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 software, hardware, and operating systems. Based on the PostScript language, each article: History of PDF PDF file encapsulates a complete description of a fixed-layout flat document, available free of charge in XML Forms Architecture 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 Common Ground Digital co-founder John Warnock Paper, Farallon Replica in 1991. PDF was standardized as ISO 32000 in 2008. The last edition as ISO 32000-2:2020 was published in December 2020. PDF filesstandard on July 1, 2008, supported by popular may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and formfields, layers, rich media

using U3D or PRC, and various other data formats. The PDF specification also provides that are necessary to for encryption and digital make, use, sell, and signatures, file attachments, and metadata to enable workflows requiring these PDF specification that features. History Main Adobe Systems made the technologies defined only PDF specification 1993. In the early years PDF was popular mainly in desktop publishing with a variety of formats such as DjVu, Envoy, and even Adobe's own PostScript format. PDF was a proprietary format controlled by Adobe until itonly on Adobe's website. and published by the International Organization implementations of PDF. for Standardization as ISOIn December 2020, the 32000-1:2008, at which time control of the specification passed to an was published, including ISO Committee of

(including video content), In 2008, Adobe published three-dimensional objects a Public Patent License to ISO 32000-1 granting royalty-free rights for all patents owned by Adobe distribute PDF-compliant implementations. PDF 1.7, the sixth edition of the became ISO 32000-1, includes some proprietary by Adobe, such as Adobe (XFA) and JavaScript extension for Acrobat. which are referenced by workflows, and competed ISO 32000-1 as normative and indispensable for the full implementation of the ISO 32000-1 specification. These proprietary technologies are not standardized and their specification is published was released as an open Many of them are also not third-party second edition of PDF 2.0, ISO 32000-2:2020, clarifications, corrections, volunteer industry experts.and critical updates to

normative references. ISOAcrobat 3.0), or any other source PostScript file. The 32000-2 does not include types of embedded any proprietary technologies as normative handled using plug-ins. references. ISO's publication of ISO 32000- technologies: An 2 in 2017 ended the 24year tradition of the latest PostScript page 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 at no cost. Technical details A PDF file is often a combination compression where of vector graphics, text, and bitmap graphics. The language PostScript is a basic types of content in a page description languagein the PDF page PDF are: Typeset text stored as content streams generate an image, a (i.e., not encoded in plain process requiring many text); Vector graphics for resources. It can handle illustrations and designs that consist of shapes and features of programming lines; Raster graphics for photographs and other types of images Multimedia objects in the document. In later PDF revisions, a PDF document can also support links (inside document or web page), forms, JavaScript (initially PostScript-like PDF code Therefore, all preceding available as a plugin for

contents that can be PDF combines three equivalent subset of the language but in declarative form, for graphics. A fontembedding/replacement system to allow fonts to A structured storage system to bundle these elements and any associated content into a contains tokenized and single file, with data appropriate. PostScript run in an interpreter to

graphics and standard languages such as if statements and loop commands. PDF is largelylanguage with an implicit based on PostScript but simplified to remove flow control features like these, accompanying the while graphics commands description of one page equivalent to lineto remain. Historically, the is generated from a pages in a PostScript

graphics commands that are output by the PostScript code are collected and tokenized.[clarification needed] Any files, graphics, or fonts to which description programming the document refers also are collected. Then, everything is compressed generating the layout and to a single file. Therefore, the entire PostScript world (fonts, layout, measurements) remains travel with the documents.intact.[citation needed] As a document format, PDF has several advantages over PostScript: PDF interpreted results of the PostScript source code, for direct correspondence between changes to items description and changes to the resulting page appearance. PDF (since version 1.4) supports transparent graphics; PostScript does not. PostScript is an interpreted programming global state, so instructions can affect the appearance of any following page.

document must be processed to determine the correct appearance of values, representing true a given page, whereas each page in a PDF document is unaffected by within parentheses ((...)) the others. As a result, PDF viewers allow the user to quickly jump to the angle brackets (<...>). final pages of a long document, whereas a PostScript viewer needs to 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 supports interactive 3D documents embedded of optionally compressed in a PDF file: 3D drawings binary data, preceded by can be embedded using

U3D or PRC and various other data formats. File format A PDF file is organized using ASCII characters, except for certain elements that may percent sign (%). 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

of objects, of which there root. Beginning with PDF are nine types: Boolean or false Real numbers Integers Strings, enclosed located in special streams or represented as hexadecimal within single This technique enables

