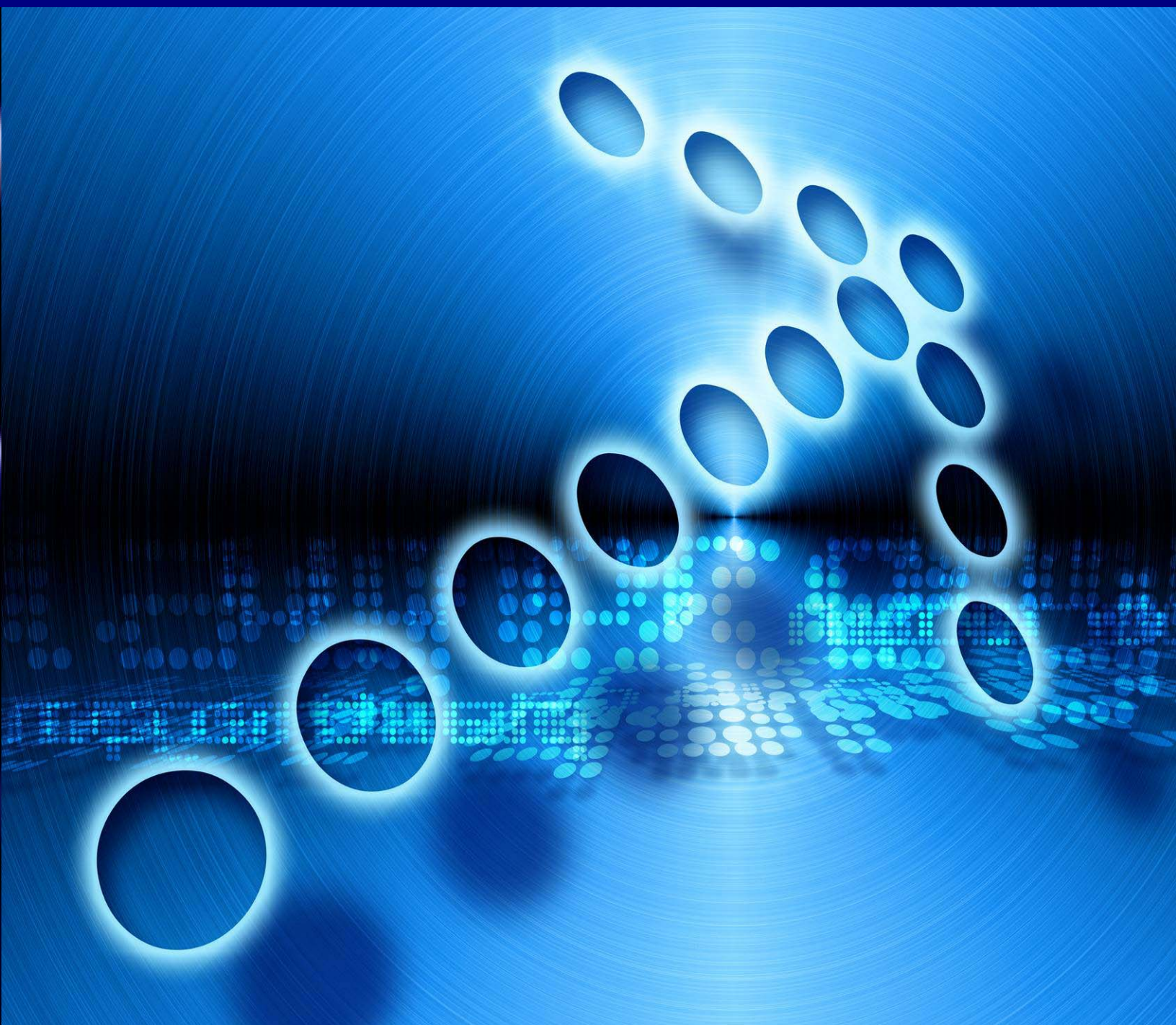


Glossary of PDF Terminology

Aquaforest Limited

www.aquaforest.com



Background

The PDF format was created by Adobe in 1993 for the purpose of portable document exchange between systems and applications. The PDF specification has been updated and published and Adobe over the years but PDF is now an open standard that was officially published on July 1, 2008 by the ISO as ISO 32000-1:2008. PDF documents involve a wide range of technical areas which this Glossary seeks to explain.

A

Accessibility



To address the needs of visually impaired and other users who must employ assistive technology in order to read, the U.S. Congress passed Section 508 in 1998, an amendment to the Rehabilitation Act.

Section 508 requires U.S. Federal government agencies to procure accessible software and to produce accessible electronic documents.

PDF includes several facilities in support of Section 508 and accessibility of documents to users with disabilities. In particular, many computer users with visual impairments use screen readers to read documents aloud. To enable proper vocalization, either through a screen reader or by some more direct invocation of a text-to-speech engine, PDF supports the following features:

Specifying the natural language used for text in a PDF document—for example, as English or Spanish, or used to hide or reveal optional content.

