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 these features. History Main 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. Paper, Farallon Replica and 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 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), threedimensional objects using U3D or PRC, and various other data

formats. The PDF specification sell, and distribute PDFalso provides for encryption and compliant implementations.

digital signatures, file attachments, and metadata to enable workflows requiring article: History of PDF Adobe Systems made the PDF specification available free of charge in 1993. In the early years PDF was popular mainly in desktop publishing workflows, and competed with aimplementation of the ISO variety of formats such as DjVu, 32000-1 specification. These

even Adobe's own PostScript format. PDF was a proprietary it was released as an open standard on July 1, 2008, and published by the International Organization for Standardization as ISO 32000- 2:2020, was published,

1:2008, at which time control of including clarifications, 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 that 2017 ended the 24-year

are necessary to make, use,

PDF 1.7, the sixth edition of the PDF specification that became ISO 32000-1, includes some proprietary technologies defined only by Adobe, such as Adobe XML Forms Architecture (XFA) and JavaScript extension for Acrobat, which are referenced by ISO 32000-1 as normative and indispensable for the full Envoy, Common Ground Digital proprietary technologies are not

standardized and their specification is published only on Adobe's website. Many of format controlled by Adobe until them are also not supported by popular third-party implementations of PDF. In December 2020, the second edition of PDF 2.0, ISO 32000corrections, and critical updates

> to normative references. ISO 32000-2 does not include any proprietary technologies as normative references. ISO's publication of ISO 32000-2 in

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 graphics and standard features appearance of any following 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 commands that are output by images Multimedia objects in the document. In later PDF revisions, a PDF document can tokenized.[clarification needed] destination page (unless the also support links (inside document or web page), forms, which the document refers also Structuring Conventions have JavaScript (initially available as are collected. Then, everything been carefully compiled and a plugin for Acrobat 3.0), or any is compressed to a single file. 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 tokenized and interpreted the layout and graphics. A font- results of the PostScript source elements that may have binary embedding/replacement system code, for direct correspondence content. The file starts with a to allow fonts to travel with the documents. A structured storage system to bundle these changes to the resulting page elements and any associated content into a single file, with

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 of programming languages such as if statements and loop commands. PDF is largely based on PostScript but simplified to remove flow controlappearance of a given page, features like these, while graphics commands equivalent document is unaffected by the to lineto remain. Historically, theothers. As a result, PDF PostScript-like PDF code is generated from a source PostScript file. The graphics the PostScript code are collected and Any files, graphics, or fonts to

Therefore, the entire PostScript supports interactive 3D world (fonts, layout, measurements) remains intact.[citation needed] As a document format, PDF has several advantages over PostScript: PDF contains

between changes to items in the PDF page description and appearance. PDF (since version 1.4) supports

transparent graphics; PostScript does not. PostScript is an interpreted programming language with an implicit global state, so instructions accompanying the description of one page can affect the page. Therefore, all preceding pages in a PostScript document must be processed to determine the correct whereas each page in a PDF viewers allow the user to

quickly jump to the final pages

of a long document, whereas a

process all pages sequentially

PostScript viewer needs to

before being able to display the optional PostScript Document included). PDF 1.6 and later documents embedded 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

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 /ObjStm). This technique 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-bitspecifying an object's characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects enclosed within square bracketslocated near the end of the file

objects indexed by names enclosed within double angle brackets (<<...>>) Streams, usually containing large amounts of optionally 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 Such a stream may be used 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 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 generation number (other than 0). An index table, also called the cross-reference table, is

([...]) Dictionaries, collections of 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 ASCII format, be marked with the xref keyword, and follow the content streams that describe

> a standard stream object, possibly with filters applied. 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

optional cross-reference

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 to the start of the cross-reference table (starting with the xref keyword) or the cross-reference stream object, followed by The %%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 access to the objects in the file, object's dictionary: A reference 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 each page, there are one or multiple main body composed of indirect the text, vector and images objects. Version 1.5 introduced being drawn on the page. The content stream is stack-based. streams, which have the form of similar to PostScript. The maximum 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, though they are 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 Paths are usually composed of filtered for compression at the start of the file. PDF files may be optimized using Adobe but can also be constructed Acrobat software or QPDF. Page dimensions are not limitedPostScript, PDF does not allow filters: ASCII85Decode, a filter by the format itself. However, Adobe Acrobat imposes a limit with lines and curves. Paths of 15 million in by 15 million in, or 225 trillion in2 (145,161 km2). Imaging model The basic Strokes and fills can use any design of how graphics are represented in PDF is very similar to that of PostScript, except for the use of transparency, which was added tiling pattern in which a piece of introduced in PDF 1.2; it can device-independent Cartesian coordinate system to describe the surface of a page. A PDF

graphical elements. A key

graphics state, which is a

collection of graphical

parameters that may be

page description can use a matrix to scale, rotate, or skew concept in PDF is that of the changed, saved, and restored (as of version 2.0) 25 graphics simplest are the axial shading

state properties, of which some (Type 2) and radial shading

of the most important are: The current transformation matrix (CTM), which determines the path The color space The alpha stream. The dictionary constant, which is a key component of transparency Black point compensation Vector graphics As in PostScript, vector graphics in PDF are constructed with paths.image.) Images are typically lines and cubic Bézier curves, from the outlines of text. Unlike following general-purpose a single path to mix text outlinesused to put the stream into 7-bit

can be stroked, filled, fill then stroked, or used for clipping. color set in the graphics state, including patterns. PDF supports several types of patterns. The simplest is the

in PDF 1.4. PDF graphics use a artwork is specified to be drawn use one of two groups of

repeatedly. This may be a colored tiling pattern, with the colors specified in the pattern object, or an uncolored tiling pattern, which defers color specification to the time the pattern is drawn. Beginning with LZWD ecode, a filter based on PDF 1.3 there is also a shading LZW Compression; it can use pattern, which draws continuously varying colors. There are seven types of by a page description. PDF has shading patterns of which the

images in PDF (called Image XObjects) are represented by coordinate system The clipping dictionaries with an associated describes the properties of the image, and the stream contains the image data. (Less control (introduced in PDF 2.0) commonly, small raster images may be embedded directly in a page description as an inline purposes. Image filters supported in PDF include the

