

# PDF – Basics – Cheat Sheet

## PDF Terminology

<b>PDF 2.0</b>	ISO 32000	Latest core PDF specification that applies to all PDF files. Fully vendor-neutral specification of every non-obsolete PDF feature since PDF 1.0.
<b>PDF/A</b>	ISO 19005	Archival. PDF for the long-term preservation of static page appearance.
<b>PDF/X</b>	ISO 15930	eXchange. PDF for graphic arts and professional printing workflows including packaging and labelling.
<b>PDF/UA</b>	ISO 14289	Universal Accessibility. PDF supporting accessibility and assistive technology such as screen readers for those with vision impairments.
<b>PDF/R</b>	ISO 23504	Raster. Simplified PDF that uses banded images and that is easy to create by low-resource embedded devices such as consumer flatbed scanners.
<b>PDF/VT</b>	ISO 16612	Variable/Transactional for high-speed variable data (VDP) and transactional printing in the graphic arts and professional printing industries. Builds on PDF/X.
<b>PDF/VCR</b>	ISO 16613	Variable Content Replacement. Templated PDF/X supporting late-stage variable content merging, such as adding batch numbers on pharmaceutical packaging.
<b>PDF/E</b>	ISO 24517	Engineering. PDF 1.6 based subset to support engineering-centric 3D PDF workflows such as aerospace and automotive engineering. Superseded by PDF/A-4.

### PDF/raster PDF/X-5pg PDF/A-4f

**PDF subset**  
 Uppercase letters = ISO  
 Lowercase = Industry

**Subset Conformance Level(s)**  
 (optional, lowercase letters)

**Version of subset**  
 (optional)

A single PDF can conform to multiple subsets and conformance levels.

## Key PDF features by version

<b>PDF 1.0</b>	1993	Pure text. Only RGB. Resolution independent graphics.
<b>PDF 1.1</b>	1996	Device-independent color. Article threads. External links. Security. Multimedia. Actions.
<b>PDF 1.2</b>	1996	Prepress features: CMYK, spot color, halftoning, overprinting, OPI (Open Prepress Interface). Flate compression. New types of annotations. Interactive forms.
<b>PDF 1.3</b>	2000	DeviceN color. 2-byte CID fonts. Smooth shadings. More types of annotations. Large media. Page labels. Digital signatures. JavaScript. Alternate images. Masked images.
<b>PDF 1.4</b>	2001	Transparency and blend modes. Improved security. More prepress features. JBIG2 images.
<b>PDF 1.5</b>	2003	JPEG 2000. Layers (optional content). Tagged PDF. Object streams and cross reference streams for better compression.
<b>PDF 1.6</b>	2004	OpenType. Ultra-large media. Watermarking. Visibility expressions. AES encryption. Interactive 3D with U3D. Measurement properties.
<b>PDF 1.7</b>	2006	Portable collections (packages). 3D enhancements. Redaction annotations. Standardized as ISO 32000-1:2008.
<b>PDF 2.0</b>	2017	ISO 32000-2. UTF-8 strings. 256-bit AES-CBC encryption. Unicode passwords. Black point compensation. Rich media annotations. PAdES. PRC for 3D. Geospatial features. Document parts. Associated Files. Metadata streams. Deprecation of older encryption and other legacy features, including XFA.
<b>PDF 2.0 Extensions</b>		256-bit AES-GCM encryption. Hash algorithm, elliptical curve, and digital signature updates. Integrity protection via MACs. STEP for 3D. Clarification on PDF 1.7 and PDF 2.0 namespaces.

## Common terms for PDF features

<b>Bookmarks</b>	Outlines which use Actions or Destinations.
<b>Comments</b>	Markup annotations.
<b>Compression</b>	Filters on streams. Object streams. Cross reference streams.
<b>Fast Web View</b>	Linearization.
<b>Files</b>	Embedded Files and File Attachment annotations.
<b>Forms</b>	Widget annotations and Fields. Also referred to as <b>AcroForm</b> .
<b>Hyperlinks</b>	Link annotations, URI actions. Actions and Destinations.
<b>JavaScript (JS)</b>	ECMAScript for PDF (ISO 21757). ECMAScript Actions.
<b>Layers</b>	Optional Content (OC), Visibility Expressions. Marked Content.
<b>Multimedia</b>	3D, Movie, Screen, and RichMedia annotations with Actions.
<b>Page size</b>	The page <b>MediaBox</b> .
<b>Portfolios</b>	Collections and Navigators with Embedded Files.
<b>Properties</b>	Document Information dictionary and XMP Metadata streams.
<b>Scanned PDF</b>	Contains images of content such as produced by a scanner or camera. Often has OCR-ed invisible text (Text render mode 3) placed on top of the image allowing text selection by users.
<b>Security</b>	Encryption, Crypt filters, and Digital Signatures.
<b>Tags</b>	Tagged PDF, including Marked Content and Logical Structure.