Providing textual descriptions for images or other items that do not translate naturally into text or replacement text for content that does translate into text but is represented in a nonstandard way (such as with a ligature or illuminated character”)

Specifying the expansion of abbreviations or acronyms

Further Information : <http://www.section508.gov/>

Adobe Acrobat



Adobe Acrobat supports creating, modifying, indexing, searching, displaying, and manipulating PDF (Portable Document Format) files.

Acrobat includes Acrobat Distiller for creating PDF files from PostScript files created from desktop applications.

Adobe Reader



Adobe Reader is a stand-alone program or Web browser plug-in from Adobe that lets you view a PDF (Portable Document Format) file in its original format and appearance.

Adobe Reader is available free and can be downloaded from Adobe Systems Inc. at <http://www.adobe.com>.

Other alternative PDF Readers are available – see the entry on PDF Readers.

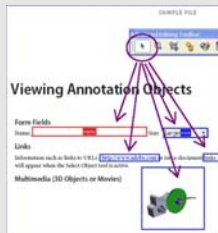
Glossary of PDF Terminology

Adobe Systems



Adobe Systems created the PDF format in 1993 and continue to create new versions of the specification.

Annotations

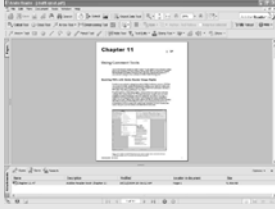


PDF includes support for a wide variety of standard annotation types. A full list is provided below along with the PDF version they were first introduced in.

- Text annotation
- Link annotation
- Free text annotation (PDF 1.3)
- Line annotation (PDF 1.3)
- Square annotation (PDF 1.3)
- Circle annotation (PDF 1.3)
- Polygon annotation (PDF 1.5)
- Polyline annotation (PDF 1.5)
- Highlight annotation (PDF 1.3)
- Underline annotation (PDF 1.3)
- Squiggly-underline annotation (PDF 1.3)
- Strikeout annotation (PDF 1.3)
- Rubber stamp annotation (PDF 1.3)
- Caret annotation (PDF 1.5)
- Ink annotation (PDF 1.3)
- Pop-up annotation (PDF 1.3)
- File attachment annotation (PDF 1.3)
- Sound annotation (PDF 1.2)
- Movie annotation (PDF 1.2)
- Widget annotation (PDF 1.2)
- Screen annotation (PDF 1.5)
- Printer's mark annotation (PDF 1.4)
- Trap network annotation (PDF 1.3)
- Watermark annotation (PDF 1.6)
- 3D annotation (PDF 1.6)
- Redact annotation (PDF 1.7)

Many of the annotation types may be displayed in either the open or the closed state. When closed, they appear on the page in some distinctive form, such as an icon, a box, or a rubber stamp, depending on the specific annotation type. When the user activates the annotation by clicking it, it exhibits its associated object, such as by opening a pop-up window displaying a text note or by playing a sound or a movie

Attachments



A file attachment “annotation” (introduced in PDF 1.3) contains a reference to a file, which typically shall be embedded in the PDF file. For example, a table of data might use a file attachment annotation to link to a spreadsheet file based on that data. Activating the annotation extracts the embedded file and gives the user an opportunity to view it or store it in the file system.

B

Bates Numbering



A Bates stamp uniquely identifies document pages with sequential numbers or numerical markings. The system is primarily used in the legal field, but also in business and medicine. These unique stamps set each document page apart, allowing them to be clearly identified by all parties involved in litigation.

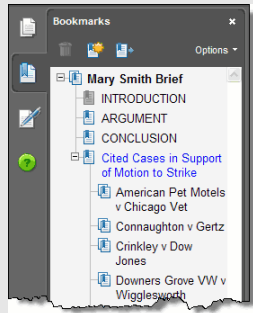
Bates numbers may include a prefix and a suffix, using any alphanumeric characters chosen by the user. For example : TIMN0053288.

Acrobat 8.0 allows users to apply Bates numbers to one document or to a whole set of documents, and they may apply more than one Bates Numbering sequence to a single or multiple documents. Document sets may also be given a unique set of numbers or prefixes which might include text such as a set number, case number, or firm name.

Although most Bates Stamps are now digitally created, the phrases dates back to the days when the identifying marks were manually stamped onto documents using stamps produced by the Bates Stamp Machine Company.

http://www.adobe.com/devnet/acrobat/pdfs/bates_numbering.pdf

Bookmark



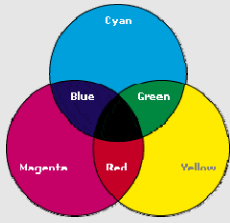
In PDF terminology, “bookmarks” are generally termed “document outlines”.

A PDF document may contain a document outline that the conforming reader may display on the screen, allowing the user to navigate interactively from one part of the document to another.

The outline consists of a tree-structured hierarchy of outline items which serve as a visual table of contents to display the document’s structure to the user.

C

CMYK



The CMYK color model (process color, four color) is a subtractive color model, used in color printing, and is also used to describe the printing process itself. CMYK refers to the four inks used in some color printing: cyan, magenta, yellow, and key black.

