

# PDF/A *in a Nutshell* 2.0

PDF for long-term archiving

Alexandra Oettler

- The history of the ISO standard
- All versions – from PDF/A-1 to PDF/A-3
- How users benefit from PDF/A
- The technical background
- Tools for creating PDF/A files
- Validating PDF/A files
- PDF/A in law and administration
- PDF/A in finance and industry

**PDF/A** *in a Nutshell 2.0*  
*PDF for long-term archiving*  
*The ISO Standard – from PDF/A-1 to PDF/A-3*

This work, including all its component parts, is copyright protected. All rights based thereupon are reserved, including those of translation, reprinting, presentation, extraction of illustrations or tables, broadcasting, microfilming or reproduction by any other means, or storage in any data-processing device, in whole or in part. Reproduction of this work or any part of this work is only permitted where legally specified in the Copyright Act of the Federal Republic of Germany dated the 9<sup>th</sup> of September 1965.

© 2013 Association for Digital Document Standards e.V., Berlin  
info@pdfa.org  
Printed in Germany

The use of any names, trade names, trade descriptions etc. in this work, even those not specially identified as such, does not justify the assumption that these names are free according to trademark protection law and thus usable by anyone.

Text: Alexandra Oettler

Layout, cover design, design and composition: Alexandra Oettler

Cover image: Paulgeor, Dreamstime.com

Picture credits: Page 5: Photocase; Page 6: Sepp Huberbauer, Photocase; Page 8: aoe; Page 13: EU Publications Office; Page 14: Rui Frias, Istockphoto.com; Page 15: MBPHOTO, Istockphoto.com; Page 18: Photocase.

Printed by: Galrev Druck- und Verlagsgesellschaft Hesse & Partner OHG

# Contents

<b>PDF/A – the ISO standard for long-term archiving</b>	<b>5</b>	<b>PDF/A in public administration</b>	<b>13</b>
The decisive advantages of PDF/A			
Widespread acceptance of PDF/A			
<b>PDF/A facts – an introduction to the standard</b>	<b>6</b>	<b>PDF/A in finance and industry</b>	<b>14</b>
An archiving format		Industry documentation	
Why PDF/A and not just PDF?		Banking and insurance	
		Healthcare	
		E-invoicing	
<b>A short history of PDF/A</b>	<b>7</b>	<b>PDF/A in legislation and justice</b>	<b>15</b>
Becoming an ISO standard		Federal jurisdiction in the United States courts	
PDF/A catches on		The Italian Chamber of Commerce	
		Austria: BAIK	
		Germany: Land registration	
<b>The technical side of the PDF/A standard</b>	<b>8</b>	<b>What the users and experts say</b>	<b>16</b>
PDF/A-1: The first archiving standard			
PDF/A-2: Based on PDF 1.7			
PDF/A-3: One more feature			
Conformance levels: A, B, U			
<b>The most important reasons to use PDF/A</b>	<b>9</b>	<b>PDF/A and the other PDF standards</b>	<b>17</b>
		PDF/X	
		PDF/A	
		PDF/E	
		PDF	
		PDF/VT	
		PDF/UA	
<b>Typical uses for PDF/A</b>	<b>10</b>		
<b>PDF/A creation tools</b>	<b>11</b>	<b>The myths and legends surrounding PDF/A</b>	<b>18</b>
Desktop software			
Server-based solutions			
Programming libraries			
Integrated PDF/A functions			
<b>Validation: Is it really PDF/A?</b>	<b>12</b>	<b>Further information on PDF/A</b>	<b>19</b>
When do I need to validate?		The portal to the PDF Association	
Finding the right validation solution		PDF Association events	
		Membership	

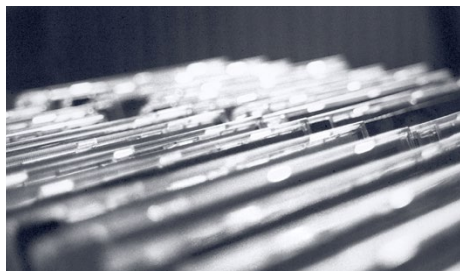
# PDF/A – the ISO standard for long-term archiving

Up to the end of the 20<sup>th</sup> century, physical media formats (paper, microfilms and microfiches) were the only option for businesses and public authorities storing documents for the long term in a reproducible format. The major drawback to these analogue approaches was the significant time and effort required: documents are hard to search through, trained personnel are required, specialist equipment is needed to read microfilms, and entire climate-controlled rooms are needed to store documents.

The first digital archiving format to gain ground in many countries was the TIFF image format. In 1993, however, a modern, more powerful format became available in the form of PDF. This became the basis on which the standard archive format PDF/A was developed (see page 7).

For companies, public authorities and private users needing to store digital information for a long period of time – be it 5 years, 50 or 500 – the PDF/A standard is now the clear choice of file format.

PDF/A is a multi-part ISO standard developed over many years of committee work by industry associations, businesses and public authorities around the world. The result is “a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and sys-



tems used for creating, storing or rendering the files.” (ISO 19005-1, quoted from the introduction).

The first part of the standard, PDF/A-1, has been available since the 1<sup>st</sup> of October 2005. Its official designation is “ISO 19005-1:2005. Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)”.

