

PDF/A *in a Nutshell*

Long Term Archiving with PDF

Olaf Drümmer, Alexandra Oettler, Dietrich von Seggern



Olaf Drümmer, Alexandra Oettler, Dietrich von Seggern

PDF/A *in a Nutshell*

Long-Term Archiving with PDF

Olaf Drümmer
o.druemmer@callassoftware.com
Alexandra Oettler
pdfakompakt@alexandra-oettler.de
Dietrich von Seggern
d.seggern@callassoftware.com

ISBN: 978-3-9811648-1-7

Bibliographic information published by *Die Deutsche Bibliothek*
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie.
Detailed bibliographic data is available at <<http://dnb.ddb.de>>.

This work and all its parts are protected by copyright. All rights, including translation, reproduction, presentation, use of illustrations and tables, radio broadcasting, microfilming, any other means of replication, and storage in data processing systems, are reserved. This also applies to extracts. Any replication of this work or of parts thereof, even in isolated cases, is only permissible in accordance with the currently valid version of the German copyright legislation of September 9th 1965. A copyright fee must always be paid. Violations fall under the prosecution act of German Copyright Law.

© 2007 callas software GmbH, Berlin
Published by Association for Digital Document Standards ADDS – PDF/A Competence Center, Berlin – www.pdfa.org
Translation: © 2007 Association for Digital Document Standards ADDS – PDF/A Competence Center, Berlin

Printed in Germany

The use of general descriptive names, trade names, trademarks, and so on, in this publication, even if not specifically identified, does not imply that these names are not protected by the relevant laws and regulations or that they can be used by anyone.

Layout, design, and composition: Alexandra Oettler; Cover design: Anja Godolt; Cover picture: Sepp Huberbauer – photocase.com/de
Printing: Galrev Druck- und Verlagsgesellschaft Hesse & Partner OHG

Preface

Our world is getting more digital by the day. A lot of information and documents only exist in digital form today, but will they still be legible „tomorrow“? That was the theme of an interesting TV show appropriately called „The Digital Disaster“. It began with cave drawings from the stone age and papyrus rolls from ancient Egypt, both of which have survived as documents for thousands of years. What documents from the 21st century will future generations be able to find and still read? But it's happening much quicker than you may realize. I always carry a 3½ inch floppy disk in my pocket, and it demonstrates a lot of the problems of long-term archiving. It begins with the hardware: where can you buy a 3½ inch floppy disk today? And even if you find one, there's a good chance that the disk is physically damaged. If these two hardware hurdles are successfully cleared, then what kind of software or document will we find on the floppy disk? Are the appropriate viewing and processing programs still available? And this example is a mere 15 years old!

My short anecdote leads us to the demand on the long-term archiving of documents. Electronic archiving is critical for businesses and organizations, because documents today often only exist in digital format. The length of time that business documents have to be archived varies from sector to sectors and country to country, but some examples can help us to get an idea. Federal laws often require an archiving period of around 10 years. Banks and insurances demand that customer dossiers be retained for more than 50 years. In the engineering branch, archival periods of 100 years are common for aircraft, bridges hopefully hold a whole lot longer.

And saving documents in proprietary formats for this length of time is really not a good idea. This leads to the second problem with the digital document world - that many users already have a real „format zoo“, which can quickly become unmanageable (if it isn't already so). Proprietary document formats

have to be migrated on a regular basis, in order that newer versions of the processing software can still read them.

Employees working on customer dossiers aren't really impressed when 10 different viewing programs are opened up at the same time. In some of the programs they might not even know how to navigate around in a document. In order to solve this problem, a document and archiving format is needed that guarantees the required long-term archiving period and offers the option of a single format type.

This is where PDF/A as an ISO standard for long-term archiving enters the stage. The „A“ stands for „Archive“ and the PDF/A standard was specifically created for long-term archiving. It envisions a single PDF/A archive for all documents in an organization, from input through to output, and includes all of the areas inbetween.

You will find many more advantages to PDF/A on the following pages, written with the aim of converting the very formal ISO standard into a form that is easily understood and enhanced with practical examples. Since PDF/A resolves a lot of the critical problems that users have, the PDF/A Competence Center was formed as an association with the aim of providing information over PDF/A, promoting the distribution of the standard, and acting as a central point of contact for your questions dealing with PDF/A. We hope that this booklet gives you a good overview and introduction to PDF/A, and also helps as a motivator for implementing the standard.



Berlin, in September, 2007
*Thomas Zellmann,
Chairman PDF/A Competence Center*

PS: a special thanks goes out to our member callas software GmbH, who initiated the German version of this booklet and provided it to the PDF/A Competence Center for translation into English and for further distribution.

Preface



Throughout history, it has always been important to preserve our past for future generations. Until the last 20 years in our paper centric world, this was a fairly easy task. One would simply take the folders of papers or other objects that were to be preserved and send them off to an archive for safe keeping or place them in a fire retardant container. With electronic documents this task is not as easily approached, which is how PDF/Archive or PDF/A came into being.

PDF/Archive addresses the growing need to electronically archive documents in a way that would ensure preservation of their contents over an extended period of time. Additionally, it ensures that the documents will be able to be retrieved and rendered with a consistent and predictable result each time they are viewed.

AIIM, the Enterprise Content Management Association, and NPES – The Association for Suppliers of Printing, Publishing and Converting Technologies were approached by numerous organizations which were being faced with the need to preserve over long periods of time, large quantities of electronic documents. After reviewing the options of maintaining this electronic history in TIFF, XML, native format or PDF, it was decided that PDF would be the best format as it would enable the accurate rendering of the document as it had been intended to be displayed. However, in order to ensure the long term preservation of the electronic documents, PDF would need to be enhanced slightly.

The joint effort of AIIM and NPES brought together the document and content management experts with the graphics experts who had already developed the PDF/X family of standards. When we announced the proposed work to develop a subset of PDF tags for long-term preservation of electronic documents, we were overwhelmed by the interest to participate from virtually every area in the world.

AIIM's expertise as an accredited standards developer and the secretariat of ISO TC 171, Document Management Applications and ISO TC 171 SC2, Document Applications, AIIM brought to the project the means for gaining ISO approval and wider adoption of the standard. ISO 19005-1, Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1) became an approved ISO standard within 22 months of introduction as a new project through the dedicated efforts of many records managers, archivists, software developers and end users.

While adoption of the standard has been a little slower than we had anticipated, we are encouraged by the continuing interest and growing adoption of the standard. This book along with the continuing efforts of AIIM and the PDF/A Competence Centre will continue to increase the adoption rate of PDF/A in the industry.

Silver Spring, in September, 2007

Betsy Fanning

AIIM, Director, Standards

Table of Contents

Durable documents with the PDF/A standard

Open files are not always complete	9
TIFF as an archive format	9
PDF data containers	10
Why PDF/A and not PDF?	11
The introduction of the PDF/A standard	11
How to create archive PDFs	12
Who stands to benefit from PDF/A?	13
Table: Comparison between PDF/A-1a and PDF/A-1b	15
Overview: Which file formats are suitable for archiving?	16
Is XPS an alternative to PDF/A?	18



PDF/A creation: Analog, digital, and mass processing

PDF/A from scanned documents	21
Scanning options in Acrobat 8 Professional	22
Converting pages that have already been scanned to PDF/A	23
The Distiller engine	25
PDF/A document generation using the Distiller	25
Office and administration	28
PDF/A in Office 2007	28
Office 2003 and the PDFMaker	29
PDF/A using the 3-Heights PDF Producer	31
PDF/A 'en masse'	32
PDF/A 'from nothing'	32
Creating PDF/A from print data streams	33



Illustrations: PixelQuelle.de



Illustrations: photocase.com/de

From PDF to PDF/A: Converting PDFs to archive PDFs

PDF/A generation with Preflight	34
---------------------------------	----

Converting PDF to PDF/A with pdfaPilot	37
--	----

Is this really a PDF/A file? PDF/A validation

Validation with Preflight	39
---------------------------	----

pdfaPilot PDF/A	41
-----------------	----

Archive PDFs in everyday life: What issues might arise?

Images	42
---------------	----

Resolution is not part of the PDF/A standard	43
--	----

Permitted and prohibited compression types	43
--	----

Transparency	44
---------------------	----

Colors	46
---------------	----

Fonts	48
--------------	----

Metadata	50
-----------------	----

PDF/A and metadata	50
--------------------	----

Accessibility	52
----------------------	----

Creating an accessible PDF file from Word	54
---	----

Interactive PDF files	56
------------------------------	----

Comments and annotations	56
--------------------------	----

Forms	58
-------	----

Embedding fonts for PDF/A forms	59
---------------------------------	----

PDF/A for design drawings	60
----------------------------------	----

Electronic signatures	61
------------------------------	----

Security levels	62
-----------------	----

Digital signatures in PDF with Acrobat	63
--	----

Challenges in practice	64
------------------------	----

The outlook: PDF/A in the future

Enhancements in PDF/A-2	65
Looking towards PDF/A-3	66
PDF/A-1 developments	66
PDF/A in one hundred years time	67

What the error messages mean

Preflight results and troubleshooting for PDF/A	68
---	----

Glossary

Explanation of terms relating to PDF/A	80
--	----

About:

The PDF/A Competence Center	86
------------------------------------	-----------

AIIM	87
-------------	-----------



Sepp Huberbauer – photocase.com/de

1.

Durable documents with the PDF/A standard

There are certain documents that people want to keep because of their sentimental value: Love letters, photographs of their first day at school, or holiday snaps, for example. Other documents have to be kept for legal reasons. These document include birth certificates, academic certificates and reports, invoices that are needed for tax purposes, insurance documents, and contracts.

In the days when everything existed on paper – in the pre-digital era – the main problem was remembering which index file, folder, or shoe box you'd used to store your letters or contracts. In today's world of digital documents, the task of archiving is fundamentally different. Thanks to search functions or database solutions, even the most forgetful of us can easily find a par-

ticular document or photo on our computers. In addition, any possible space problems can be solved simply by purchasing additional RAM. However, there are certain risks and uncertainties that might influence the shelf life of digital documents. These risks do not only arise from the physical durability of the data carriers used although it is clear that magnetic tape, CD-ROMs, and DVDs will not necessarily last any longer than paper and ink. However, photographic prints dating from 1900 still exist today. Still, it's debatable whether or not we will similarly be able to view the millions of digital snapshots being taken and stored on mobile phone memory cards all over the world in, for example, 2107.

In addition to the restrictions imposed by the limited lifetime of data carriers, the



Markus Imorde – photocase.com/de

document format and software used also present a considerable challenge for the durability of electronic documents. Yesterday's, today's, and tomorrow's software

It's a common problem: Opening old documents in brand-new programs doesn't always work. The rate of success for the opposite direction (new documents in old programs) is even less encouraging. Software developers do try to achieve backward compatibility that enables files that are, say, five years old to be opened using a current program release. However, this can change the layout and page rendering, meaning that not everything is displayed exactly as it ought to be. More recent software tends to generate documents with additional features that older versions may not be able to display. In some cases, it is not even possible to open current files in previous versions of a program. For example, whereas a Microsoft Word 95 file can normally be opened in Word 2003, it is not possible to open a Word 2003 document in Word 95.

Because software production cycles are becoming ever shorter – one major release per year is not unusual – the challenge that arises from new pro-

gram developments is greater than that caused by the aging of storage media. The successful long-term archiving of digital files is at least as threatened by the constant rollout of new program versions as by damaged data or data carriers.

Open files are not always complete

File formats are not all equally suitable for the long-term, secure archiving of content. If it is not possible to store all the elements required for the complete display of content in a file format – graphics and fonts as well as text – then the possibility of stumbling blocks when it is attempted to use the file later on cannot be ruled out. If, for example, the program used cannot find linked external images, a page cannot be

displayed as required. Instead, the frame where the image should appear displays only a rough preview of the image or a question mark. The problem of open files for which not all illustrations and fonts are available has been causing irritating delays for printers and their suppliers for a long time. However, the introduction of PDF, a format that can store all the components required for a printed document, has greatly simplified work in this area. In addition, layout files such as XPress or InDesign are now becoming increasingly less common in printers' archives. Instead, printers are storing the actual PDF documents that were used for the printing task.

TIFF as an archive format

For a long time, many public authorities and companies that need to store large quantities of correspondence, records, invoices, contracts, and similar information

in digital archives have been using the pixel image format TIFF (Tagged Image File Format). This format digitalizes templates containing text and images pixel by pixel. TIFF is an estab-

lished image file format that has both advantages and disadvantages. Pixel-based formats store the appearance of templates. Problems with missing graphics and fonts do not occur, since the format stores all of the template elements as an image. Since TIFF is widespread and is subject to few file handling complications when upgrading to a new program version, many users believe that the future of the format is guaranteed. However, while TIFF may indeed be a de facto standard, it is not an official norm for safe archiving. Other disadvantages include the relatively large file size and the fact that scanned texts cannot be searched without OCR (text recognition), since this format converts them to image elements. →

TIFF-G4 – a black and white TIFF variant that works with a compression method developed for fax technology – is commonly used for archiving.



One of the reasons for the global success of PDF must be considered to be the free availability of the Adobe Reader. This PDF viewer is available for download from Adobe Systems' Web site in many language versions for numerous commonly used platforms:
www.adobe.com/products/acrobat/readstep2.html

PDF data containers

The development of PDF (Portable Document Format), which Adobe Systems has been promoting since 1993, has significantly simplified data management and exchange for a great number of users from completely different fields. PDF allows obstacles that can arise during the transmission or storage of files to be neatly avoided.

■ PDF files can be opened on all established operating systems. Free PDF readers are available for all of the important platforms including Windows, Apple, Linux, and mobile devices.

■ With this format, the document layout is true to the original. Since PDF can incorporate different types of content such as text (and the relevant fonts), images, and graphics, nasty surprises relating to miss-

ing illustrations or incorrect fonts – like those that occur when Word documents are opened on another computer – are not usually a problem.

■ PDF is an open format. This means that companies other than Adobe Systems (who invented PDF) can develop software for creating or displaying PDF. The 'release' of PDF by Adobe has brought independence for both users and developers and, as a result, there is a high probability that there will still be programs for generating and displaying PDFs in decades to come.

