents can be one page or thousands of pages, very simple or extremely complex with a rich use of fonts, graphics, colour, and images. PDF is an open standard, and anyone may write applications that can read or write PDFs royalty-free.

## History

When PDF first came out, in the early 1990s, it was slow to catch on. At the time, not only did the only PDF creation tools of the time (Acrobat) cost money, but so did the software to view and print PDF files. Additionally, there were competing formats such as Envoy, Common Ground Digital Paper, DjVu and even Adobe's own PostScript file format (.ps). Adobe started distributing the Acrobat Reader program at no cost, and continued to support PDF through its slow multi-year ramp-up. Competing formats eventually died out, and PDF became a well-accepted standard.

## Technology

PDF is primarily the combination of three technologies:

- a cut-down form of PostScript for generating the layout and graphics,
- a font-embedding/replacement system to allow fonts to travel with the documents, and
- a structured storage system to bundle these elements into a single file, with data compression where appropriate.

## Comparison between PDF and HTML

PDF and HTML are not equivalent technologies, but are both commonly found on the Web.

HTML is a method for describing the content of a webpage in a manner that is open to interpretation by the browser which renders it on the user's screen. This permits content to be rendered to suit the viewer rather than the content provider, and also means that an HTML file will not necessarily look exactly the same in different browsers. PDF on the other hand is strictly concerned with describing the content of a document such that the original layout and typesetting are fully preserved.

Since many content providers do not like the fluid nature of HTML rendering, PDF has become widespread to force a particular layout. With HTML the same can be achieved by using a raster graphics (or recently, SVG, a vector graphics standard) image to present text, but then the text can not be copied as such, nor can a subtext be searched within it.

Use of images also leads to larger file sizes. (Sometimes the same is done in a PDF file, and the same disadvantages apply.)

A typical example of the differences this leads to for the viewer is with zooming:

- Enlarging a PDF document magnifies the text but preserves the original layout and spacing; a practical limit on zooming follows from the requirement to keep a text column within the width of the screen (otherwise horizontal scrolling would be needed during and after reading each line, which would be very cumbersome).
- With HTML, a larger font size is used and lines re-wrap accordingly to fit the browser window. Note, though, that tagged PDFs can be reflowed quite well, keeping columns of text within the viewable area.

## PostScript

PostScript is a computer language – or more precisely a page description language – that is run in an interpreter to generate an image. This process requires a fair amount of resources.

PDF is a subset of those PostScript language elements that define the graphics, and only requires a very simple interpreter. For instance, flow control commands like if and loop are removed, while graphics commands such as lineto remain.

That means that the process of turning PDF back into a graphic is a matter of simply reading the description, rather than running a program in the PostScript interpreter. However the entire PostScript world in terms of fonts, layout and measurement remains intact.

Often the PostScript-like PDF code is generated from a source PostScript file. The graphics commands that the PS code outputs are collected and tokenized, any files, graphics or fonts the document references are also collected, and then everything is compressed into a single file.

As a document format, PDF has several advantages over PostScript. One is that a document resides in a single file, whereas the same document in PostScript may span multiple files (graphics, etc.) and probably occupies more space. In addition, PDF contains already-interpreted results of the PostScript source code, so it is less computation-intensive and faster to open, and there is a more direct correspondence between changes to items in the PDF page description and changes to the resulting appearance of the page. Finally, if displayed with Adobe Reader, a font-substitution strategy ensures the document will be readable even if the end-user does not have the "proper" fonts installed.