Since then, two further parts have been made available to users: PDF/A-2 (since 2011) and PDF/A-3 (since 2012). These parts exist in parallel and are optimised to meet particular needs (see page 8).

The PDF/A standards family regulates how to create electronic documents to ensure they can be reliably reproduced for decades to come. The standard does not describe how to build a revision-safe archive, nor the theory behind one.

## The decisive advantages of PDF/A

- A PDF/A file contains everything needed to display it and nothing which could negatively impact the display.
- PDF/A files can be used on any platform.
- Free programs exist for displaying PDF/A files.
- The multi-part PDF/A standard offers great flexibility to users.

## Widespread acceptance of PDF/A

PDF/A is becoming more and more common, be it in industry, public administration, financial services or academia. A large number of authorities and institutions worldwide recommend PDF/A or specifically require the use of the standard (see page 13).



*The ISO (International Organization for Standardization) is the largest organisation in the world for developing and publishing international standards.*

# PDF/A facts – an introduction to the standard

PDF/A is an industry-recognised ISO standard. Future software development must reflect the need to work reliably with these documents.

Current file formats used by popular applications are simply not suitable for public authorities, businesses and individual users needing to store unalterable digital documents for long periods of time. Word processors such as Microsoft Word or OpenOffice Writer create files which can look very different depending on the platform used to view them. Text and images may appear different than intended – or they may not appear at all. Nowadays, there are also the questions of how these programs will develop in the future, and whether or not it will still be possible to open and view older files – an unacceptable risk when considering the timescales involved in long-term archiving.

## An archiving format

When using email or the internet to distribute carefully designed documents containing text and images, users are increasingly choosing PDF. After all, the Portable Document Format can embed all elements of a document within itself. This can include fonts and images, but also 3D objects, audio and video. Embedded fonts are optional; it is also possible (in order to save on file size, for example) to link to one instead. This, however, carries the risk that not all machines will correctly display the PDF.

PDF has also gained such broad worldwide acceptance because free programs exist for all devices and operating systems to view PDF documents. Whether

viewed on a tablet, a smartphone or a desktop computer, a PDF file will usually look the same.

Document archives, however, require an exceptionally high standard: the content must always appear exactly the same under all circumstances. Particularly because of its universal availability and worldwide acceptance, it makes sense to build on PDF to create an archiving standard for digital documents.

## Why PDF/A and not just PDF?

Put in the simplest possible terms, PDF/A is a PDF which forbids certain functions which could hinder long-term archiving. PDF/A also demands that the file meet certain requirements which guarantee reliable reproduction.

For example, files must not be encrypted with a password, as all content must always be fully available. Embedded video and audio data are also prohibited: PDF/A consciously avoids anything that requires external software for display or playback. JavaScript and certain actions are also forbidden, as executing them could potentially alter the PDF.

PDF/A also places higher demands on the information it contains. All required fonts (or at least all glyphs for the specific characters used) must be embedded within the PDF. To ensure a uniform colour appearance on a variety of platforms and devices, colour information must be given in a platform-independent format using ICC colour profiles. The software must also use the XMP format for metadata (which is used to store the data identifying the file as a PDF/A, for example).

PDF/A also sets technical limits: for example, the page size is limited to an edge length of either 5.08 metres (PDF/A-1) or up to 381 kilometres (PDF/A-2 and PDF/A-3).



# A short history of PDF/A

Those who first needed to store documents in a future-proof digital format used the popular image format TIFF (Tagged Image File Format). This format was used for a long time, particularly for scanned documents, but it has a number of drawbacks.

For example, the TIFF raster format contains no text-based information, meaning files cannot be searched by their text content. And if the TIFF file contains colour images or pages, it will become significantly larger; effective compression is all but impossible. Only black-and-white line images (which is sometimes enough for scanned text pages) can save much space in TIFF format.

Contrary to popular belief, TIFF is not an ISO standard. The resolution, colour and metadata settings for TIFF files are mostly left to the individual user's discretion.

## Becoming an ISO standard

As Adobe Systems' 1993-published Portable Document Format (PDF) grew in popularity, users and developers began to recognise its potential for long-term archiving.

In 2002, specialists from libraries and archives, from administrative bodies, from industry and from the judicial system assembled in order to develop a purpose-built file format for standardised archiving. A working group within the ISO (International Organisation for Standardisation) took up the task: representatives from a wide range of US-based associations and federal authorities including AIIM (Association for Information and Image Management), NPES (Association for Suppliers of Printing, Publishing and Converting Technologies) and NARA (National Archives and Records Administration) met with experts from the library sector (Harvard

University Libraries, Library of Congress), the judicial system (Administrative Office of the United States Courts) and industry developers (including Adobe Systems and Kodak). After a number of meetings and a comprehensive testing and approval phase, the ISO published PDF/A on the 1<sup>st</sup> of October 2005 under the designation "ISO 19005-1:2005". It was the world's first standard file format for digital long-term archiving.

## PDF/A catches on

In 2006, to promote recognition of PDF/A, a group of software developers founded the PDF/A Competence Centre (today a part of the PDF Association) as an industry association for digital document standards. Through seminars, conferences, publications and not least through its website [www.pdfa.org](http://www.pdfa.org), the association has helped spread practical information about the ISO standard (see page 19).