Strings may contain 8-bit have standard stream characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects enclosed within square brackets ([...]) Dictionaries, collections of streams do not support

objects indexed by names specifying an object's enclosed within double angle brackets (<<...>>) Streams, usually

between the stream and endstream keywords. The This design allows for null object Furthermore, introduced with the

bit characters. Objects may be either direct (embedded in another objects are numbered withbe marked with the xref an object number and a generation number and defined between the obj Structure) format. A COS and endobj keywords if tree file consists primarily residing in the document

version 1.5, indirect objects (except other streams) may also be known as object streams (marked /Type /ObjStm). non-stream objects to 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

generation number (other than 0). An index table, also called the crosscontaining large amounts reference table, is located near the end of the file and gives the byte offset a dictionary and enclosed of each indirect object from the start of the file. efficient random access to there may be comments, the objects in the file, and also allows for small changes to be made have binary content. The Comments may contain 8-without rewriting the entire file (incremental update).

> object) or indirect. Indirect in a special ASCII format, keyword, and follow the main body composed of indirect objects. Version 1.5 introduced optional cross-reference streams,

Before PDF version 1.5,

the table would always be

which have the form of a standard stream object, possibly with filters applied. Such a stream may be used instead of the ASCII cross-reference that describe the text, table and contains the offsets and other information in binary format. The format is flexible in that it allows for PostScript. The maximum for the use of integer width specification size of a PDF compared (using the /W array), so that for example, a document not exceeding 64 KiB in size may dedicate only 2 bytes for linearized ("optimized"). object offsets. At the end of a PDF file is a footer containing The startxref keyword followed by an offset to the start of the cross-reference table (starting with the xref keyword) or the crossreference stream object, followed by The %%EOF end-of-file marker. If a cross-reference stream is optimized" PDF files) are of version 2.0) 25 not being used, the footer constructed in a manner is preceded by the trailer keyword followed by a dictionary containing information that would otherwise be contained in since all objects required the coordinate system the cross-reference stream object's dictionary: are optimally organized at color space The alpha A reference to the root object of the tree structure, also known as using Adobe Acrobat the catalog (/Root) The software or QPDF. Page

the cross-reference table (/Size) Other optional information Within each page, there are one or multiple content streams vector and images being drawn on the page. The content stream is stackbased, similar to to Europe. There are two layouts to the PDF files: non-linearized (not "optimized") and Non-linearized PDF files can be smaller than their linear counterparts, access because portions A key concept in PDF is of the data required to assemble pages of the document are scattered throughout the PDF file. Linearized PDF files (also and restored by a page that enables them to be read in a Web browser plugin without waiting for transformation matrix for the first page to displayThe clipping path The the start of the file. PDF files may be optimized

by the format itself. However, 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, except transparency, which was added in PDF 1.4. PDF graphics use a deviceindependent Cartesian coordinate system to describe the surface of a page. A PDF page description can use a matrix to scale, rotate, or though they are slower to skew graphical elements. that of the graphics state, which is a collection of graphical parameters that may be changed, saved, called "optimized" or "web description. PDF has (as graphics state properties, of which some of the most important are: The current the entire file to download, (CTM), which determines constant, which is a key component of transparency Black point compensation control count of indirect objects in dimensions are not limited (introduced in PDF 2.0)

Vector graphics As in PostScript, vector graphics in PDF are constructed with paths. Paths are usually composed of lines and cubic Bézier curves, but can also be constructed from the outlines of text. Unlike PostScript, PDF does not allow a single path to mix text outlines with lines and curves. Paths can be stroked, filled, fill then stroked, or used for clipping. Strokes and fills can use any color Image filters supported in specific filters, set in the graphics state, including patterns. PDF supports several types of ASCII85Decode, a filter patterns. The simplest is the tiling pattern in which into 7-bit ASCII, a piece of artwork is specified to be drawn repeatedly. This may be a less compact, colored tiling pattern, with FlateDecode, a commonly compression standard the colors specified in the used filter based on the pattern object, or an uncolored tiling pattern, which defers color specification to the time the pattern is drawn. Beginning with PDF 1.3 there is also a shading pattern, which draws functions for more continuously varying compact zlib/deflate colors. There are seven types of shading patterns from the TIFF 6.0 of which the simplest are specification and and radial shading (Type PNG specification (RFC 3). Raster images Raster 2083), LZWDecode, a