(Type 3). Raster images Raster

ASCII, ASCIIHexDecode, similar to ASCII85Decode but 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),

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 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 regular, italic, bold, and bold lossless filter based on the JPEG 2000 standard, introduced in PDF 1.5. Normally Helvetica (v3) (in regular, 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. Text Text in PDF is represented guaranteed to be available in by text elements in page content streams. A text element display correctly if the system specifies that characters should has them installed. Fonts may 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 (integers) that map to glyphs in location. In PDF 1.4 the imaging a digital typeface. It may either the current font using an

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 (Although the WinAnsi and based on widely used standard MacRoman encodings are digital font formats: Type 1 (and derived from the historical 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, differences to a predefined or have a special significance in PDF documents: Times (v3) (in recommended with TrueType italic) Courier (in regular, oblique, bold and bold oblique) designed for Type 1 fonts, and oblique, bold and bold oblique) TrueType fonts are complex. 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 the reader, and may only be substituted if they are not embedded in a PDF. Within text object drawn on the page strings, characters are shown using character codes describe the characteristics of a encoding. There are several

WinAnsi, MacRoman, and

many encodings for East Asian languages and a font can have its own built-in encoding. 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 font's built-in encoding or provide a lookup table of built-in encoding (not fonts). The encoding mechanisms in PDF were the rules for applying them to For large fonts or fonts with non-standard 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 about the characters is to be preserved. Transparency The original imaging model of PDF was, like PostScript's, opaque: each completely replaced anything previously marked in the same model was extended to allow transparency. When predefined encodings, including transparency is used, new objects interact with previously

marked objects to produce blending effects. The addition of includes document structure transparency to PDF was done and semantics information to 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, butof standard structure types and (OCGs), each describing a set 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 were relatively vague in ISO 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 is anticipated to facilitate further their own encryption systems concept of a transparency group in PDF specification is independent of existing notions targeted at accessibility, of "group" or "layer" in applications such as Adobe Illustrator. Those groupings reflect logical relationships among objects that are meaningful when editing those more formally known as objects, but they are not part of Optional Content Groups the imaging model. Additional (OCGs), refer to sections of features Logical structure and accessibility A "tagged" PDF

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 Optional Content Groups 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 and signatures A PDF file may 32000-1, support for tagged PDF among consuming devices, including assistive technology (AT), is uneven as of 2021. ISO 32000-2, however, for PDF 2.0 files. The PDF includes an improved discussion of tagged PDF whichthat third parties can define adoption. An ISO-standardized for PDF. PDF files may be subset of PDF specifically PDF/UA, was first published in details on implementing digital (layers) With the introduction of ISO 32000-2. PDF files may PDF version 1.5 (2003) came the concept of Layers. Layers,

can be selectively viewed or

(see clause 14.8 in ISO 32000) 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 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 OCGs. Encryption be encrypted, for security, in which case a password is needed to view or edit the contents. PDF 2.0 defines 256bit AES encryption as standard Reference also defines ways digitally signed, to provide secure authentication; complete 2012. Optional Content Groups signatures in PDF is provided in also contain embedded DRM restrictions that provide further controls that limit copying, editing, or printing. These restrictions depend on the content in a PDF document that reader software to obey them, so the security they provide is

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 authority. For example, it can be restricted even when the document is decrypted, which can include modifying, printing, with a modified form or or copying text and graphics outannotation data Import form of the document, or adding or modifying text notes and AcroForm fields. The user password encrypts the file, while the owner password does Instantiate new pages from not, instead relying on client software to respect these restrictions. An owner passworddigital signature form field 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 cannot be assured once the file cryptography. Adobe Reader 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 receiver, the information the to print or make copy of excerpts of the text as if the document were not limited by password protection. Beginning the document has been signed with PDF 1.5, Usage rights (UR) signatures are used to Advanced Electronic

limited. The standard security

enable additional interactive features that are not available by default in a particular PDF viewer application. The signature is used to validate that the permissions have been ETSI as TS 102 778. File be used to allow a user: To save the PDF document along data files in FDF, XFDF, and text (CSV/TSV) formats Export Document Information form data files in FDF and XFDF formats Submit form data fields such as author, title, named page templates Apply a dates. This is optional and is digital signature to existing Create, delete, modify, copy, import, and export annotations For example, Adobe Systems grants permissions to enable additional features in Adobe Reader, using public-key verifies that the signature uses a certificate from an Adobeauthorized certificate authority. used in other file formats. PDF Any PDF application can use this same mechanism for its own purposes. Under specific circumstances including nonpatched systems of the receiver of a digital signed document sees can be manipulated by the sender after documents can also contain by the signer. PAdES (PDF

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 granted by a bona fide granting attachments PDF files can have file attachments which processors may access and open or save to a local filesystem. Metadata PDF files can contain two types of metadata. The first is the Dictionary, a set of key/value subject, creation and update 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 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 to the document catalog), using an extensible schema. PDF 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 free with disabilities. The content Adobe Reader cannot remove these settings. Accessibility PDF files can be created specifically to be accessible for be displayed by software that people with disabilities. PDF file does not fully support the tags' 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, technology. Multimedia Rich Hal, and Kurzweil 1000 and 3000 can read tagged PDF. Moreover, tagged PDFs can be that can be embedded or linked interactive form fields to a 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 significant challenges with PDF added with images and links to XFDF flags): HTML Form accessibility is that PDF documents have three distinct views, which, depending on the document. Forms Interactive document's creation, can be inconsistent with each other. The three views are (i) the physical view, (ii) the tags view, different methods for integrating simpler than PDF since the and (iii) the content view. The physical view is displayed and printed (what most people consider a PDF document). The (also known as Acrobat forms), defined in the PDF specification tags view is what screen readers and other assistive

