Portable Document Format (PDF), standardized as ISO 32000, is a file format developed by Adobe in 1992 to present documents, including text formatting and images, in a (including video content), three- Organization for manner independent of application software, hardware, and operating systems. Based on the PostScript language, each PDF file encapsulates a complete description of a fixedlayout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. Systems made the PDF PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was standardized as ISO 32000 in 2008. The last edition as ISO 32000-2:2020 was published in December 2020. PDF files may Paper, Farallon Replica and contain a variety of content

besides flat text and graphics including logical structuring elements, interactive elements it was released as an open such as annotations and form- standard on July 1, 2008, and fields, layers, rich media formats. The PDF specification the specification passed to an also provides for encryption and ISO Committee of volunteer

digital signatures, file attachments, and metadata to enable workflows requiring these features. History Main article: History of PDF Adobe specification available free of charge in 1993. In the early in desktop publishing workflows, and competed with a ISO 32000-1, includes some

even Adobe's own PostScript

format controlled by Adobe until published by the International dimensional objects using U3D Standardization as ISO 32000or PRC, and various other data 1:2008, at which time control of 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 PDFcompliant implementations. years PDF was popular mainly PDF 1.7, the sixth edition of the PDF specification that became variety of formats such as DjVu, proprietary technologies defined Envoy, Common Ground Digital only by Adobe, such as Adobe XML Forms Architecture (XFA)

and JavaScript extension for

format. PDF was a proprietary

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 plain text); Vector graphics for standardized and their specification is published only on Adobe's website. Many of them are also not supported by photographs and other types of based on PostScript but popular third-party implementations of PDF. In December 2020, the second edition of PDF 2.0, ISO 32000- also support links (inside 2:2020, was published, including clarifications, corrections, and critical updates a plugin for Acrobat 3.0), or any PostScript file. The graphics to normative references. ISO

32000-2 does not include any proprietary technologies as normative references. ISO's publication of ISO 32000-2 in 2017 ended the 24-year tradition of the latest PDF specification being freely available from Adobe. Starting in April, 2023, to provide PDF developers and stakeholders with their accustomed level of access, the PDF Association and its sponsors made ISO 32000-2 available for download content into a single file, with at no cost. Technical details A

of vector graphics, text, and bitmap graphics. The basic types of content in a PDF are: Typeset text stored as content streams (i.e., not encoded in illustrations and designs that consist of shapes and lines; Raster graphics for images Multimedia objects in the document. In later PDF document or web page), forms, PostScript-like PDF code is JavaScript (initially available as generated from a source

other types of embedded contents that can be handled using plug-ins. PDF combines three technologies: An equivalent subset of the PostScript page description programming language but in declarative form, for generating is compressed to a single file. embedding/replacement systemworld (fonts, layout, to allow fonts to travel with the documents. A structured storage system to bundle these document format, PDF has elements and any associated data compression where PDF file is often a combination appropriate. PostScript

language 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 statements and loop commands. PDF is largely simplified to remove flow control features like these, while revisions, a PDF document can graphics commands equivalent to lineto remain. Historically, the commands that are output by the PostScript code are collected and tokenized.[clarification needed] Any files, graphics, or fonts to which the document refers also are collected. Then, everything the layout and graphics. A font- Therefore, the entire PostScript measurements) remains intact.[citation needed] As a several advantages over

PostScript: PDF contains

tokenized and interpreted

results of the PostScript source

code, for direct correspondence documents embedded in a PDF amounts of optionally between changes to items in the PDF page description and changes to the resulting page appearance. PDF (since version 1.4) supports transparent graphics; PostScriptcharacters, except for certain does not. PostScript is an interpreted programming language with an implicit global header containing a magic state, so instructions accompanying the description of one page can affect the appearance of any following format is a subset of a COS page. Therefore, all preceding ("Carousel" Object Structure) pages in a PostScript document format. A COS tree file consists between the obj and endobj 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 ordered collections of objects destination page (unless the optional PostScript Document Structuring Conventions have been carefully compiled and included). PDF 1.6 and later supports interactive 3D

file: 3D drawings can be embedded using U3D or PRC and various other data formats. enclosed between the stream File format A PDF file is organized using ASCII elements that may have binary content. The file starts with a number (as a readable string) and the version of the format, for example %PDF-1.7. The

primarily of objects, of which there are nine types: Boolean values, representing true or false Real numbers Integers Strings, enclosed within parentheses ((...)) or represented as hexadecimal within single angle brackets

enclosed within square bracketsnumbers of small indirect objects indexed by names enclosed within double angle brackets (<<...>>) Streams, usually containing large

characters. Names, starting

with a forward slash (/) Arrays,

compressed binary data, preceded by a dictionary and and endstream keywords. 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 keywords if residing in the document root. Beginning with PDF version 1.5, indirect objects (except other streams) may also be located in special streams known as object streams (marked /Type /ObjStm). This technique (<...>). Strings may contain 8-bitenables non-stream objects to

applied to them, reduces the size of files that have large ([...]) Dictionaries, collections of objects and is especially useful for Tagged PDF. Object streams do not support

have standard stream filters

specifying an object's generation number (other than 0). An index table, also called the cross-reference table, is located near the end of the file and gives the byte offset of each indirect object from the start of the file. This design allows for efficient random access to the objects in the file, %%EOF end-of-file marker. If a Linearized PDF files (also and also allows for small changes to be made without rewriting the entire file (incremental update). Before PDF version 1.5, the table would always be in a special ASCII format, be marked with the xref keyword, and follow the object's dictionary: A reference main body composed of indirect to the root object of the tree objects. Version 1.5 introduced structure, also known as the optional cross-reference streams, which have the form of indirect objects in the crossa standard stream object, possibly with filters applied. Such a stream may be used instead of the ASCII crossreference table and contains the offsets and other information in binary format. The format is flexible in that it allows for integer width specification (using the /W array), so that for example, a document not exceeding 64 KiB non-linearized (not "optimized") in PDF 1.4. PDF graphics use a in size may dedicate only 2 bytes for object offsets. At the

end of a PDF file is a footer containing The startxref keyword followed by an offset toslower to access because the start of the cross-reference portions of the data required to table (starting with the xref keyword) or the cross-reference document are scattered stream object, followed by The throughout the PDF file. cross-reference stream is not being used, the footer is preceded by the trailer keyword constructed in a manner that followed by a dictionary containing information that would otherwise be contained inwaiting for the entire file to the cross-reference stream catalog (/Root) The count of reference table (/Size) Other optional information Within eachby the format itself. However, page, there are one or multiple content streams that describe the text, vector and images being drawn on the page. The content stream is stack-based. similar to PostScript. The maximum size of a PDF compared to Europe. There are except for the use of two layouts to the PDF files: and linearized ("optimized").

smaller than their linear counterparts, though they are assemble pages of the called "optimized" or "web optimized" PDF files) are enables them to be read in a Web browser plugin without download, since all objects required for the first page to display are optimally organized at the start of the file. PDF files may be optimized using Adobe Acrobat software or QPDF. Page dimensions are not limited Adobe Acrobat imposes a limit of 15 million in by 15 million in, or 225 trillion in 2 (145,161 km2). Imaging model The basic design of how graphics are represented in PDF is very similar to that of PostScript, transparency, which was added device-independent Cartesian Non-linearized PDF files can be coordinate system to describe

the surface of a page. A PDF page description can use a matrix to scale, rotate, or skew graphical elements. A key concept in PDF is that of the graphics state, which is a collection of graphical parameters that may be changed, saved, and restored by a page description. PDF has object, or an uncolored tiling (as of version 2.0) 25 graphics state properties, of which some specification to the time the of the most important are: The current transformation matrix (CTM), which determines the coordinate system The clipping continuously varying colors. path The color space The alpha There are seven types of constant, which is a key component of transparency Black point compensation control (introduced in PDF 2.0) Vector graphics As in PostScript, vector graphics in PDF are constructed with paths.dictionaries with an associated Paths are usually composed of lines and cubic Bézier curves, but can also be constructed from the outlines of text. Unlike the image data. (Less with lines and curves. Paths can be stroked, filled, fill then

stroked, or used for clipping.