So, can users who want to keep documents such as contracts or invoices for long periods of time trust in PDF to make sure that their documents will work just as well in ten, fifteen, or one hundred years time as they do today? It might well be the case that PDF files created today will still work without any significant problems in 2017. How-

PDF specifications:

Since it was introduced at the start of the 1990s, the PDF file format has been in a state of constant development. The current PDF specification is version 1.7, which was introduced with Acrobat 8. Today, it is extremely rare to come across PDF files with a version number lower than 1.3, and modern PDF generation programs only have backward compatibility to version 1.3 at the most.

With each PDF version, Adobe Systems publishes a reference that describes the features and functions of the version in detail. The specification history contains 'milestones' – important features that were introduced with the new version. Some of these milestones are listed below.

Acrobat 1 (1993, PDF 1.0): PDF 1.0 incorporates most of the functions offered by the page description language PostScript Level 2. All basic functions for text, vector graphics, and raster graphics are available.

Acrobat 2 (1994, PDF 1.1): This version supports the Lab color space and CalRGB. It also supports TrueType fonts.

Acrobat 3 (1996, PDF 1.2): This version enables color separation and supports Unicode and CID fonts (Chinese, Japanese, and Korean). It also supports ZIP compression.

Acrobat 4 (1999, PDF 1.3): PDF 1.3 contains the complete PostScript Level 3 graphics model. It enables multi-channel color spaces (DeviceN) and supports ICC profiles for the reliable reproduction of colors. It introduces smooth shades and page geometry boxes, which are useful for prepress processes (TrimBox, CropBox, and BleedBox).

Acrobat 5 (2001, PDF 1.4): From this version, PDF files can contain transparency. This version also introduces 'tagged PDF' (= structured PDF), which enables content accessibility. The security options are enhanced with this version. In addition, the image compression type JBIG2 is supported.

Acrobat 6 (2003, PDF 1.5): With this version, PDF documents can contain layers (also called 'optional content'). JPEG2000 image compression is supported.

Acrobat 7 (2004, PDF 1.6): This version supports OpenType fonts. With this version, 3D content can be inserted. Users can create virtual page sizes with edges of up to 381 km in length.

Acrobat 8 (2006, PDF 1.7): Unicode path specifications simplify the correct specification of links, even across international language systems. The new Acrobat 'PDF packages' function allows several independent PDF documents to be forwarded in a single file. The recipient requires Acrobat or Reader 8.

ever, only the new PDF/A standard can guarantee that users will be able to view exactly the same content as when their documents were created. This format brings the kind of legal certainty that can be decisive in many business and administrative contexts.

Why PDF/A and not PDF?

Why has a special PDF standard now been defined for archiving documents? Are traditional PDF documents not 'good enough' for long-term archiving? PDF has some excellent characteristics that lend themselves to the creation of archive documents. Like a container, a PDF can incorporate completely different elements such as text, images, and fonts. In addition, it reproduces layouts that are true to the original and is cross-platform capable. However, certain requirements must be met in order to enable the exact reproduction of content.

Required: One 'must' is that users require full access to all elements belonging to a document. For example, fonts must be embedded – a link to the font in question is not sufficient. This means that if, in 10 years time, a user who tries to open a document does not have a required font on his or her computer, special characters or symbols will not be displayed correctly.

Prohibited: In addition, some PDF features must be avoided. Such elements are prohibited because they would undermine the required document durability, and include interactive elements and PDF layers. These features inhibit the unambiguity that is required from an effective PDF/A file. For example, in the case of a PDF document with layers, users printing it out in 50 years time might well ask themselves which layers are valid and which are not. This kind of decision needs to be made now – when the PDF is created.

A PDF/A document is basically a traditional PDF document that fulfills precisely defined specifications. In order to prevent

users from repeatedly having to test and discuss the best appearance of a well-functioning archive PDF, industry experts decided in 2002 to work together to develop the PDF/A standard.

The introduction of the PDF/A standard

The PDF/A standard for long-term archiving was adopted by ISO (International Organization for Standardization) in autumn 2005. The PDF/A standard was published with the number 'ISO 19005-1:2005' and is based on PDF specification 1.4. An additional part, PDF/A-2, is currently being prepared. This part shall refer to PDF Version 1.7.

The PDF/A standard aims to enable the creation of PDF documents whose visual appearance will remain the same over the course of time. These files should be software-independent and unrestricted by the systems used to create, store, and reproduce them. As far as PDF/A is concerned, practice soon caught up with theory. While Acrobat Professional 7 contained only 'draft' PDF/A functions, Acrobat 8, which has been available since the end of 2006, now offers creation and verification features that comply with the adopted standard.

Many new PDF/A tools and solutions for creating and verifying files have entered the market since the introduction of the standard – from 'small' tools for individual users who want to create PDF/A documents every now and again to extensive server solutions that can create a hundred thousand archive documents from databases in just a few hours time. →



ISO is an international organization for standardization, active primarily in technical and electronic fields. The PDF/A standard was developed by industry and development experts.



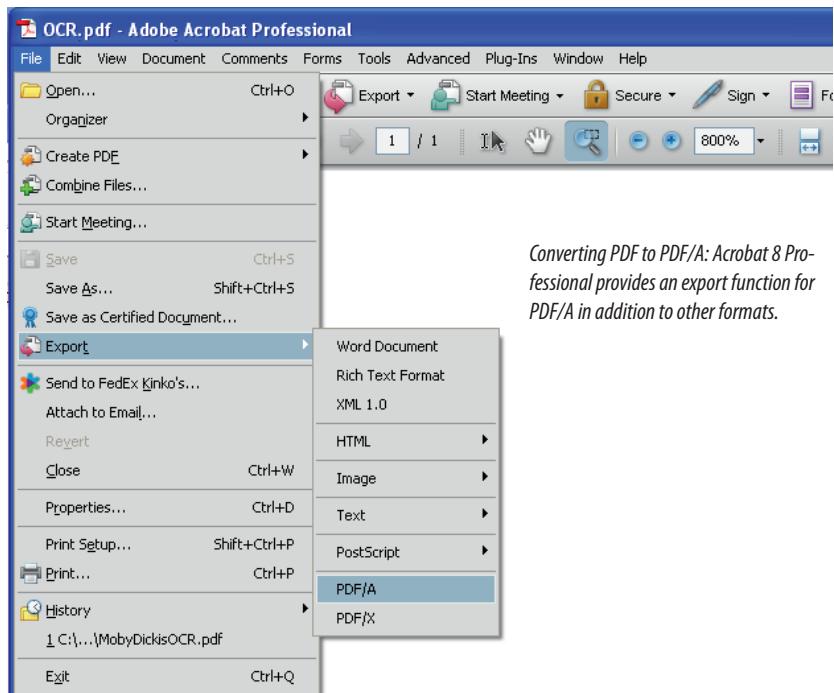
International companies and experts from the field of PDF technology have joined forces to form the PDF/A Competence Center. It aims to promote the exchange of information and experiences relating to long-term archiving. Users can visit www.pdfa.org for up-to-date advice and background information as well as a discussion forum on PDF/A.

PDF/A has two levels of compliance:

PDF/A-1a (Level A) applies to semantic correctness and structure. Each character must have a Unicode equivalent. The structure is expressed by tags.

PDF/A-1b (Level B) applies to visual integrity.

Any file that meets the requirements for PDF/A-1a will also comply with PDF/A-1b, which is less strict.



Converting PDF to PDF/A: Acrobat 8 Professional provides an export function for PDF/A in addition to other formats.

How to create archive PDFs

There are many different conditions that might be encountered when creating PDF/A files. The process differs depending on whether existing PDF documents are already available or whether they need to be

Converting paper documents to PDF/A: Scanned documents can be automatically given searchable text following digitalization. Text recognition software is used for this.

generated from working files such as Word or PowerPoint files.

■ **PDF/A files from files or data:** This field relates to newly created PDF files from applications including word processing, image editing, and layout programs. The process can be realized by means of a PDF export from the source program, Acrobat Professional, Distiller, or other PDF converters. For the mass conversion of content to PDF/A, there are program modules that can convert database content or print data streams to the PDF/A format.

■ **Converting scanned paper documents to PDF/A:** Often, documents that exist only on paper, such as contracts, invoices, and books, need to be digitalized using a scanner. Over the past years, the results of the scanning process have usually been stored as Bitmap TIFFs. However, PDF is increasingly being used for scanned documents, and before long the majority of scanned files will probably be stored directly as PDF/A files. For example, users can scan paper documents using Acrobat 8 Professional and save them as PDF/A files. It is often possible to make the text in a PDF/A file searchable using the text recognition function (OCR). Images and histori-



Dirk Herold – photocase.com/de

cal documents can also be scanned for conversion to PDF/A. Solutions and services for mass processing are available for users who wish to scan a large number of pages or documents.

■ Creating PDF/A from PDF: Many users already have PDF documents that are not PDF/A-compliant. It is often not possible to recreate such documents from the source program because, for example, they were not created locally but were sent to the user in question by e-mail. There are several methods for converting PDFs to PDF/As. Acrobat 8 Professional is one of the applications that can be used. However, Adobe is not the only company to market software for this particular task. There are many different products on the market, ranging from single-user solutions to systems for high throughput.

■ Is this really a PDF/A file? When working with PDF/A on a daily basis, file verification is also important. Is it sensible to believe the sender of a PDF document when

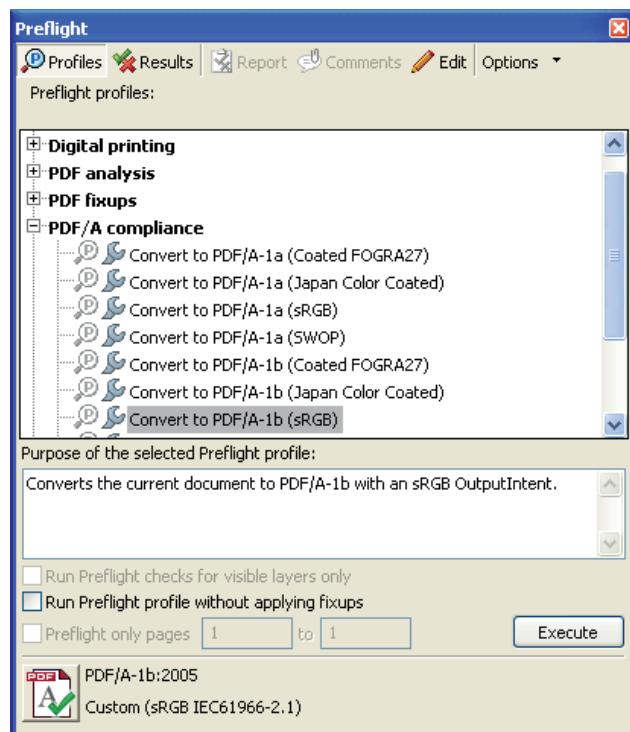
he or she says that it's a PDF/A file? Before received files are saved in an archive, they must be checked to make sure that they are PDF/A-compliant. There are various tools that enable file verification: In addition to Acrobat 8 Professional, there are other applications including Berlin-based callas software's pdfaPilot, which enables the verification and creation of PDF/A files as well as providing some additional functions.

Who stands to benefit from PDF/A?

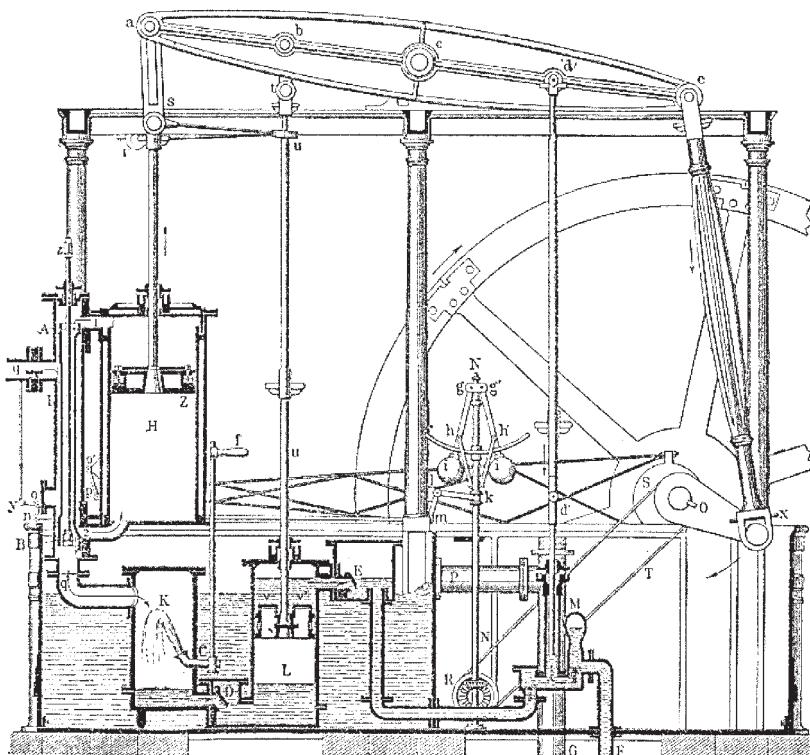
Many sectors and professions have been waiting for a PDF standard for archiving. It is useful not only for archives, administrative departments, industry, and commerce but also for research and teaching. Many different types of content can be saved as PDF/A files. Below are a few randomly selected examples from various fields.

■ Saving e-mails as PDF/A: Today, more and more correspondence, some of it of a contractual nature, is being sent by e-mail. Anyone who has switched from one e-mail program to another knows the difficulties involved in transferring old mail to the new system. Since PDF/A is a safe format, it makes sense to save e-mail archives on back-up media in the form of PDF/A at regular intervals.

■ Saving brochures, manuals, and information sheets as PDF/A: Many companies and public authorities already make a large quantity of information available in the form of PDF downloads. Why not create these documents in the future-proof PDF/A format straight away and distribute them as PDF/As? →



PDF/A validation with Preflight: The Preflight validation and correction tool is part of Acrobat 8 Professional. It generates PDF/A files and checks existing PDF/A documents to make sure that they comply with the standard.