technologies use to deliver high-quality navigation and reading experience to users view is based on the physical order of objects within the PDF's content stream and may with AcroForms. XFA was 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 JavaScript). Alongside the published as ISO 14289-1 in 2012 and establishes normative interactive forms (AcroForms) language for accessible PDF Media PDF is a PDF file including interactive content within the file. It can contain images, audio, video content, orlocator (URL). Interactive form buttons. For example, if the interactive PDF is a digital catalog for an E-commerce business, products can be listedsettings of the action's on the PDF pages and can be 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

specifications. XML Forms Architecture (XFA) forms, introduced in the PDF 1.5 format specification. Adobe XFA Forms are not compatible deprecated from PDF with PDF 2.0. AcroForms were introduced in the PDF 1.2 format. AcroForms permit using objects (e.g. text boxes, Radio buttons, etc.) and some code (e.g. standard PDF action types, support submitting, resetting, and importing data. The "submit" action transmits the names and values of selected specified uniform resource field names and values may be submitted in any of the following formats, (depending on the ExportFormat, SubmitPDF, and format HTML 4.01 Specification since PDF 1.5; HTML 2.0 since 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much body of an FDF document consists of only one required object. Forms Data Format is (since PDF 1.2). The Forms Data Format can be used when

included in all later PDF

submitting form data to a server, receiving the response, XFDF conforms to the XML 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] 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 19444-1:2019 - Document in PDF 1.4) the XML version of management — XML Forms Forms Data Format, but the XFDF implements only a subsetISO 32000-2 (XFDF 3.0). This of FDF containing forms and annotations. Some entries in the FDF dictionary do not have The entire document can be XFDF equivalents – such as the submitted rather than individual standardized as ISO 32000, is a 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 external files may use Forms data; as can be done when using an FDF file. The XFDF specification is referenced (but files. The usage rights (UR) not included) in PDF 1.5 specification (and in later versions). It is described separately in XML Forms Data formats, and export form data Format Specification. The PDF files in FDF and XFDF formats. information needed to display it. 1.4 specification allowed form submissions in XML format, but introduced a proprietary format Camelot Project" initiated by this was replaced by submissions in XFDF format in Architecture (XFA). Adobe XFA Warnock in 1991. PDF was

the PDF 1.5 specification. standard. XFDF can be used in feature, and most PDF the same way as FDF; e.g., form data is submitted to a server, modifications are made, is referenced from ISO 32000then sent back and the new form data is imported in an interactive form. It can also be used to export form data to imported back into the corresponding PDF interactive form. As of August 2019, XFDF Anyone may create applications 3.0 is an ISO/IEC standard under the formal name ISO Data Format — Part 1: Use of standard is a normative reference of ISO 32000-2. PDF specification. Portable fields and values, as was defined in PDF 1.4. AcroForms in 1992 to present documents, can keep form field values in external stand-alone files containing key-value pairs. The independent of application Data Format (FDF) and XML Forms Data Format (XFDF) signatures define rights for import form data files in FDF, XFDF, and text (CSV/TSV) In PDF 1.5, Adobe Systems for forms; Adobe XML Forms

Forms are not compatible with ISO 32000's AcroForms 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 2.0). Split and merge PDF files can be splitted and merged, using applications. Licensing 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 Document Format (PDF), file format developed by Adobe including text formatting and images, in a manner 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 PDF has its roots in "The Adobe co-founder John

standardized as ISO 32000 in 2008. The last edition as ISO 32000-2:2020 was published in all patents owned by Adobe that32000-2 available for download December 2020. PDF files may are necessary to make, use, contain a variety of content besides flat text and graphics including logical structuring elements, interactive elements PDF specification that became such as annotations and formfields, layers, rich media (including video content), three- only by Adobe, such as Adobe dimensional objects using U3D XML Forms Architecture (XFA) illustrations and designs that or PRC, and various other data and JavaScript extension for formats. The PDF specification Acrobat, which are referenced also provides for encryption andby ISO 32000-1 as normative digital signatures, file 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 in desktop publishing workflows, and competed with a December 2020, the second variety of formats such as DjVu, edition of PDF 2.0, ISO 32000- equivalent subset of the Envoy, Common Ground Digital 2:2020, was published, Paper, Farallon Replica and even Adobe's own PostScript format. PDF was a proprietary format controlled by Adobe until 32000-2 does not include any it was released as an open standard on July 1, 2008, and published by the International Organization for Standardization as ISO 32000- tradition of the latest PDF 1:2008, at which time control of specification being freely the specification passed to an ISO Committee of volunteer industry experts. In 2008, Adobe published a Public

Patent License to ISO 32000-1 access, the PDF Association granting royalty-free rights for sell, and distribute PDFcompliant implementations. PDF 1.7, the sixth edition of the bitmap graphics. The basic ISO 32000-1, includes some proprietary technologies definedstreams (i.e., not encoded in and indispensable for the full implementation of the ISO 32000-1 specification. These proprietary technologies are not also support links (inside standardized and their specification is published only on Adobe's website. Many of them are also not supported by other types of embedded popular third-party implementations of PDF. In including clarifications, to normative references. ISO proprietary technologies as normative references. ISO's publication of ISO 32000-2 in 2017 ended the 24-year available from Adobe. Starting in April, 2023, to provide PDF developers and stakeholders with their accustomed level of

and its sponsors made ISO at no cost. Technical details A PDF file is often a combination of vector graphics, text, and types of content in a PDF are: Typeset text stored as content plain text); Vector graphics for consist of shapes and lines; Raster graphics for 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 a plugin for Acrobat 3.0), or any contents that can be handled using plug-ins. PDF combines three technologies: An PostScript page description programming language but in corrections, and critical updates 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 content into a single file, with 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 appearance of any following of programming languages such as if statements and loop commands. PDF is largely based on PostScript but simplified to remove flow controlappearance of a given page, features like these, while graphics commands equivalent document is unaffected by the to lineto remain. Historically, theothers. As a result, PDF PostScript-like PDF code is generated from a source PostScript file. The graphics