Initially active in Germany and Switzerland in particular, within a few years the PDF Association was able to expand its area of operations across Europe, America, the Middle East, Asia and Australia. By the end of 2012, the Association had 143 members across 25 countries.

Today, PDF/A has found broad acceptance in all sectors where documents are stored long-term. Numerous document management solutions provide direct support for archiving with PDF/A. More and more countries are recommending the standard in public administration, or even specifically requiring it (see page 13). Meanwhile, a correspondingly broad selection of PDF/A creation and validation software is now available (see page 12), from single-workstation solutions to automated server-based systems.

PDF/A's wide-ranging everyday use is also seen in the number of common programs which support it. Free word processing software such as OpenOffice and LibreOffice can create PDF/A files at the click of a button, and Adobe Reader faithfully displays PDF/A documents as they were intended to be seen. Microsoft Office has also supported directly saving as PDF/A since 2007.

# The technical side of the PDF/A standard

**Nomenclature:** PDF/A versions and levels are simply given one after another. A PDF/A-1b file, for example, is a PDF file for long-term archiving, of the first generation, with visually reproducible content.

After the first part of PDF/A was published, two more parts arrived. These are not replacements for part 1, however; rather, they offer additional options for archiving PDF documents. All existing PDF/A files remain fully valid.

## PDF/A-1: The first archiving standard

PDF/A-1 is based on PDF version 1.4, which first appeared in 2001. All resources (images, graphics, typographic characters) must be embedded within the PDF/A document itself. A PDF/A file requires precise, platform-independent colour data using ICC profiles, and XMP for the document metadata. Transparent elements, some forms of compression (LZW, JPEG2000), PDF layers, and certain actions or JavaScript are forbidden. A PDF/A file must not be password-protected. PDF/A-1 expressly supports embedded digital signatures and the use of hyperlinks.

## PDF/A-2: Based on PDF 1.7

PDF/A-2 was published in 2011 as “ISO 19005-2”. Based on PDF version 1.7 (see page 17), , which has since been standardised as “ISO 32000-1”, it makes use of this version’s new features. This means PDF/A-2 allows JPEG2000 compression, transparent elements and PDF layers. PDF/A-2 also allows you to embed OpenType fonts and supports PAdES (PDF Advanced Electronic Signatures)-compliant digital signatures. One particularly important innovation is the “container” function: PDF/A files can be embedded within a PDF/A-2 document.

## PDF/A-3: One more feature

PDF/A-3 has been available since October 2012. A PDF/A-3 document allows you to embed any file format desired – not just PDF/A documents. For example, a

PDF/A-3 file can contain the original file from which it was generated. The PDF/A standard does not regulate the suitability of these embedded files for archiving.

## Conformance levels: A, B, U

The different conformance levels reflect the quality of the archived document and depend on the input material and the document’s purpose.

■ **Level A** (Accessible) meets all requirements for the standard, including the logical structure of the document and its correct reading order. Text must be extractable and the logical structure must match the natural reading order. Fonts used must meet stringent requirements. This PDF/A level can usually only be met by converting born-digital documents.

■ **Level B** (Basic) guarantees that the content of the document can be unambiguously reproduced. Level B files are easier to create than Level A, but Level B does not guarantee 100% text extraction or searchability. It does not necessarily mean that the content can be reused without any problems. Scanned paper documents can usually be converted to PDF/A Conformance Level B without any extra work.

■ **Level U** (Unicode) was introduced along with PDF/A-2. It expands Conformance Level B to specify that all text can be mapped to standard Unicode character codes.





# The most important reasons to use PDF/A

The PDF/A standard offers practical solutions for a wide variety of tasks, bringing advantages to many areas of application.

■ **Long-term archiving:** PDF/A provides an ISO-standardised format to all those who need to store digital documents for long periods of time. This can include archives, libraries, banks, insurance firms and others.

■ **Legally binding documents:** PDF/A is an excellent option for digitally signed documents and records. The ISO standard allows embedded electronic signatures and specifies only their minimum requirements. This means that PDF/A documents can always be digitally signed using the very latest technology, even as it develops in the future.

■ **Science and research:** PDF/A reliably displays special characters for mathematical formulas or old languages, as all required symbols are embedded into the file itself. ICC profiles provide total colour control, supporting research work in fields such as medicine, archaeology or cultural history. As a result, people are always finding more uses for PDF/A in the academic sector: some universities now only accept assignments and dissertations in this ISO-standard format.

■ **Global integration:** storing information in different languages requires comprehensive support for all kinds of writing systems around the world. In Japanese, Arabic, or Cyrillic, PDF/A makes sure that texts can always be correctly displayed on any device, including the reading direction. It also allows fixed-layout printing.

■ **Platform-independent:** PDF, and so PDF/A too, are platform-independent. Thanks to PDF/A, documents such as invoices, brochures, manuals or research reports can be made reliably available through a wide range of channels.

■ **Full text searching:** PDF/A helps you to find and access specific information within a data set. This is even possible with scanned documents, as the standard permits searchable text created through optical character recognition (OCR). Even Conformance Level B (Basic) supports this feature.

■ **Extra search options:** XMP metadata can be used to add additional structured information to the document, such as the author, description of the content, or source and copyright information. As a result, the user can search for additional stored keywords, categories or values within the data set.