Design drawings: Historical, hand-drawn plans can survive for centuries. A standard such as PDF/A is required to make sure that digital plans are also viable in the long term.

■ Plans, maps, and design drawings:

Digital maps, architectural drawings, and construction plans can also be used by future generations if their creator uses PDF/A. The law stipulates that certain engineering plans, such as construction plans for bridges, must be kept for several decades.

■ **Signed digital contracts:** An increasing amount of business correspondence is sent electronically. PDF/A documents can be digitally signed to enable legally effective contracts to be concluded using only digital means.

■ Correct colors in image documents:

PDF/A also enables the accurate display of colors, an important advantage when working with digital image data. For example, if a museum or gallery already has digital images of exhibits, they can be converted to PDF/A files and benefit from the advantages of the archiving format. Digital imaging procedures are becoming increasingly important in medical diagnostics, meaning that PDF/A could be useful in this field, too.

■ **Accessible PDF files:** In America, accessibility in the digital world has been an issue for a long time – especially for the Internet. Enabling the accessibility of information to visually impaired members of society is now also on the agenda in Europe. Since PDF/A specifically supports structured content in PDF documents, it is ideal for processing accessible PDF documents that can be read out by screen readers.

■ **Storing print documents as PDF/A:** Printers and prepress companies will be relieved to hear that the PDF/X standard that is widespread in their industry sectors is completely compatible with the new PDF/A standard. A PDF document can be simultaneously PDF/X and PDF/A-compliant.

Although the PDF/A standard was intended for long-term archiving, it also has advantages that can be appreciated immediately. Anyone who passes information and documents on to customers, readers, or partners can benefit from PDF/A.

PDF/A guarantees data consistency. This rules out problems such as non-embedded fonts, which can result in gobbledegook. Color management functions mend images that are too pale or too bright. In addition, PDF/A prevents many processing problems that can occur with password-protected PDF documents or when printing files.

The quick brown fox jumps over the lazy dog
The qu ck brwnfx j umps vrthl zay dg

Font problems caused by non-embedded fonts: PDF/A prevents problems with fonts such as the inconsistent tracking illustrated in the second line above.

Consequently, people who use PDF/A are not only doing themselves a favor – they are also helping file recipients, since PDF/A eliminates many of the problems that can otherwise be encountered right from the start.

Table: Comparison between PDF/A-1a and PDF/A-1b

ISO has split the PDF/A standard into two levels of conformity. PDF/A-1a (Level A) requires documents to be capable of being reproduced without visual ambiguity, demands that text can be depicted in Uni-

code, and stipulates that the content structure must be correct. PDF/A-1b-compliance (Level B) only requires the reproduction of documents without visual ambiguity.

	ISO 19005-1:2005: PDF/A-1a (Level A)	ISO 19005-1:2005: PDF/A-1b (Level B)
Conformance	Complete PDF/A conformance	Restricted PDF/A conformance
Aim	To produce archive PDFs with full access to all content	To produce archive PDFs that only guarantee visual reproducibility
PDF version	PDF 1.4	
PDF/A identification	Users are obliged to state the PDF/A identifier and the level of compliance	
Metadata	Specifications such as author, document title, creation data, and source program must be XMP-compliant	
Logical structure	Structure and accessibility must be realized using tags and alternate image descriptions and by stating the language used	There are no explicit logical structure requirements
Encryption	Security settings are prohibited. It must be possible to open/process the PDF file in question without requiring a password.	
Colors	All colors must be identified. Device-dependent color spaces must be identified by output intent.	
Transparency	Not permitted	
PDF layers	Not permitted	
Compression	LZW compression not permitted JPEG2000 compression not permitted	
Fonts	All fonts must (at least as subsets) be available (embedded) directly in the PDF document in question Mapping of character codes to glyphs must be unambiguous Each letter must have a Unicode equivalent	–
Annotations	Comments that take the form of sound or movies are not permitted. Traditional text/label-style annotations are permitted.	
Referenced content	Referenced (non-embedded) images or page content are not permitted	
Alternate images	Alternate images (for lower-resolution screen display) are not permitted	
Programming languages	Embedded JavaScript is not permitted	
Actions	Certain actions, such as opening movies or sound files or sending or resetting forms, are not permitted	
Forms	Permitted, but with restrictions	

Overview: Which file formats are suitable for archiving?

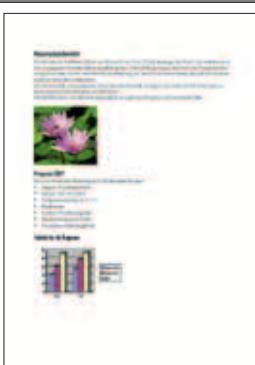
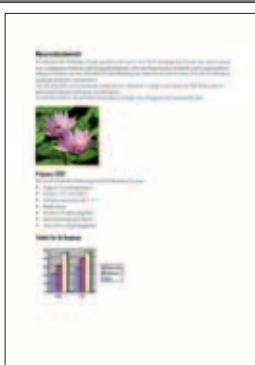
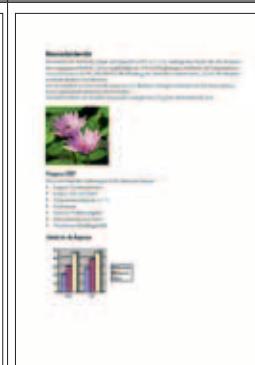
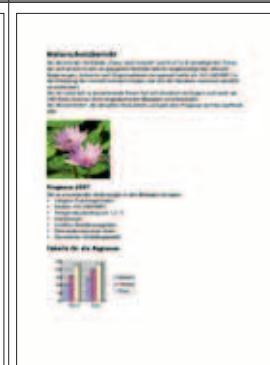
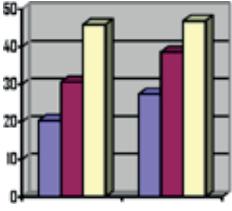
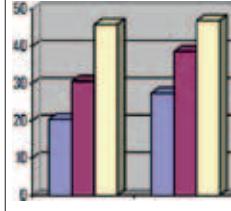
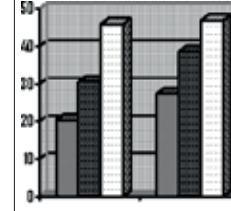
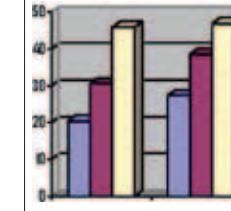
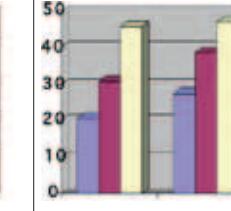
How are documents normally archived in private home offices, company offices, public authorities, and organizations?

Many users simply store the original documents (for example, Word, Excel, or PowerPoint files). This can lead to nasty surprises with regard to reliable displayability and the future usability of the files. This method of archiving is therefore not recommended.

TIFF-G4 has been the de facto standard for numerous companies and administrative departments for many years. TIFF-G4 files are monochrome, black and white TIFFs (bitmaps) that can be archived in a way that saves space thanks to compression

features based on fax technology. The JPEG image format is commonly used for color documents. This format produces relatively small file sizes in spite of the colors. The PDF/A standard and XPS (XML Paper Specification), which was developed at the end of 2006 by Microsoft, are relatively new formats. Both of these formats offer facsimile quality as well as supporting structured content, thereby allowing the reliable and complete indexing of text. PDF/A is already standardized, but XPS still has to stand the test of time.

The table below gives an overview of the long-term archiving features offered by both these formats.

PDF/A	XPS	TIFF-G4	JPEG	DOC (Word)
				
96 KB	112 KB	56 KB	88 KB	120 KB
345.560.000 €	345.560.000 €	345.560.000 €	345.560.000 €	345.560.000 ⚡
				
				

	PDF/A	XPS	TIFF-G4	JPEG	DOC (Word)
ISO standard for archiving	✓ Yes	✗ No	● De facto standard, but not an official standard	✗ No	✗ No
Facsimile quality	✓ Yes	✓ Yes	✓ Yes	✓ Yes	✗ No
Font security	✓ Yes Maximum level of security thanks to strictly defined specifications in the PDF/A standard	✓ Yes Fonts are embedded	● No fonts exist, since the files are pixel images.	● No fonts exist, since the files are pixel images.	✗ No The display of fonts can vary on different computers. Users are not told about missing fonts. Users are not told about replaced fonts.
Searchable text	✓ Yes Enabled by means of OCR, even for text on scanned pages	✓ Yes	● Possible if searchable text is generated using OCR. There is no standard procedure for storing the text in the TIFF-G4 file.	✗ No	✓ Yes
Consistent colors	✓ Yes Consistent colors are required by the standard	● Possible	✗ No Produces black and white bitmaps	● Possible	✗ No
Images and graphics are fixed document parts	✓ Yes	✓ Yes	✓ Yes They are incorporated into the pixel image	✓ Yes They are incorporated into the pixel image	✗ No Cannot always be handled safely
Structured data	✓ Yes – with PDF/A-1a with tagged PDF	✓ Yes With XML	✗ No	✗ No	● Possible
Cross-platform capable	✓ Yes	✗ No	✓ Yes	✓ Yes	● Only with restrictions (font problems)
Free viewer	✓ Yes PDF/A is always displayed in exactly the same way	✓ Yes (currently only for Windows)	✓ Yes (but not widely used)	✓ Yes Most Internet browsers can display JPEGs	✓ Yes There are free alternatives to Office but documents may be displayed differently

What's to be done with JPEG and TIFF-G4 archives?

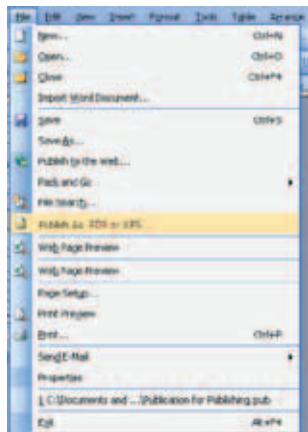
There are basically two options for converting large document archives that currently use TIFF-G4 or JPEG to PDF/A along with their existing inventory: Permanently or temporarily.

If the number of documents handled is not too high and regular access to the data is required, converting the image files to

PDF/A is worthwhile. This involves using mass conversion solutions that package pixel information into PDF and can enable text searching using text recognition.

However, if users only need to call up data from an archive every now and then, 'on the fly' solutions can be used to generate a PDF/A file from a particular original image file. ■

Is XPS an alternative to PDF/A?



Documents can be 'published as PDF or XPS' from Office programs.

The current version of Internet Explorer also serves as an XPS viewer (as yet, only on Windows).

The screenshot shows a Microsoft Internet Explorer window with the title "C:\Users\Administrator\Desktop\nutshell\The PDF/A Standard.pdf - Windows Internet Explorer - [Working Offline]". The page content discusses the ISO 19005-1 standard for PDF/A, mentioning its goal of reproducibility and self-containment. It also compares PDF vs PDF/A, noting that PDF is not guaranteed to be reproducible. The browser interface includes standard toolbar buttons and a status bar at the bottom.

Since the end of 2006 and especially in Microsoft circles, it has been argued that XPS might be a future alternative to both PDF and PDF/A. However, it is still necessary to clarify whether or not XPS can offer at least the same level of technical expertise as PDF/A. If it can do so, its special features must also be defined.

For the moment, let's put aside the fact that PDF has existed since 1993 and that PDF/A was adopted as an ISO standard in autumn 2005, whereas XPS has been available since the end of 2006/start of 2007 as a part of Microsoft's Vista operating system. Instead, let's look at the nature of PDF / PDF/A and XPS. XPS can be described as a format that statically depicts the content that a monitor would or should display or a print device would or should print out. In this respect, it could be called an up-to-date 'spool' format, comparable with PCL, AFP, or PostScript but more modern. On Microsoft Vista at least, users can display a document in XPS spool format in Internet Explorer.

PDF can also be used as a 'spool' format. The Californian company Apple has been using it as such for years for its Mac OS X operating system. However, for a long time PDF has not been used just as the final stop for static digital documents – instead, it has become the basis of document-based process flows.

The best-known usage of PDF in this context is for electronic forms; recently, PDF has increasingly been used for collecting information with sophisticated comment functions that can be used by multiple users for PDF documents with ensured traceability. The many established possibilities offered by digital signing round off the use of PDF in various document workflows, even including certified and qualified digital signatures (cf. the section on digital signatures on page 61).

All in all, PDF builds a bridge between the different electronic document formats (even if no PDF export function is available, anything that can be printed out can also be 'printed' to a PDF) while also being a document format in its own right with its own, separate fields of application.

As far as forms, comments, and signatures in PDF documents are concerned, files containing such elements can be adequately archived as PDF/A. In this respect, both PDF and the PDF/A ISO 19005-1 go far beyond the conceptual and actual features of XPS.

Comparison of XPS and PDF/A

Even in the subarea where PDF/A and XPS can be directly compared with each other, weaknesses and restrictions for XPS quickly emerge, making it unlikely that XPS will ever completely usurp the position of PDF/A. Some examples are presented below.