the PostScript code are collected and tokenized.[clarification needed] destination page (unless the Any files, graphics, or fonts to which the document refers also Structuring Conventions have are collected. Then, everything been carefully compiled and is compressed to a single file. Therefore, the entire PostScript supports interactive 3D world (fonts, layout,

commands that are output by

measurements) remains intact.[citation needed] As a document format, PDF has several advantages over PostScript: PDF contains tokenized and interpreted results of the PostScript source elements that may have binary

between changes to items in the PDF page description and changes to the resulting page appearance. PDF (since version 1.4) supports

does not. PostScript is an interpreted programming language with an implicit global there are nine types: Boolean state, so instructions

accompanying the description of one page can affect the page. Therefore, all preceding

must be processed to determine the correct whereas each page in a PDF

viewers allow the user to quickly jump to the final pages of a long document, whereas a brackets (<<...>>) Streams, PostScript viewer needs to process all pages sequentially before being able to display the compressed binary data,

optional PostScript Document included). PDF 1.6 and later

file: 3D drawings can be embedded using U3D or PRC and various other data formats. another object) or indirect. File format A PDF file is organized using ASCII characters, except for certain

code, for direct correspondence 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

transparent graphics; PostScript("Carousel" Object Structure) format. A COS tree file consists /ObjStm). This technique primarily of objects, of which values, representing true or

false Real numbers Integers Strings, enclosed within parentheses ((...)) or represented as hexadecimal pages in a PostScript document within single angle brackets

> (<...>). Strings may contain 8-bit characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects enclosed within square brackets ([...]) Dictionaries, collections of objects indexed by names enclosed within double angle usually containing large amounts of optionally 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 (%). documents embedded in a PDF Comments may contain 8-bit characters. Objects may be either direct (embedded in Indirect objects are numbered with an object number and a generation number and defined 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 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 the start of the cross-reference enables them to be read in a 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, object's dictionary: A reference of 15 million in by 15 million in, 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 content streams that describe main body composed of indirect the text, vector and images objects. Version 1.5 introduced being drawn on the page. The optional cross-reference streams, which have the form of similar to PostScript. The 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 document are scattered in size may dedicate only 2 bytes for object offsets. At the end of a PDF file is a footer called "optimized" or "web

containing The startxref keyword followed by an offset toconstructed in a manner that table (starting with the xref keyword) or the cross-reference waiting for the entire file to stream object, followed by The download, since all objects %%EOF end-of-file marker. If a required for the first page to cross-reference stream is not being used, the footer is followed by a dictionary containing information that would otherwise be contained inby the format itself. However, the cross-reference stream 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 each except for the use of content stream is stack-based. maximum size of a PDF compared to Europe. There are graphical elements. A key two layouts to the PDF files: non-linearized (not "optimized") graphics state, which is a and linearized ("optimized"). Non-linearized PDF files can be parameters that may be smaller than their linear counterparts, though they are slower to access because assemble pages of the throughout the PDF file. Linearized PDF files (also

optimized" PDF files) are Web browser plugin without display are optimally organized at the start of the file. PDF files preceded by the trailer keyword may be optimized using Adobe Acrobat software or QPDF. Page dimensions are not limited Adobe Acrobat imposes a limit 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, page, there are one or multiple transparency, which was added in PDF 1.4. PDF graphics use a 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 concept in PDF is that of the collection of graphical changed, saved, and restored by a page description. PDF has (as of version 2.0) 25 graphics portions of the data required to 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

constant, which is a key component of transparency Black point compensation control (introduced in PDF 2.0) commonly, small raster images lossless bi-level (black/white) Vector graphics As in PostScript, vector graphics in PDF are constructed with paths.image.) Images are typically Paths are usually composed of filtered for compression lines and cubic Bézier curves, but can also be constructed from the outlines of text. Unlike following general-purpose PostScript, PDF does not allow filters: ASCII85Decode, a filter 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 introduced in PDF 1.2; it can artwork is specified to be drawn use one of two groups of repeatedly. This may be a

colored tiling pattern, with the colors specified in the pattern object, or an uncolored tiling pattern, which defers color specification to the time the pattern is drawn. Beginning with LZWDecode, a filter based on

PDF 1.3 there is also a shading LZW Compression; it can use 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 RunLengthDecode, a simple images in PDF (called Image XObjects) are represented by

dictionaries with an associated

stream. The dictionary

describes the properties of the specific filters, DCTDecode, a image, and the stream contains lossy filter based on the JPEG the image data. (Less may be embedded directly in a filter based on the Group 3 or page description as an inline purposes. Image filters supported in PDF include the

ASCII, ASCIIHexDecode, similar to ASCII85Decode but less compact, FlateDecode, a commonly used filter based on all image content in a PDF is the deflate algorithm defined in RFC 1951 (deflate is also used allows image data to be stored 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 specification (RFC 2083),

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 typeface, or it may include an PNG specification, compression method for

streams with repetitive data using the run-length encoding algorithm and the image-

standard, CCITTFaxDecode, a 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 a single path to mix text outlinesused to put the stream into 7-bit 1.4, and JPXDecode, a lossy or lossless filter based on the JPEG 2000 standard, introduced in PDF 1.5. Normally embedded in the file. But PDF 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. Text Text in PDF is represented by text elements in page content streams. A text element predictors (filters) from the PNG 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 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 derived from the historical 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, differences to a predefined or have a special significance in PDF documents: Times (v3) (in recommended with TrueType regular, italic, bold, and bold italic) Courier (in regular, oblique, bold and bold oblique) Helvetica (v3) (in regular, oblique, bold and bold oblique) TrueType fonts are complex. 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 embedded in a PDF. Within text object drawn on the page strings, characters are shown using character codes the current font using an encoding. There are several predefined encodings, including transparency is used, new WinAnsi, MacRoman, and many encodings for East Asian marked objects to produce its own built-in encoding. (Although the WinAnsi and MacRoman encodings are