Strokes and fills can use any

color set in the graphics state, including patterns. PDF supports several types of patterns. The simplest is the tiling pattern in which a piece of ASCII, ASCIIHexDecode, artwork is specified to be drawn similar to ASCII85Decode but repeatedly. This may be a colored tiling pattern, with the colors specified in the pattern pattern, which defers color pattern is drawn. Beginning withintroduced in PDF 1.2; it can PDF 1.3 there is also a shading use one of two groups of pattern, which draws shading patterns of which the simplest are the axial shading (Type 2) and radial shading (Type 3). Raster images Raster LZWDecode, a filter based on images in PDF (called Image XObjects) are represented by stream. The dictionary describes the properties of the from the TIFF 6.0 specification image, and the stream contains and predictors (filters) from the PostScript, PDF does not allow commonly, small raster images RunLengthDecode, a simple a single path to mix text outlinesmay be embedded directly in a compression method for page description as an inline image.) Images are typically filtered for compression purposes. Image filters

supported in PDF include the following general-purpose filters: ASCII85Decode, a filter used to put the stream into 7-bit less compact, FlateDecode, a commonly used filter based on the deflate algorithm defined in RFC 1951 (deflate is also used in the gzip, PNG, and zip file formats among others); predictor functions for more compact zlib/deflate compression: Predictor 2 from the TIFF 6.0 specification and predictors (filters) from the PNG specification (RFC 2083), LZW Compression; it can use one of two groups of predictor functions for more compact LZW compression: Predictor 2 PNG specification, streams with repetitive data using the run-length encoding algorithm and the image-

specific filters, DCTDecode, a

lossy filter based on the JPEG standard, CCITTFaxDecode, a embedded font file. The latter lossless bi-level (black/white) filter based on the Group 3 or Group 4 CCITT (ITU-T) fax compression standard defined in ITU-T T.4 and T.6, JBIG2Decode, a lossy or lossless bi-level (black/white) filter based on the JBIG2 standard, introduced in PDF 1.4, and JPXDecode, a lossy or Additionally PDF supports the lossless filter based on the JPEG 2000 standard, introduced in PDF 1.5. Normally described by PDF graphic all image content in a PDF is embedded in the file. But PDF allows image data to be stored in external files by the use of external streams or Alternate Images. Standardized subsets of PDF, including PDF/A and PDF/X, prohibit these features. Helvetica (v3) (in regular, Text Text in PDF is represented oblique, bold and bold oblique) provide a lookup table of by text elements in page content streams. A text element fonts are sometimes called the specifies that characters should base fourteen fonts. These be drawn at certain positions. The characters are specified using the encoding of a selected font resource. A font object in PDF is a description of guaranteed to be available in

typeface, or it may include an case is called an embedded font while the former is called an unembedded font. The font based on widely used standard the current font using an digital font formats: Type 1 (and encoding. There are several its compressed variant CFF), TrueType, and (beginning with WinAnsi, MacRoman, and PDF 1.6) OpenType. Type 3 variant in which the components of the font are operators. Fourteen typefaces, known as the standard 14 fonts, properties of the Windows and have a special significance in PDF documents: Times (v3) (in fonts using these encodings regular, italic, bold, and bold italic) Courier (in regular, oblique, bold and bold oblique) Symbol Zapf Dingbats These fonts, or suitable substitute fonts with the same metrics, should be available in most PDF readers, but they are not a digital typeface. It may either the reader, and may only describe the characteristics of a display correctly if the system

has them installed. Fonts may be substituted if they are not embedded in a PDF. Within text strings, characters are shown using character codes files that may be embedded are (integers) that map to glyphs in predefined encodings, including many encodings for East Asian languages and a font can have its own built-in encoding. (Although the WinAnsi and MacRoman encodings are derived from the historical Macintosh operating systems, work equally well on any platform.) PDF can specify a predefined encoding to use, the font's built-in encoding or differences to a predefined or built-in encoding (not recommended with TrueType fonts). The encoding mechanisms in PDF were designed for Type 1 fonts, and the rules for applying them to TrueType fonts are complex. For large fonts or fonts with non-standard glyphs, the

special encodings Identity-H (for horizontal writing) and Identity-V (for vertical) are used.alpha. The model is closely With such fonts, it is necessary aligned with the features of to provide a ToUnicode table if Adobe Illustrator version 9. The content (text, graphics, and semantic information about the characters is to be preserved. Transparency The original imaging model of PDF was, like the PDF 1.4 specification was PostScript's, opaque: each object drawn on the page completely replaced anything previously marked in the same location. In PDF 1.4 the imagingconcept of a transparency model was extended to allow transparency. When transparency is used, new objects interact with previously marked objects to produce blending effects. The addition of reflect logical relationships transparency to PDF was done among objects that are by means of new extensions that were designed to be ignored in products written to PDF 1.3 and earlier specifications. As a result, files that use a small amount of transparency might view acceptably by older viewers, butand semantics information to files making extensive use of transparency could be viewed incorrectly by an older viewer. The transparency extensions are based on the key concepts builds on the logical structure

of transparency groups, blending modes, shape, and blend modes were based on those used by Adobe Photoshop at the time. When published, the formulas for calculating blend modes were kept secret by Adobe. They have since been published. The were relatively vague in ISO group in PDF specification is independent of existing notions devices, including assistive of "group" or "layer" in applications such as Adobe Illustrator. Those groupings meaningful when editing those objects, but they are not part of subset of PDF specifically the imaging model. Additional features Logical structure and accessibility A "tagged" PDF includes document structure enable reliable text extraction and accessibility. Technically speaking, tagged PDF is a stylized use of the format that

framework introduced in PDF 1.3. Tagged PDF defines a set of standard structure types and attributes that allow page images) to be extracted and reused for other purposes. Tagged PDF is not required in situations where a PDF file is intended only for print. Since the feature is optional, and since the rules for Tagged PDF 32000-1, support for tagged PDF among consuming technology (AT), is uneven as of 2021. ISO 32000-2, however, includes an improved discussion of tagged PDF which is anticipated to facilitate further adoption. An ISO-standardized targeted at accessibility, PDF/UA, was first published in 2012. Optional Content Groups (see clause 14.8 in ISO 32000) (layers) With the introduction of PDF version 1.5 (2003) came the concept of Layers. Layers, more formally known as **Optional Content Groups** (OCGs), refer to sections of content in a PDF document that can be selectively viewed or

hidden by document authors or ISO 32000-2. PDF files may viewers. This capability is usefulalso contain embedded DRM in CAD drawings, layered artwork, maps, multi-language documents, etc. Basically, it consists of an Optional Content restrictions depend on the Properties Dictionary added to the document root. This dictionary contains an array of Optional Content Groups (OCGs), each describing a set of information and each of which may be individually displayed or suppressed, plus a file and prevents opening, and set of Optional Content Configuration Dictionaries, which give the status (Displayed or Suppressed) of the given OCGs. Encryption and signatures A PDF file may be encrypted, for security, in which case a password is needed to view or edit the contents. PDF 2.0 defines 256- password encrypts the file, bit AES encryption as standard for PDF 2.0 files. The PDF Reference also defines ways that third parties can define their own encryption systems for PDF. PDF files may be digitally signed, to provide secure authentication; complete restrictions that a document

details on implementing digital

restrictions that provide further controls that limit copying, editing, or printing. These reader software to obey them, so the security they provide is limited. The standard security provided by PDF consists of two different methods and two different passwords: a user password, which encrypts the an owner password, which specifies operations that should (UR) signatures are used to be restricted even when the document is decrypted, which can include modifying, printing, by default in a particular PDF or copying text and graphics outviewer application. The of the document, or adding or modifying text notes and AcroForm fields. The user while the owner password does be used to allow a user: To not, instead relying on client software to respect these restrictions. An owner passwordannotation data Import form can easily be removed by software, including some free online services. Thus, the use

author places on a PDF

signatures in PDF is provided in document are not secure, and

cannot be assured once the file is distributed; this warning is displayed when applying such restrictions using Adobe Acrobat software to create or edit PDF files. Even without removing the password, most freeware or open source PDF readers ignore the permission "protections" and allow the user to print or make copy of excerpts of the text as if the document were not limited by password protection. Beginning with PDF 1.5, Usage rights enable additional interactive features that are not available signature is used to validate that the permissions have been granted by a bona fide granting authority. For example, it can save the PDF document along with a modified form or data files in FDF, XFDF, and text (CSV/TSV) formats Export form data files in FDF and XFDF formats Submit form data Instantiate new pages from

named page templates Apply a

digital signature form field Create, delete, modify, copy, import, and export annotations For example, Adobe Systems grants permissions to enable additional features in Adobe Reader, using public-key cryptography. Adobe Reader verifies that the signature uses a certificate from an Adobeauthorized certificate authority. Any PDF application can use this same mechanism for its own purposes. Under specific circumstances including nonpatched systems of the receiver, the information the receiver of a digital signed document sees can be manipulated by the sender after 2.0 allows metadata to be the document has been signed attached to any object in the by the signer. PAdES (PDF Advanced Electronic Signatures) is a set of restrictions and extensions to PDF and ISO 32000-1 making itto the document catalog), using significant challenges with PDF suitable for advanced electronic an extensible schema. PDF signatures. This is published by documents can also contain ETSI as TS 102 778. File attachments PDF files can have page display layout and zoom file attachments which processors may access and

open or save to a local

digital signature to existing

filesystem. Metadata PDF files can contain two types of metadata. The first is the **Document Information** Dictionary, a set of key/value fields such as author, title, subject, creation and update dates. This is optional and is referenced from an Info key in the trailer of the file. A small set captions, audio descriptions, of fields is defined and can be extended with additional text values if required. This method PDFs, but this feature is not is deprecated in PDF 2.0. In PDF 1.4, support was added for Leading screen readers, Metadata Streams, using the **Extensible Metadata Platform** (XMP) to add XML standardsbased extensible metadata as used in other file formats. PDF document, such as information about embedded illustrations. fonts, and images, as well as the whole document (attaching display settings, including the level in a Viewer Preferences object. Adobe Reader uses