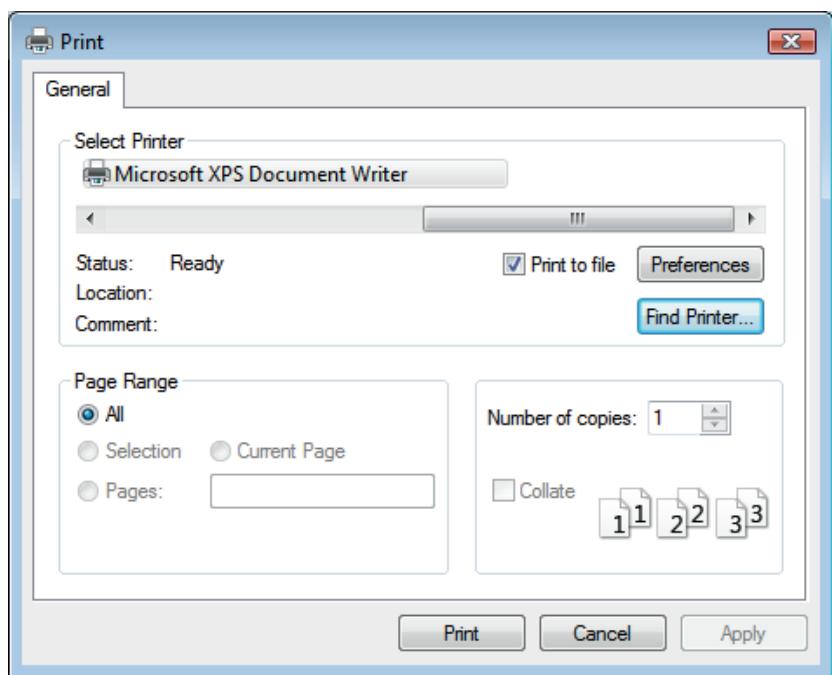
One of the first things to point out is that the graphics models of PDF and XPS are very similar. However, since XPS lacks certain constructs it can be difficult or even impossible to convert PDF documents to XPS without losing data (conversion from XPS to PDF is always possible, thanks to the greater scope of func-

tions offered by PDF). Below is a detailed list of XPS properties in comparison to PDF/A:

- It does not support an overprinting function.
- User-defined, colorable image masks are not permitted.
- The JBIG2 compression procedure, which is an extremely powerful procedure for scanned documents, is not supported.
- Multi-channel color spaces with more than eight components are not permitted.
- JPEG2000 (to be supported from PDF/A-2) is also missing from XPS's range of features.
- XPS only has one transparency blending mode – PDF has 16 (but note that transparency is only permitted in PDF/A documents from PDF/A-2).
- Finally, XPS lacks the bookmarks, notes, and comments that have now become important for so many users.
- On the other hand, PDF/A does not support Windows Media Photo format or the corresponding image compression procedure, but this is not surprising since it was only released by Microsoft in the third quarter of 2006. In any case, XPS Specification 1.0 of October 18th 2006 still talks about Windows Media Photo even though Microsoft has now renamed the format as HD Photo. This indicates that the format is still not entirely consolidated.

From Microsoft's point of view, this fact is probably considered to be a plus rather than a disadvantage. At the end of the day, Microsoft applications can make do with what XPS can display.

However, even in this respect Microsoft could be accused of not putting all of its efforts into XPS: Although Publisher 2007 permits the use of CMYK and spot colors, it saves them in XPS format as RGB, which



unfortunately casts doubt upon the applications claim to be a serious publishing tool. Users do not even see a warning, which, considering the 'high-quality printing' XPS export style, must be considered as misguiding at the very least. Microsoft's own PDF export process suffers from the same restriction.

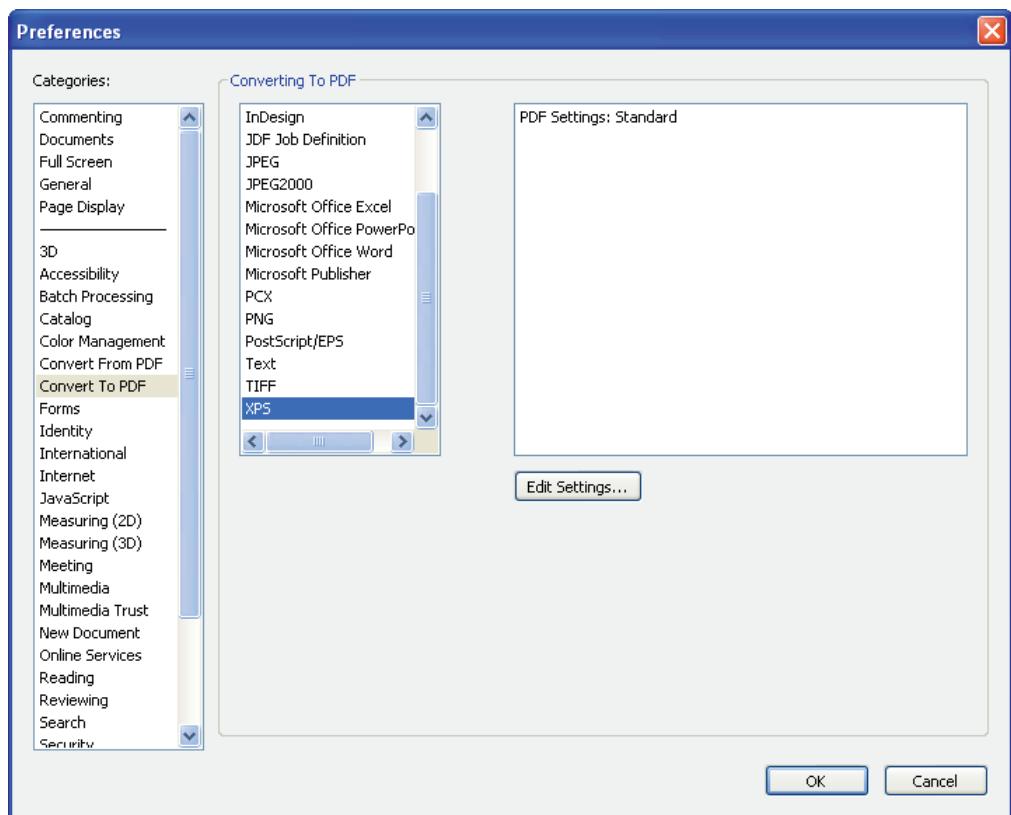
It's no secret that the performance of XPS, however impressive, can only be guaranteed if the program from which a file is being saved in XPS format expressly supports XPS (or the Windows Presentation Foundation architecture).

Over the course of time, many applications are certain to implement such support, since otherwise they will eventually be unable to claim that they support Vista. However, it will take years for the required updates to find their way onto Windows computers, and there will probably be a whole range of applications that will either not support XPS at all or will only introduce XPS support later on. If a program does not expressly support XPS, it is still possible to create an XPS document on Vista using the XPS printer driver. This involves a transformation of the original GDI protocol into XPS notation. →

The Print function can be used to generate XPS files from any application (including Photoshop or CorelDraw), but the quality is less than perfect.

Is XPS an alternative to PDF/A?

*Adobe Acrobat 8 can convert XPS to PDF.
However, this only works on Windows.*



Acrobat can handle a whole range of document formats. Users can now use the 'Open...' command in Acrobat 8 to open XPS documents as PDF files.

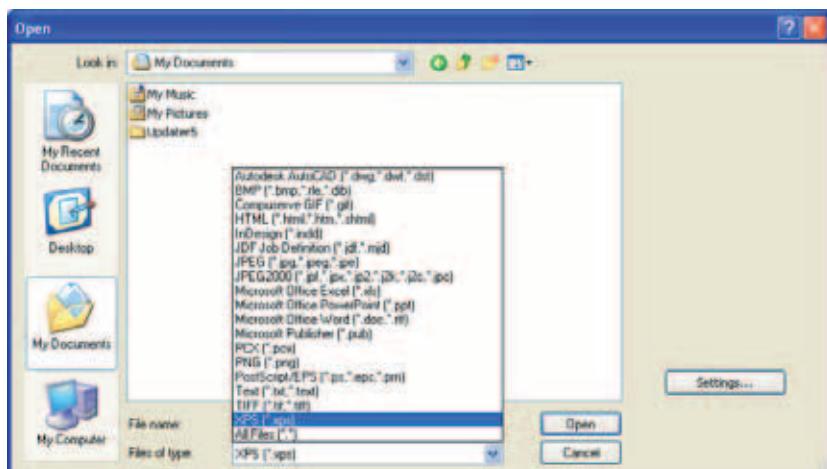
With applications that use PostScript for high-quality printing – including all professional publishing applications such as Adobe PageMaker, Quark XPress, CorelDraw, and Adobe Photoshop – the user is confronted with nothing more than a crutch: He or she ends up with a file export that has the quality of a screenshot.

Even if Microsoft and third-party suppliers manage to iron out some of these issues during coming years, XPS will re-

main a format that, while being extremely well-suited to depicting typical Office documents, cannot depict other document types or can only store them with unnecessary restrictions. In this respect, XPS does not offer the universal support of all document types that has made PDF such a powerful and popular format.

The best thing about XPS is therefore that it can be used to create a much higher quality PDF from applications that support it than previous methods such as GDI, PCL, or PostScript printer drivers. From Version 8, Adobe Acrobat offers an import filter for XPS. It enables users to easily open an XPS and simply save it as a PDF.

There is, of course, a suspicion that Microsoft intends to overstretch the capabilities of XPS by using its market power to position the format not just as a spool format or printer language but also as a universal exchange format – but it certainly doesn't fulfill the requirements for the latter as well as PDF. PDF should remain the more reliable format for many years to come.



PDF/A creation: Analog, digital, and mass processing

2.

PDF/A is always the destination, but the point of departure can differ greatly from user to user. This chapter concentrates on three main tasks: Converting paper documents to PDF/A, exporting Microsoft Office files and other documents in a way that allows them to be archived, and mass processing PDF archive files. The special process flow for converting existing PDF files to PDF/A is explained in detail later on.

PDF/A from scanned documents

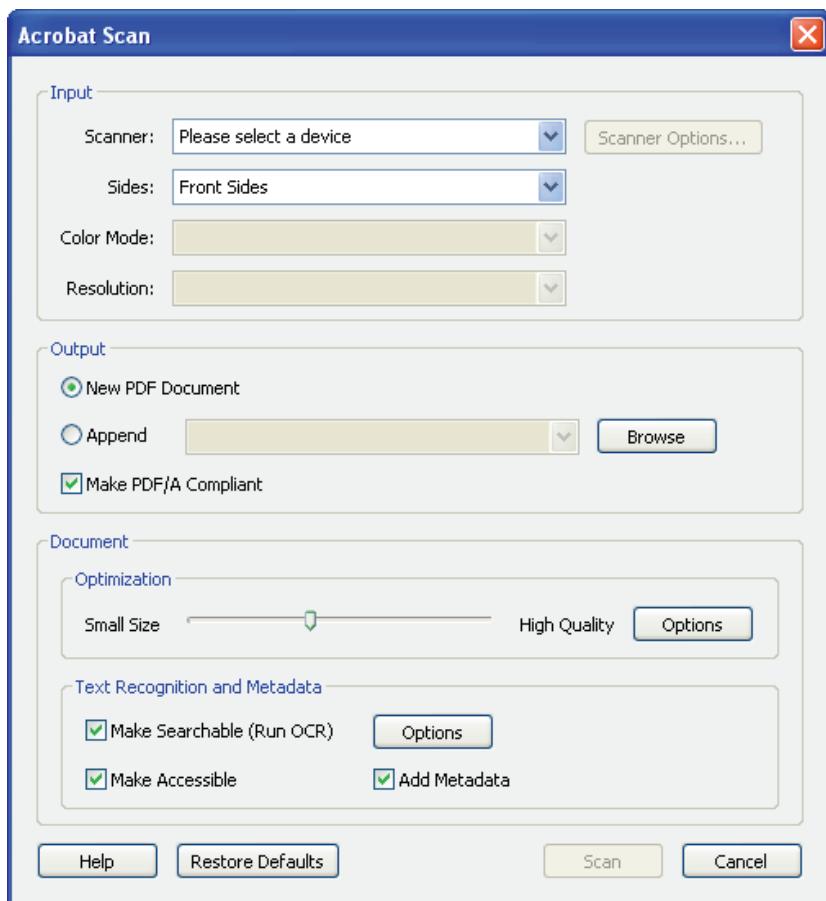
'Analog to digital' conversion is normally required when users have received the documents that need to be archived as printed pages rather than creating them

themselves. For example, telephone bills are often sent as printouts by mail. In some cases, documents that need to be retained are only available as printouts because the digital originals have been deleted from users' computers. In addition, many documents were created by typewriter or by hand in the days before computerization.

In such cases, the only way to digitize document pages is by using a document scanner. As well as the type of scanner (flatbed scanner or a device with bypass feed), the scope of features provided by the software also has a bearing on whether or not the digitalization process can create a faultless PDF/A and dictates which additional features can be used to enhance the usability of the document (for example, OCR for full-text searching). →



Roufoto – photocase.com/de



Acrobat Scan: Users can use a checkbox to define that a digitalized PDF document is to be standardized in line with PDF/A. If required OCR (text recognition), accessibility, and metadata options can be activated.

Incidentally, all modern scanners support the use of PDF as the initial format (in addition to image formats such as JPEG or TIFF). However, not all scanners are currently able to generate PDF/A. This restriction is sure to change as the standard becomes more prevalent. Due to space restrictions, it is impossible to mention all established scan pro-

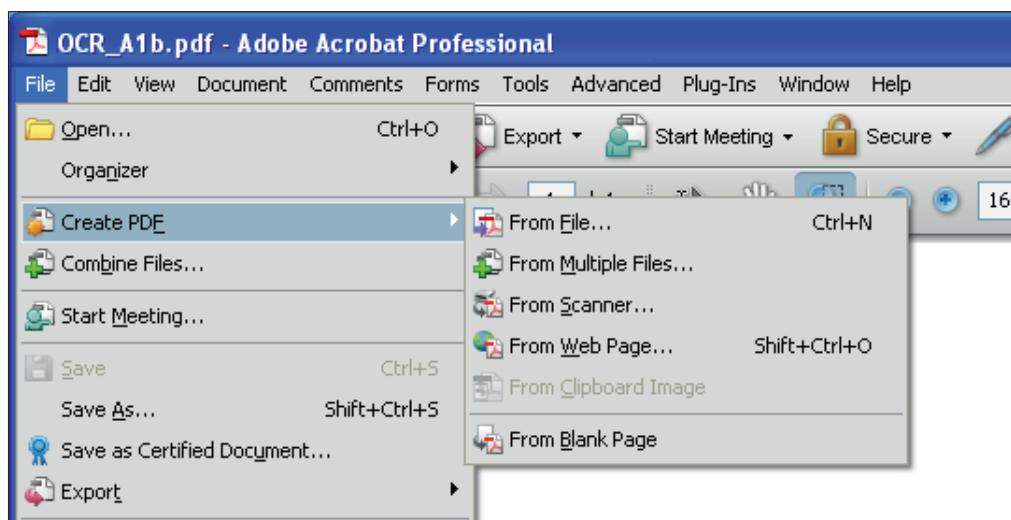
grams in this publication; we are therefore going to concentrate on Acrobat Professional's Scan To PDF function, available from Version 8 for the production of standard-compliant PDF/A documents.