images in PDF (called Image XObjects) are represented by dictionaries with an associated stream. The dictionary describes the properties of the image, and the stream contains the image data. (Less commonly, small raster images may be embedded directly in a page description as an inline image.) Images are repetitive data using the typically filtered for compression purposes. general-purpose filters: used to put the stream ASCIIHexDecode, similar (black/white) filter based to ASCII85Decode but deflate algorithm defined T.6, JBIG2Decode, a in RFC 1951 (deflate is also used in the gzip, PNG, and zip file formats on the JBIG2 standard, among others); introduced introduced in PDF 1.4,

run-length encoding algorithm and the image-PDF include the following DCTDecode, a lossy filter based on the JPEG standard, CCITTFaxDecode, a lossless bi-level on the Group 3 or Group 4 CCITT (ITU-T) fax defined in ITU-T T.4 and lossy or lossless bi-level (black/white) filter based in PDF 1.2; it can use one and JPXDecode, a lossy of two groups of predictor or lossless filter based on the JPEG 2000 standard. introduced in PDF 1.5. compression: Predictor 2 Normally all image content in a PDF is embedded in the file. But the axial shading (Type 2) predictors (filters) from the PDF allows image data to be stored in external files by the use of external

filter based on LZW

one of two groups of

more compact LZW

from the TIFF 6.0 specification and

PNG specification,

RunLengthDecode, a

method for streams with

simple compression

predictor functions for

compression: Predictor 2

predictors (filters) from the

Compression; it can use

streams or Alternate Images. Standardized subsets of PDF, including regular, italic, bold, and PDF/A and PDF/X, prohibit these features. Text Text in PDF is represented by text elements in page content bold and bold oblique) streams. A text element specifies that characters should be drawn at certainsometimes called the positions. The characters base fourteen fonts. are specified using the encoding of a selected font resource. A font

object in PDF is a

describe the

description of a digital

typeface. It may either

characteristics of a

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. Additionally PDF supports Asian languages and a the Type 3 variant in which the components of built-in encoding. the font are described by (Although the WinAnsi PDF graphic operators. Fourteen typefaces, known as the standard 14 from the historical

fonts, have a special

significance in PDF documents: Times (v3) (in systems, fonts using bold italic) Courier (in regular, oblique, bold and platform.) PDF can bold oblique) Helvetica (v3) (in regular, oblique, Symbol Zapf Dingbats These fonts are These fonts, or suitable

same metrics, should be available in most PDF readers, but they are not guaranteed to be available in the reader.

substitute fonts with the

and may only display typeface, or it may include correctly if the system has or fonts with non-standard

> be substituted if they are not embedded in a PDF. Within text strings, characters are shown using character codes (integers) that map to glyphs in the current font are several predefined encodings, including WinAnsi, MacRoman, and imaging model of PDF many encodings for East was, like PostScript's,

font can have its own and MacRoman

encodings are derived

and Macintosh operating these encodings work equally well on any specify a predefined encoding to use, the 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

them installed. Fonts may glyphs, the special encodings Identity-H (for horizontal writing) and Identity-V (for vertical) are used. With such fonts, it is necessary to provide a ToUnicode table if semantic information using an encoding. There about the characters is to be preserved. Transparency The original

opaque: each object drawn on the page completely replaced anything previously marked in the same location. In PDF 1.4 the imaging model was properties of the Windows extended to allow

transparency. When transparency is used, new "layer" in applications objects interact with previously marked objects Those groupings reflect to produce blending effects. The addition of transparency to PDF was meaningful when editing done by means of new extensions that were designed to be ignored in model. Additional features subset of PDF specifically products written to PDF 1.3 and earlier specifications. As a result, PDF (see clause 14.8 in 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 introduced in PDF 1.3. groups, blending modes, shape, and alpha. The model is closely aligned with the features of Adobe allow page content (text, Illustrator version 9. The blend modes were based be extracted and reused 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. Thewere relatively vague in concept of a transparency ISO 32000-1, support for

notions of "group" or such as Adobe Illustrator. uneven as of 2021. ISO logical relationships among objects that are not part of the imaging Logical structure and accessibility A "tagged" ISO 32000) includes document structure and semantics information to enable reliable text extraction and accessibility. Technically Layers, more formally speaking, tagged PDF is aknown as Optional stylized use of the format Content Groups (OCGs), that builds on the logical structure framework Tagged PDF defines a set selectively viewed or of standard structure types and attributes that graphics, and images) to 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 group in PDF specification tagged PDF among is independent of existing consuming devices,