these settings to override the

user's default settings when opening the document. The free Adobe Reader cannot remove these settings. Accessibility PDF files can be created specifically to be accessible for people with disabilities. PDF file formats in use as of 2014 can include tags, text equivalents, and more. Some software can automatically produce tagged always enabled by default. including JAWS, Window-Eyes, Hal, and Kurzweil 1000 and 3000 can read tagged PDF. Moreover, tagged PDFs can be re-flowed and magnified for readers with visual impairments. Adding tags to older PDFs and those that are generated from scanned documents can present some challenges. One of the accessibility is that PDF documents have three distinct views, which, depending on the document's creation, can be inconsistent with each other. The three views are (i) the physical view, (ii) the tags view,

and (iii) the content view. The physical view is displayed and printed (what most people consider a PDF document). The document. Forms Interactive tags view is what screen readers and other assistive technologies use to deliver high-quality navigation and reading experience to users with disabilities. The content view is based on the physical order of objects within the PDF's content stream and may be displayed by software that does not fully support the tags' view, such as the Reflow feature in Adobe's Reader. PDF/UA, the International Standard for accessible PDF based on ISO 32000-1 was first XFA Forms are not compatible published as ISO 14289–1 in 2012 and establishes normative deprecated from PDF with PDF Data Format can be used when language for accessible PDF technology. Multimedia Rich

Media PDF is a PDF file

including interactive content

within the file. It can contain

buttons. For example, if the

interactive PDF is a digital

catalog for an E-commerce

on the PDF pages and can be

added with images and links to interactive form fields to a the website and buttons to order directly from the Forms is a mechanism to add forms to the PDF file format.

PDF currently supports two data and PDF forms. Both formats today coexist in the PDF specification: AcroForms introduced in the PDF 1.2 format specification and included in all later PDF specifications. XML Forms Architecture (XFA) forms, introduced in the PDF 1.5

with AcroForms. XFA was

format specification. Adobe

in the PDF 1.2 format. AcroForms permit using objects and incorporating it into the (e.g. text boxes, Radio buttons, interactive form. It can also be

that can be embedded or linked etc.) and some code (e.g. JavaScript). Alongside the images, audio, video content, orstandard PDF action types, interactive forms (AcroForms)

support submitting, resetting, and importing data. The business, products can be listed"submit" action transmits the

names and values of selected

specified uniform resource locator (URL). Interactive form field names and values may be submitted in any of the following formats, (depending on the settings of the action's different methods for integrating ExportFormat, SubmitPDF, and XFDF flags): HTML Form format HTML 4.01 Specification since PDF 1.5; HTML 2.0 since (also known as Acrobat forms), 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much

simpler than PDF since the body of an FDF document consists of only one required object. Forms Data Format is defined in the PDF specification (since PDF 1.2). The Forms 2.0. AcroForms were introduced submitting form data to a server, receiving the response, used to export form data to

> imported back into the corresponding PDF interactive form. FDF was originally defined in 1996 as part of ISO 32000-2:2017.[citation needed]

stand-alone files that can be

XML Forms Data Format

(XFDF) (external XML Forms Data Format Specification, Version 2.0; supported since PDF 1.5; it replaced the "XML" form submission format defined interactive form. It can also be in PDF 1.4) the XML version of Forms Data Format, but the XFDF implements only a subsetimported back into the of FDF containing forms and annotations. Some entries in the FDF dictionary do not have 3.0 is an ISO/IEC standard XFDF equivalents – such as the under the formal name ISO Status, Encoding, JavaScript, Page's keys, EmbeddedFDFs, Differences, and Target. In addition, XFDF does not allow the spawning, or addition, of new pages based on the given data; as can be done when using an FDF file. The XFDF specification is referenced (but not included) in PDF 1.5 specification (and in later versions). It is described separately in XML Forms Data Format Specification. The PDF 1.4 specification allowed form submissions in XML format, but Forms Data Format (XFDF) this was replaced by submissions in XFDF format in signatures define rights for the PDF 1.5 specification. XFDF conforms to the XML standard. XFDF can be used in formats, and export form data the same way as FDF; e.g.,

form data is submitted to a then sent back and the new form data is imported in an used to export form data to stand-alone files that can be corresponding PDF interactive 19444-1:2019 - Document management — XML Forms Data Format — Part 1: Use of ISO 32000-2 (XFDF 3.0). This standard is a normative The entire document can be submitted rather than individual files without having to pay fields and values, as was defined in PDF 1.4. AcroForms Adobe holds patents to PDF, can keep form field values in external stand-alone files containing key-value pairs. The complying with its PDF external files may use Forms Data Format (FDF) and XML files. The usage rights (UR) import form data files in FDF, XFDF, and text (CSV/TSV) files in FDF and XFDF formats. software, hardware, and

In PDF 1.5, Adobe Systems server, modifications are made, introduced a proprietary format for forms; Adobe XML Forms Architecture (XFA). Adobe XFA Forms are not compatible with ISO 32000's AcroForms feature, and most PDF processors do not handle XFA content. The XFA specification form. As of August 2019, XFDF is referenced from ISO 32000-1/PDF 1.7 as an external proprietary specification and was entirely deprecated from PDF with ISO 32000-2 (PDF 2.0). Split and merge PDF files can be splitted and merged, using applications. Licensing reference of ISO 32000-2. PDF Anyone may create applications that can read and write PDF royalties to Adobe Systems; but licenses them for royaltyfree use in developing software specification. Portable Document Format (PDF), standardized as ISO 32000, is a file format developed by Adobe in 1992 to present documents, including text formatting and images, in a manner independent of application

operating systems. Based on the PostScript language, each PDF file encapsulates a layout flat document, including raster images and other information needed to display it. even Adobe's own PostScript PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was standardized as ISO 32000 in 2008. The last edition as ISO contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements Adobe published a Public such as annotations and formfields, layers, rich media dimensional objects using U3D are necessary to make, use, or PRC, and various other data sell, and distribute PDFformats. The PDF specification compliant implementations. also provides for encryption and PDF 1.7, the sixth edition of the developers and stakeholders digital signatures, file attachments, and metadata to ISO 32000-1, includes some enable workflows requiring these features. History Main article: History of PDF Adobe

Systems made the PDF

specification available free of

charge in 1993. In the early by ISO 32000-1 as normative years PDF was popular mainly and indispensable for the full implementation of the ISO in desktop publishing complete description of a fixed- workflows, and competed with a 32000-1 specification. These variety of formats such as DjVu, proprietary technologies are not the text, fonts, vector graphics, Envoy, Common Ground Digital standardized and their Paper, Farallon Replica and format. PDF was a proprietary format controlled by Adobe until popular third-party it was released as an open standard on July 1, 2008, and published by the International Organization for 32000-2:2020 was published in Standardization as ISO 32000- including clarifications, December 2020. PDF files may 1:2008, at which time control of corrections, and critical updates the specification passed to an ISO Committee of volunteer industry experts. In 2008, Patent License to ISO 32000-1 granting royalty-free rights for (including video content), three- all patents owned by Adobe thattradition of the latest PDF

PDF specification that became

only by Adobe, such as Adobe

and JavaScript extension for

Acrobat, which are referenced

specification is published only on Adobe's website. Many of them are also not supported by implementations of PDF. In December 2020, the second edition of PDF 2.0, ISO 32000-2:2020, was published, to normative references. ISO 32000-2 does not include any proprietary technologies as normative references. ISO's publication of ISO 32000-2 in 2017 ended the 24-year specification being freely available from Adobe. Starting in April, 2023, to provide PDF with their accustomed level of access, the PDF Association proprietary technologies defined and its sponsors made ISO 32000-2 available for download XML Forms Architecture (XFA) at no cost. Technical details A PDF file is often a combination

of vector graphics, text, and

bitmap graphics. The basic types of content in a PDF are: Typeset 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 based on PostScript but images Multimedia objects in the document. In later PDF also support links (inside document or web page), forms, PostScript-like PDF code is JavaScript (initially available as generated from a source a plugin for Acrobat 3.0), or any PostScript file. The graphics other types of embedded contents that can be handled using plug-ins. PDF combines three technologies: An equivalent subset of the PostScript page description programming language but in declarative form, for generating is compressed to a single file. embedding/replacement systemworld (fonts, layout, to allow fonts to travel with the documents. A structured intact.[citation needed] As a storage system to bundle these document format, PDF has elements and any associated several advantages over content into a single file, with PostScript: PDF contains data compression where tokenized and interpreted appropriate. PostScript language PostScript is a page

description language run in an between changes to items in interpreter to generate an the PDF page description and image, a process requiring changes to the resulting page many resources. It can handle appearance. PDF (since graphics and standard features version 1.4) supports of programming languages transparent graphics; PostScript does not. PostScript is an such as if statements and loop commands. PDF is largely interpreted programming language with an implicit global simplified to remove flow controlstate, so instructions features like these, while accompanying the description revisions, a PDF document can graphics commands equivalent of one page can affect the to lineto remain. Historically, the appearance of any following commands that are output by the PostScript code are collected and tokenized.[clarification needed] document is unaffected by the Any files, graphics, or fonts to which the document refers also viewers allow the user to are collected. Then, everything the layout and graphics. A font- Therefore, the entire PostScript PostScript viewer needs to measurements) remains