If you are preparing documents solely for the web, then for the same number of image samples RGB is negligibly faster with possibly smaller file size. If you are preparing documents for print, then CMYK is better since it preserves spot colors, and separating RGB may be unsatisfying or not possible.

If you are preparing documents for both web and print CMYK is the better choice. CMYK documents coexist quite nicely in the RGB world, but the converse does not always hold true. And there are ways to convert high-resolution PDF images into low-res for the web.

Color Profiles



The PDF format supports objects specified in a number of device dependent, device independent, and special color spaces.

Device color spaces are those used by the creation device. For images, this would often be the color space of the data from the scanner. However, it could also be the color space applied in Photoshop or Quark, or the result of color conversion by the printer driver, whichever made the last modification.

Adobe, Apple and a few other companies proposed an industry standard for device independent color. Initial implementations included the CIE-based device independent color in PostScript Level 2 and Apple's ColorSync. The group, and the effort, eventually became the International Color Consortium (ICC).

The premise of ICC color is that you can use a profile to map source colors into a mathematically-rigorous absolute color space and have other profiles to map from this absolute color space to destination colors.

ICC color also recognizes that the same colors can be applied different ways. This has led to the notion of "rendering intents" for different objects. Four standard rendering intents are defined but others are possible.

<http://www.color.org>

D

Digital Signatures



A digital signature (introduced in PDF 1.3) may be used to authenticate the identity of a user and the document's contents. It stores information about the signer and the state of the document when it was signed.

The signature may be purely mathematical, such as a public/private-key encrypted document digest, or it may be a biometric form of identification, such as a handwritten signature, fingerprint, or retinal scan.

<http://www.adobe.com/security/digsig.html>

http://learn.adobe.com/wiki/download/attachments/52658564/Acrobat_Digital_Signatures_in_PDF.pdf?version=1

Distiller



Distiller is part of the Adobe Acrobat family of products. Distiller is the main engine for turning PostScript files into PDF documents. Although there are other programs for generating PDF files, Distiller is the primary one. Some page layout programs can generate PDF files from within the program, but are often simply acting as a frontend to Distiller,.

Adobe Acrobat Distiller works with postscript files, converting them to PDF and allowing higher quality so that it reproduces graphics and complex documents more accurately.

E

Encodings

Characters:	A	é	Ń
ASCII	65 (41)	n/a	n/a
ISO-8859-1 (Latin 1)	65 (41)	233 (89)	n/a
Unicode (UCS)	65 (41)	233 (89)	1488 (9500)
7-bit ASCII	41	n/a	n/a
ISO-8859-1	41	E9	n/a
UTF-8	41	C3 A9	D7 90
UTF-16	00 41	00 E9	05 D0
UTF-32	00 00 00 41	00 00 E9	00 00 D0

The PDF specification defines support for a number of different encodings including Adobe Standard Encoding, Windows, Macintosh and other encodings. An encoding is a set of characters in a given order. The associated index number or numerical code assigned to each character is used as a means for accessing that character.

The advantage of using the Adobe standard character set and encoding is that the font developer can use a single character set and encoding for multiple operating systems (e.g. Macintosh, Windows, and UNIX). This works because ATM software (where present) and printer drivers know to re-encode fonts with the Adobe Standard Encoding to a variety of other encodings.

For more detailed information see :

http://www.adobe.com/devnet/opentype/archives/char_sets.html

Encryption



A PDF document can be encrypted (PDF 1.1) to protect its contents from unauthorized access. Encryption can be applied to the majority of strings and streams in the document's PDF file.

F

fdf



File conforming to the Forms Data Format containing form data or annotations that may be imported into a PDF file

fdf can be used when submitting form data to a server, receiving the response, and incorporating it into the interactive form. It can also be used to export form data to stand-alone files that can be stored, transmitted electronically, and imported back into the corresponding PDF interactive form. In addition, beginning in PDF 1.3, fdf can be used to define a container for annotations that are separate from the PDF document to which they apply.

fdf Toolkit :

<http://www.adobe.com/devnet/acrobat/fdf toolkit.html>

File Structure



The general structure of a PDF file is composed of the following code components: header, body, cross-reference (xref) table, and trailer.

The header contains just one line that identifies the version of PDF. Example: %PDF-1.6

The trailer contains pointers to the xref table and to key objects contained in the trailer dictionary. It ends with %%EOF to identify end of file.

The xref table contains pointers to all the objects included in the PDF file. It identifies how many objects are in the table, where the object begins (the offset), and its length in bytes.

The body contains all the object information — fonts, images, words, bookmarks, form fields, and so on.

When you perform a Save operation on a PDF file, the new, incremental information is appended to the original structure (see figure 2); that is, a new body, xref table, and trailer are added to the original PDF file.