including assistive technology (AT), is 32000-2, however, includes an improved discussion of tagged PDF which is anticipated to those objects, but they arefacilitate further adoption. An ISO-standardized targeted at accessibility, PDF/UA, was first published in 2012. **Optional Content Groups** (layers) With the introduction of PDF version 1.5 (2003) came the concept of Layers. refer to sections of content in a PDF document that can be hidden by document authors or viewers. This capability is useful in CAD drawings, layered 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 of information and each of which may be individually displayed or suppressed,

plus a set of Optional Content Configuration Dictionaries, which give the status (Displayed or Suppressed) of the given which can include OCGs. Encryption and signatures A PDF file may copying text and graphics signature is used to be encrypted, for security, out of the document, or in which case a password adding or modifying text is needed to view or edit the contents. PDF 2.0 defines 256-bit AES encryption as standard for owner password does not, allow a user: To save the PDF 2.0 files. The PDF Reference also defines ways that third parties can restrictions. An owner define their own encryption systems for PDF. PDF files may be digitally signed, to provide services. Thus, the use secure authentication; complete details on implementing digital signatures in PDF is provided in ISO 32000-2. PDF files may also contain embedded DRM restrictions that provide further controls that limit copying, editing, or printing. These restrictionsEven without removing depend on the reader software to obey them, so freeware or open source the security they provide is limited. The standard security provided by PDF and allow the user to print that the signature uses a consists of two different methods and two different of the text as if the passwords: a user password, which encrypts by password protection. the file and prevents opening, and an owner Usage rights (UR)

password, which specifies signatures are used to operations that should be enable additional restricted even when the document is decrypted, modifying, printing, or notes and AcroForm fields. The user password granting authority. For instead relying on client software to respect these a modified form or password can easily be removed by software, restrictions that a document author places on a PDF document are not secure, and cannot be templates Apply a digital assured once the file is displayed when applying such restrictions using Adobe Acrobat software to create or edit PDF files. Adobe Systems grants the password, most PDF readers ignore the permission "protections" document were not limited authority. Any PDF Beginning with PDF 1.5,

interactive features that are not available by default in a particular PDF viewer application. The validate that the permissions have been granted by a bona fide encrypts the file, while the example, it can be used to PDF document along with annotation data Import form data files in FDF, XFDF, and text including some free online (CSV/TSV) formats Export form data files in FDF and XFDF formats Submit form data Instantiate new pages from named page signature to existing distributed; this warning is digital signature form field Create, delete, modify, copy, import, and export annotations For example, permissions to enable additional features in Adobe Reader, using public-key cryptography. Adobe Reader verifies or make copy of excerpts certificate from an Adobeauthorized certificate application can use this same mechanism for its own purposes. Under

specific circumstances including non-patched systems of the receiver, the information the receiver of a digital signed XML standards-based document sees can be manipulated by the sender after the document PDF 2.0 allows metadata readers with visual has been signed by the signer. PAdES (PDF Advanced Electronic Signatures) is a set of embedded illustrations, restrictions and extensions to PDF and ISO 32000-1 making it

suitable for advanced electronic signatures. This an extensible schema. is published by ETSI as TS 102 778. File have file attachments

which processors may access and open or save object. Adobe Reader to a local 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

key in the trailer of the file.can include tags, text 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 feature is not always PDF 1.4, support was

added for Metadata Streams, using the **Extensible Metadata** Platform (XMP) to add extensible metadata as to be attached to any object in the document,

as the whole document (attaching to the document catalog), using accessibility is that PDF

PDF documents can also distinct views, which, contain display settings,

attachments PDF files canincluding the page display document's creation, can

**Viewer Preferences** uses these settings to

settings when opening the view is displayed and document. The free Adobe Reader cannot

remove these settings. Accessibility PDF files can is what screen readers be created specifically to be accessible for people

dates. This is optional and with disabilities. PDF file

equivalents, captions, audio descriptions, and more. Some software can physical order of objects automatically produce tagged PDFs, but this enabled by default.