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 files making extensive use of font's built-in encoding or provide a lookup table of built-in encoding (not fonts). The encoding mechanisms in PDF were designed for Type 1 fonts, and the rules for applying them to For large fonts or fonts with non-standard glyphs, the special encodings Identity-H (for horizontal writing) and Identity-V (for vertical) are used.calculating blend modes were With such fonts, it is necessary kept secret by Adobe. They to provide a ToUnicode table if semantic information about the concept of a transparency characters is to be preserved. Transparency The original imaging model of PDF was, like of "group" or "layer" in PostScript's, opaque: each completely replaced anything previously marked in the same (integers) that map to glyphs in location. In PDF 1.4 the imagingmeaningful when editing those model was extended to allow transparency. When objects interact with previously languages and a font can have blending effects. The addition of includes document structure transparency to PDF was done and semantics information to 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, but 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 blend modes were based on those used by Adobe Photoshop at the time. When the PDF 1.4 specification was published, the formulas for have since been published. The group in PDF specification is independent of existing notions applications such as Adobe Illustrator. Those groupings reflect logical relationships among objects that are objects, but they are not part of the imaging model. Additional features Logical structure and accessibility A "tagged" PDF (see clause 14.8 in ISO 32000) 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 of standard structure types and (OCGs), each describing a set 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 and signatures A PDF file may 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, for PDF 2.0 files. The PDF includes an improved discussion of tagged PDF whichthat third parties can define is anticipated to facilitate further their own encryption systems adoption. An ISO-standardized for PDF. PDF files may be subset of PDF specifically targeted at accessibility, PDF/UA, was first published in details on implementing digital 2012. Optional Content Groups signatures in PDF is provided infreeware or open source PDF (layers) With the introduction of ISO 32000-2. PDF files may 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 reader software to obey them, can be selectively viewed or hidden by document authors or limited. The standard security viewers. This capability is usefulprovided by PDF consists of in CAD drawings, layered artwork, maps, multi-language documents, etc. Basically, it

consists of an Optional Content file and prevents opening, and Properties Dictionary added to an owner password, which the document root. This dictionary contains an array of **Optional Content Groups** of information and each of which may be individually displayed or suppressed, plus a modifying text notes and set of Optional Content Configuration Dictionaries, which give the status (Displayed or Suppressed) of the given OCGs. Encryption be encrypted, for security, in which case a password is needed to view or edit the contents. PDF 2.0 defines 256- restrictions that a document bit AES encryption as standard author places on a PDF Reference also defines ways digitally signed, to provide secure authentication; complete edit PDF files. Even without also contain embedded DRM restrictions that provide further controls that limit copying, editing, or printing. These restrictions depend on the so the security they provide is two different methods and two different passwords: a user password, which encrypts the

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 AcroForm fields. The user password encrypts the file, while the owner password does 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 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 removing the password, most 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 (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

that the permissions have been ETSI as TS 102 778. File granted by a bona fide granting attachments PDF files can have PDF files can be created 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 Document Information form data files in FDF and XFDF formats Submit form data fields such as author, title, Instantiate new pages from named page templates Apply a dates. This is optional and is 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 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 documents can also contain the document has been signed by the signer. PAdES (PDF Advanced Electronic Signatures) is a set of restrictions and extensions to PDF and ISO 32000-1 making ituser's default settings when

file attachments which processors may access and open or save to a local filesystem. Metadata PDF files can contain two types of metadata. The first is the Dictionary, a set of key/value subject, creation and update referenced from an Info key in the trailer of the file. A small set 3000 can read tagged PDF. of fields is defined and can be extended with additional text values if required. This method readers with visual is deprecated in PDF 2.0. In PDF 1.4, support was added for older PDFs and those that are 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 to the document catalog), using and (iii) the content view. The an extensible schema. PDF display settings, including the page display layout and zoom level in a Viewer Preferences object. Adobe Reader uses these settings to override the

signatures. This is published by Adobe Reader cannot remove

these settings. Accessibility specifically to be accessible for people with disabilities. PDF file 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, Hal, and Kurzweil 1000 and Moreover, tagged PDFs can be re-flowed and magnified for impairments. Adding tags to generated from scanned documents can present some challenges. One of the significant challenges with PDF 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, physical view is displayed and printed (what most people consider a PDF document). The tags view is what screen readers and other assistive technologies use to deliver high-quality navigation and reading experience to users suitable for advanced electronic opening the document. The freewith disabilities. The content

view is based on the physical

order of objects within the PDF's content stream and may with AcroForms. XFA was 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 JavaScript). Alongside the published as ISO 14289–1 in 2012 and establishes normative interactive forms (AcroForms) language for accessible PDF technology. Multimedia Rich Media PDF is a PDF file including interactive content that can be embedded or linked interactive form fields to a within the file. It can contain images, audio, video content, orlocator (URL). Interactive form buttons. For example, if the interactive PDF is a digital catalog for an E-commerce business, products can be listedsettings of the action's on the PDF pages and can be added with images and links to XFDF flags): HTML Form 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 simpler than PDF since the 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 stand-alone files that can be deprecated from PDF with PDF corresponding PDF interactive 2.0. AcroForms were introducedform. FDF was originally in the PDF 1.2 format. AcroForms permit using objects 32000-2:2017.[citation needed] (e.g. text boxes, Radio buttons, XML Forms Data Format etc.) and some code (e.g. standard PDF action types, support submitting, resetting, and importing data. The "submit" action transmits the names and values of selected specified uniform resource field names and values may be XFDF equivalents – such as the submitted in any of the followingStatus, Encoding, JavaScript, formats, (depending on the ExportFormat, SubmitPDF, and addition, XFDF does not allow format HTML 4.01 Specification new pages based on the given since PDF 1.5; HTML 2.0 since data; as can be done when 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much specification (and in later body of an FDF document consists of only one required object. Forms Data Format is defined in the PDF specification submissions in XML format, but (since PDF 1.2). The Forms Data Format can be used when submissions in XFDF format in submitting form data to a server, receiving the response, XFDF conforms to the XML and incorporating it into the interactive form. It can also be used to export form data to