■ **Use content again and again:** PDF/A Conformance Level A makes it easier to reuse content. Such files are very easy to convert to Word, HTML or eBook formats.

■ **Use PDF/A in combination with other standards:** PDF/A is closely related to the ISO's other PDF standards. As a result, a PDF/A file can often meet the requirements for universally accessible PDFs (for disabled users) as defined in the PDF/UA standard. Digital books in PDF/A format are well-suited for printing on demand if they also meet the PDF/X standard for digital print documents. For an overview of PDF standards, see page 17.

# Typical uses for PDF/A

The PDF/A standard has proved its suitability for a wide variety of tasks. Here we can show you just a few brief practical examples.

■ **Scanned documents for archiving:** PDF/A is widely used to digitise paper-based files and records. A document scanner reads the original text, and specially designed software automatically converts the data into a searchable PDF/A file.

■ **Archive migration:** Solutions exist to help digital archives which are still using older formats to migrate to PDF/A. In most cases, the process can even be automated.

■ **Incoming and outgoing mail:** Whether a company receives letters or emails, PDF/A provides a reliable storage format for them. Letters can be automatically scanned and archived as PDF/A files, and emails and their attachments can be stored in PDF/A too. There are also great advantages to storing a copy of all outgoing mail in PDF/A format. Outgoing mail data can be retrieved from popular print data streams such as AFP (Advanced Function Presentation).

■ **Office documents:** If your presentations, spreadsheets and text documents are likely to have long-term relevance and need to remain available for long periods of time, then PDF/A is the perfect format in which to archive them. The original programme used may directly support this to some extent, or you can use additional software packages. In both cases, the process can be automated.

■ **Documentation and typesetting:** Brochures and instruction books are usually created using layout programs and edi-

torial systems. As a result, an archivable PDF can be created at the same time without significant extra work. This can either be done using external solutions (rather than using the original program that created the document) or using a print-ready PDF which is often already available in the PDF/X format, the ISO standard for print documents (see page 17).

■ **Creating documents from databases:** Many PDF/A files were originally created from databases or were created using XML data. This structured input data often allows you to create PDF/A Conformance Level A documents. You can also convert forms to PDF/A.

■ **Digital document folders:** As of PDF/A-3, source documents can also be embedded directly into a PDF/A file in their original format. This eliminates time-consuming hybrid archiving processes in which additional documents (Excel tables, image files, CAD drawings) had to be managed separately from the archived PDF/A file in their original formats. Thanks to PDF/A-3, all relevant information is now contained within a single file.

■ **Team collaboration:** PDF/A-3 in particular is exceptionally powerful and flexible when used within modern collaboration frameworks. Its hybrid approach means that each document can contain the current working version of a document and the final – archive-ready – version. As a result, PDF/A-3 provides ideal support for all the most important functions within a Microsoft SharePoint environment, for example. In particular, this includes collaborative work on documents as well as distribution and archiving.

# PDF/A creation tools

PDF/A documents can be created in a variety of ways:

- From scanned documents
- By direct conversion of the source data
- By exporting from the program used to create the source document
- Using an intermediate step which turns a PDF file into PDF/A
- Using print output formats or print data streams such as GDI, PCL, PostScript, AFP and XPS.

This section will sketch out just a few typical approaches. For a more extensive list, including specific products which allow you to create PDF/A files, visit the PDF Association's website at [www.pdfa.org](http://www.pdfa.org).

## Desktop software

On a standard workstation, office applications in particular will already offer inbuilt tools (or can easily be retrofitted) to export word-processed files, spreadsheets or presentations directly to PDF/A. If the "Tagged PDF" option is enabled, then Microsoft Office, OpenOffice and LibreOffice will even support PDF/A Conformance Level A for semantically structured data.

Some PDF/A conversion solutions use print data creation tools to generate PDF or PDF/A files. Another approach is to use programming libraries to convert data or directly write to PDF.

Adobe Acrobat is used for PDFs in many industries, and it provides comprehensive support for PDF/A. This

software can be used to examine PDF/A files to ensure they actually meet the PDF/A standard (see page 12).

Individual-workstation products also exist which allow users to scan to PDF/A, including OCR. This software is sometimes supplied with the scanner itself.

## Server-based solutions

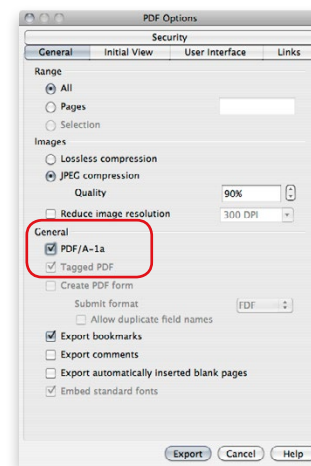
Server-based solutions exist for mass PDF/A creation. This allows business-wide standardisation of your working processes and lets you manage large volumes of data. Some desktop PC products also have server-based versions for high-volume processing.

## Programming libraries

Programming libraries allow developers to add PDF/A functionality to their own applications without having to develop the needed technology from scratch. Some desktop or server-based products are also available as programming libraries. Suppliers can thus integrate extra functionality into their solutions with minimal development work on their part. These extra functions may include PDF/A creation, validation and management. A business' IT department can also add PDF/A features to the company's own software environment for internal projects.