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 others. As a result, PDF quickly jump to the final pages of a long document, whereas a process all pages sequentially before being able to display the destination page (unless the optional PostScript Document Structuring Conventions have been carefully compiled and included). PDF 1.6 and later results of the PostScript source supports interactive 3D code, for direct correspondence documents embedded in a PDF file: 3D drawings can be embedded using U3D or PRC and various other data formats. enclosed between the stream File format A PDF file is organized using ASCII characters, except for certain elements that may have binary content. The file starts with a header containing a magic number (as a readable string) and the version of the format, for example %PDF-1.7. The format is a subset of a COS ("Carousel" Object Structure) format. A COS tree file consists between the obj and endobj primarily of objects, of which there are nine types: Boolean values, representing true or false Real numbers Integers Strings, enclosed within parentheses ((...)) or represented as hexadecimal within single angle brackets (<...>). Strings may contain 8-bitenables non-stream objects to characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects enclosed within square bracketsnumbers of small indirect ([...]) Dictionaries, collections of objects and is especially useful objects indexed by names

enclosed within double angle

brackets (<<...>>) Streams,

usually containing large

amounts of optionally

compressed binary data, preceded by a dictionary and and endstream keywords. 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 ASCII format, be marked with keywords if residing in the document root. Beginning with PDF version 1.5, indirect objects (except other streams) may also be located in special streams known as object streams (marked /Type /ObjStm). This technique have standard stream filters applied to them, reduces the size of files that have large for Tagged PDF. Object streams do not support

specifying an object's

generation number (other than

0). An index table, also called

the cross-reference table, is located near the end of the file and gives the byte offset of each indirect object from the start of the file. This design allows for efficient random access to the objects in the file, and also allows for small changes to be made without rewriting the entire file (incremental update). Before PDF version 1.5, the table would always be in a special the xref keyword, and follow the main body composed of indirect objects. Version 1.5 introduced optional cross-reference streams, which have the form of a standard stream object, possibly with filters applied. Such a stream may be used instead of the ASCII crossreference table and contains the offsets and other information in binary format. The format is flexible in that it allows for integer width specification (using the /W array), so that for example, a document not exceeding 64 KiB in size may dedicate only 2 bytes for object offsets. At the end of a PDF file is a footer

containing The startxref keyword followed by an offset toslower to access because the start of the cross-reference portions of the data required to table (starting with the xref keyword) or the cross-reference document are scattered stream object, followed by The throughout the PDF file. %%EOF end-of-file marker. If a Linearized PDF files (also cross-reference stream is not being used, the footer is preceded by the trailer keyword constructed in a manner that followed by a dictionary containing information that would otherwise be contained inwaiting for the entire file to the cross-reference stream object's dictionary: A reference required for the first page to to the root object of the tree structure, also known as the catalog (/Root) The count of indirect objects in the crossreference table (/Size) Other optional information Within eachby the format itself. However, page, there are one or multiple Adobe Acrobat imposes a limit PostScript, vector graphics in content streams that describe the text, vector and images or 225 trillion in 2 (145,161 being drawn on the page. The content stream is stack-based, similar to PostScript. The maximum size of a PDF compared to Europe. There are except for the use of two layouts to the PDF files: non-linearized (not "optimized") in PDF 1.4. PDF graphics use a can be stroked, filled, fill then and linearized ("optimized"). Non-linearized PDF files can be coordinate system to describe smaller than their linear

counterparts, though they are assemble pages of the called "optimized" or "web optimized" PDF files) are enables them to be read in a Web browser plugin without download, since all objects display are optimally organized at the start of the file. PDF files constant, which is a key may be optimized using Adobe component of transparency Acrobat software or QPDF.

km2). Imaging model The basic lines and cubic Bézier curves, design of how graphics are represented in PDF is very similar to that of PostScript, transparency, which was added with lines and curves. Paths device-independent Cartesian stroked, or used for clipping. Strokes and fills can use any the surface of a page. A PDF color set in the graphics state,

page description can use a matrix to scale, rotate, or skew graphical elements. A key concept in PDF is that of the graphics state, which is a collection of graphical parameters that may be changed, saved, and restored by a page description. PDF has (as of version 2.0) 25 graphics state properties, of which some of the most important are: The current transformation matrix (CTM), which determines the coordinate system The clipping path The color space The alpha Black point compensation Page dimensions are not limited control (introduced in PDF 2.0) Vector graphics As in of 15 million in by 15 million in, PDF are constructed with paths. Paths are usually composed of but can also be constructed from the outlines of text. Unlike PostScript, PDF does not allow a single path to mix text outlines including patterns. PDF supports several types of patterns. The simplest is the tiling pattern in which a piece of ASCII, ASCIIHexDecode, artwork is specified to be drawn similar to ASCII85Decode but repeatedly. This may be a colored tiling pattern, with the colors specified in the pattern object, or an uncolored tiling

specification to the time the pattern is drawn. Beginning withintroduced in PDF 1.2; it can PDF 1.3 there is also a shading use one of two groups of

pattern, which defers color

pattern, which draws continuously varying colors. There are seven types of shading patterns of which the simplest are the axial shading (Type 2) and radial shading (Type 3). Raster images Raster LZWDecode, a filter based on images in PDF (called Image XObjects) are represented by dictionaries with an associated stream. The dictionary describes the properties of the from the TIFF 6.0 specification image, and the stream contains and predictors (filters) from the the image data. (Less commonly, small raster images RunLengthDecode, a simple may be embedded directly in a page description as an inline image.) Images are typically filtered for compression

purposes. Image filters

supported in PDF include the

following general-purpose filters: ASCII85Decode, a filter used to put the stream into 7-bit filter based on the Group 3 or less compact, FlateDecode, a

commonly used filter based on JBIG2Decode, a lossy or the deflate algorithm defined in lossless bi-level (black/white) RFC 1951 (deflate is also used filter based on the JBIG2 in the gzip, PNG, and zip file formats among others);

predictor functions for more compact zlib/deflate compression: Predictor 2 from the TIFF 6.0 specification and predictors (filters) from the PNG in external files by the use of specification (RFC 2083),

LZW Compression; it can use

one of two groups of predictor functions for more compact LZW compression: Predictor 2 PNG specification, compression method for streams with repetitive data using the run-length encoding algorithm and the imagespecific filters, DCTDecode, a lossy filter based on the JPEG

standard, CCITTFaxDecode, a lossless bi-level (black/white) Group 4 CCITT (ITU-T) fax compression standard defined in ITU-T T.4 and T.6, standard, introduced in PDF 1.4, and JPXDecode, a lossy or lossless filter based on the JPEG 2000 standard. introduced in PDF 1.5. Normally all image content in a PDF is embedded in the file. But PDF allows image data to be stored external streams or Alternate Images. Standardized subsets of PDF, including PDF/A and PDF/X, prohibit these features. Text Text in PDF is represented by text elements in page content streams. A text element specifies that characters should be drawn at certain positions. The characters are specified using the encoding of a selected font resource. A font object in PDF is a description of a digital typeface. It may either

describe the characteristics of a

typeface, or it may include an

embedded font file. The latter case is called an embedded font while the former is called an unembedded font. The font files that may be embedded are (integers) that map to glyphs in semantic information about the based on widely used standard the current font using an digital font formats: Type 1 (and encoding. There are several its compressed variant CFF), TrueType, and (beginning with PDF 1.6) OpenType. Additionally PDF supports the Type 3 variant in which the components of the font are described by PDF graphic operators. Fourteen typefaces, known as the standard 14 fonts, properties of the Windows and have a special significance in PDF documents: Times (v3) (in fonts using these encodings regular, italic, bold, and bold italic) Courier (in regular, oblique, bold and bold oblique) Helvetica (v3) (in regular, oblique, bold and bold oblique) Symbol Zapf Dingbats These fonts are sometimes called the base fourteen fonts. These fonts, or suitable substitute fonts with the same metrics, should be available in most PDF readers, but they are not guaranteed to be available in the reader, and may only display correctly if the system has them installed. Fonts may

be substituted if they are not strings, characters are shown using character codes WinAnsi, MacRoman, and many encodings for East Asian object drawn on the page languages and a font can have completely replaced anything its own built-in encoding. (Although the WinAnsi and MacRoman encodings are derived from the historical Macintosh operating systems, work equally well on any platform.) PDF can specify a predefined encoding to use, the by means of new extensions font's built-in encoding or provide a lookup table of differences to a predefined or built-in encoding (not recommended with TrueType fonts). The encoding mechanisms in PDF were designed for Type 1 fonts, and the rules for applying them to TrueType fonts are complex. For large fonts or fonts with non-standard glyphs, the special encodings Identity-H