Scanning options in Acrobat 8 Professional

In the majority of cases, users can process documents with Adobe Acrobat Professional as well as with the original software regardless of which scanner they use. This application enables the production of PDF files for completely different requirements and diverse uses. From Version 8, there is a checkbox that allows you to select PDF/A compliance.

Important settings

Once the scanner is connected up and switched on, the user can trigger the creation of a PDF in Acrobat by choosing 'File' → 'Create PDF' → 'From Scanner' in the menu bar. In the dialog box that then appears, the scanner being used can be selected from the list of devices and the user can define whether the application is to scan only the front side of the document or both the front and back sides. In the 'Output' area, users can decide whether the current scan process should generate a new PDF document or append the scanned material to an existing PDF. The 'Make PDF/A Compliant' checkbox is especially useful here, and should be selected. Quality settings for the PDF document can be made using a slider or in more detail using the 'Options' button.



Scanning and creating PDFs with Acrobat 8: PDF files can also be produced from scanned pages with Adobe software.

Text recognition, accessibility, and metadata functions can be used to give the new PDF additional features.

The text recognition function creates searchable text (otherwise, the scanned



Text Recognition and Metadata: These options give the PDF additional features such as searchable text and metadata. These features are not PDF/A-relevant, but they do enhance the functionality of the PDF file.

pages in the PDF simply take the form of pixels). Accessibility (enabling access to content for the visually impaired) gives the PDF structural information that describes the order in which it is to be read. This is required to optimize the use of screen readers. Metadata can be considered as a kind of digital label that contains information such as the title of the document, copyright, keywords, and author. This is useful for administrating digital archives.

Detailed description of text recognition

Additional settings for text recognition can be made using the 'Options' button. This



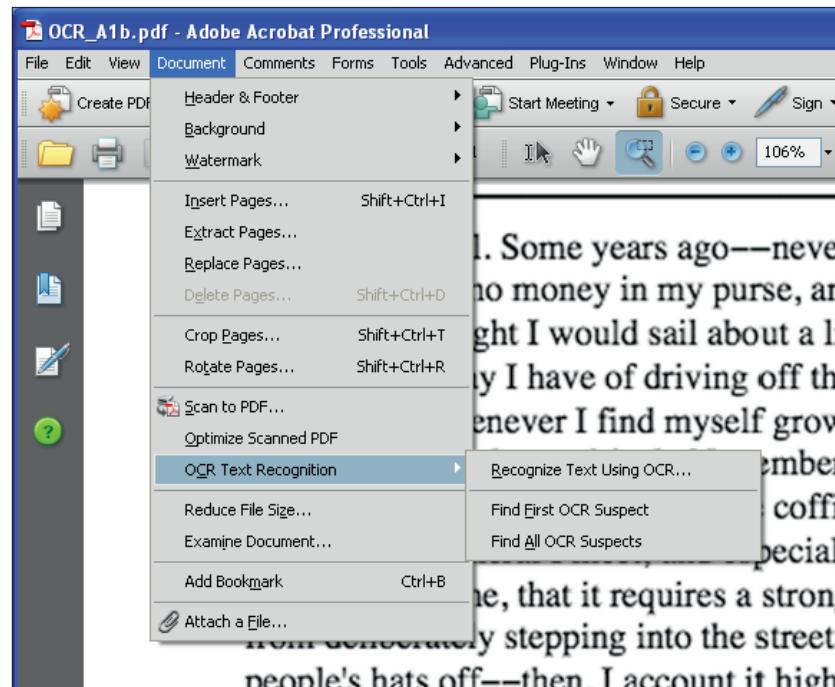
Recognize Text – Settings: The 'PDF Output Style' field contains options for generating a simple PDF image with searchable text or a more complex PDF file with separate areas for text and graphics where possible.

area includes a language setting and other fine-tuning settings for text recognition. For example, the user can define whether the scan output should be a searchable image or formatted text with graphics. However, note the following: Even the second, superior option is no guarantee of PDF/A-

1a-compliance, since errors can still occur when structures are reconstructed. This is why the restricted version, PDF/A-1b, is used here, too.

Converting pages that have already been scanned to PDF/A

The procedure used to convert scanned documents that already exist in the form of pixel data to PDF documents in Acrobat



Professional is rather different. First, the image file (TIFF or JPEG) is imported by choosing 'File' → 'Create PDF' → 'From File' from the menu. It is then converted to a PDF file. The 'Document' menu contains the 'Optimize Scanned PDF' function. Once the document has been converted into a PDF, the user can use this function to improve it before subjecting it to text recognition.

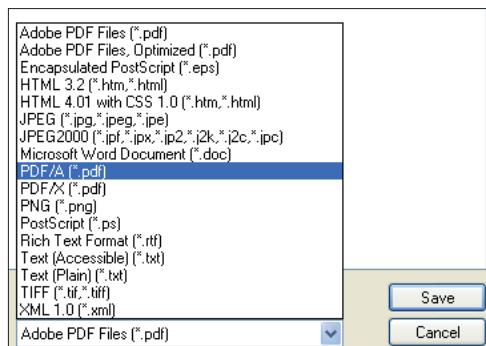
Text recognition is also called from the 'Document' menu item. It is triggered with the 'OCR Text Recognition' → 'Recognize Text Using OCR' command.

The user can then check that the process worked correctly: Clicking 'Find All OCR Suspects' triggers a search for image elements that could not be converted to text. →

Document optimization and text recognition: The 'Optimize Scanned PDF' function can be used to enhance the source material for text recognition, e.g. by removing edge shadows. Following this process, the user can use Adobe's 'Recognize Text Using OCR' function to generate searchable text.

Saving or exporting the document as a PDF/A

The PDF document must now be converted to PDF/A. This can be achieved in just a few steps using the Export function or with the 'Save As' command. Both methods involve



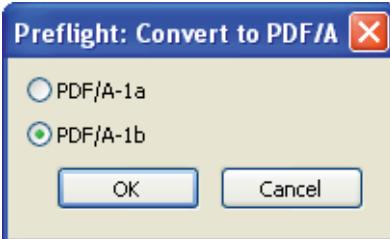
A multitude of formats for saving documents: This long list of formats includes the PDF/A standard.

the use of the integrated Acrobat Professional Preflight engine, which carries out the conversion to PDF/A. Regardless of

whether the user chooses the Export option or 'Save As', only the PDF/A-1b level in the 'Settings' will be successful.

Even after text recognition, metadata input, and the integration of structural information for accessibility, scanned documents do not automatically have advanced PDF/A-1a features.

When the user clicks 'OK', Acrobat generates a PDF/A file from the PDF document. ■



Scanned documents are always converted to PDF/A-1b: To ensure a successful PDF/A conversion, the preset PDF/A-1b-compliance specification must not be changed.

Saving space when generating PDF files from scanned documents:

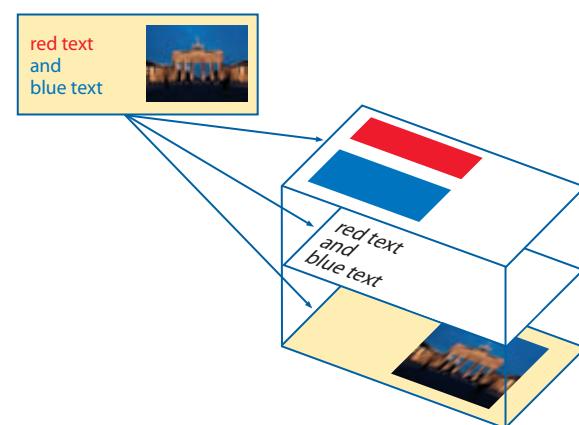
The generation of PDF files from digitalized paper documents has a disadvantage – the image data for such files normally requires more memory capacity than digital pages of text. A PDF generated from a Word document will be considerably smaller than a PDF file that is generated from a Word printout using a scanner.

This comparatively high file size is particularly cumbersome if a large number of documents with many pages need to be archived. There is a big difference between 10,000 x 40 KB and 10,000 x 400 KB: 400 MB will still fit on a CD-ROM; 4 GB will not.

An important factor for determining the size of PDF files is whether the document is read in black and white (line scan), grayscale, or in color – color data consists of much more information than bitonal data and the resulting data quantity is therefore also larger.

Various image compression types have been developed over the past years to enable users to save memory space when storing image data. The best known of these methods is JPEG compression. PDF/A permits compression, but not all types. JPEG and JBIG2 are permitted, but JPEG2000 is not. In addition to the type of compression, the compression level is also important for a scanned text. This is for readability reasons. Higher compression levels can render the image/text progressively less clearly.

Berlin-based LuraTech has been working on effective image compression for digitalized company documents for years. During the course of the develop-



ment of PDF/A, LuraTech has enhanced the product and service scope of scan-to-image and scan-to-PDF solutions by adding a scan-to-PDF/A function. The JBIG2 compression used has been improved by a type of layer technology that enables color documents to be digitalized in a legible manner while using relatively low amounts of memory.



In addition to compression, there are various text recognition functions and options for integrating metadata into PDF/A files.

More information on the Internet at:
www.luratech.com

The Distiller engine

For a long time, the Distiller was the only recommended way of producing faultless PDF files for tasks such as professional printing from certain programs. In the light of constantly improving PDF creation functions such as those used in new versions of widely used programs like Microsoft Office and InDesign or at operating-system level (for instance, Mac OS X), the Distiller is losing some of its importance. However, it is still an important component of the Acrobat package.

The Distiller uses a slightly specious method to convert various file formats into the PDF format: It uses the PostScript format that is generated temporarily when printing files. Because PostScript and PDF are related to each other both with regard to development and on a structural level, conversion from PostScript to PDF is normally easy to achieve – as long as the appropriate printer drivers and a PostScript to PDF converter, like Adobe Distiller, are available.

Since all popular programs have print functions, the generation of PDFs using a

combination of print data and the Distiller is an all-purpose method. In addition, the related format EPS (Encapsulated PostScript) can also be directly ‘distilled’.

When is the Distiller useful? The Distiller is the appropriate tool for creating PDFs from applications that do not offer a PDF export function or an option for saving files as PDFs. However, the Distiller has more features. Watched folders enable the creation of PDFs to be automated and standardized – a useful feature for many usage environments.

PDF/A document generation using the Distiller

With Acrobat 8, Adobe has implemented presets for standard-compliant conversion to PDF/A. The Distiller can only be used to create PDF/A-1b-compliant documents; PDF/A-1a documents cannot be created for technical reasons, since the required structural information is not generated or passed using the PostScript method.

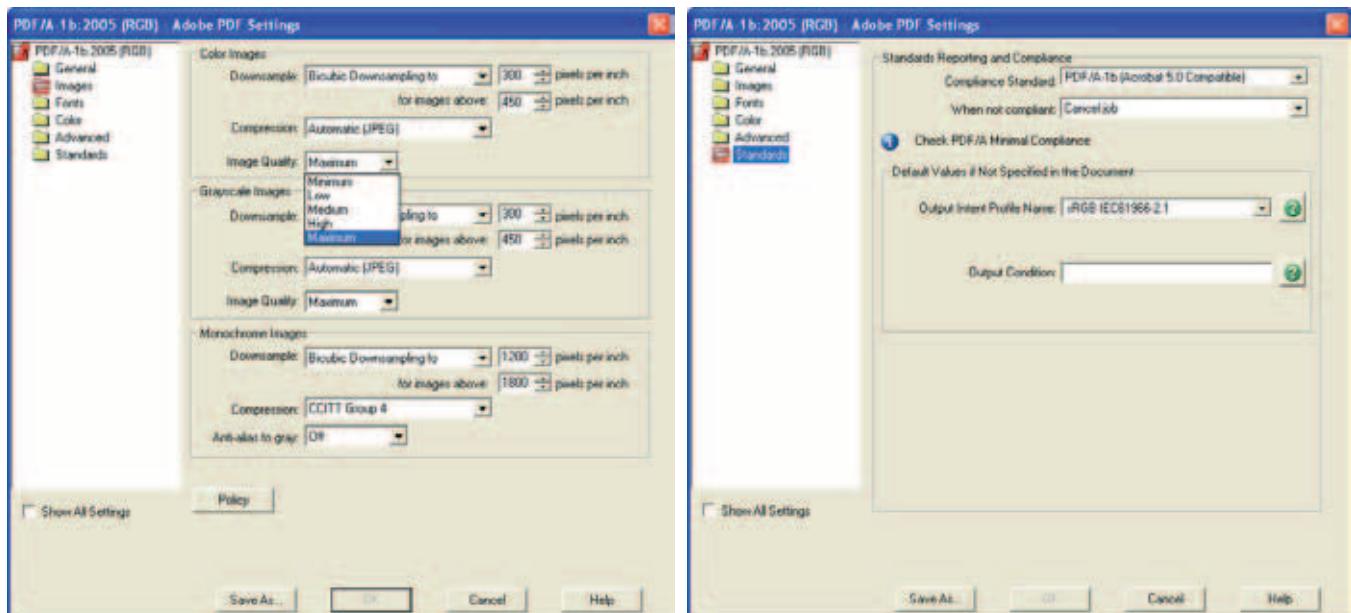


Acrobat Distiller

Acrobat Distiller 8: The Distiller is a program that is contained in the Acrobat package. Even Acrobat 1 was shipped with a Distiller. Reliable PDF/A creation functions were introduced with Distiller 8.

PostScript and EPS files can also be converted to PDF by dragging them into the Acrobat Distiller window or onto the Distiller icon. This means that there is no need to use the ‘Open...’ command.

PDF/A using Distiller 8: There are two PDF/A settings, each with a different color space. This enables the creation of PDF/A files in RGB (for displaying on monitors) and in CMYK (for printing).



PDF/A settings: Users can change the compression level and resolution in the 'Images' section. The compression type can be changed to 'ZIP'. The preset sRGB output intent in the Standards section is the generally recommended intent for RGB. In the case of CMYK, the preset US profile can be changed to a profile more suited to the European market.