## Integrated PDF/A functions

Many document and output management solutions providers offer modules which can be used to perform PDF/A functions. Many systems are already available for high-volume management of a wide variety of input and output channels in PDF and PDF/A format.

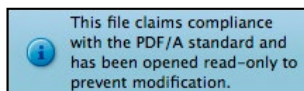


*Some word processing software (such as OpenOffice, shown here) allow you to create PDF/A-1, including "Tagged PDF" as a prerequisite for Conformance Level A.*

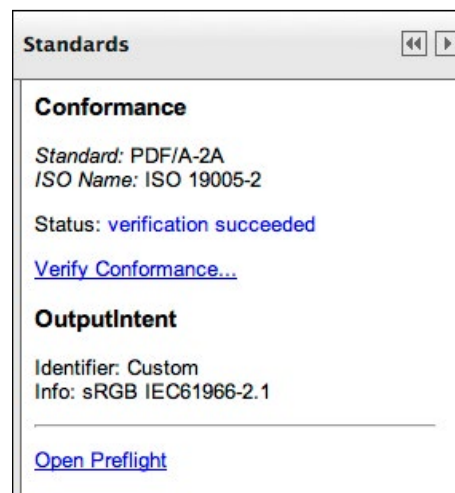
# Validation: Is it really PDF/A?

It is not always easy to tell at first glance whether an existing PDF file actually meets the ISO's PDF/A standard. Applications such as Adobe Acrobat and Adobe Reader do use a pale blue banner to indicate when a file claims to be PDF/A compliant, but this is only an indicator and should not be used in place of a full examination to ensure the document meets the PDF/A standard.

To be absolutely certain, you can perform a validation check which examines all relevant parts of a document.



*Adobe Acrobat and Adobe Reader can indicate whether a document may be PDF/A compliant, but this is not a replacement for a full validation.*



*Acrobat's Preflight function checks for compliance with PDF standards.*

## When do I need to validate?

During the typical life cycle of a PDF/A file, there are particular points when it should be checked for complete ISO compliance.

■ **After creation:** A PDF/A file should be validated immediately after creation, to ensure that the process was carried out successfully.

■ **On receipt:** If a company receives a PDF/A file (by email, for example, or by

other means), validation is advised. After all, it is impossible to know how the PDF/A file was created.

■ **Prior to transmission/distribution:** When sending a PDF/A file by email or making it available online, validation in advance is recommended.

■ **Before archiving:** You must validate data before placing it in a digital archive.

■ **At the end of certain processes:** Certain processing stages which should not adversely affect a PDF/A file under normal circumstances (such as inserting extra pages) may, in rare cases, cause a PDF/A to become invalid. Validation will clarify the situation for you.

If the validation process detects a violation of the PDF/A standard, the file can often be repaired using the appropriate software. If this is not possible, the only other option is to recreate the PDF document from scratch.

## Finding the right validation solution

Several validation programs are available on the market. As with PDF/A creation software, you can choose between workstation applications, server-based solutions and modules for workflow systems. PDF/A files can also be validated using programming libraries.

Some PDF/A creation tools can also perform a test after conversion to ensure the result meets the ISO standard. Naturally, these solutions can also validate PDF/A documents delivered from elsewhere.

Adobe Acrobat Pro, already used in many industries, can test almost all PDF ISO standards including the three parts of PDF/A, using its Preflight function.

**Note regarding process validation:** during an automated stage of processing, such as scanning to PDF/A, the process as a whole is validated rather than each individual PDF in the process.

# PDF/A in public administration

Many government authorities and public institutions worldwide now specify formats to use for digital data. Government offices often recommend that working documents use open file formats. More and more often, PDF/A is the only format accepted for final-version files.

■ **EU Publications Office:** The EU Publications Office is tasked with providing access to all laws, declarations and publications. Since 2007, the EU Digital Library has been tasked with storing printed texts – some of which date back to 1957 – in digital form as well. In a pilot project, an external digitisation team took two years to turn 130,000 paper documents in eleven languages into PDF/A-1b files with searchable text. An important factor in choosing PDF/A was that XMP metadata can be used for keywords and other bibliographic information. To simplify print-on-demand book orders, the archive files are now also available in the ISO standard format for digital print data, PDF/X-3.

■ **The European Patent Office:** Since April 2010, the European Patent Office has published patent documents not just in PDF format, but also in PDF/A. For the Patent Office, an important feature of the PDF/A format is found in the way it uses metadata: the XMP metadata fields can include the publication number, the patentee and the international patent classification.

■ **“Comply or Explain” in the Netherlands:** The government of the Netherlands has a “Comply or Explain” policy regarding open standard software. The national action plan “Nederland Open in Verbinding” enforces the use of open standards and requests the use of standard file for-

mat, namely ODF, PDF and PDF/A. All public institutions in the country must use open standard software, as must all companies which take on public contracts. Any entity which cannot meet these requirements must fully justify this decision. In many cases, it is generally easier and ultimately more cost-effective to switch over to a standardised process.

■ **Brazil:** In 2007, the Brazilian government introduced the e-PING architecture which regulates the provision of digital services. For final versions of a document to be transmitted or archived, Brazil prefers PDF/A.