imported back into the defined in 1996 as part of ISO (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 of FDF containing forms and annotations. Some entries in the FDF dictionary do not have Page's keys, EmbeddedFDFs, Differences, and Target. In the spawning, or addition, of using an FDF file. The XFDF specification is referenced (but not included) in PDF 1.5 versions). It is described separately in XML Forms Data Format Specification. The PDF 1.4 specification allowed form this was replaced by the PDF 1.5 specification. standard. XFDF can be used in the same way as FDF; e.g., form data is submitted to a

server, modifications are made, is referenced from ISO 32000then 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 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 specification. Portable The entire document can be fields and values, as was defined in PDF 1.4. AcroForms in 1992 to present documents, can keep form field values in external stand-alone files containing key-value pairs. The independent of application 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 raster images and other files in FDF and XFDF formats. information needed to display it. 1:2008, at which time control of In PDF 1.5, Adobe Systems introduced a proprietary format Camelot Project" initiated by for forms: Adobe XML Forms Architecture (XFA). Adobe XFA Warnock in 1991. PDF was Forms are not compatible with ISO 32000's AcroForms feature, and most PDF processors do not handle XFA December 2020. PDF files may are necessary to make, use, content. The XFA specification contain a variety of content

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 form. As of August 2019, XFDF Anyone may create applicationsformats. The PDF specification 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 article: History of PDF Adobe complying with its PDF Document Format (PDF), submitted rather than individual standardized as ISO 32000, is ayears PDF was popular mainly file format developed by Adobe in desktop publishing including text formatting and images, in a manner software, hardware, and

PDF file encapsulates a

PDF has its roots in "The

Adobe co-founder John

operating systems. Based on the PostScript language, each layout flat document, including standardized as ISO 32000 in 2008. The last edition as ISO sell, and distribute PDF-

besides flat text and graphics 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 also provides for encryption and digital signatures, file attachments, and metadata to enable workflows requiring these features. History Main Systems made the PDF specification available free of charge in 1993. In the early workflows, and competed 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 it was released as an open complete description of a fixed- standard on July 1, 2008, and published by the International the text, fonts, vector graphics, Organization for Standardization as ISO 32000the 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 32000-2:2020 was published in all patents owned by Adobe that

compliant implementations. PDF 1.7, the sixth edition of the bitmap graphics. The basic PDF specification that became ISO 32000-1, includes some proprietary technologies definedstreams (i.e., not encoded in only by Adobe, such as Adobe XML Forms Architecture (XFA) and JavaScript extension for 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 also support links (inside standardized and their specification is published only on Adobe's website. Many of them are also not supported by other types of embedded popular third-party implementations of PDF. In December 2020, the second edition of PDF 2.0, ISO 32000- equivalent subset of the 2:2020, was published, including clarifications, corrections, and critical updates declarative form, for generating tokenized and interpreted 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 graphics and standard features appearance of any following at no cost. Technical details A PDF file is often a combination such as if statements and loop

of vector graphics, text, and types of content in a PDF are: Typeset text stored as content plain text); Vector graphics for illustrations and designs that consist of shapes and lines; Raster graphics for photographs and other types of commands that are output by images Multimedia objects in the document. In later PDF a plugin for Acrobat 3.0), or any is compressed to a single file. contents that can be handled using plug-ins. PDF combines three technologies: An PostScript page description programming language but in to allow fonts to travel with the documents. A structured storage system to bundle these changes to the resulting page elements and any associated content into a single file, with 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 of programming languages

commands. PDF is largely based on PostScript but simplified to remove flow control features like these, while graphics commands equivalent to lineto remain. Historically, the PostScript-like PDF code is generated from a source PostScript file. The graphics the PostScript code are collected and revisions, a PDF document can tokenized.[clarification needed] Any files, graphics, or fonts to document or web page), forms, which the document refers also JavaScript (initially available as are collected. Then, everything Therefore, the entire PostScript world (fonts, layout, measurements) remains intact.[citation needed] As a document format, PDF has several advantages over PostScript: PDF contains the layout and graphics. A font- results of the PostScript source embedding/replacement systemcode, for direct correspondence between changes to items in the PDF page description and appearance. PDF (since version 1.4) supports transparent graphics; PostScript does not. PostScript is an interpreted programming language with an implicit global state, so instructions accompanying the description of one page can affect the 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 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 brackets (<<...>>) Streams, PostScript viewer needs to process all pages sequentially before being able to display the compressed binary data, 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 in a PDF Comments may contain 8-bit file: 3D drawings can be embedded using U3D or PRC and various other data formats. another object) or indirect. 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 /ObjStm). This technique 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-bitspecifying an object's characters. Names, starting with a forward slash (/) Arrays, ordered collections of objects ([...]) Dictionaries, collections of and gives the byte offset of objects indexed by names enclosed within double angle usually containing large amounts of optionally 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 (%). characters. Objects may be either direct (embedded in Indirect objects are numbered with an object number and a generation number and defined Such a stream may be used 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 enables non-stream objects to have standard stream filters applied to them, reduces the size of files that have large numbers of small indirect for Tagged PDF. Object streams do not support

the cross-reference table, is enclosed within square bracketslocated near the end of the file 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 ASCII format, be marked with 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. 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 to objects and is especially useful the start of the cross-reference table (starting with the xref keyword) or the cross-reference