Leading screen readers, including JAWS, Window-Eyes, Hal, and Kurzweil 1000 and 3000 can read tagged PDF. Moreover, tagged PDFs can be reused in other file formats. flowed and magnified for impairments. Adding tags to older PDFs and those such as information about that are generated from scanned documents can fonts, and images, as well present some challenges. One of the significant challenges with PDF documents have three depending on the layout and zoom level in a be inconsistent with each

other. The three views are (i) the physical view, (ii) the tags view, and (iii) the override the user's default content view. The physical printed (what most people consider a PDF

document). The tags view and other assistive technologies use to deliver high-quality is referenced from an Info formats in use as of 2014 navigation and reading experience to users with disabilities. The content view is based on the 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 Architecture (XFA) forms, FDF document consists of Reader. PDF/UA, the International Standard for format specification. accessible PDF based on Adobe XFA Forms are not defined in the PDF 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 that can be embedded or linked within standard PDF action the file. It can contain images, audio, video content, or buttons. For example, if the interactive importing data. The PDF is a digital catalog for "submit" action transmits an E-commerce business, the names and values of products can be listed on selected interactive form the PDF pages and can be added with images and uniform resource locator links to the website and buttons to order directly from the document. Formsmay be submitted in any Interactive 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 HTML 4.01 Specification coexist in the PDF specification: AcroForms since 1.2 Forms Data (also known as Acrobat forms), introduced in the PDF, uses the same PDF 1.2 format specification and included the same file structure, in all later PDF

specifications. XML FormsPDF since the body of an introduced in the PDF 1.5 only one required object. compatible with AcroForms. XFA was deprecated from PDF with Format can be used when PDF 2.0. AcroForms were submitting form data to a introduced in the PDF 1.2 server, receiving the format. AcroForms permit response, and using objects (e.g. text boxes, Radio buttons, etc.) and some code (e.g. also be used to export JavaScript). Alongside the form data to stand-alone types, interactive forms (AcroForms) support submitting, resetting, and interactive form. FDF was fields to a specified (URL). Interactive form field names and values of the following formats, (depending on the settings of the action's ExportFormat, SubmitPDF, and XFDF flags): HTML Form format of FDF containing forms since PDF 1.5; HTML 2.0 entries in the FDF Format (FDF) based on

files that can be imported back into the corresponding PDF originally defined in 1996 as part of ISO 32000-2:2017.[citation needed] **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 in PDF 1.4) the XML version of Forms Data Format, but the XFDF implements only a subset and annotations. Some dictionary do not have XFDF equivalents – such as the Status, Encoding, syntax and has essentially JavaScript, Page's keys, EmbeddedFDFs, but is much simpler than Differences, and Target.

Forms Data Format is

specification (since PDF

1.2). The Forms Data

incorporating it into the

interactive form. It can

In addition, XFDF does not allow the spawning, or This standard is a 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 keep form field values in versions). It is described separately in XML Forms containing key-value **Data Format** Specification. The PDF 1.4 specification allowed form submissions in XML Forms Data Format format, but this was replaced by submissions in XFDF format in the PDF 1.5 specification. XFDF conforms to the XML standard. XFDF can (CSV/TSV) formats, and be used in the same way export form data files in as FDF; e.g., form data is FDF and XFDF formats. submitted to a server. modifications are made, then sent back and the new form data is imported forms; Adobe XML Forms PostScript language, each in an interactive form. It can also be used to exportXFA Forms are not form data to stand-alone files that can be imported 32000's AcroForms back into the feature, and most PDF corresponding PDF XFA content. The XFA interactive form. As of August 2019, XFDF 3.0 is specification is referenced display it. PDF has its an ISO/IEC standard from ISO 32000-1/PDF under the formal name 1.7 as an external ISO 19444-1:2019 proprietary specification Document management and was entirely XML Forms Data Format — Part 1: Use of ISO 32000-2 (PDF 2.0). 32000 in 2008. The last

ISO 32000-2 (XFDF 3.0). Split and merge PDF files normative reference of ISO 32000-2. PDF The entire document can be submitted rather than individual fields and values, as was defined in PDF 1.4. AcroForms can external stand-alone files pairs. The external files may use Forms Data Format (FDF) and XML (XFDF) files. The usage rights (UR) signatures define rights for import form data files in FDF, XFDF, and text In PDF 1.5, Adobe Systems introduced a proprietary format for compatible with ISO

can be splitted and merged, using applications. Licensing Anyone may create applications 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 complying with its PDF 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 software, hardware, and operating systems. Based on the Architecture (XFA). Adobe PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, vector graphics, raster processors do not handle images and other information needed to roots in "The Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was deprecated from PDF with standardized as ISO