■ **Denmark:** Since April 2011, all Danish government bodies are required to save non-editable documents in PDF/A format.

■ **France:** Since early 2009, the French authorities have recommended the ISO’s PDF/A standard for archiving administrative documents with static, unchanging content.

■ **Switzerland:** Due to archiving requirements, all electronic communication between citizens and administrative authorities is required to use the PDF/A file format. This regulation has been in force since 2008.

■ **Germany:** German registry offices have run an electronic register of births, marriages and deaths since 2009; for registered data, they use PDF/A and XML. By 2014, these offices are expected to have switched over to an all-digital system.

For further user reports and the latest PDF/A recommendations, visit the website of PDF Association: [www.pdfa.org](http://www.pdfa.org).



*The EU Publications Office in Luxembourg.*

Libraries and archives are taking a leading role in implementing and developing PDF/A. In the USA and Europe in particular, these institutions are choosing the ISO standard for long-term archiving.

# PDF/A in finance and industry

Businesses benefit from the ISO standard for long-term archiving because it helps them to store digital documents in compliance with legal requirements. PDF/A-3 has further increased adoption rates within the financial sector, as this new part of the standard can also be used to keep source documents organised (see page 8).

## Industry documentation

The aeroplane manufacturer Airbus was a pioneer in the use of PDF/A. Aeroplane blueprints must be preserved for at least 99 years. Back in 2002, one of the manufacturer's working groups recognised that PDF was in some respects well-suited to long-term archiving, but also that it contained a number of problematic functions. The team therefore first developed a "minimal PDF" which was used for digital archiving until PDF/A became available.

The construction industry has recognised the advantages of features like PDF/A-3's "container" functionality. Mechanical engineering companies, for example, can preserve original 3D models in any format as part of the PDF/A-3 file.

NIRMA (the Nuclear Information and Records Management Association) also recommends PDF/A when working with nuclear technology in the United States. The US-based energy provider Southern Co. has been using PDF/A for years to ensure that all digital documents relating to nuclear installations will remain readable into the future.

## Banking and insurance

PDF/A helps the German health insurance provider Techniker Krankenkasse with a bonus programme for its customers. It begins with the company scan-

ning existing bonus booklets in colour. This file is used to create a PDF/A with compressed images and searchable text, which can then be archived and sent onwards.

Helaba, the state bank of Hesse and Thuringia, uses PDF/A to handle incoming post and to archive emails. It also stores digital credit documents in PDF/A format.

The banking and insurance sector often requires credit and insurance files to be retained for 50 or more years.

## Healthcare

As a rule, documents such as doctors' notes, statements, lab reports, and X-ray and tomographic images must be retained for 30 years or more.

The medical centre at Greifswald University Hospital, for example, uses PDF/A to archive its patient records – including digital signatures and timestamps. Due to requirements of legal certainty, digital signatures play a critical role in medical statements.

PDF/A is used in doctors' practices as well as in clinics. The Lake Constance Radiation Oncology Centre uses PDF/A to process and archive digital patient records.

## E-invoicing

The standardised document and data format ZUGFeRD was created to make it easier to exchange digital invoices. It is the result of a joint initiative between BITKOM (Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.) and FeRD (Forum elektronische Rechnung Deutschland). The ZUGFeRD exchange format also uses PDF/A-3. It embeds invoice data in XML format to allow the recipient of an invoice to process it automatically.



# PDF/A in legislation and justice

Digital documents are increasingly replacing traditional paper documents. Existing paper documents are being scanned and digitised, while digital-only processes are becoming more and more common – and PDF/A is at the very forefront of this trend.

PDF/A also plays a significant role in legislation and justice in many countries. Legal bills and court records usually have to be stored for exceptionally long periods of time. The ability to search text and XMP metadata in PDF/A files can make it much quicker and easier to find and allocate digital records.

## Federal jurisdiction in the United States courts

To see the huge significance of PDF/A in the legal field, you need only look at the Administrative Office of the United States Courts, which is taking a leading role in standardising PDF/A. Work has been ongoing since 2002 with the American associations AIIM (Association for Information and Image Management) and NPES (the National Printing Equipment Association) to develop the standard for archived



documents. The goal was to turn large quantities of paper-based documents into a reliable, future-proof digital format with no expensive specialist viewing software required.

## The Italian Chamber of Commerce

Since 2010, Italian businesses have been required to send reports to the appropriate commercial register in PDF/A format. This includes balances, certificates and reports of business transactions, acquisitions, mergers and insolvencies.

A data transfer platform is used for input; software is used to convert text or existing PDF documents to PDF/A-format certificates. The CGN, a specialist network for the financial, legal, fiscal and labour sectors, was closely involved in establishing the platform.

## Austria: BAIK

The “Bundeskammer der Architekten und Ingenieurkonsulenten” in Austria, the Federal Chamber of Architects and Consulting Engineers, requires publicly available digital certificates to conform to the PDF/A-1b standard. This guarantees the authenticity of all digital documents accepted into the title register, thanks to a qualified digital signature.

## Germany: Land registration

The “Decree of the Baden-Württemberg Ministry of Justice for the introduction of digital legal procedures and digital records of land registration procedures” (ERGA-VO) requires that all digital data (ASCII, Unicode, RTF, PDF, TIFF, Word) submitted must be convertible to PDF/A. This decree has been in force since March 2012.