Users can choose between two default settings in the main window of Acrobat Distiller. PDF/A in color space RGB is mainly suited for use on computer screens. CMYK PDF/A is intended for printing out with either an office printer or with professional four color printing on an offset printer.



lio – photocase.com/de

PDF/A settings in detail

Changes to the preset default settings for the generation of PDF/A should only be made after due consideration to avoid creating non-compliant documents. These settings can be modified by choosing 'Settings' → 'Edit Adobe PDF Settings'.

Settings that influence the resolution and compression of images can be made in the 'Images' section. Files with lower resolution and higher compression values are smaller, but this can worsen the display quality. However, the compression type can be changed to ZIP, which does not impair the image quality.

When creating PDF/A with the CMYK color space, European users should take a look at the 'Standards' section. The Output Intent presetting here is intended for the US market. (The term 'output intent' comes from the color management field and refers to the regulation of color settings for printing.) In this area, users can select an output intent that is more suited for use in Europe, such as the European ICC profile 'ISO Coated FOGRA27', which is contained in the Acrobat 8 scope of delivery.

If a change is made to a default profile, the changed profile is saved as a copy; Distiller default settings cannot be overwritten.

Additional throughput with watched folders

PDF settings (the settings that specify how PDF files are to be ‘distilled’) can also be appended to file-system folders. Such folders are called ‘watched folders’ or ‘hot folders’.

Hot folders can be set up in just a few steps in the Distiller. The user specifies which file-system folders are to be watched, selects the required post processing setting – in this case, the setting for PDF/A – and the Distiller then creates two new folders (‘In’ and ‘Out’) in each hot folder, as well as a ‘joboptions’ instruction file.

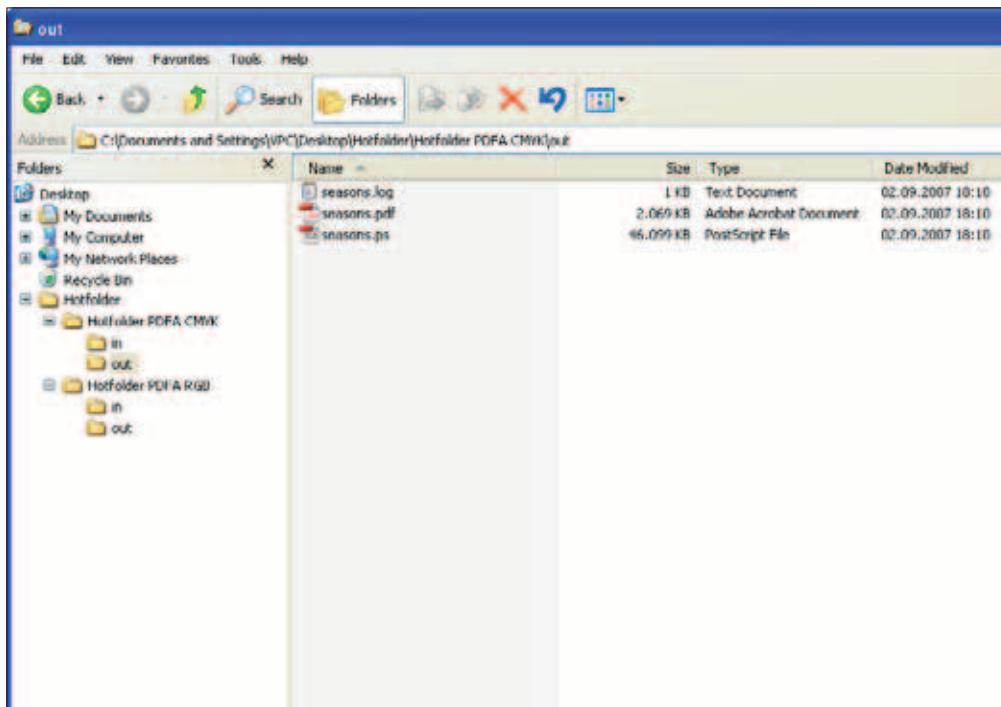
If a print file is now saved in this folder, the job specifications are implemented automatically without user intervention. Users can save multiple documents in a watched folder for processing. In addition to the fact that this process can be carried out automatically, another advantage is that the quality of the files remains constant.

However, as far as licensing is concerned, the Distiller is not intended to be used to enable entire departments to access watched folders on the server. Adobe markets a server version of the Distiller for the mass creation of PDFs in compa-



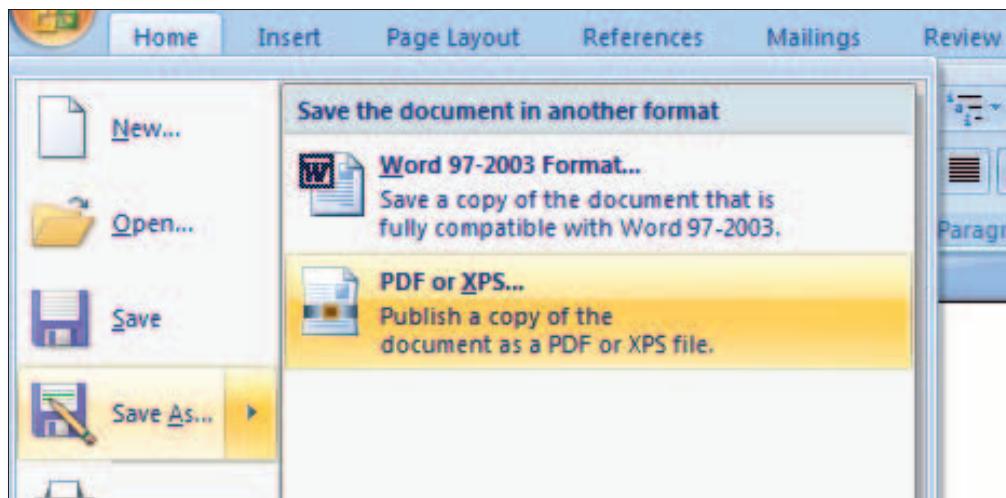
nies. This high-throughput solution has now acquired a different naming and is marketed as the ‘LiveCycle PDF Generator PostScript’.

Watched folders: The Distiller enables watched folders to be set up. Why not create separate ‘hot folders’ for PDF/A (RGB) and PDF/A (CMYK) in order to create PDF files more effectively?



In and Out: PostScript print files are sent to the ‘In’ folder. The Distiller processes each file in accordance with the settings in the job options (in this case, the setting for PDF/A files in the RGB color space). The new PDF/A files are sent to the ‘Out’ folder. The process creates log files containing information on the process flow.

Only available as an add-in: Users can only benefit from the function that enables documents to be published as PDF files in Office 2007 after downloading and installing a free add-in. Internet: www.microsoft.com



Office and administration

Many users all over the world use Microsoft Office programs to create their documents. Frequently, working files in DOC, PPT, or Excel format are used for internal and external communication and for storing files in archives. This process can sometimes cause problems for recipients and is not optimum for long-term storage. PDF – or, even better, PDF/A – minimizes if not totally eliminates problems that occur when exchanging and archiving files. The

creation of PDFs in the new Microsoft Office 2007 is slightly different from the process in previous versions of Office.

PDF/A in Office 2007

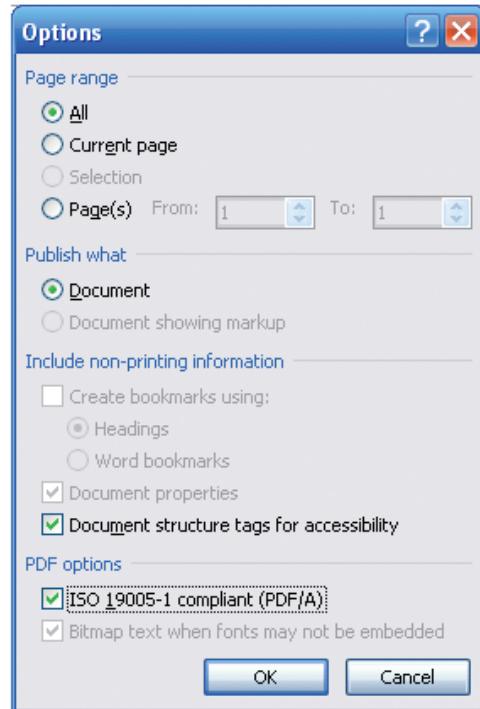
Unlike with the previous versions of Microsoft Office, Office 2007 enables the export of PDF/A files without requiring the use of Acrobat or the Distiller.

Before the Office 2007 package was rolled out at the end of 2006, discussions took place on PDF features. Adobe Systems and Microsoft disagreed about the integration of a direct PDF output function in Office 2007 programs. As a solution to the dispute, users must now download a separate 'Save As PDF or XPS' add-in from the Microsoft Web site and install it in the application package later on. The following Office 2007 programs benefit from the new export function: Access, Excel, InfoPath, OneNote, PowerPoint, Publisher, Visio, and Word.

Once the PDF export add-in has been installed, an extra option is added to the Save As command that enables Office documents to be saved as PDFs. The 'Options' dialog box in the 'Publish As' area is a useful feature. Users can select a checkbox here to create PDF files in accordance with the PDF/A standard: 'ISO 19005-1 compliant (PDF/A)'. When the user clicks 'OK', the program creates a PDF/A-1b-compliant file.

If users wish to proceed as in Office 2003 and use a connection to the PDF/Distiller

XPS is a device-independent document format developed by Microsoft. The abbreviation stands for 'XML Paper Specification'.



PDF/A from Office 2007: The Options in the PDF export dialog include the 'ISO 19005-1 compliant (PDF/A)' setting, which enables the generation of a level B PDF/A.

settings to export PDFs from Office 2007, they should expect to experience problems in conjunction with Acrobat 8.0. According to the manufacturers, this is due to the fact that the rollout dates of the two software solutions were so close together. An Acrobat update to version 8.1 should solve these incompatibility issues.

Office 2003 and the PDFMaker

It is only possible to generate PDF/A documents from Office 2003 using the PDF-Maker add-in and a connection to Acrobat (or the Adobe Distiller). Acrobat 8 Professional provides current conversion settings for PDF/A. Users can create both PDF/A-1a-compliant and PDF/A-1b-compliant files from Office programs.

Settings for PDF/A-1b

The Office application menu (for example, in Word) has an 'Adobe PDF' entry that enables the triggering of PDF generation and access to the presettings. The 'Change Conversion Settings' command opens a dialog box where users can select options and make additional settings.

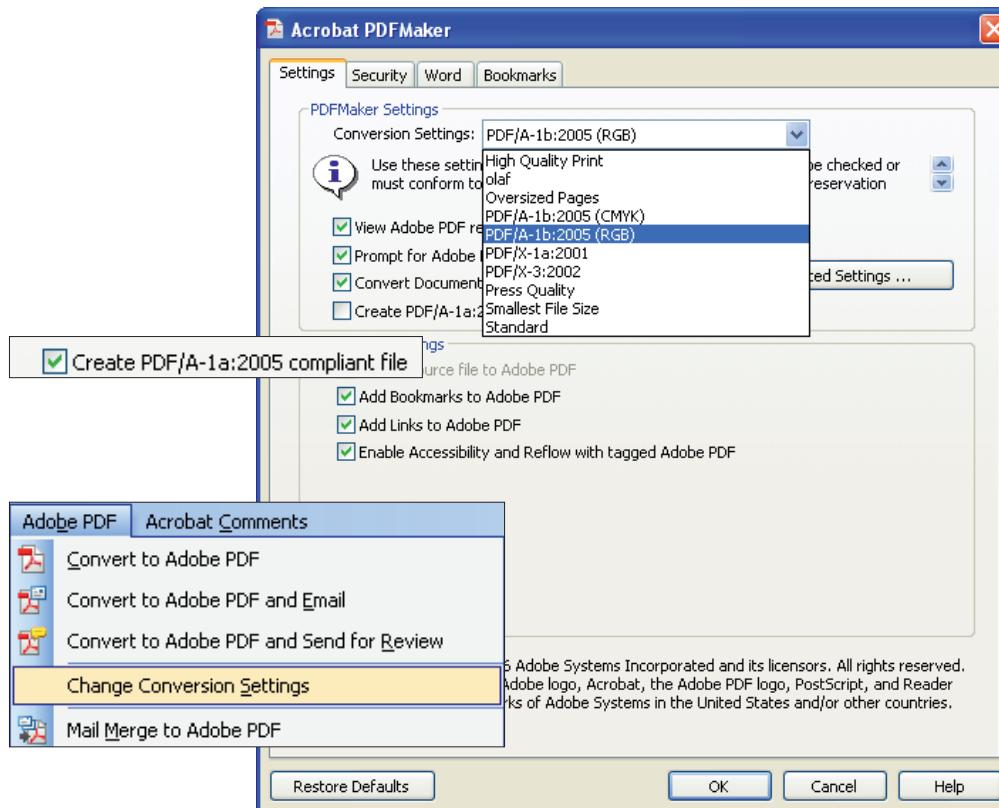
The 'Settings' tab consists of a dropdown menu with various options delivered with Adobe Distiller. There are two PDF/A-1b variants here – one for four-color CMYK output, and one for RGB monitor output. In this example, the RGB variant is used. Clicking the 'Advanced Settings ...' button opens detailed Adobe PDF settings. Users can change the image resolution and compression type here, but it is important to take care not to make changes that could endanger the PDF/A-compliance of files (for example, for Acrobat compatibility). However, let us return to the conversion settings tabs in the Microsoft application.

Acrobat 7 offered support for preliminary versions of the PDF/A standard.