generation number (other than

0). An index table, also called

stream object, followed by The download, since all objects %%EOF end-of-file marker. If a required for the first page to cross-reference stream is not being used, the footer is followed by a dictionary containing information that would otherwise be contained inby the format itself. However, the cross-reference stream 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 eachexcept for the use of 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 graphical elements. A key two layouts to the PDF files: non-linearized (not "optimized") graphics state, which is a and linearized ("optimized"). Non-linearized PDF files can be parameters that may be smaller than their linear counterparts, though they are slower to access because portions of the data required to state properties, of which some (Type 2) and radial shading 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

at the start of the file. PDF files lines and cubic Bézier curves, preceded by the trailer keyword may be optimized using Adobe but can also be constructed Acrobat software or QPDF. Adobe Acrobat imposes a limit with lines and curves. Paths object's dictionary: A reference of 15 million in by 15 million in, can be stroked, filled, fill then or 225 trillion in 2 (145,161 km2). Imaging model The basic Strokes and fills can use any design of how graphics are represented in PDF is very similar to that of PostScript,

> 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 concept in PDF is that of the collection of graphical changed, saved, and restored by a page description. PDF has shading patterns of which the (as of version 2.0) 25 graphics of the most important are: The current transformation matrix (CTM), which determines the path The color space The alpha stream. The dictionary constant, which is a key component of transparency Black point compensation control (introduced in PDF 2.0) commonly, small raster images

Vector graphics As in

PDF are constructed with paths. display are optimally organized Paths are usually composed of from the outlines of text. Unlike Page dimensions are not limitedPostScript, PDF does not allow a single path to mix text outlines stroked, or used for clipping. color set in the graphics state, including patterns. PDF supports several types of patterns. The simplest is the page, there are one or multiple transparency, which was added tiling pattern in which a piece of in PDF 1.4. PDF graphics use a artwork is specified to be drawn repeatedly. This may be a colored tiling pattern, with the colors specified in 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 continuously varying colors. There are seven types of simplest are the axial shading (Type 3). Raster images Raster images in PDF (called Image XObjects) are represented by coordinate system The clipping dictionaries with an associated describes the properties of the image, and the stream contains the image data. (Less

may be embedded directly in a

PostScript, vector graphics in

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 ASCII, ASCIIHexDecode, similar to ASCII85Decode but 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); 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 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 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 used to put the stream into 7-bit 1.4, and JPXDecode, a lossy or regular, italic, bold, and bold lossless filter based on the JPEG 2000 standard. introduced in PDF 1.5. Normally Helvetica (v3) (in regular, all image content in a PDF is embedded in the file. But PDF allows image data to be stored fonts are sometimes called the 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. PDF readers, but they are not Text Text in PDF is represented guaranteed to be available in by text elements in page content streams. A text element display correctly if the system predictors (filters) from the PNG specifies that characters should has them installed. Fonts may 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 (integers) that map to glyphs in

> describe the characteristics of a encoding. There are several 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 (Although the WinAnsi and based on widely used standard MacRoman encodings are digital font formats: Type 1 (and derived from the historical its compressed variant CFF), TrueType, and (beginning with Macintosh operating systems, 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, have a special significance in PDF documents: Times (v3) (in italic) Courier (in regular, oblique, bold and bold oblique) oblique, bold and bold oblique) Symbol Zapf Dingbats These base fourteen fonts. These fonts, or suitable substitute fonts with the same metrics, should be available in most the reader, and may only be substituted if they are not embedded in a PDF. Within text strings, characters are shown using character codes a digital typeface. It may either the current font using an predefined encodings, including WinAnsi, MacRoman, and many encodings for East Asian languages and a font can have its own built-in encoding. properties of the Windows and fonts using these encodings

work equally well on any

predefined encoding to use, the files making extensive use of 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 Identity-V (for vertical) are used.calculating blend modes were With such fonts, it is necessary kept secret by Adobe. They to provide a ToUnicode table if have since been published. The is anticipated to facilitate further semantic information about the concept of a transparency characters is to be preserved. Transparency The original imaging model of PDF was, like of "group" or "layer" in PostScript's, opaque: each object drawn on the page completely replaced anything previously marked in the same location. In PDF 1.4 the imagingmeaningful when editing those model was extended to allow transparency. When transparency is used, new objects interact with previously marked objects to produce blending effects. The addition of includes document structure transparency to PDF was done and semantics information to 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 builds on the logical structure that use a small amount of framework introduced in PDF transparency might view

platform.) PDF can specify a

acceptably by older viewers, butof standard structure types and transparency could be viewed incorrectly by an older viewer. The transparency extensions are based on the key concepts Tagged PDF is not required in of transparency groups, blending modes, shape, and alpha. The model is closely aligned with the features of Adobe Illustrator version 9. The were relatively vague in ISO blend modes were based on those used by Adobe Photoshop at the time. When the PDF 1.4 specification was published, the formulas for group in PDF specification is independent of existing notions targeted at accessibility, applications such as Adobe Illustrator. Those groupings reflect logical relationships among objects that are objects, but they are not part of Optional Content Groups the imaging model. Additional features Logical structure and accessibility A "tagged" PDF enable reliable text extraction and accessibility. Technically speaking, tagged PDF is a stylized use of the format that

attributes that allow page content (text, graphics, and images) to be extracted and reused for other purposes. 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 devices, including assistive technology (AT), is uneven as of 2021. ISO 32000-2, however, includes an improved discussion of tagged PDF which adoption. An ISO-standardized subset of PDF specifically 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. Layers, more formally known as (OCGs), refer to sections of content in a PDF document that can be selectively viewed or (see clause 14.8 in ISO 32000) 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 1.3. Tagged PDF defines a set Optional Content Groups