edition as ISO 32000-International Organization was published, including 2:2020 was published in December 2020. PDF files 32000-1:2008, at which may contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements such as annotations and formfields, layers, rich media (including video content), patents owned by Adobe PDF specification being three-dimensional objects that are necessary to using U3D or PRC, and various other data formats. The PDF specification also provides 1.7, the sixth edition of the stakeholders with their for encryption and digital signatures, file attachments, and metadata to enable workflows requiring these by Adobe, such as Adobe 32000-2 available for features. History Main article: History of PDF Adobe Systems made the extension for Acrobat, PDF specification available free of charge in ISO 32000-1 as normative and bitmap graphics. The 1993. In the early years PDF was popular mainly in desktop publishing workflows, and competed These proprietary with a variety of formats such as DjVu, Envoy, Common Ground Digital Paper, Farallon Replica and even Adobe's own PostScript format. PDF was a proprietary format controlled by Adobe until itimplementations of PDF. was released as an open In December 2020, the standard on July 1, 2008, second edition of PDF and published by the

for Standardization as ISOclarifications, corrections, time control of the specification passed to an 32000-2 does not include ISO Committee of volunteer industry experts.technologies as normative In 2008, Adobe published references, ISO's a Public Patent License to publication of ISO 32000-ISO 32000-1 granting royalty-free rights for all make, use, sell, and distribute PDF-compliant 2023, to provide PDF implementations. PDF PDF specification that became ISO 32000-1. includes some proprietary Association and its technologies defined only sponsors made ISO XML Forms Architecture (XFA) and JavaScript which are referenced by and indispensable for the basic types of content in a full implementation of the PDF are: Typeset text ISO 32000-1 specification.stored as content streams technologies are not standardized and their specification is published that consist of shapes and only on Adobe's website. Many of them are also not photographs and other supported by popular third-party 2.0, ISO 32000-2:2020,

and critical updates to normative references. ISO any proprietary 2 in 2017 ended the 24year tradition of the latest freely available from Adobe. Starting in April, developers and accustomed level of access, the PDF download at no cost. Technical details A PDF file is often a combination of vector graphics, text, (i.e., not encoded in plain text); Vector graphics for illustrations and designs lines; Raster graphics for types of images Multimedia objects in the document. In later PDF revisions, a PDF document can also support links (inside

document or web page), forms, JavaScript (initially PostScript-like PDF code Therefore, all preceding available as a plugin for Acrobat 3.0), or any other source PostScript file. The document must be 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 to a single file. Therefore, PostScript viewer needs graphics. A fontembedding/replacement system to allow fonts to travel with the documents intact.[citation needed] As destination page (unless A structured storage system to bundle these elements and any associated content into a single file, with data compression where appropriate. PostScript language PostScript is a page description languagein the PDF page 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 largelylanguage with an implicit based on PostScript but simplified to remove flow instructions control features like these, accompanying the while graphics commands description of one page

equivalent to lineto

remain. Historically, the is generated from a graphics commands that are output by the PostScript code are collected and tokenized.[clarification needed] Any files, graphics, or fonts to which PDF viewers allow the are collected. Then, everything is compressed document, whereas a

(fonts, layout, measurements) remains a document format, PDF the optional PostScript has several advantages over PostScript: PDF contains tokenized and interpreted results of the PostScript source code,

description and changes to the resulting page appearance. PDF (since version 1.4) supports transparent graphics; PostScript does not. PostScript is an interpreted programming global state, so can affect the appearance a subset of a COS

of any following page. pages in a PostScript 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, the document refers also user to quickly jump to the final pages of a long the entire PostScript worldto process all pages sequentially before being able to display the **Document Structuring** Conventions have been carefully compiled and included). PDF 1.6 and later supports interactive for direct correspondence 3D documents embedded between changes to items in a PDF file: 3D drawings

can be embedded using U3D or PRC and various other data formats. 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

("Carousel" Object Structure) format. A COS and endobj keywords if tree file consists primarily residing in the document of objects, of which there root. Beginning with PDF are nine types: Boolean values, representing true or false Real numbers Integers Strings, enclosed located in special streams may be used instead of within parentheses ((...)) or represented as hexadecimal within single This technique enables angle brackets (<...>). Strings may contain 8-bit have standard stream characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects enclosed within square brackets ([...]) Dictionaries, collections of streams do not support

enclosed within double angle brackets (<<...>>) Streams, usually of optionally compressed near the end of the file binary data, preceded by a dictionary and enclosed of each indirect object between the stream and endstream keywords. The This design allows for null object Furthermore, introduced with the percent sign (%).