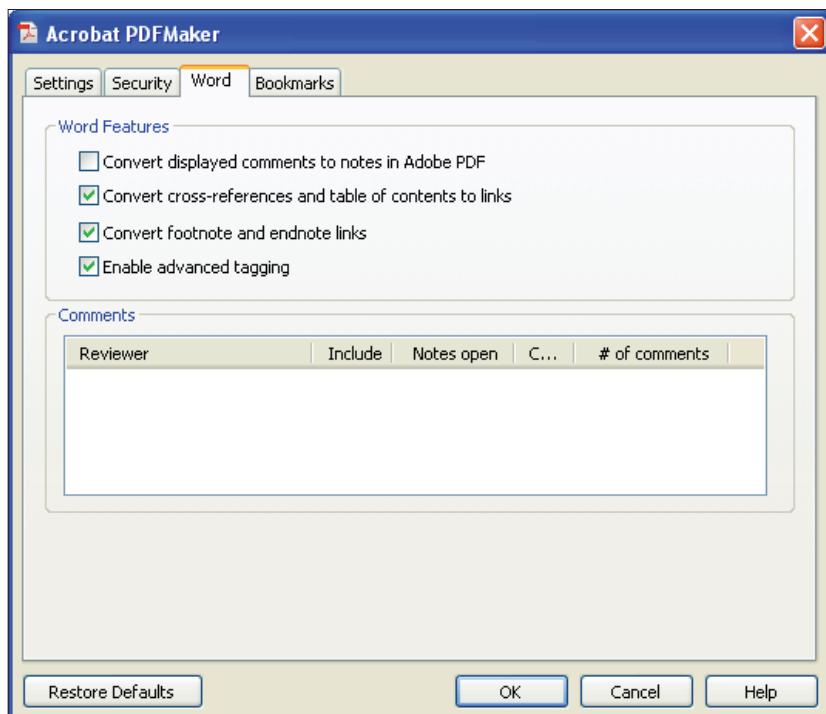
As of Acrobat 8, full support of the final PDF/A standard is offered.

Be careful with the security settings

Because security settings – passwords for opening, printing, or changing PDF files – are not permitted in PDF/A files, users should not make any changes on the 'Security' tab. Users who wish to protect their PDF/A files must protect the storage location of these files. This can be achieved by implementing password protection for a folder or drive, for example. →

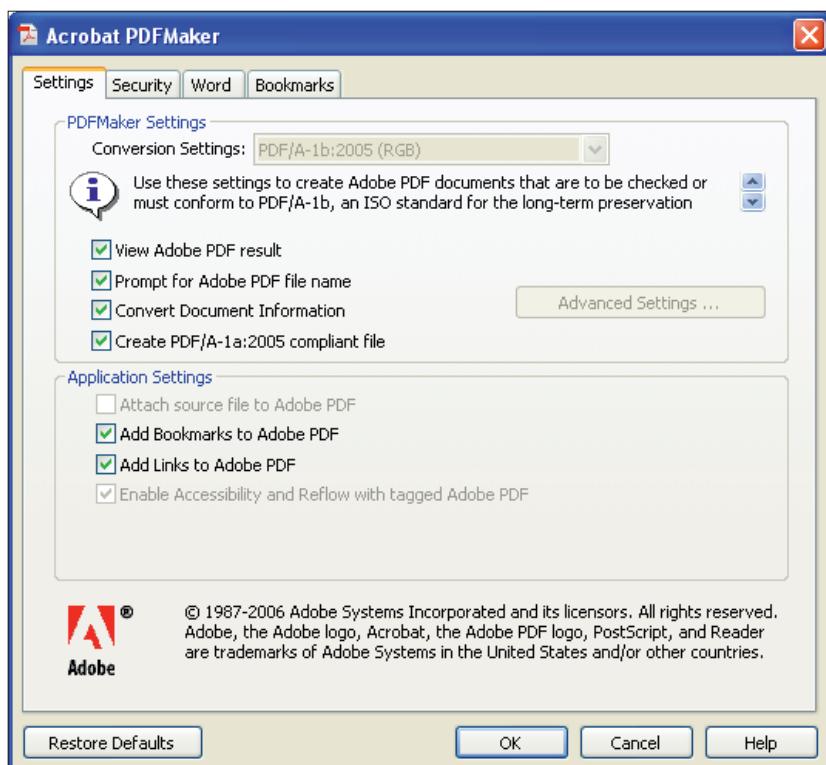


Adobe PDF in Word 2003: The Conversion Settings are essential tools for successfully producing PDF/A files. The settings for PDF/A-1b produce files that are suitable for long-term archiving. PDF documents in RGB color mode are intended to be displayed on monitors; those in CMYK mode are primarily intended to be printed. The conversion setting for PDF/A-1a can be activated by selecting the relevant checkbox.



The 'Word' tab contains the 'Enable advanced tagging' checkbox, which is useful for users who want to generate structured PDFs.

PDF/A-1a: This PDF conversion setting is activated by selecting a checkbox. It activates a function that can convert the advanced features of the higher compliance level, such as fonts and structure, from Office documents into the resulting PDF files.



Options for Word

The settings on the 'Word' tab relate to issues such as comments, tables of contents, and the 'Enable advanced tagging' function.

All of these features help users to produce structured PDF files (tagged PDFs). However, it only makes sense to adopt tags when carrying out a PDF conversion if the source Word document is already completely and consistently structured using formats. (For more information, see the 'Accessible PDF files' chapter on page 52.)

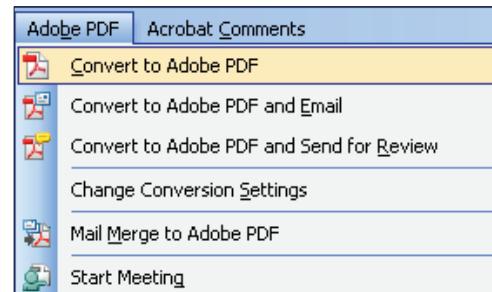
Nevertheless, it is possible to successfully create PDF/A-1b-compliant files without using such structural elements.

Bookmarks

Users can choose to use Word formats for the generation of PDF bookmarks. Bookmarks are permitted for PDF/A. Users can make personal specifications for styles, headings, or Word bookmarks.

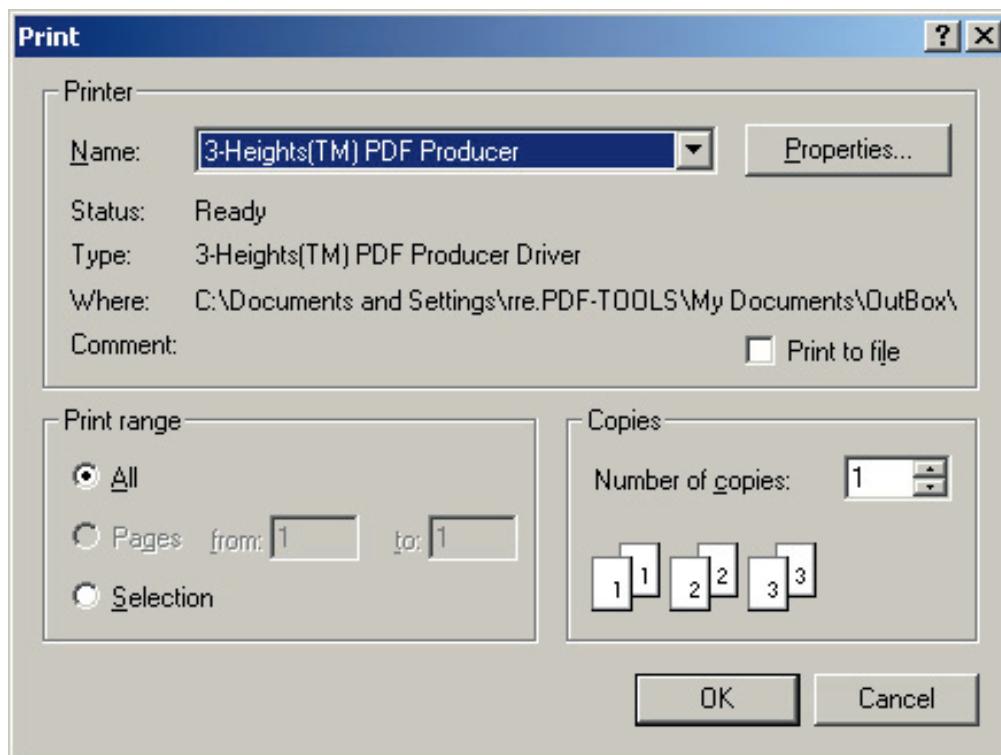
So how do you create a PDF/A-1a-compliant file?

The conversion setting for PDF/A-1a takes the form of a checkbox in the PDF-Maker Settings. If this checkbox is activated, the settings in the 'Advanced Settings' pulldown menu are locked to prevent users from making conflicting settings.



Starting the conversion: This button is used to trigger PDF conversion using PDFMaker. It uses the current default settings to do so.

This constitutes the entire setup procedure for the PDF/A generation. To create a future-proof PDF, the user now only has to click 'Convert to Adobe PDF'. ■



Printer selection: Selecting the 3-Heights PDF Producer as the printer enables the generation of PDF documents from any Windows program.

PDF/A using the 3-Heights PDF Producer

Exporting PDF from Window applications is not only a facility that is offered in more recent Office versions or in conjunction with the Adobe Distiller – there is a whole range of converters that can generate PDF documents. However, only a few products are capable of handling PDF/A.

PDF Tools AG's 3-Heights PDF Producer produces PDF/A-compliant files for long-term archiving. This tool is capable of creating PDF documents that meet various

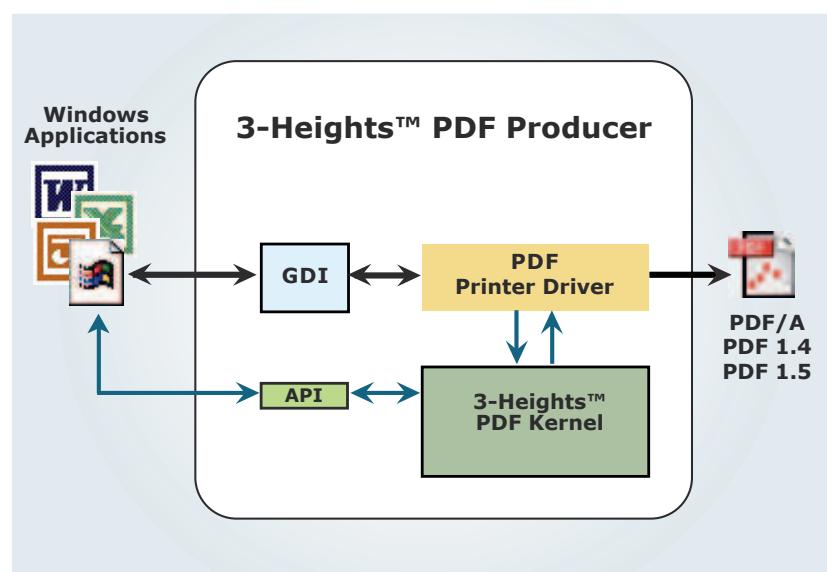
for redistribution on clients and multi-user servers. Swiss-based PDF Tools AG provides a whole host of tools and libraries for the creation and processing of PDFs. The company's products can be purchased directly or via OEM partners. A free test version of the 3-Heights Producer Developer Kit (SDK) is available on the manufacturer's Web site: www.pdf-tools.com.

3-Heights PDF Producer: This solution latches on to Windows' print functions to deliver different types of PDFs, including PDF/A.

pdf-tools.com
Premium PDF Technology

specifications (including PDF/A-compliance) from any Windows program using GDI printer drivers. The 3-Heights PDF Producer offers both synchronous and parallel generation of PDF documents. The tool also supports both client-side and server-side PDF generation.

In addition to a software developer kit for application development, runtime packages are also available as installation kits



PDF/A ‘en masse’

In some cases, instead of needing to archive single documents or hundreds of documents per day as PDF/A, users need to archive large datasets consisting of tens of thousands of documents. The number of invoices, contractual documents, and receipts regularly generated by companies working in telecommunications, energy supply, or public administration, can be extremely substantial. Since these documents are normally personalized (that is, addressed to certain recipients), databases or structured data often come into play when creating them.

PDFlib for high-volume PDF/A generation

The Munich-based company PDFlib supplies tools for developers. The PDFlib program family is used to produce and process PDFs, and enables PDF documents to be generated from structure data (text from databases, XML) using a library ('lib' stands for 'library'). The new PDF files that are created in this way can be filled with variable content if required. This might include adding different names for invoice forms or business cards.

PDFlib products for the automatic generation of PDFs in high-throughput conditions are used in business and prepress workflows and in the Web2Print field. The library has supported the important PDF/X prepress standard for years. As of PDFlib 7, it also supports the high-volume generation of PDF/A-1a and PDF/A-1b documents.

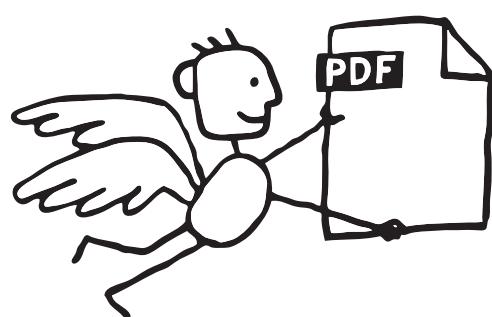
The PDFlib product range offers PDF/A support for various application areas.

- PDF/A documents can be created from scratch. The process can draw on material stored in a database.

- Scanned documents or other pixel-based image files can be converted to PDF/A.

- Existing PDF/A documents can be subjected to further processing in an automated workflow. For instance, they can be merged or split.

- In addition, the PDFlib can create PDF/A-1a files that contain all required structural information.



For more information on PDFlib solutions, see www.pdflib.com on the Internet.

PDF/A ‘from nothing’

This term refers to PDF files for which there is no fully-formed source document. Instead, they are generated ‘on-the-fly’ from variable elements. Example: An Internet supplier provides a password-protected area where customers can download current invoicing documents. Variable data such as names, addresses, customer numbers, and invoice details are delivered from a database. The page layout, company logo, and a current advertisement are often compiled from databases in accordance with the design specifications of the company’s designers. More rarely, fixed page backgrounds are used and the personalized specifications are added to them.

Solutions that are capable of generating PDF documents ‘en masse’ from database-supported content have been on the market for a long time. However, PDF/A-compliance is a relatively new feature. It was introduced immediately after the adoption of the PDF/A standard.