For full details of the PDF file structure internals refer to the PDF Specification document :

http://www.adobe.com/devnet/pdf/pdf_reference.html

F (continued)

Filters



The PDF Specification supports a number of filters to enable encoding of streams. For example, this provides support for compressing the text portions of PDF files and compressed image formats such as JPEG. The current list of support filters is shown below :

- ASCIIHexDecode
- ASCII85Decode
- LZWDecode
- FlateDecode (PDF 1.2)
- RunLengthDecode
- CCITTFaxDecode
- JBIG2Decode (PDF 1.4)
- DCTDecode (JPEG)
- JPXDecode (JPEG 2000 - PDF 1.5)
- Crypt (PDF 1.5)

Fonts

Arial
Arial Black
Comic Sans MS
Courier New
Georgia
Impact
Times New Roman
Trebuchet MS
Verdana

A font defines glyphs for a particular character set. A character is an abstract symbol, whereas a glyph is a specific graphical rendering of a character. For example, the glyphs **A**, **A**, and **A** are renderings of the abstract "A" character. The Helvetica and Times fonts define glyphs for a set of standard Latin characters.

In PDF, the term font refers to a font dictionary, a PDF object that identifies the font program and contains additional information about it. There are several different font types, identified by the Subtype entry of the font dictionary. For most font types, the font program shall be defined in a separate font file, which may be either embedded in a PDF stream object or obtained from an external source. The font program contains glyph descriptions that generate glyphs. When creating a PDF, the creator will have a choice on how font embedding is managed for the document.

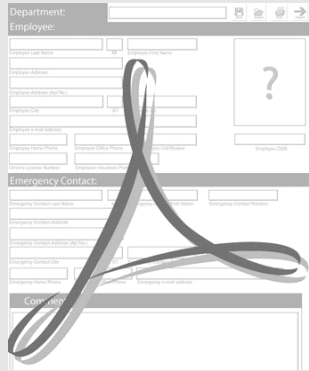
Full Font Embedding will result in a larger file size but the recipient doesn't need the same font to view or edit the file

Subset Font Embedding will just embed the subset of the font required so the recipient does not need the same font to view the document, but may do to edit.

No Font Embedding will result in the smallest font size but the recipient needs to have same fonts installed.

F (continued)

Forms

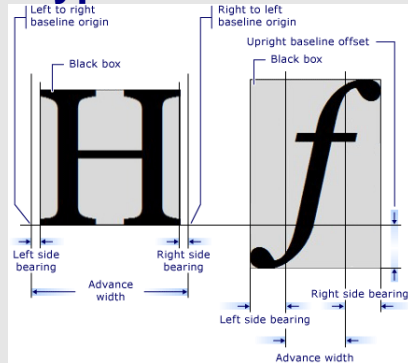


An interactive form (PDF 1.2)—sometimes referred to as an AcroForm—is a collection of fields for gathering information interactively from the user.

A PDF document may contain any number of fields appearing on any combination of pages, all of which make up a single, global interactive form spanning the entire document. Arbitrary subsets of these fields can be imported or exported from the document.

G

Glyph



A glyph is a specific graphical rendering of a character.

EXAMPLE 1 The glyphs **A**, *A*, and **A** are renderings of the abstract “A” character.

I

Image-Only PDF



A PDF that consists solely of scanned images – like an electronic photocopy. Running the PDF file through a PDF OCR process can create a searchable PDF.

See http://www.aquaforest.com/files/searchable_pdf.pdf

ISO 32000

Although Adobe had a history of making the PDF specification publicly available, it wasn't until early 2007 that Adobe Systems Incorporated announced its intention to release the full Portable Document Format (PDF) 1.7 specification to ANSI and AIIM for the purpose of publication by the International Organization for Standardization (ISO).

ISO 32000-1:2008 specifies a digital form for representing electronic documents to enable users to exchange and view electronic documents independent of the environment in which they were created or the environment in which they are viewed or printed. It is intended for the developer of software that creates PDF files (conforming writers), software that reads existing PDF files and interprets their contents for display and interaction (conforming readers) and PDF products that read and/or write PDF files for a variety of other purposes (conforming products).

http://www.adobe.com/devnet/pdf/pdf_reference.html

J

JavaScript <script>

PDF documents can be enhanced with JavaScript for Acrobat which is an extension of core JavaScript, version 1.5 of ISO-16262.

JavaScript was originally created to offload web page processing from a server onto a client in web-based applications. Acrobat extends the core language by adding new objects and their accompanying methods and properties, to the JavaScript language. These Acrobat-specific objects enable a developer to manage document security, communicate with a database, handle file attachments, manipulate a PDF file so that it behaves as an interactive, web-enabled form, and so on. Because the Acrobat-specific objects are added on top of core JavaScript, developers still have access to its standard classes, including Math, String, Date, and RegExp.

PDF documents have great versatility since they can be displayed both within the Acrobat software as well as a web browser. Therefore, it is important to be aware of the differences between JavaScript used in a PDF file and JavaScript used in a web page:

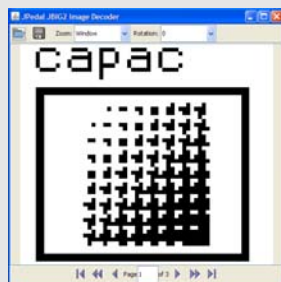
- JavaScript in a PDF file does not have access to objects within an HTML page. Similarly, JavaScript in a web page cannot access objects within a PDF file.