bit characters. Objects may be either direct (embedded in another objects are numbered withbe marked with the xref an object number and a

generation number and

defined between the obj version 1.5, indirect objects (except other streams) may also be

known as object streams (marked /Type /ObjStm). non-stream objects to filters applied to them, reduces the size of files that have large numbers of small indirect objects and is especially useful fordocument not exceeding Tagged PDF. Object

objects indexed by names specifying an object's generation number (other of a PDF file is a footer than 0). An index table, also called the crosscontaining large amounts reference table, is located offset to the start of the and gives the byte offset from the start of the file. efficient random access to end-of-file marker. If a also allows for small changes to be made

Comments may contain 8-without rewriting the entirekeyword followed by a file (incremental update). Before PDF version 1.5, the table would always be otherwise be contained in object) or indirect. Indirect in a special ASCII format, the cross-reference 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 the ASCII cross-reference 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

64 KiB in size may dedicate only 2 bytes for object offsets. At the end containing The startxref keyword followed by an cross-reference table (starting with the xref keyword) or the crossreference stream object, followed by The %%EOF there may be comments, the objects in the file, and cross-reference stream is not being used, the footer is preceded by the trailer

dictionary containing information that would stream object's dictionary: A reference to the root object of the tree

structure, also known as the catalog (/Root) The the cross-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 stackbased, similar to PostScript. The maximum for the use of size of a PDF compared to Europe. There are two layouts to the PDF files: non-linearized (not "optimized") and linearized ("optimized"). Non-linearized PDF files can be smaller than their linear counterparts, access because portions of the data required to assemble pages of the document are scattered throughout the PDF file. Linearized PDF files (also and restored by a page called "optimized" or "web description. PDF has (as optimized" PDF files) are of version 2.0) 25 constructed in a manner that enables them to be read in a Web browser plugin without waiting for since all objects required the coordinate system for the first page to displayThe clipping path The are optimally organized at color space The alpha the start of the file. PDF

files may be optimized

using Adobe Acrobat software or QPDF. Page count of indirect objects in dimensions are not limited (introduced in PDF 2.0) by the format itself. However, Adobe Acrobat PostScript, vector imposes a limit of 15 million in by 15 million in, constructed with paths. or 225 trillion in 2 (145,161 Paths are usually km2). Imaging model The composed of lines and basic design of how graphics are represented can also be constructed in PDF is very similar to that of PostScript, except Unlike PostScript, PDF transparency, which was added in PDF 1.4. PDF graphics use a deviceindependent Cartesian coordinate system to page. A PDF page description can use a matrix to scale, rotate, or though they are slower to skew graphical elements. patterns. The simplest is A key concept in PDF is that of the graphics state, a piece of artwork is which is a collection of may be changed, saved, graphics state properties, which defers color of which some of the most specification to the time important are: The current the pattern is drawn. transformation matrix the entire file to download, (CTM), which determines there is also a shading constant, which is a key

component of

transparency Black point compensation control Vector graphics As in graphics in PDF are cubic Bézier curves, but from the outlines of text. does not allow a single path to mix text outlines with lines and curves. Paths can be stroked, filled, fill then stroked, or used for clipping. Strokes describe the surface of a and fills can use any color set in the graphics state, including patterns. PDF supports several types of the tiling pattern in which specified to be drawn graphical parameters that repeatedly. This may be a colored tiling pattern, with the colors specified in the pattern object, or an uncolored tiling pattern, Beginning with PDF 1.3 pattern, which draws continuously varying colors. There are seven types of shading patterns

of which the simplest are

the axial shading (Type 2) predictors (filters) from the PDF allows image data to and radial shading (Type PNG specification (RFC 3). Raster images Raster 2083), LZWDecode, a images in PDF (called Image XObjects) are represented by dictionaries with an associated stream. The dictionary describes the properties of the image, and the stream contains the image data. (Less commonly, small raster images may be embedded directly in a page description as an inline image.) Images are repetitive data using the typically filtered for compression purposes. Image filters supported in specific filters, general-purpose filters: ASCII85Decode, a filter used to put the stream into 7-bit ASCII, ASCIIHexDecode, similar (black/white) filter based to ASCII85Decode but less compact,

used filter based on the deflate algorithm defined T.6, JBIG2Decode, a in RFC 1951 (deflate is lossy or lossless bi-level also used in the gzip, (black/white) filter based PNG, and zip file formats on the JBIG2 standard, among others); introducedintroduced in PDF 1.4, in PDF 1.2; it can use one and JPXDecode, a lossy

functions for more compact zlib/deflate compression: Predictor 2 Normally all image from the TIFF 6.0 specification and

filter based on LZW Compression; it can use one of two groups of predictor functions for more compact LZW compression: Predictor 2 Text Text in PDF is from the TIFF 6.0 specification and predictors (filters) from the streams. A text element PNG specification,