(OCGs), each describing a set can include modifying, printing, with a modified form or of information and each of which may be individually displayed or suppressed, plus a modifying text notes 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- restrictions that a document bit AES encryption as standard author places on a PDF 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 edit PDF files. Even without details on implementing digital signatures in PDF is provided infreeware or open source PDF 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 authority. For example, it can be restricted even when the document is decrypted, which

or copying text and graphics outannotation data Import form of the document, or adding or AcroForm fields. The user password encrypts the file, while the owner password does Instantiate new pages from not, instead relying on client software to respect these restrictions. An owner passworddigital signature form field can easily be removed by software, including some free online services. Thus, the use document are not secure, and cannot be assured once the file cryptography. Adobe Reader is distributed; this warning is displayed when applying such restrictions using Adobe Acrobat software to create or removing the password, most readers ignore the permission "protections" and allow the user receiver, the information the to print or make copy of excerpts of the text as if the document were not limited by password protection. Beginning the document has been signed 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 that the permissions have been ETSI as TS 102 778. File granted by a bona fide granting attachments PDF files can have be used to allow a user: To save the PDF document along

data files in FDF, XFDF, and text (CSV/TSV) formats Export form data files in FDF and XFDF formats Submit form data named page templates Apply a digital signature to existing Create, delete, modify, copy, import, and export annotations For example, Adobe Systems grants permissions to enable additional features in Adobe Reader, using public-key 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 of a digital signed document sees can be manipulated by the sender after by the signer. PAdES (PDF 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 file attachments which processors may access and open or save to a local

filesystem. Metadata PDF files include tags, text equivalents, 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 3000 can read tagged PDF. 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 older PDFs and those that are 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 to the document catalog), using and (iii) the content view. The 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 free with disabilities. The content Adobe Reader cannot remove these settings. Accessibility PDF files can be created specifically to be accessible for be displayed by software that people with disabilities. PDF file does not fully support the tags' formats in use as of 2014 can view, such as the Reflow

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, technology. Multimedia Rich Hal, and Kurzweil 1000 and re-flowed and magnified for readers with visual impairments. Adding tags to 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. Forms Interactive document's creation, can be inconsistent with each other. The three views are (i) the physical view is displayed and printed (what most people tags view is what screen readers and other assistive technologies use to deliver high-quality navigation and reading experience to users view is based on the physical order of objects within the PDF's content stream and may with AcroForms. XFA was

feature in Adobe's Reader. PDF/UA, the International Standard for accessible PDF based on ISO 32000-1 was first published as ISO 14289-1 in 2012 and establishes normative language for accessible PDF Media PDF is a PDF file including interactive content Moreover, tagged PDFs can be that can be embedded or linked within the file. It can contain images, audio, video content, or buttons. For example, if the interactive PDF is a digital catalog for an E-commerce business, products can be listed on the PDF pages and can be significant challenges with PDF added with images and links to 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 physical view, (ii) the tags view, different methods for integrating data and PDF forms. Both formats today coexist in the PDF specification: AcroForms consider a PDF document). The (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 deprecated from PDF with PDF 2.0. AcroForms were introduced

in the PDF 1.2 format.

AcroForms permit using objects 32000-2:2017.[citation needed] stand-alone files that can be (e.g. text boxes, Radio buttons, XML Forms Data Format 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 field names and values may be XFDF equivalents – such as the submitted rather than individual submitted in any of the followingStatus, Encoding, JavaScript, formats, (depending on the settings of the action's ExportFormat, SubmitPDF, and addition, XFDF does not allow XFDF flags): HTML Form format HTML 4.01 Specification new pages based on the given external files may use Forms since PDF 1.5; HTML 2.0 since data; as can be done when 1.2 Forms Data Format (FDF) based on PDF, uses the same syntax and has essentially the same file structure, but is much specification (and in later 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 submissions in XML format, but introduced a proprietary format (since PDF 1.2). The Forms Data Format can be used when submissions in XFDF format in Architecture (XFA). Adobe XFA submitting form data to a server, receiving the response, XFDF conforms to the XML 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

(XFDF) (external XML Forms Data Format Specification, Version 2.0; supported since PDF 1.5; it replaced the "XML" form submission format defined 19444-1:2019 - Document in PDF 1.4) the XML version of management — XML Forms Forms Data Format, but the XFDF implements only a subsetISO 32000-2 (XFDF 3.0). This of FDF containing forms and annotations. Some entries in the FDF dictionary do not have The entire document can be Page's keys, EmbeddedFDFs, Differences, and Target. In the spawning, or addition, of using an FDF file. The XFDF specification is referenced (but files. The usage rights (UR) not included) in PDF 1.5 versions). It is described separately in XML Forms Data Format Specification. The PDF 1.4 specification allowed form this was replaced by the PDF 1.5 specification. standard. XFDF can be used in feature, and most PDF the same way as FDF; e.g., form data is submitted to a server, modifications are made, is referenced from ISO 32000then sent back and the new form data is imported in an interactive form. It can also be used to export form data to

imported back into the corresponding PDF interactive form. As of August 2019, XFDF 3.0 is an ISO/IEC standard under the formal name ISO Data Format — Part 1: Use of standard is a normative reference of ISO 32000-2. PDF fields and values, as was defined in PDF 1.4. AcroForms can keep form field values in external stand-alone files containing key-value pairs. The Data Format (FDF) and XML Forms Data Format (XFDF) 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 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

2.0). Split and merge PDF files can be splitted and merged, using applications. Licensing Anyone may create applications PDF file encapsulates a 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 PDF has its roots in "The complying with its PDF specification. Portable Document Format (PDF), standardized as ISO 32000, is astandardized as ISO 32000 in file format developed by Adobe 2008. The last edition as ISO 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 layout flat document, including raster images and other

Camelot Project" initiated by Adobe co-founder John Warnock in 1991. PDF was 32000-2:2020 was published in specification available free of December 2020. PDF files may charge in 1993. In the early contain a variety of content besides flat text and graphics

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

ELEMENT BELOW