- In HTML, JavaScript is able to manipulate such objects as Window. JavaScript for Acrobat cannot access this particular object but it can manipulate PDF-specific objects.

JavaScript for PDF:

<http://www.adobe.com/devnet/acrobat/javascript.html>

JBIG2

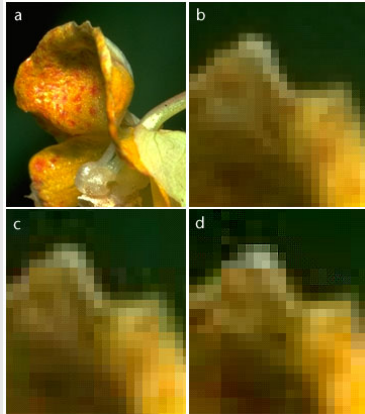


JBIG2 is an image compression standard for bi-level images, developed by the Joint Bi-level Image Experts Group. It is suitable for both lossless and lossy compression and produces smaller files than the CCITT Group 3 and Group 4 compression schemes also supported by PDF.

<http://www.jpeg.org/jbig/jbigpt2.html>

J (Continued)

JPEG



A commonly used method of lossy compression for photographic images. The degree of compression can be adjusted, allowing a selectable tradeoff between storage size and image quality. JPEG typically achieves 10:1 compression with little perceptible loss in image quality.

<http://www.jpeg.org/jpeg/index.html>

JPEG2000



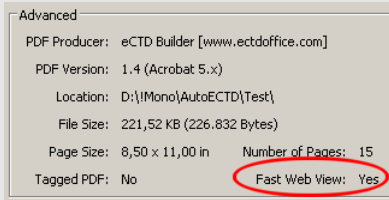
JPEG 2000 is a wavelet-based image compression standard and coding system. As well as some enhanced compression, the main advantage offered by JPEG 2000 is the significant flexibility of the codestream. The codestream obtained after compression of an image with JPEG 2000 is scalable in nature, meaning that it can be decoded in a number of ways; for instance, by truncating the codestream at any point, one may obtain a representation of the image at a lower resolution, or signal-to-noise ratio. Another difference, in comparison with JPEG, is in terms of visual artifacts: JPEG 2000 produces ringing artifacts, manifested as blur and rings near edges in the image, while JPEG produces ringing artifacts and 'blocking' artifacts, due to its 8×8 blocks.

JPEG 2000 has been published as an ISO standard, ISO/IEC 15444. As of 2009[update], JPEG 2000 is not widely supported in web browsers, and hence is not generally used on the World Wide Web.

<http://www.jpeg.org/jpeg2000/>

L

Linearization



Linearization (“Fast Web View”) of PDF is an optional feature available introduced in PDF 1.2 that enables efficient incremental access of the file in a network environment.

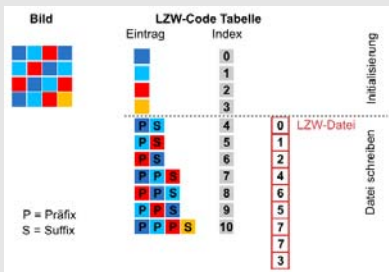
The primary goal for a linearized PDF file is to achieve the following behaviour for documents of arbitrary size and so that the total number of pages in the document should have little or no effect on the user-perceived performance of viewing any particular page:

When a document is opened, display the first page as quickly as possible. The first page to be viewed may be an arbitrary page of the document, not necessarily page 0 (though opening at page 0 is most common).

When the user requests another page of an open document (for example, by going to the next page or by following a link to an arbitrary page), display that page as quickly as possible.

When data for a page is delivered over a slow channel, display the page incrementally as it arrives. To the extent possible, display the most useful data first.

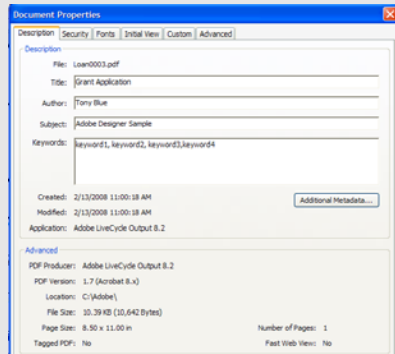
LZW



GIF images are compressed using the Lempel-Ziv-Welch (LZW) lossless data compression technique to reduce the file size without degrading the visual quality. This compression technique was patented in 1985. Controversy over the licensing agreement between the patent holder, Unisys, and CompuServe in 1994 spurred the development of the Portable Network Graphics (PNG) standard; since then all the relevant patents have expired.

M

Metadata



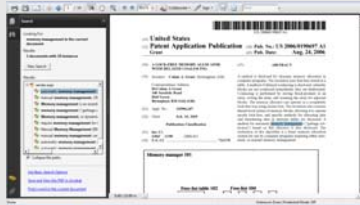
PDF metadata can be added in the form of document properties. Typically, these document properties include things like search keywords, title, author, and subject but can be extended to include custom fields.

Adding this type of information about a PDF document allows for faster, more efficient archival and document retrieval. This metadata is stored in the PDF document information dictionary.

More sophisticated metadata structures can be managed using XMP. For more details see the XMP entry.