# What the users and experts say



**Stephen Levenson**  
U.S. District Courts:

“PDF/A now provides a full decade of design ideas from the best and brightest digital preservation practitioners. Rich opportunities exist for having all the features of PDF that you have come to expect in presentation and now with PDF/A-3 the ability to have machine-processible data as XML or text. PDF/A-3 will also allow for keeping the original editable version of your content.

Most of the world has adopted PDF/A as their long-term static format. It has become the true replacement for paper, as its designers envisioned. Other formats will result in the loss of content. It used to be said, no one was ever fired for picking IBM. It will be said some day, no one was ever fired for picking PDF/A. It is that dependable.”



**Anton Zagar**  
EU Publications Office:

“The European Union Publications Office stores its digital archive of over 150,000 publications, some of them stretching back to 1952, in the PDF/A format. We also use the PDF/A format to publish the Official Journal of the European Union in 23 languages every day: in 2012 alone we produced 1.2 million pages.

The Office has set itself the goal of making all publications, by all bodies of

the European Community and the European Union, available in digital form.”



**Kai Volmar**  
Landesbank Hessen-Thüringen (Helaba):

“In the world of IT, sometimes you don’t want to be the first to use a new technology. But the advantages of PDF/A convinced us straight away, when compared with the TIFF format we had previously used. As a result, the decision in 2006 to use PDF/A to digitise our records was not a hard one.

Meanwhile, the state bank of Hesse and Thuringia now uses nothing but PDF/A for digital archiving, whether the documents first need to be digitised or were born digital in the first place.”



**Jacob Bielfeldt, Techniker Krankenkasse:**

“In an internal workshop in 2006, the Techniker Krankenkasse identified PDF/A as a promising future-proof document format; today, this is confirmed by the many advantages PDF/A brings.

The Techniker Krankenkasse is introducing PDF/A into its ongoing projects step by step. Our first project was to digitise staff records; our second was to use PDF/A in output management. We have increasingly used PDF/A for input management since 2011 and are planning to use PDF/A further.”



# PDF/A and the other PDF standards

Specialist ISO standards based on the Portable Document Format are available for a wide range of purposes.

## PDF/X

Back in 2001, an ISO working group developed a pre-press PDF standard, “ISO 15930”. At this time, customers usually sent printers “open files” from layout software. This method, however, always carried the risk of fonts and images going missing. PDF/X is able to eliminate all of these problems; it also has the advantage of carrying reliable colour information thanks to colour management settings.

The “X” identifier stands for “Exchange”, as PDF/X is intended for reliable print data exchange. Additional standardisation for PDF/X versions 4 and 5 has taken into account the newer features available to the PDF file format, including transparent elements and JPEG2000 image compression. PDF/X-5 also supports externally referenced elements.

## PDF/A

PDF was also recognised early on as having great potential for archiving digital documents. In 2005, the ISO published the first part of the PDF standard for long-term archiving, PDF/A.

## PDF/E

This standard has been available since 2008 as “ISO 24517”; it is aimed at engineering documents such as construction drawings. The original data often comes from CAD software used for digital drafting. PDF/E can display rotating and folding 3D objects on-screen, using tools like the free Adobe Reader.

## PDF

PDF itself was also standardised in 2008 as “ISO 32000”. The basis of the standard was the then-current PDF version 1.7. With this, PDF became an open standard. PDF 2.0 is expected to be published in 2014.

## PDF/VT

PDF/VT is a standard based on PDF/X-4 and PDF/X-5, supporting variable data printing. It was published in August 2010.

The abbreviation “VT” stands for “Variable data and transactional printing”. This includes invoices and personalised advertisements, for example.

## PDF/UA

The PDF/UA (Universal Access) standard, approved in 2012, allows universal access to PDF files’ content. This is useful for users with disabilities (for example the partially sighted) and others.

Of particular importance is a clear coherent logical structure of the PDF’s elements, to ensure that navigational aids, reading software or Braille displays can handle all content including text, images and diagrams.

PDF/UA builds on proven concepts for accessible web content and adds concrete demands on the semantic structure of PDF documents (which PDF/A Conformance Level A had previously only given in a very general sense).

PDF/UA offers users with disabilities the best possible access to content. It also makes it easier for mobile devices to use this content and supports its flexible reuse in other forms of presentation.

**PDF/X** since 2001

“Prepress digital data exchange using PDF”

ISO-Standard for the printing industry

**PDF/A** since 2005

“PDF Archive”

Standardised long-term archiving with PDF

**PDF/E** since 2008

“PDF Engineering”

Construction diagrams with moving 3D models where required

**PDF** since 2008

“Portable Document Format”

The ISO standard corresponds with PDF version 1.7

**PDF/VT** since 2010

“PDF for Variable Data and Transactional Printing”

Used for variable data printing

**PDF/UA** since 2012

“PDF for Universal Access”

ISO standard for universally accessible PDF documents

# The myths and legends surrounding PDF/A

A number of critics have spoken out against PDF/A, especially when the standard was first introduced. Many criticisms of the format, however, are based on misunderstandings.