(for horizontal writing) and embedded in a PDF. Within text Identity-V (for vertical) are used With such fonts, it is necessary to provide a ToUnicode table if characters is to be preserved. Transparency The original predefined encodings, including imaging model of PDF was, like PostScript's, opaque: each previously marked in the same location. In PDF 1.4 the imaging model was extended to allow transparency. When transparency is used, new objects interact with previously marked objects to produce blending effects. The addition of transparency to PDF was done that were designed to be ignored in products written to PDF 1.3 and earlier specifications. As a result, files that use a small amount of transparency might view acceptably by older viewers, but files making extensive use of transparency could be viewed incorrectly by an older viewer. The transparency extensions are based on the key concepts of transparency groups,

blending modes, shape, and alpha. The model is closely aligned with the features of Adobe Illustrator version 9. The content (text, graphics, and blend modes were based on those used by Adobe Photoshop at the time. When the PDF 1.4 specification was published, the formulas for calculating blend modes were kept secret by Adobe. They have since been published. The were relatively vague in ISO concept of a transparency group in PDF specification is independent of existing notions devices, including assistive of "group" or "layer" in applications such as Adobe Illustrator. Those groupings reflect logical relationships among objects that are meaningful when editing those objects, but they are not part of subset of PDF specifically the imaging model. Additional features Logical structure and accessibility A "tagged" PDF includes document structure and semantics information to enable reliable text extraction and accessibility. Technically speaking, tagged PDF is a stylized use of the format that builds on the logical structure framework introduced in PDF

1.3. Tagged PDF defines a set viewers. This capability is useful of standard structure types and in CAD drawings, layered attributes that allow page images) to be extracted and reused for other purposes. Tagged PDF is not required in situations where a PDF file is intended only for print. Since the feature is optional, and since the rules for Tagged PDF of information and each of 32000-1, support for tagged PDF among consuming technology (AT), is uneven as of 2021. ISO 32000-2, however, (Displayed or Suppressed) of includes an improved discussion of tagged PDF whichand signatures A PDF file may is anticipated to facilitate further be encrypted, for security, in adoption. An ISO-standardized which case a password is targeted at accessibility, PDF/UA, was first published in bit AES encryption as standard 2012. Optional Content Groups for PDF 2.0 files. The PDF (see clause 14.8 in ISO 32000) (layers) With the introduction of Reference also defines ways PDF version 1.5 (2003) came the concept of Layers. Layers, more formally known as **Optional Content Groups** (OCGs), refer to sections of content in a PDF document that details on implementing digital can be selectively viewed or

artwork, maps, multi-language documents, etc. Basically, it consists of an Optional Content Properties Dictionary added to the document root. This dictionary contains an array of **Optional Content Groups** (OCGs), each describing a set which may be individually displayed or suppressed, plus a set of Optional Content Configuration Dictionaries, which give the status the given OCGs. Encryption needed to view or edit the contents. PDF 2.0 defines 256that third parties can define their own encryption systems for PDF. PDF files may be digitally signed, to provide secure authentication; complete signatures in PDF is provided in hidden by document authors or ISO 32000-2. PDF files may

also contain embedded DRM restrictions that provide further controls that limit copying, editing, or printing. These restrictions depend on the reader software to obey them, so the security they provide is limited. The standard security provided by PDF consists of two different methods and two different passwords: a user password, which encrypts the file and prevents opening, and an owner password, which specifies operations that should (UR) signatures are used to be restricted even when the document is decrypted, which can include modifying, printing, or copying text and graphics outviewer application. The of the document, or adding or modifying text notes and AcroForm fields. The user password encrypts the file, while the owner password does be used to allow a user: To not, instead relying on client software to respect these restrictions. An owner passwordannotation data Import form can easily be removed by software, including some free online services. Thus, the use restrictions that a document author places on a PDF document are not secure, and

is distributed; this warning is displayed when applying such restrictions using Adobe Acrobat software to create or edit PDF files. Even without removing the password, most freeware or open source PDF readers ignore the permission "protections" and allow the user verifies that the signature uses to print or make copy of excerpts of the text as if the document were not limited by password protection. Beginning this same mechanism for its with PDF 1.5, Usage rights enable additional interactive features that are not available by default in a particular PDF signature is used to validate that the permissions have been the document has been signed granted by a bona fide granting by the signer. PAdES (PDF authority. For example, it can save the PDF document along with a modified form or data files in FDF, XFDF, and text (CSV/TSV) formats Export ETSI as TS 102 778. File form data files in FDF and XFDF formats Submit form data file attachments which Instantiate new pages from named page templates Apply a open or save to a local cannot be assured once the file digital signature to existing

digital signature form field Create, delete, modify, copy, import, and export annotations For example, Adobe Systems grants permissions to enable additional features in Adobe Reader, using public-key cryptography. Adobe Reader a certificate from an Adobeauthorized certificate authority. Any PDF application can use own purposes. Under specific circumstances including nonpatched systems of the receiver, the information the receiver of a digital signed document sees can be manipulated by the sender after Advanced Electronic Signatures) is a set of restrictions and extensions to PDF and ISO 32000-1 making it suitable for advanced electronic signatures. This is published by attachments PDF files can have processors may access and filesystem. Metadata PDF files

can contain two types of metadata. The first is the Document Information Dictionary, a set of key/value fields such as author, title, subject, creation and update dates. This is optional and is referenced from an Info key in the trailer of the file. A small set captions, audio descriptions, of fields is defined and can be extended with additional text values if required. This method PDFs, but this feature is not is deprecated in PDF 2.0. In PDF 1.4, support was added for Leading screen readers, Metadata Streams, using the

Extensible Metadata Platform (XMP) to add XML standardsbased extensible metadata as used in other file formats. PDF 2.0 allows metadata to be attached to any object in the document, such as information about embedded illustrations, fonts, and images, as well as the whole document (attaching an extensible schema. PDF documents can also contain display settings, including the page display layout and zoom level in a Viewer Preferences object. Adobe Reader uses

these settings to override the

user's default settings when

opening the document. The freephysical view is displayed and Adobe Reader cannot remove these settings. Accessibility PDF files can be created specifically to be accessible for readers and other assistive people with disabilities. PDF file technologies use to deliver formats in use as of 2014 can include tags, text equivalents,

and more. Some software can automatically produce tagged always enabled by default. including JAWS, Window-Eyes, view, such as the Reflow

Hal, and Kurzweil 1000 and

3000 can read tagged PDF. Moreover, tagged PDFs can be Standard for accessible PDF re-flowed and magnified for readers with visual impairments. Adding tags to older PDFs and those that are generated from scanned documents can present some challenges. One of the accessibility is that PDF documents have three distinct

document's creation, can be

inconsistent with each other.

and (iii) the content view. The

The three views are (i) the

printed (what most people consider a PDF document). The tags view is what screen high-quality navigation and reading experience to users with disabilities. The content view is based on the physical order of objects within the PDF's content stream and may be displayed by software that does not fully support the tags' feature in Adobe's Reader. PDF/UA, the International based on ISO 32000-1 was first published as ISO 14289-1 in 2012 and establishes normative language for accessible PDF technology. Multimedia Rich Media PDF is a PDF file including interactive content to the document catalog), using significant challenges with PDF that can be embedded or linked within the file. It can contain images, audio, video content, or views, which, depending on the buttons. For example, if the interactive PDF is a digital catalog for an E-commerce business, products can be listed physical view, (ii) the tags view, on the PDF pages and can be

added with images and links to

the website and buttons to order directly from the document. Forms Interactive Forms is a mechanism to add forms to the PDF file format. PDF currently supports two different methods for integrating ExportFormat, SubmitPDF, and XFDF implements only a subset data and PDF forms. Both formats today coexist in the PDF specification: AcroForms (also known as Acrobat forms), introduced in the PDF 1.2 format specification and included in all later PDF specifications. XML Forms Architecture (XFA) forms, introduced in the PDF 1.5 format specification. Adobe XFA Forms are not compatible with AcroForms. XFA was deprecated from PDF with PDF Data Format can be used when not included) in PDF 1.5 2.0. AcroForms were introducedsubmitting form data to a in the PDF 1.2 format. AcroForms permit using objects and incorporating it into the (e.g. text boxes, Radio buttons, interactive form. It can also be etc.) and some code (e.g. JavaScript). Alongside the standard PDF action types, interactive forms (AcroForms) support submitting, resetting, and importing data. The "submit" action transmits the names and values of selected interactive form fields to a

specified uniform resource locator (URL). Interactive form Version 2.0; supported since field names and values may be PDF 1.5; it replaced the "XML" submitted in any of the followingform submission format defined formats, (depending on the settings of the action's XFDF flags): HTML Form format HTML 4.01 Specification annotations. Some entries in since PDF 1.5; HTML 2.0 since the FDF dictionary do not have 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much Differences, and Target. In simpler than PDF since the body of an FDF document consists of only one required object. Forms Data Format is defined in the PDF specification using an FDF file. The XFDF (since PDF 1.2). The Forms server, receiving the response, versions). It is described used to export form data to stand-alone files that can be imported back into the corresponding PDF interactive form. FDF was originally defined in 1996 as part of ISO 32000-2:2017.[citation needed] standard. XFDF can be used in XML Forms Data Format (XFDF) (external XML Forms