O

OCR PDF

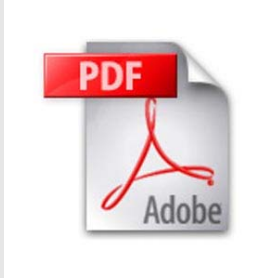


This describes the processing of taking an Image-Only PDF, running an OCR process on the images in the document. Typically an updated version of the PDF is produced that includes a transparent text layer thus making the PDF fully searchable.

See http://www.aquaforest.com/files/searchable_pdf.pdf

P

PDF Specification



Adobe has historically made the PDF specification available to the public at this location :

http://www.adobe.com/devnet/pdf/pdf_reference.html

Since ISO 32000 is equivalent to Adobe's PDF 1.7, Adobe is not producing a PDF 1.8 Reference. However, Adobe is publishing a document specifying what extended features for PDF, beyond ISO 32000-1 (PDF 1.7), are supported in its newly released products. This makes use of the extensibility features of PDF as documented in ISO 32000 in Annex E.

PDF Viewer



A PDF viewer is a software that renders a view of a PDF document. There are several PDF viewer applications other than Adobe Reader such as :

Foxit Reader

http://www.foxitsoftware.com/pdf/rd_intro.php

PDF Escape

<http://www.pdfescape.com/>

PDF Hammer

<http://www.pdfhammer.com/>

CoolPDF

<http://www.softpedia.com/get/Office-tools/PDF/Cool-PDF-Reader.shtml>

Sumatra

<http://blog.kowalczyk.info/software/sumatrapdf/index.html>

PDF-XChange

<http://www.docu-track.com/product/pdf-xchange-viewer>

Preview (Mac)

<http://support.apple.com/kb/HT2506>

P (continued)

PDF/A

PDF/A - **ISO 19005** - is a file format for the long-term archiving of electronic documents. It is based on the PDF Reference Version 1.4 from Adobe Systems Inc. (implemented in Adobe Acrobat 5 and latest versions) and is defined by ISO 19005-1:2005, an ISO Standard that was published on October 1, 2005:

Document Management - Electronic document file format for long term preservation - Part 1: Use of PDF 1.4 (PDF/A-1)

A new version of PDF/A based on PDF 1.7 - ISO 32000-1 is currently under development (ISO/DIS 19005-2).[1]

PDF/A is in fact a subset of PDF, obtained by leaving out PDF features not suited to long-term archiving.

<http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>

A new version of PDF/A based on PDF 1.7 - ISO 32000-1 is currently under development (ISO/DIS 19005-2).

PDF/E

The PDF/E (PDF/Engineering) format is a PDF subset whose purpose is to handle engineering and design documents. The PDF/E format has widespread use in the creation and exchange of various engineering and technical documents

See the entry for ISO 24517 for more details.

P (continued)

PDF/X

PDF/X - **ISO 15930** - is a collection of standards defining a number of conformance levels, all of them targeted at ensuring predictable and consistent printing in a professional print environment.

PDF/X is an umbrella term for several ISO standards that define a subset of the PDF standard eliminates many of the color, font, and trapping variables that lead to printing problems.

The purpose of PDF/X is to facilitate graphics exchange, and it therefore has a series of printing related requirements, which do not apply to standard PDF files.

For example, in PDF/X-1a all fonts need to be embedded and all images need to be CMYK or spot colors.

PDF/X-2 and PDF/X-3 accept calibrated RGB and CIELAB colors, while retaining most of the other restrictions of PDF/X-1a.

<http://pdf.editme.com/PDFX>

PDF/UA

PDF/Universal Accessibility - ISO 14289

Document management applications -- Electronic document file format enhancement for accessibility (PDF/UA). The mission of PDF/UA is to develop technical and other standards for the authoring, remediation and validation of PDF content to ensure accessibility for people that use assistive technology such as screen readers for users who are blind. This standard is currently under development.

<http://pdf.editme.com/pdfua>

P (continued)

Portfolios



A PDF Portfolio contains multiple files assembled into an integrated PDF unit. The files in a PDF Portfolio can be in different formats and created in different applications. For example, suppose you have a project that includes text documents, email messages, spreadsheets, CAD drawings, and PowerPoint presentations. You could combine all of these documents into a PDF Portfolio. The original files retain their individual identities but are assembled into one PDF Portfolio file. Users can open, read, edit, and format each component file independently of the other component files in the PDF Portfolio.