## Glossary

<b>Action</b>	PDF feature enabling automatic behaviours triggered by a user interaction or event (e.g., change to a different page when a bookmark is clicked, follow a URL hyperlink, etc.).
<b>AT</b>	Assistive Technology. Associated with PDF/UA and Tagged PDF.
<b>BBox</b>	Bounding Box. A common key name.
<b>Conformance level</b>	Represented by letter designators after a PDF subset acronym, (e.g., PDF/A-1b, PDF/X-5pg, PDF/VT-2s). Each Conformance Level has its own specialized set of rules and requirements.
<b>COS</b>	Carousel Object Syntax. The syntax used by PDF and FDF files. "Carousel" was the codename for Acrobat 1.0.
<b>Cross-reference stream</b>	(PDF 1.5 and later only) Cross-reference information stored in a stream instead of a standard cross-reference <code>xref</code> table. Trailer dictionary entries are in cross reference stream dictionary.
<b>Destination</b>	An object defining a view of a document, comprising a page, the location of the document window on that page, and zoom factor.
<b>Direct object</b>	PDF object that occurs inline where it is defined and does not have its own object identifier (object number and generation number pair).
<b>FDF</b>	Forms Data Field file format to store interactive form data
<b>Hybrid-reference PDF</b>	(PDF 1.5 and later only). PDF containing objects referenced by conventional cross-reference tables in addition to objects in object streams referenced by cross-reference streams.
<b>Linearized PDF</b>	Commonly referred to as Fast Web View.
<b>obj</b>	Object abbreviation. A reserved PDF keyword.
<b>Object stream</b>	(PDF 1.5 and later only). A stream in which indirect objects may be stored, as an alternative to being stored in PDF body sections.
<b>OCG</b>	Optional Content Group. A selectable "layer" of page content.
<b>Owner Password</b>	Password with full (owner) access, including ability to change passwords and access permissions of the PDF document.
<b>PAdES</b>	PDF Advanced Electronic Signatures. ETSI standard EN 319 142.
<b>Page Label</b>	Optional descriptive text for referring to pages that can be shown on-screen (e.g., i, ii, iii, ..., Chapter 1, Chapter 2, etc). This contrasts with the zero-based integer page index used internally in PDF files.
<b>startxref</b>	Reserved PDF keyword that occurs just before the <code>%%EOF</code> end-of-file comment marker along with the byte offset to the cross-reference data for the PDF file (expressed as an integer in ASCII).
<b>trailer</b>	The trailer dictionary is required in every PDF and defines special objects (e.g., largest object number, the Document Catalog root). Also keyword.
<b>User Password</b>	Password with restricted access permissions (as set by an author).
<b>Widget</b>	A subtype of PDF annotation used with interactive forms that represent the GUI "widgets" through which data entry is done.
<b>XFA</b>	XML Forms Architecture. Proprietary XML-based specifications supporting dynamic forms. Deprecated in PDF 2.0.
<b>XFDF</b>	XML-based version of FDF defined by ISO 19444-1.
<b>XMP</b>	eXtensible Metadata Platform. XML-based metadata standard (ISO 16684) used by many file formats. Required by PDF subsets and PDF 2.0.
<b>xref</b>	Reserved PDF keyword that indicates the start of a standard cross-reference table. Often shorthand for "cross reference table".

## MIME types

<b>application/pdf</b>	Official MIME type for all PDF files. See RFC 8118.
<b>application/fdf</b>	Official MIME type for FDF (Forms Data Field) files. See ISO 32000 for FDF file specification.
<b>application/xfdf</b>	Official MIME type for XFDF (XML Forms Data Field) files. See ISO 19444-1 for XFDF file format specification.