Data Format Specification, in PDF 1.4) the XML version of Forms Data Format, but the of FDF containing forms and XFDF equivalents - such as the Status, Encoding, JavaScript, Page's keys, EmbeddedFDFs, addition, XFDF does not allow the spawning, or addition, of new pages based on the given data; as can be done when specification is referenced (but specification (and in later separately in XML Forms Data Format Specification. The PDF 1.4 specification allowed form submissions in XML format, but this was replaced by submissions in XFDF format in the PDF 1.5 specification. XFDF conforms to the XML the same way as FDF; e.g., form data is submitted to a

server, modifications are made, introduced a proprietary format the PostScript language, each then sent back and the new form data is imported in an interactive form. It can also be used to export form data to stand-alone files that can be imported back into the corresponding PDF interactive form. As of August 2019, XFDF is referenced from ISO 32000-3.0 is an ISO/IEC standard under the formal name ISO 19444-1:2019 - Document management — XML Forms Data Format — Part 1: Use of ISO 32000-2 (XFDF 3.0). This standard is a normative reference of ISO 32000-2. PDF Anyone may create applications besides flat text and graphics The entire document can be submitted rather than individual files without having to pay fields and values, as was defined in PDF 1.4. AcroForms Adobe holds patents to PDF, can keep form field values in external stand-alone files containing key-value pairs. The complying with its PDF external files may use Forms Data Format (FDF) and XML Forms Data Format (XFDF) files. The usage rights (UR) signatures define rights for import form data files in FDF, XFDF, and text (CSV/TSV) formats, and export form data files in FDF and XFDF formats. In PDF 1.5, Adobe Systems

for forms; Adobe XML Forms Forms are not compatible with ISO 32000's AcroForms feature, and most PDF processors do not handle XFA content. The XFA specification 1/PDF 1.7 as an external proprietary specification and was entirely deprecated from PDF with ISO 32000-2 (PDF can be splitted and merged, using applications. Licensing that can read and write PDF royalties to Adobe Systems; but licenses them for royaltyspecification. Portable Document Format (PDF),

in 1992 to present documents,

including text formatting and

independent of application

operating systems. Based on

software, hardware, and

images, in a manner

PDF file encapsulates a Architecture (XFA). Adobe XFA complete description of a fixedlayout flat document, including the text, fonts, vector graphics, raster images and other information needed to display it. PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was standardized as ISO 32000 in 2008. The last edition as ISO 2.0). Split and merge PDF files 32000-2:2020 was published in December 2020. PDF files may contain a variety of content including logical structuring elements, interactive elements such as annotations and formfields, layers, rich media (including video content), threefree use in developing software dimensional objects using U3D or PRC, and various other data formats. The PDF specification also provides for encryption and standardized as ISO 32000, is adigital signatures, file file format developed by Adobe attachments, and metadata to enable workflows requiring these features. History Main article: History of PDF Adobe Systems made the PDF specification available free of

charge in 1993. In the early

years PDF was popular mainly and indispensable for the full in desktop publishing implementation of the ISO workflows, and competed with a 32000-1 specification. These variety of formats such as DjVu, proprietary technologies are not plain text); Vector graphics for Envoy, Common Ground Digital standardized and their Paper, Farallon Replica and even Adobe's own PostScript format. PDF was a proprietary format controlled by Adobe until popular third-party it was released as an open standard on July 1, 2008, and published by the International Organization for Standardization as ISO 32000- including clarifications, 1:2008, at which time control of corrections, and critical updates a plugin for Acrobat 3.0), or any 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 thattradition of the latest PDF are necessary to make, use, sell, and distribute PDFcompliant implementations. PDF 1.7, the sixth edition of the developers and stakeholders PDF specification that became ISO 32000-1, includes some proprietary technologies definedand its sponsors made ISO only by Adobe, such as Adobe XML Forms Architecture (XFA) and JavaScript extension for Acrobat, which are referenced by ISO 32000-1 as normative

specification is published only on Adobe's website. Many of implementations of PDF. In December 2020, the second edition of PDF 2.0, ISO 32000- also support links (inside 2:2020, was published, to normative references. ISO 32000-2 does not include any proprietary technologies as normative references. ISO's publication of ISO 32000-2 in 2017 ended the 24-year specification being freely available from Adobe. Starting in April, 2023, to provide PDF with their accustomed level of access, the PDF Association 32000-2 available for download content into a single file, with at no cost. Technical details A PDF file is often a combination of vector graphics, text, and bitmap graphics. The basic

types of content in a PDF are: Typeset text stored as content streams (i.e., not encoded in illustrations and designs that consist of shapes and lines; Raster graphics for them are also not supported by photographs and other types of images Multimedia objects in the document. In later PDF revisions, a PDF document can document or web page), forms, JavaScript (initially available as other types of embedded contents that can be handled using plug-ins. PDF combines three technologies: An equivalent subset of the PostScript page description programming language but in declarative form, for generating the layout and graphics. A fontembedding/replacement system to allow fonts to travel with the documents. A structured storage system to bundle these elements and any associated data compression where appropriate. PostScript language 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 version 1.4) supports of programming languages such as if statements and loop commands. PDF is largely based on PostScript but simplified to remove flow controlstate, so instructions features like these, while graphics commands equivalent of one page can affect the to lineto remain. Historically, the appearance of any following PostScript-like PDF code is page. Therefore, all preceding generated from a source

PostScript file. The graphics commands that are output by the PostScript code are collected and tokenized.[clarification needed] document is unaffected by the Any files, graphics, or fonts to which the document refers also viewers allow the user to are collected. Then, everything is compressed to a single file. Therefore, the entire PostScript PostScript viewer needs to world (fonts, layout, measurements) remains

several advantages over PostScript: PDF contains tokenized and interpreted results of the PostScript source supports interactive 3D code, for direct correspondence documents embedded in a PDF amounts of optionally between changes to items in file: 3D drawings can be

intact.[citation needed] As a

document format, PDF has

the PDF page description and changes to the resulting page appearance. PDF (since transparent graphics; PostScriptcharacters, except for certain does not. PostScript is an interpreted programming language with an implicit global header containing a magic accompanying the description

must be processed to determine the correct appearance of a given page, whereas each page in a PDF others. As a result, PDF quickly jump to the final pages of a long document, whereas a process all pages sequentially before being able to display the ordered collections of objects destination page (unless the optional PostScript Document Structuring Conventions have been carefully compiled and included). PDF 1.6 and later

embedded using U3D or PRC and various other data formats. File format A PDF file is organized using ASCII elements that may have binary content. The file starts with a number (as a readable string) and the version of the format, for example %PDF-1.7. The format is a subset of a COS ("Carousel" Object Structure) pages in a PostScript document format. A COS tree file consists primarily of objects, of which there are nine types: Boolean values, representing true or false Real numbers Integers Strings, enclosed within parentheses ((...)) or represented as hexadecimal within single angle brackets (<...>). Strings may contain 8-bit characters. Names, starting with a forward slash (/) Arrays, enclosed within square brackets ([...]) Dictionaries, collections of objects indexed by names enclosed within double angle brackets (<<...>>) Streams, usually containing large

compressed binary data,

preceded by a dictionary and enclosed between the stream and endstream keywords. 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 ASCII format, be marked with between the obj and endobj keywords if residing in the document root. Beginning with PDF version 1.5, indirect objects (except other streams) may also be located in special streams known as object streams (marked /Type ObjStm). This technique enables non-stream objects to have standard stream filters applied to them, reduces the size of files that have large numbers of small indirect objects and is especially useful for Tagged PDF. Object streams do not support specifying an object's generation number (other than 0). An index table, also called the cross-reference table, is

located near the end of the file and gives the byte offset of each indirect object from the start of the file. This design allows for efficient random and also allows for small changes to be made without rewriting the entire file (incremental update). Before PDF version 1.5, the table would always be in a special

streams, which have the form of indirect objects in the crossa standard stream object, possibly with filters applied. Such a stream may be used instead of the ASCII crossreference table and contains the offsets and other information in binary format. The format is flexible in that it allows for integer width specification (using the /W array), so that for example, a document not exceeding 64 KiBnon-linearized (not "optimized") in size may dedicate only 2 bytes for object offsets. At the end of a PDF file is a footer

containing The startxref

optional cross-reference

objects. Version 1.5 introduced structure, also known as the

keyword followed by an offset to the start of the cross-reference table (starting with the xref keyword) or the cross-reference stream object, followed by The access to the objects in the file, %%EOF end-of-file marker. If a cross-reference stream is not being used, the footer is preceded by the trailer keyword followed by a dictionary containing information that would otherwise be contained in the cross-reference stream the xref keyword, and follow the object's dictionary: A reference main body composed of indirect to the root object of the tree