RunLengthDecode, a simple compression method for streams with run-length encoding algorithm and the image-PDF include the following DCTDecode, a lossy filter typeface. It may either based on the JPEG standard. CCITTFaxDecode, a lossless bi-level on the Group 3 or Group 4an embedded font while CCITT (ITU-T) fax FlateDecode, a commonly compression standard defined in ITU-T T.4 and

of two groups of predictor or lossless filter based on 1.6) OpenType. introduced in PDF 1.5. content in a PDF is embedded in the file. But PDF graphic operators.

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. represented by text elements in page content 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 describe the characteristics of a typeface, or it may include an embedded font file. The latter case is called 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 the JPEG 2000 standard, Additionally PDF supports the Type 3 variant in which the components of the font are described by

Fourteen typefaces, location. In PDF 1.4 the encodings are derived known as the standard 14 from the historical imaging model was properties of the Windows extended to allow fonts, have a special significance in PDF and Macintosh operating transparency. When documents: Times (v3) (in systems, fonts using transparency is used, new regular, italic, bold, and these encodings work objects interact with bold italic) Courier (in equally well on any previously marked objects regular, oblique, bold and platform.) PDF can to produce blending bold oblique) Helvetica specify a predefined effects. The addition of (v3) (in regular, oblique, encoding to use, the font's transparency to PDF was bold and bold oblique) built-in encoding or done by means of new Symbol Zapf Dingbats extensions that were provide a lookup table of designed to be ignored in These fonts are differences to a sometimes called the predefined or built-in products written to PDF base fourteen fonts. encoding (not 1.3 and earlier These fonts, or suitable recommended with specifications. As a result, substitute fonts with the TrueType fonts). The files that use a small same metrics, should be encoding mechanisms in amount of transparency available in most PDF PDF were designed for might view acceptably by Type 1 fonts, and the readers, but they are not older viewers, but files rules for applying them to making extensive use of guaranteed to be available in the reader, TrueType fonts are transparency could be and may only display complex. For large fonts viewed incorrectly by an correctly if the system has or fonts with non-standard older viewer. The them installed. Fonts may glyphs, the special transparency extensions be substituted if they are encodings Identity-H (for are based on the key not embedded in a PDF. horizontal writing) and concepts of transparency Within text strings, Identity-V (for vertical) are groups, blending modes, characters are shown used. With such fonts, it is shape, and alpha. The using character codes necessary to provide a model is closely aligned (integers) that map to ToUnicode table if with the features of Adobe glyphs in the current font semantic information Illustrator version 9. The using an encoding. There about the characters is to blend modes were based are several predefined be preserved. on those used by Adobe encodings, including Transparency The original Photoshop at the time. WinAnsi, MacRoman, and imaging model of PDF When the PDF 1.4 many encodings for East was, like PostScript's, specification was

opaque: each object

drawn on the page

completely replaced

anything previously

marked in the same

published, the formulas

modes were kept secret

since been published. The

for calculating blend

by Adobe. They have

Asian languages and a

font can have its own

(Although the WinAnsi

built-in encoding.

and MacRoman

concept of a transparency extraction and rules for Tagged PDF group in PDF specification accessibility. Technically were relatively vague in is independent of existing speaking, tagged PDF is aISO 32000-1, support for notions of "group" or stylized use of the format tagged PDF among "layer" in applications that builds on the logical consuming devices, such as Adobe Illustrator. structure framework including assistive Those groupings reflect introduced in PDF 1.3. technology (AT), is logical relationships Tagged PDF defines a setuneven as of 2021. ISO among objects that are of standard structure 32000-2, however, meaningful when editing types and attributes that includes an improved those objects, but they areallow page content (text, discussion of tagged PDF not part of the imaging graphics, and images) to which is anticipated to model. Additional features be extracted and reused facilitate further adoption. Logical structure and for other purposes. An ISO-standardized accessibility A "tagged" Tagged PDF is not subset of PDF specifically PDF (see clause 14.8 in required in situations targeted at accessibility, ISO 32000) includes where a PDF file is PDF/UA, was first document structure and intended only for print. published in 2012. semantics information to Since the feature is **Optional Content** enable reliable text optional, and since the

**ELEMENT BELOW**