PostScript



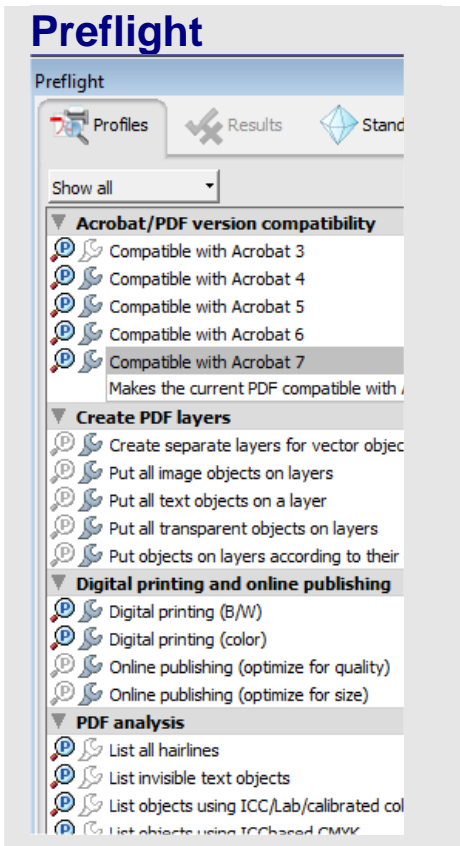
PostScript (PS) is best known for its use as a page description language in the electronic and desktop publishing areas.

An EPS file is a PostScript program, saved as a single file that includes a low-resolution preview "encapsulated" inside of it, allowing some programs to display a preview on the screen

Adobe PDF has the same imaging capabilities as PostScript because they share the underlying architecture of Adobe Imaging Model.

P (continued)

Preflight

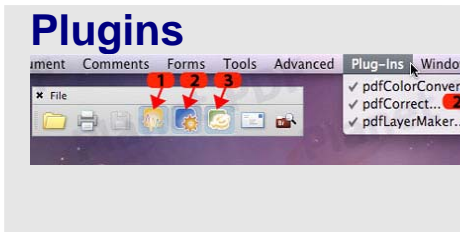


Preflighting means analyzing a file for suitability in printing or for other specific purposes such as PDF/A-1b compliance.

A preflight assessment might examine a file for the proper color mode of images, whether images are compressed, whether fonts are accessible either embedded or accessible to the operating system, or any number of other conditions that might interfere with successfully printing a job.

The tools used to preflight files might be stand-alone applications or features built into programs used for printing to commercial printing equipment. Preflighting is built into Adobe Acrobat from version 6 onwards.

Plugins



Adobe Acrobat has a Plugin API which allows third party developers to create plugins such as stamping tools or tools to provide additional support for CAD drawings

<http://www.adobe.com/devnet/acrobat/?view=overview>

R

Raster (Bitmap) Images



In computer graphics, a raster graphics image or bitmap is a data structure representing a generally rectangular grid of pixels, or points of color, viewable via a monitor, paper, or other display medium. Raster images are stored in image files with varying formats (see Comparison of graphics file formats).

Raster graphics are resolution dependent. They cannot scale up to an arbitrary resolution without loss of apparent quality. This property contrasts with the capabilities of vector graphics, which easily scale up to the quality of the device rendering them. Raster graphics deal more practically than vector graphics with photographs and photo-realistic images, while vector graphics often serve better for typesetting or for graphic design.

Rasterization

Rasterization or Rasterisation is the task of taking an image described in a vector graphics format (shapes) and converting it into a raster image (pixels or dots) for output on a video display or printer, or for storage in a bitmap file format.

It may be appropriate to perform this on a PDF file where the file has complex Vector graphics or there is a need to produce a rendering of a PDF file as a bitmap image.

Reader Extensions

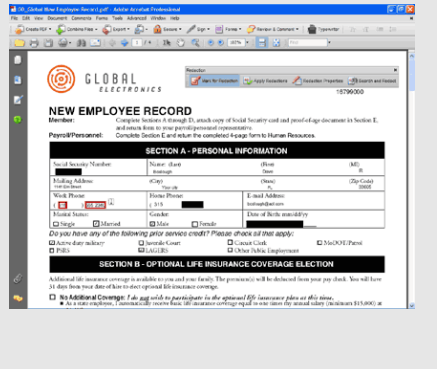
The Adobe LiveCycle Reader Extensions is software that users can use to enable some features, in Acrobat Reader (now Adobe Reader) 5.1 and later on a per-file basis. These are features otherwise found in the full licensed product Adobe Acrobat.

For example, Adobe Reader cannot normally save filled in forms or apply digital signatures. If LiveCycle Reader Extensions is purchased with suitable options, it can prepare files that Reader can save or sign.

<http://www.adobe.com/products/livecycle/readerextensions/>

R (continued)

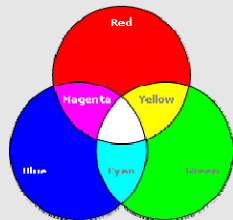
Redaction



Redaction refers to methods for permanently removing content and replacing it with a space holder of some type, such as a colored box, default text or code, or blank space.

Adobe Acrobat 8 Professional includes a set of redaction tools, and third party tools are also available.

RGB

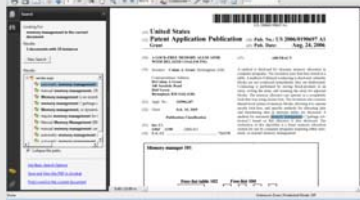


RGB is an additive color model widely use for displaying imaging on screens and can be contrasted with CMYK for printing.

By mixing Red, Green and Blue, a large percentage of the visible color spectrum can be represented.