> reference table (/Size) Other optional information Within each page, there are one or multiple content streams that describe the text, vector and images being drawn on the page. The content stream is stack-based. similar to PostScript. The maximum size of a PDF compared to Europe. There are two layouts to the PDF files: and linearized ("optimized"). Non-linearized PDF files can be smaller than their linear counterparts, though they are

catalog (/Root) The count of

slower to access because portions of the data required to assemble pages of the document are scattered throughout the PDF file. Linearized PDF files (also called "optimized" or "web optimized" PDF files) are constructed in a manner that enables them to be read in a Web browser plugin without waiting for the entire file to download, since all objects required for the first page to display are optimally organized at the start of the file. PDF files may be optimized using Adobe Acrobat software or QPDF. by the format itself. However, Adobe Acrobat imposes a limit of 15 million in by 15 million in, or 225 trillion in2 (145,161 km2). Imaging model The basic lines and cubic Bézier curves, design of how graphics are represented in PDF is very similar to that of PostScript, except for the use of transparency, which was added with lines and curves. Paths in PDF 1.4. PDF graphics use a can be stroked, filled, fill then

device-independent Cartesian

coordinate system to describe

the surface of a page. A PDF

page description can use a

matrix to scale, rotate, or skew graphical elements. A key concept in PDF is that of the graphics state, which is a collection of graphical parameters that may be changed, saved, and restored by a page description. PDF has object, or an uncolored tiling (as of version 2.0) 25 graphics state properties, of which some specification to the time the of the most important are: The current transformation matrix (CTM), which determines the coordinate system The clipping continuously varying colors. path The color space The alpha There are seven types of constant, which is a key component of transparency Black point compensation Vector graphics As in PostScript, vector graphics in Paths are usually composed of stream. The dictionary but can also be constructed from the outlines of text. Unlike the image data. (Less

stroked, or used for clipping.

Strokes and fills can use any

color set in the graphics state,

including patterns. PDF

supports several types of patterns. The simplest is the tiling pattern in which a piece of artwork is specified to be drawn repeatedly. This may be a colored tiling pattern, with the colors specified in the pattern pattern, which defers color pattern is drawn. Beginning with PDF 1.3 there is also a shading pattern, which draws shading patterns of which the simplest are the axial shading (Type 2) and radial shading Page dimensions are not limitedcontrol (introduced in PDF 2.0) (Type 3). Raster images Raster images in PDF (called Image XObjects) are represented by PDF are constructed with paths.dictionaries with an associated describes the properties of the image, and the stream contains PostScript, PDF does not allow commonly, small raster images a single path to mix text outlinesmay be embedded directly in a page description as an inline image.) Images are typically

filtered for compression purposes. Image filters supported in PDF include the following general-purpose

filters: ASCII85Decode, a filter lossless bi-level (black/white) used to put the stream into 7-bit filter based on the Group 3 or ASCII, ASCIIHexDecode, similar to ASCII85Decode but less compact, FlateDecode, a commonly used filter based on the deflate algorithm defined in lossless bi-level (black/white) RFC 1951 (deflate is also used filter based on the JBIG2 in the gzip, PNG, and zip file formats among others); introduced in PDF 1.2; it can use one of two groups of predictor functions for more compact zlib/deflate compression: Predictor 2 from the TIFF 6.0 specification and predictors (filters) from the PNG in external files by the use of specification (RFC 2083), LZWDecode, a filter based on LZW Compression; it can use one of two groups of predictor functions for more compact LZW compression: Predictor 2 from the TIFF 6.0 specification and predictors (filters) from the PNG specification, RunLengthDecode, a simple compression method for streams with repetitive data using the run-length encoding algorithm and the imagespecific filters, DCTDecode, a lossy filter based on the JPEG standard, CCITTFaxDecode, a

Group 4 CCITT (ITU-T) fax compression standard defined in ITU-T T.4 and T.6, JBIG2Decode, a lossy or standard, introduced in PDF 1.4, and JPXDecode, a lossy or Additionally PDF supports the lossless filter based on the JPEG 2000 standard, introduced in PDF 1.5. Normally described by PDF graphic all image content in a PDF is embedded in the file. But PDF allows image data to be stored external streams or Alternate Images. Standardized subsets of PDF, including PDF/A and PDF/X, prohibit these features. Helvetica (v3) (in regular, Text Text in PDF is represented oblique, bold and bold oblique) by text elements in page content streams. A text element fonts are sometimes called the specifies that characters should base fourteen fonts. These be drawn at certain positions. The characters are specified using the encoding of a selected font resource. A font object in PDF is a description of guaranteed to be available in a digital typeface. It may either the reader, and may only describe the characteristics of a display correctly if the system typeface, or it may include an embedded font file. The latter

case is called an embedded font while the former is called an unembedded font. The font files that may be embedded are based on widely used standard digital font formats: Type 1 (and its compressed variant CFF), TrueType, and (beginning with PDF 1.6) OpenType. Type 3 variant in which the components of the font are operators. Fourteen typefaces, known as the standard 14 fonts, have a special significance in PDF documents: Times (v3) (in regular, italic, bold, and bold italic) Courier (in regular, oblique, bold and bold oblique) Symbol Zapf Dingbats These fonts, or suitable substitute fonts with the same metrics, should be available in most PDF readers, but they are not has them installed. Fonts may

be substituted if they are not

embedded in a PDF. Within text Identity-V (for vertical) are used.alpha. The model is closely strings, characters are shown using character codes the current font using an encoding. There are several WinAnsi, MacRoman, and many encodings for East Asian object drawn on the page languages and a font can have completely replaced anything its own built-in encoding. (Although the WinAnsi and MacRoman encodings are derived from the historical properties of the Windows and Macintosh operating systems, fonts using these encodings work equally well on any platform.) PDF can specify a predefined encoding to use, the by means of new extensions font's built-in encoding or provide a lookup table of differences to a predefined or built-in encoding (not recommended with TrueType fonts). The encoding mechanisms in PDF were designed for Type 1 fonts, and the rules for applying them to TrueType fonts are complex. For large fonts or fonts with non-standard glyphs, the special encodings Identity-H (for horizontal writing) and

With such fonts, it is necessary aligned with the features of to provide a ToUnicode table if Adobe Illustrator version 9. The (integers) that map to glyphs in semantic information about the blend modes were based on characters is to be preserved. Transparency The original predefined encodings, including imaging model of PDF was, like the PDF 1.4 specification was PostScript's, opaque: each previously marked in the same location. In PDF 1.4 the imagingconcept of a transparency model was extended to allow transparency. When transparency is used, new objects interact with previously marked objects to produce blending effects. The addition of reflect logical relationships transparency to PDF was done among objects that are that were designed to be ignored in products written to PDF 1.3 and earlier specifications. As a result, files accessibility A "tagged" PDF that use a small amount of transparency might view acceptably by older viewers, butand semantics information to files making extensive use of transparency could be viewed incorrectly by an older viewer. The transparency extensions are based on the key concepts builds on the logical structure of transparency groups, blending modes, shape, and

those used by Adobe Photoshop at the time. When published, the formulas for calculating blend modes were kept secret by Adobe. They have since been published. The group in PDF specification is independent of existing notions of "group" or "layer" in applications such as Adobe Illustrator. Those groupings meaningful when editing those objects, but they are not part of the imaging model. Additional features Logical structure and (see clause 14.8 in ISO 32000) includes document structure enable reliable text extraction and accessibility. Technically speaking, tagged PDF is a stylized use of the format that framework introduced in PDF 1.3. Tagged PDF defines a set

of standard structure types and in CAD drawings, layered attributes that allow page content (text, graphics, and images) to be extracted and reused for other purposes. Tagged PDF is not required in situations where a PDF file is intended only for print. Since the feature is optional, and since the rules for Tagged PDF of information and each of were relatively vague in ISO 32000-1, support for tagged PDF among consuming devices, including assistive technology (AT), is uneven as of 2021. ISO 32000-2, however, (Displayed or Suppressed) of includes an improved discussion of tagged PDF whichand signatures A PDF file may is anticipated to facilitate further be encrypted, for security, in adoption. An ISO-standardized which case a password is subset of PDF specifically targeted at accessibility, PDF/UA, was first published in bit AES encryption as standard while the owner password does 2012. Optional Content Groups for PDF 2.0 files. The PDF (layers) With the introduction of Reference also defines ways PDF version 1.5 (2003) came the concept of Layers. Layers, more formally known as Optional Content Groups (OCGs), refer to sections of content in a PDF document that details on implementing digital can be selectively viewed or hidden by document authors or ISO 32000-2. PDF files may

viewers. This capability is usefulalso contain embedded DRM

artwork, maps, multi-language documents, etc. Basically, it consists of an Optional Content restrictions depend on the Properties Dictionary added to the document root. This dictionary contains an array of **Optional Content Groups** (OCGs), each describing a set which may be individually displayed or suppressed, plus a file and prevents opening, and set of Optional Content Configuration Dictionaries, which give the status the given OCGs. Encryption needed to view or edit the contents. PDF 2.0 defines 256- password encrypts the file, that third parties can define their own encryption systems for PDF. PDF files may be digitally signed, to provide secure authentication; complete restrictions that a document