Report errata at <https://github.com/pdf-association/pdf-issues/>

## Lexical Rules

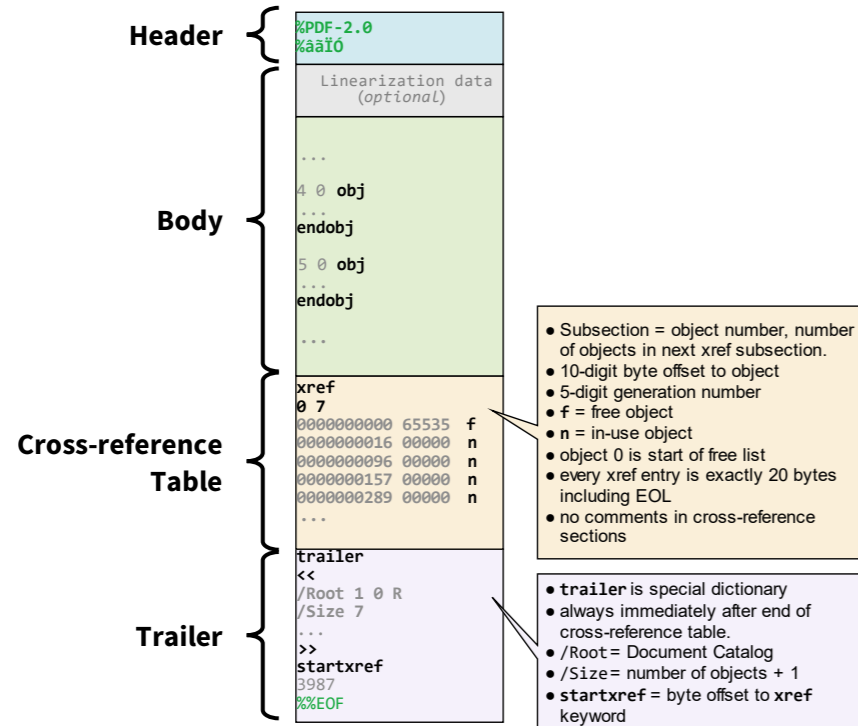
End-of-Line (EOL) Sequences	
0x0D	Carriage Return (CR) only
0x0A	Line Feed (LF) only
0x0D 0x0A	Carriage Return (CR) followed by Line Feed (LF) only

White space	
0x00	Null byte
0x09	Horizontal Tab (HT)
0x0C	Form Feed (FF)
0x20	SPACE
EOL	Any End-of-Line sequence (see above)
% ...	PDF comments (starting from % to EOL) are treated as single white space

Token Delimiter symbols	
(	Literal string start token
)	Literal string end token
<, <<	Hex string start token / dictionary start (<<) token
>, >>	Hex string end token / dictionary end (>>) token
[	Array start token
]	Array end token
/	PDF name
%	Comment to end-of-line (outside of a string or inside a content stream)
{	Only in Type 4 PostScript calculator functions
}	Only in Type 4 PostScript calculator functions

## File Structure

(when not using cross-reference streams (PDF 1.5))



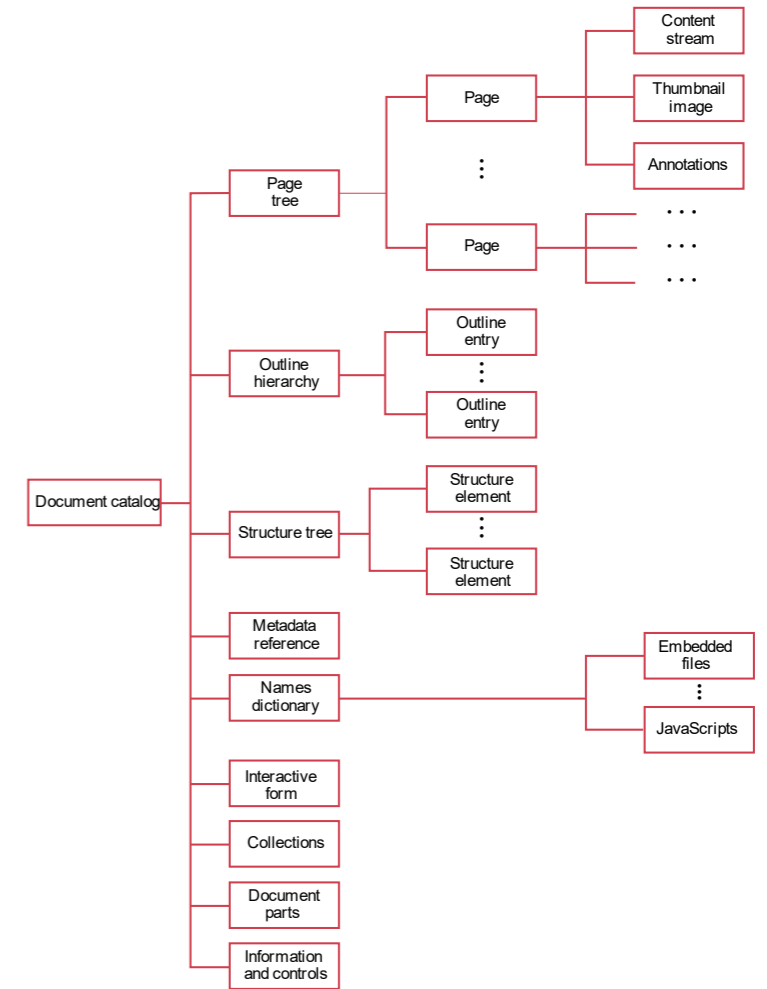
Multiple **Incremental Updates** append additional **Body**, **Cross-reference Table** and **Trailer** sections to a PDF file, allowing edits and changes without rewriting the full PDF. Link to previous PDF state is via **Prev** entry in the trailer dictionary to previous **xref**.