S

Searchable PDF



PDF Searchable Image is a PDF Image Only document with the addition of a text layer beneath the image. This approach retains the look of the original page while enabling text searchability.

A document created in PDF Searchable Image offers the best of both worlds—an exact replica of the original document that is also fully searchable. PDF Searchable Image files contain two layers: a bitmapped (image) layer and a hidden text layer. The bitmapped layer maintains the visual representation of the original document.

See http://www.aquaforest.com/files/searchable_pdf.pdf

Security

The PDF document format supports a range of security options. Document can be encrypted and require a password to read the file and another password to edit the file. Printing and copying access can be limited as can a range of other capabilities.

http://help.adobe.com/en_US/Acrobat/9.0/Standard/WS_D012A4E1-51D1-4bcd-BA9F-EF03C6F20BB6.html

<http://www.adobe.com/devnet/acrobat/security.html>

sRGB

sRGB is a RGB color space proposed by HP and Microsoft because it approximates the color gamut of the most common computer display devices. Since sRGB serves as a "best guess" for how another person's monitor produces color, it has become the standard color space for displaying images on the internet. sRGB's color gamut encompasses just 35% of the visible colors specified by CIE (see section on color spaces). Although sRGB results in one of the narrowest gamuts of any working space, sRGB's gamut is still considered broad enough for most color applications.

Adobe RGB 1998 was designed (by Adobe Systems, Inc.) to encompass most of the colors achievable on CMYK printers, but by using only RGB primary colors on a device such as your computer display. The Adobe RGB 1998 working space encompasses roughly 50% of the visible colors specified by CIE-- improving upon sRGB's gamut primarily in cyan-greens.

<http://en.wikipedia.org/wiki/SRGB>

T

Tagged PDF

Tagged PDF defines a set of rules for representing text in the page content so that characters, words, and text order can be determined reliably.

- All text shall be represented in a form that can be converted to Unicode.
- Word breaks shall be represented explicitly.
- Actual content shall be distinguished from artifacts of layout and pagination.
- Content shall be given in an order related to its appearance on the page, as determined by the conforming writer.

Tagged PDF properly reflows when saving back to Microsoft Word (.doc) or Rich Text Format (.RTF) and on devices using Windows CE, Blackberry and so forth.

http://blogs.adobe.com/acrolaw/2006/01/understanding_t_1.html

Transparency

Transparency in PDF files (introduced in PDF 1.4) refers to objects on a page, such as images or text, which are transparent or 'show through'. Transparency is typically used for shadow effects, to lighten (parts of) images so that the text on top remains readable, to make objects fade into another object or to create a tint of a certain color.

U

Unicode



A a B b C c Ď d Ě
Ń Ŧ Ÿ Δ ≡ Ж ѣ џ ѡ
س ع ك あ に 𐄂
𐄃 𐄄 𐄅 𐄆 𐄇 𐄈

Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language. It is supported in many operating systems, all modern browsers, and many other products and is supported in PDF files. The emergence of the Unicode Standard, and the availability of tools supporting it, are among the most significant recent global software technology trends.

<http://www.unicode.org/>

V

Vector Graphics



Vector graphics is the use of geometrical primitives such as points, lines, curves, polygons and is to be contrasted with Raster Graphics. PDF documents can include Vector Graphics.

Version

PDF-1.6

The PDF file format has changed several times and continues to evolve, as new versions of Adobe Acrobat were released. There have been nine versions of PDF with corresponding Acrobat releases.

(1993) – PDF 1.0 / Acrobat 1.0

(1994) – PDF 1.1 / Acrobat 2.0

(1996) – PDF 1.2 / Acrobat 3.0

(1999) – PDF 1.3 / Acrobat 4.0

(2001) – PDF 1.4 / Acrobat 5.0

(2003) – PDF 1.5 / Acrobat 6.0

(2005) – PDF 1.6 / Acrobat 7.0

(2006) – PDF 1.7 / Acrobat 8.0

(2008) – PDF 1.7, Adobe Extension Level 3-Acrobat 9.0

(2009) – PDF 1.7, Adobe Extension Level 5-Acrobat 9.1

X

XPDF

file conforming to the XML Forms Data Format 2.0 specification, which is an XML transliteration of Forms Data Format (FDF)

XMP



Adobe's Extensible Metadata Platform (XMP) is a labeling technology that allows you to embed data about a file, known as metadata, into the file itself.

Extensible Metadata Platform (XMP) is an XML-based format modeled by Adobe after W3C's RDF (Resource Description Framework) which forms the foundation of the semantic Web initiative. Adobe makes the XMP specification freely available, and offers an open-source XMP toolkit for software developers.

XMP metadata travels with the file, and can be embedded in many common file formats including PDF, TIFF, and JPEG. Metadata properties are grouped in schemas. Each schema is identified by a unique namespace URI and holds an arbitrary number of properties. While namespace URIs look very similar to the familiar Web addresses (actually, they often look the same), it's important to note that they do not identify a particular Web page. In fact, namespace URIs are not required to point to any resource – they are simply unique identifiers for some entity used in XMP.

<http://www.adobe.com/products/xmp/>

<http://www.adobe.com/devnet/xmp/>