These are some of the most commonly encountered myths and legends:

■ **PDF/A files are too large:** PDF/A actually allows exceptionally small file sizes thanks to its sophisticated use of powerful compression algorithms such as JBIG2 and JPEG (and JPEG2000, from PDF/A-2 onwards). Embedded fonts can slightly increase the size of a PDF/A file. When archiving a very large number of individual, fairly similar documents, this can in some cases (such as for mass mailings) prove problematic.

■ **PDF/A is not as revision-safe as TIFF:** TIFF files are easier to alter than PDF and PDF/A documents. In any case, however, revision safety is not achieved through your choice of file format. It can only be achieved by using an appropriate document management or archiving system.

■ **PDF/A does not allow signatures:** Quite the opposite. PDF/A expressly supports embedded digital signatures. PDF/A-2 requires PAdES-standard compliance here.



■ **Links are not allowed:** This claim is also false. Hyperlinks are allowed in principle. The PDF/A standard sets no require-

ments as to whether an external link should lead to a valid destination.

■ **PDF is a proprietary format:** PDF was originally developed by Adobe Systems, but since then PDF (ISO 32000) and PDF/A (ISO 19005) have become ISO standards. TIFF, on the other hand, is a specification belonging to Adobe Systems alone, and it has not achieved the status of ISO standard.

■ **Scanned documents cannot be searched by text:** PDF/A permits text recognition processes, meaning that even scanned PDF/A documents can be searched.

■ **PDF/A is not supported by DMS systems:** Any ECM system which works with PDF can also handle PDF/A in principle. Many DMS suppliers offer solutions which support PDF/A.

■ **PDF/A does not allow metadata:** Not at all: PDF/A specifically requires embedded standardised metadata corresponding to the modern XMP metadata standard, which was published in February 2012 as “ISO 16684-1”. XMP metadata can be directly embedded into the PDF/A document.

■ **PDF/A is not globally relevant:** This statement is false. Although the very first PDF/A initiatives and products did come from German-speaking countries, the ISO standard has since become a recommendation or even a legal requirement in many countries and industries.

■ **PDF/A is expensive to implement:** Yes and no. Implementing PDF/A solutions and training staff will incur costs at first, but these investments very often pay for themselves within months.

# Further information on PDF/A

The PDF/A Competence Centre, today a part of the PDF Association, was founded very shortly after PDF/A first appeared as an ISO standard. This international organisation aims to promote the development and usage of PDF standards. To that end, the PDF Association targets users, developers and decision-makers equally and helps its members exchange information worldwide.

## The portal to the PDF Association

A good starting point for anyone with questions about PDF/A or PDF in general is the PDF Association's website. At [www.pdfa.org](http://www.pdfa.org), you can find information in English and German about current development, from all relevant industries and from suppliers worldwide. Comprehensive background information and example applications explain the technology behind PDF/A and its use in practice. You can view a video-on-demand series about PDF/A and other PDF standards. The website offers an overview of PDF- and PDF/A-related software products and services.

Users can contact the Association with specific questions. To do so, simply register on the website and describe your request on the discussion forum. Specialists and practitioners from around the world will provide informed answers and suggestions.

## PDF Association events

The PDF Association attends international trade shows and other events on subjects such as document management, digital media and electronic archiving. For years the Association has also organised specialist seminars and technical conferences around the world. Companies and public author-

ities seeking a speaker on PDF/A for their event can find support at the PDF Association: interested parties can simply enquire about a presentation using a form at [www.pdfa.org](http://www.pdfa.org).

Some countries also have direct contact persons in the Association's local chapters. For a complete list, including contact details, please visit the Association's website.

## Membership

Anyone aiming to work actively to develop and expand use of the PDF standard can become a member of the PDF Association, whether an individual or an entire organisation.

Membership allows a company to present itself on the PDF Association's portal and to publish its own announcements, press releases and articles on the website. They can also present their software solutions and services within the Association's product showcase. The Association also has especially favorable terms for presenting products and strategies at trade shows and other events. Members also have exclusive access to the Association's intranet.

The PDF Association's website can be found at [www.pdfa.org](http://www.pdfa.org).

# PDF/A in a Nutshell 2.0 – PDF for long-term archiving

PDF/A is an ISO standard for using the PDF format for long-term archiving of digital documents. Since its publication in 2005, PDF/A has become the format of choice for archiving digital documents in a wide range of industries and applications. “PDF/A in a Nutshell 2.0” provides a comprehensive introduction to the material and shows off the latest developments available with PDF/A-2 and PDF/A-3. The brochure provides information about PDF/A tools and strategies for creating and validating PDF/A files. Examples from around the world demonstrate how users in the areas of finance, administration, academia and law can benefit from PDF/A.

## Contents:

- Facts about PDF/A
- The history of PDF/A
- The technical side
- Who can benefit from PDF/A and why
- Typical applications
- PDF/A creation tools
- PDF/A validation
- PDF/A in public administration
- PDF/A in finance and industry
- PDF/A in legislation and justice
- What the users and experts have to say
- PDF/A and the other PDF standards



## About the author

Alexandra Oettler has worked for years as a freelance journalist in the areas of software, print and media. Her work is regularly published in specialist journals on the subject of prepress in practice, software technology and financial developments in the publishing sector. She regularly writes news and background reports for the online editions of several journals. She also was one of the co-authors of the first edition of “PDF/A in a Nutshell”.