## PDF Basics

Reserved Keywords (case sensitive)	
endobj, endstream, f, false, n, null, obj, R, startxref, stream, trailer, true, xref	

Objects																						
<b>Boolean</b>	true false	<ul style="list-style-type: none"> <li>Case sensitive keywords.</li> </ul>																				
<b>Integer</b>	0 123 +45 -67890	<ul style="list-style-type: none"> <li>Signed decimal integer.</li> <li>No hexadecimal or octal integers.</li> </ul>																				
<b>Real Number</b>	1.23 -45.6 +7.8 -.9 0.	<ul style="list-style-type: none"> <li>Signed decimal floating-point numbers.</li> <li>No exponential or scientific formats.</li> <li>Integers can be used for real numbers.</li> </ul>																				
<b>String</b>	(literal string) (balanced () ok) (unbalanced \() (line \ break) (line \nbreak) (octal \234 code)  <hex-string> <48656c6c66> <41424> % 0 added	<ul style="list-style-type: none"> <li>Encrypted PDFs encrypt string objects.</li> <li>Unicode strings with byte order markers.</li> <li>Backslash escape sequences for literal strings:</li> </ul> <table border="1"> <thead> <tr> <th>Sequence</th> <th>Meaning</th> </tr> </thead> <tbody> <tr> <td>\n</td> <td>LF (0x0A)</td> </tr> <tr> <td>\r</td> <td>CR (0x0D)</td> </tr> <tr> <td>\t</td> <td>Horizontal Tab (0x09)</td> </tr> <tr> <td>\b</td> <td>Backspace (0x08)</td> </tr> <tr> <td>\f</td> <td>Formfeed (0x0C)</td> </tr> <tr> <td>\)</td> <td>Left parenthesis</td> </tr> <tr> <td>\)</td> <td>Right parenthesis</td> </tr> <tr> <td>\\</td> <td>Backslash</td> </tr> <tr> <td>\ddd</td> <td>Octal code. 1-3 digits</td> </tr> </tbody> </table> <p>string types</p> <ul style="list-style-type: none"> <li>text string</li> <li>ASCII string</li> <li>byte string</li> </ul> <p>PDFDocEncoding UTF-16BE UTF-8 (BoM: 0xFE 0xFF) (BoM: 0xEF 0xBB 0xBF)</p>	Sequence	Meaning	\n	LF (0x0A)	\r	CR (0x0D)	\t	Horizontal Tab (0x09)	\b	Backspace (0x08)	\f	Formfeed (0x0C)	\)	Left parenthesis	\)	Right parenthesis	\\	Backslash	\ddd	Octal code. 1-3 digits
Sequence	Meaning																					
\n	LF (0x0A)																					
\r	CR (0x0D)																					
\t	Horizontal Tab (0x09)																					
\b	Backspace (0x08)																					
\f	Formfeed (0x0C)																					
\)	Left parenthesis																					
\)	Right parenthesis																					
\\	Backslash																					
\ddd	Octal code. 1-3 digits																					
<b>Name</b>	/CaseSensitive12 /HashSign#23	<ul style="list-style-type: none"> <li>Start with / SOLIDUS (0x2F).</li> <li>Can use # followed by 2 hex digits.</li> </ul>																				
<b>Array</b>	[/AName true null -1.23 10 0 R ... ] [] % empty array	<ul style="list-style-type: none"> <li>One dimensional ordered collection with zero or more elements.</li> <li>Array elements can be any type of object.</li> </ul>																				
<b>Dictionary</b>	<< /KeyName value ... >>	<ul style="list-style-type: none"> <li>Associative unordered table containing <b>key/value pairs</b> known as an <b>entry</b>.</li> <li>Keys must be unique direct name objects.</li> <li>If value is <b>null</b> then same as if key does not exist.</li> </ul>																				
<b>Stream</b>	10 0 obj << /Length int ... >> stream ...stream data... endstream endobj	<ul style="list-style-type: none"> <li>Contains zero or more stream data bytes.</li> <li>Always need a stream dictionary.</li> <li>Stream data can be compressed and/or encrypted using <b>Filters</b>.</li> <li>Always an indirect object.</li> <li>Cannot be in object streams.</li> <li>Encrypted PDFs encrypt stream data.</li> </ul>																				
<b>Null object</b>	null	<ul style="list-style-type: none"> <li>Case sensitive reserved keyword.</li> </ul>																				
<b>Indirect Reference</b>	10 0 R	<ul style="list-style-type: none"> <li>Object number then generation number.</li> <li>Method to refer to another object.</li> </ul>																				

## Document Structure



## Painting Graphics