restrictions that provide further controls that limit copying, editing, or printing. These reader software to obey them, so the security they provide is limited. The standard security provided by PDF consists of two different methods and two different passwords: a user password, which encrypts the an owner password, which specifies operations that should be restricted even when the document is decrypted, which can include modifying, printing, or copying text and graphics out of the document, or adding or modifying text notes and AcroForm fields. The user not, instead relying on client software to respect these restrictions. An owner password can easily be removed by software, including some free online services. Thus, the use author places on a PDF signatures in PDF is provided in document are not secure, and cannot be assured once the file is distributed; this warning is

displayed when applying such restrictions using Adobe Acrobat software to create or edit PDF files. Even without removing the password, most freeware or open source PDF readers ignore the permission "protections" and allow the user verifies that the signature uses to print or make copy of excerpts of the text as if the document were not limited by password protection. Beginning this same mechanism for its with PDF 1.5, Usage rights (UR) signatures are used to enable additional interactive features that are not available by default in a particular PDF viewer application. The signature is used to validate granted by a bona fide granting by the signer. PAdES (PDF authority. For example, it can be used to allow a user: To save the PDF document along with a modified form or annotation data Import form data files in FDF, XFDF, and text (CSV/TSV) formats Export form data files in FDF and XFDF formats Submit form data file attachments which Instantiate new pages from processors may access and named page templates Apply a open or save to a local filesystem. Metadata PDF files digital signature to existing digital signature form field can contain two types of

Create, delete, modify, copy, import, and export annotations For example, Adobe Systems grants permissions to enable additional features in Adobe Reader, using public-key cryptography. Adobe Reader a certificate from an Adobeauthorized certificate authority. Any PDF application can use own purposes. Under specific circumstances including nonpatched systems of the receiver, the information the receiver of a digital signed document sees can be manipulated by the sender after 2.0 allows metadata to be that the permissions have been the document has been signed attached to any object in the **Advanced Electronic** Signatures) is a set of restrictions and extensions to PDF and ISO 32000-1 making it to the document catalog), using suitable for advanced electronic an extensible schema. PDF signatures. This is published by documents can also contain ETSI as TS 102 778. File

metadata. The first is the **Document Information** Dictionary, a set of key/value fields such as author, title, subject, creation and update dates. This is optional and is referenced from an Info key in the trailer of the file. A small set of fields is defined and can be extended with additional text values if required. This method is deprecated in PDF 2.0. In PDF 1.4, support was added for Metadata Streams, using the **Extensible Metadata Platform** (XMP) to add XML standardsbased extensible metadata as used in other file formats. PDF document, such as information about embedded illustrations, fonts, and images, as well as the whole document (attaching display settings, including the attachments PDF files can have page display layout and zoom level in a Viewer Preferences object. Adobe Reader uses these settings to override the user's default settings when opening the document. The free Adobe Reader cannot remove these settings. Accessibility PDF files can be created specifically to be accessible for readers and other assistive people with disabilities. PDF file technologies use to deliver formats in use as of 2014 can include tags, text equivalents, captions, audio descriptions, and more. Some software can automatically produce tagged PDFs, but this feature is not always enabled by default. Leading screen readers, including JAWS, Window-Eyes, view, such as the Reflow Hal, and Kurzweil 1000 and 3000 can read tagged PDF. Moreover, tagged PDFs can be Standard for accessible PDF re-flowed and magnified for readers with visual impairments. Adding tags to older PDFs and those that are generated from scanned documents can present some challenges. One of the accessibility is that PDF documents have three distinct views, which, depending on the buttons. For example, if the document's creation, can be inconsistent with each other. The three views are (i) the physical view, (ii) the tags view, on the PDF pages and can be and (iii) the content view. The physical view is displayed and the website and buttons to

printed (what most people consider a PDF document). The document. Forms Interactive tags view is what screen high-quality navigation and reading experience to users with disabilities. The content view is based on the physical order of objects within the PDF's content stream and may introduced in the PDF 1.2 be displayed by software that does not fully support the tags' feature in Adobe's Reader. PDF/UA, the International based on ISO 32000-1 was first XFA Forms are not compatible published as ISO 14289-1 in 2012 and establishes normative deprecated from PDF with PDF language for accessible PDF technology. Multimedia Rich Media PDF is a PDF file including interactive content significant challenges with PDF that can be embedded or linked etc.) and some code (e.g. within the file. It can contain images, audio, video content, orstandard PDF action types, interactive PDF is a digital catalog for an E-commerce business, products can be listed "submit" action transmits the

order directly from the Forms is a mechanism to add forms to the PDF file format. PDF currently supports two different methods for integrating data and PDF forms. Both formats today coexist in the PDF specification: AcroForms (also known as Acrobat forms), format specification and included in all later PDF specifications. XML Forms Architecture (XFA) forms, introduced in the PDF 1.5 format specification. Adobe with AcroForms. XFA was 2.0. AcroForms were introduced in the PDF 1.2 format. AcroForms permit using objects (e.g. text boxes, Radio buttons, JavaScript). Alongside the interactive forms (AcroForms) support submitting, resetting, and importing data. The names and values of selected added with images and links to interactive form fields to a

specified uniform resource

locator (URL). Interactive form Version 2.0; supported since formats, (depending on the settings of the action's XFDF flags): HTML Form format HTML 4.01 Specification annotations. Some entries in 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much Differences, and Target. In simpler than PDF since the body of an FDF document consists of only one required object. Forms Data Format is defined in the PDF specification using an FDF file. The XFDF (since PDF 1.2). The Forms Data Format can be used when not included) in PDF 1.5 submitting form data to a server, receiving the response, versions). It is described and incorporating it into the interactive form. It can also be used to export form data to stand-alone files that can be imported back into the corresponding PDF interactive form. FDF was originally defined in 1996 as part of ISO 32000-2:2017.[citation needed] standard. XFDF can be used in formats, and export form data XML Forms Data Format (XFDF) (external XML Forms Data Format Specification,

field names and values may be PDF 1.5; it replaced the "XML" form data is imported in an submitted in any of the followingform submission format defined interactive form. It can also be in PDF 1.4) the XML version of used to export form data to Forms Data Format, but the ExportFormat, SubmitPDF, and XFDF implements only a subsetimported back into the of FDF containing forms and since PDF 1.5; HTML 2.0 since the FDF dictionary do not have 3.0 is an ISO/IEC standard XFDF equivalents – such as the under the formal name ISO Status, Encoding, JavaScript,

Page's keys, EmbeddedFDFs, addition, XFDF does not allow the spawning, or addition, of new pages based on the given data; as can be done when specification is referenced (but fields and values, as was specification (and in later separately in XML Forms Data Format Specification. The PDF 1.4 specification allowed form submissions in XML format, but Forms Data Format (XFDF) this was replaced by submissions in XFDF format in signatures define rights for the PDF 1.5 specification. XFDF conforms to the XML the same way as FDF; e.g., form data is submitted to a server, modifications are made, introduced a proprietary format

then sent back and the new stand-alone files that can be corresponding PDF interactive form. As of August 2019, XFDF 19444-1:2019 - Document management — XML Forms Data Format — Part 1: Use of ISO 32000-2 (XFDF 3.0). This standard is a normative reference of ISO 32000-2. PDF The entire document can be submitted rather than individual defined in PDF 1.4. AcroForms can keep form field values in external stand-alone files

containing key-value pairs. The

external files may use Forms

Data Format (FDF) and XML

files. The usage rights (UR)

import form data files in FDF, XFDF, and text (CSV/TSV)

files in FDF and XFDF formats.

In PDF 1.5, Adobe Systems

for forms; Adobe XML Forms Architecture (XFA). Adobe XFA Document Format (PDF), Forms are not compatible with ISO 32000's AcroForms feature, and most PDF processors do not handle XFA content. The XFA specification images, in a manner is referenced from ISO 32000-1/PDF 1.7 as an external proprietary specification and was entirely deprecated from PDF with ISO 32000-2 (PDF 2.0). Split and merge PDF files can be splitted and merged, using applications. Licensing Anyone may create applications raster images and other that can read and write PDF

files without having to pay royalties to Adobe Systems; Adobe holds patents to PDF, but licenses them for royaltyfree use in developing software standardized as ISO 32000 in complying with its PDF

specification. Portable

file format developed by Adobe besides flat text and graphics in 1992 to present documents, including text formatting and independent of application software, hardware, and operating systems. Based on the PostScript language, each PDF file encapsulates a layout flat document, including

PDF has its roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was 2008. The last edition as ISO

standardized as ISO 32000, is acontain a variety of content including logical structuring elements, interactive elements such as annotations and formfields, layers, rich media (including video content), threedimensional objects using U3D or PRC, and various other data formats. The PDF specification complete description of a fixed- also provides for encryption and digital signatures, file the text, fonts, vector graphics, attachments, and metadata to enable workflows requiring information needed to display it these features. History Main article: History of PDF Adobe Systems made the PDF specification available free of charge in 1993. In the early

vears

32000-2:2020 was published in

December 2020. PDF files may