Creating PDF/A from print data streams

Structured data or databases do not constitute the only starting point for the high-volume generation of PDF/A documents – print data streams can also be used to create a large number of PDF documents for archiving. Print streams are used for batch printing output. The print data can be converted in order to generate formats that are suitable for archiving, such as TIFF or PDF/A.



Apfelholz – photocase.com/de



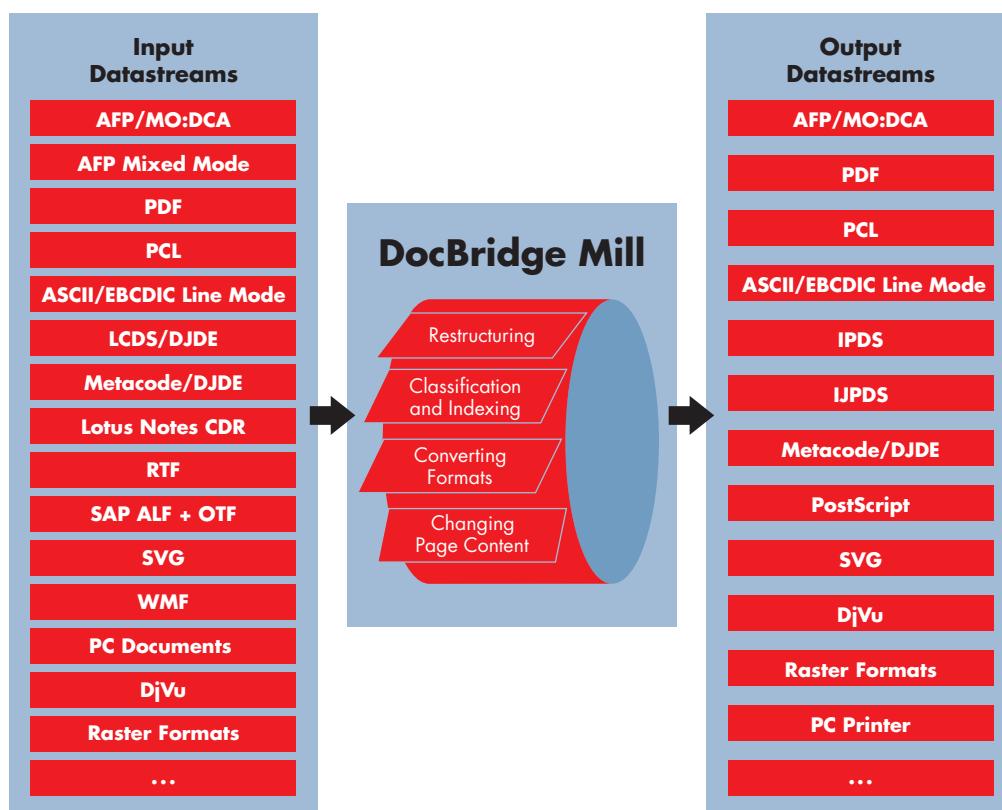
DocBridge by Compart

Compart, which is based in Böblingen, Germany, develops solutions for document management and high-volume printing. Medium-sized and large companies from various industries use this supplier's programs and services to automatically process large amounts of data traffic.

DocBridge, a modular solution constructed from several components, contains the Doc Bridge Mill – a tool for processing a whole range of file formats.

PDF has been part of Compart's development scope as both an input and output format for a long time. In the light of the adoption of PDF/A as a standard for long-term archiving, the company has added an option for PDF/A-compliant output to its products.

For more information on Compart, visit www.compart.net on the Internet. ■



Compart DocBridge Mill: As well as structuring, changing content, and creating indexes, this solution can convert input data streams to PDF/A.

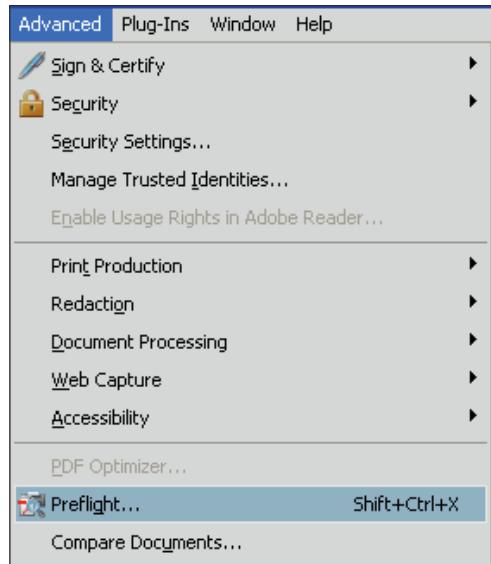
3.

From PDF to PDF/A: Converting PDFs to archive PDFs

Many users already use PDF to store documents in digital archives in companies, public authorities, or privately. Now that the PDF/A standard has been adopted, they have the opportunity to create archive documents from their existing files, thereby ensuring that they can be used in the long term. In addition, recipients of traditional PDF files that need to be retained but are not yet available as PDF/A can now convert them to archive PDF documents. In order to do so, they need to know the answer to the following question: How do you create PDF/A documents from PDF files?

PDF/A generation with Preflight

When Acrobat Professional (Version 8 or higher) is used to convert PDF files to PDF/A, the ‘engine’ that carries out the ac-



Starting the Preflight tool: The command for opening the tool is located on the 'Advanced' menu.

tual conversion is the integrated Preflight plug-in. Even if the conversion is triggered using the Acrobat 8 export function or by

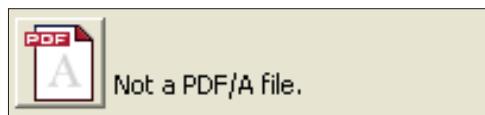


Karoline Swiezynski – photocase.com/de

clicking ‘Save As’, the Preflight module is responsible for converting the file.

The Preflight module is opened from the Acrobat ‘Advanced’ menu or by pressing Shift+Ctrl+X.

The lower section of the main Preflight window immediately provides information on the status of the opened PDF file with regard to the PDF standard: Is the document PDF/A and/or PDF/X-compliant? (PDF/X is a prepress standard.) If the PDF



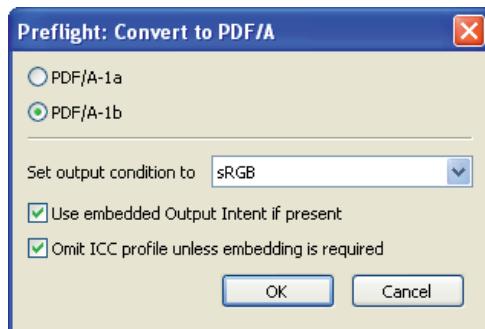
Preflight: The PDF/A icon is also a pushbutton that triggers conversion to PDF/A.

file was not created as a PDF/A, the user receives a message telling him or her that the file is ‘not a PDF/A file’. If the user now wants to trigger PDF/A conversion, he or she can simply click the PDF icon.

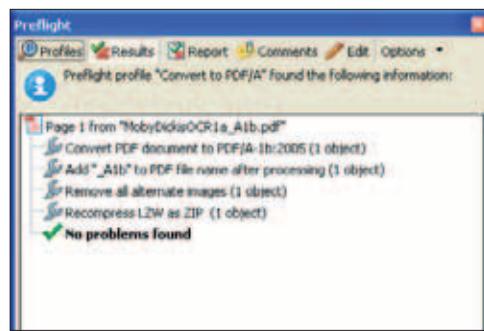
The Preflight tool uses a dialog box to ask the user whether the existing PDF files should be converted to PDF/A-1a or to a restricted PDF/A-1b version.

Conversion to PDF/A-1b

In the first scenario, the user selects the ‘PDF/A-1b’ standard and sets the output condition to ‘sRGB’ in the dialog box. This indicates that the PDF in question is destined to be displayed on a monitor. Since the PDF file quite possibly already contains an output intent, the tool provides a checkbox that specifies that the present intent is to be used. In addition, another checkbox prevents the embedding of the ICC color



Preflight: The PDF/A conversion options relating to the level of the PDF/A standard (1b in this example) and the output intent.



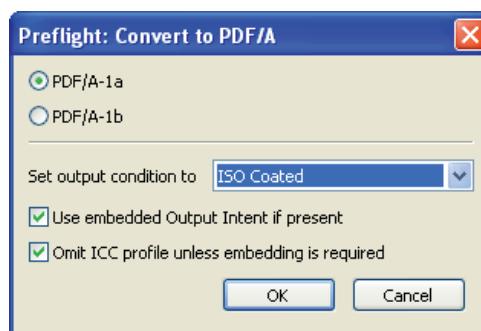
Following the conversion: The Results windows shows the steps that were carried out and informs the user that the conversion was successful.

profile if it is not required. This reduces the resulting file size.

When the user clicks the ‘OK’ button, the Preflight tool searches the existing PDF document to see whether it meets the prerequisites for successful conversion to PDF/A. If the prerequisites are met, the conversion takes place. The green tick in this example shows that no problems occurred during the conversion. Details on the conversion process are shown in the Results window in the form of a list. The list contains information such as the fact that the tool added the file name suffix ‘_A1b’ to the source document.

Conversion to PDF/A-1a

The second scenario describes the conversion of a PDF file to PDF/A-1a. The proce-



PDF/A-1a: Conversion settings with an output intent for professional offset printing.

dure is the same as for scenario 1 except that the compliance level ‘1a’ is chosen along with an output condition that is suitable for four color printing (for example, ISO Coated).

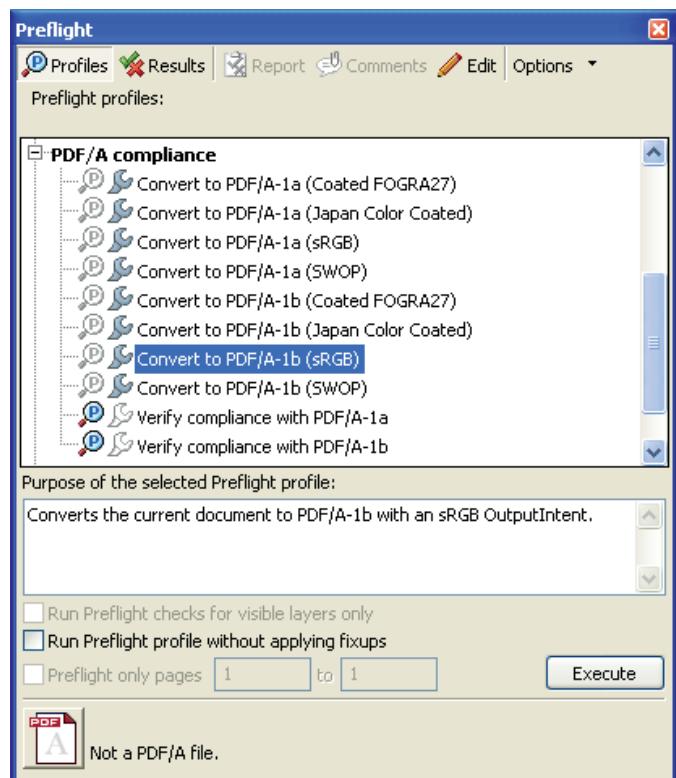
Again, the Preflight tool checks that the relevant document meets the prerequisites for the conversion. →

In this second example, a red X clearly indicates that the conversion cannot be carried out successfully. Preflight uses the Results window to inform the user of the problems that occurred. An additional area below the list explains why the problems that occurred prevent the document from being successfully converted to PDF/A-1a.

For more information on integrating this structural information via tags either before or after conversion, see the 'Accessibility' chapter on page 50.

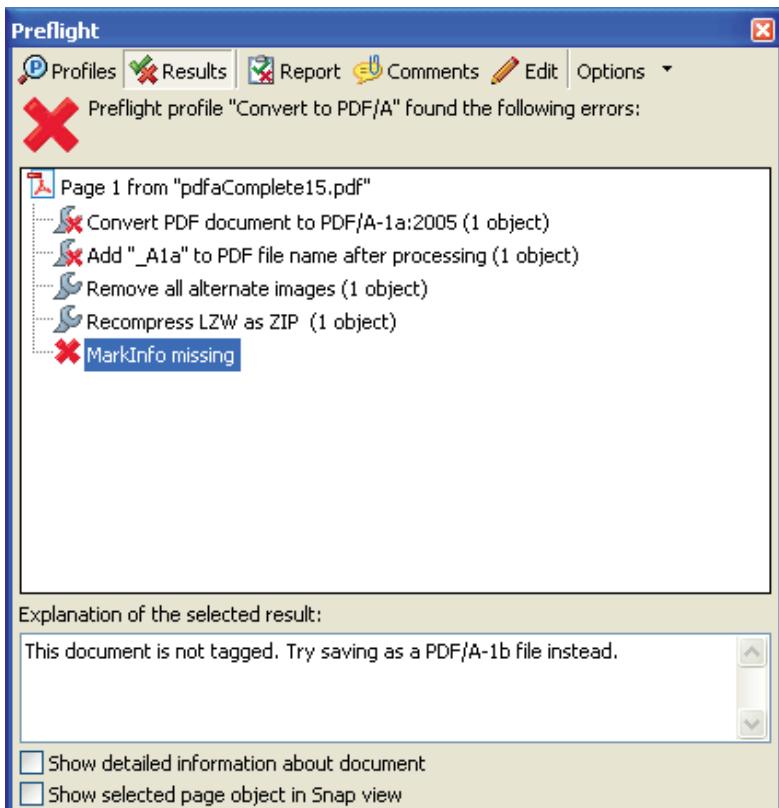
Conversion not possible: If the PDF file in question does not meet the prerequisites for conversion PDF/A, Preflight terminates the conversion process and provides the user with detailed information on the reasons for the failure of the process. For an extensive overview of these error messages, see the appendix.

The file does not have the required MarkInfo entry. This error message is relatively common if the person generating the PDF has not structured the content of the document using tags beforehand. This structural information is one of the things required in order to define the text flow order for document layouts that have multiple columns, images, and captions.



Profile list in Preflight: Both verification and conversion can be carried out using the delivered PDF/A profiles.

In this example, the source PDF document must be re-exported from the source program either using different preparation methods/settings or repaired.



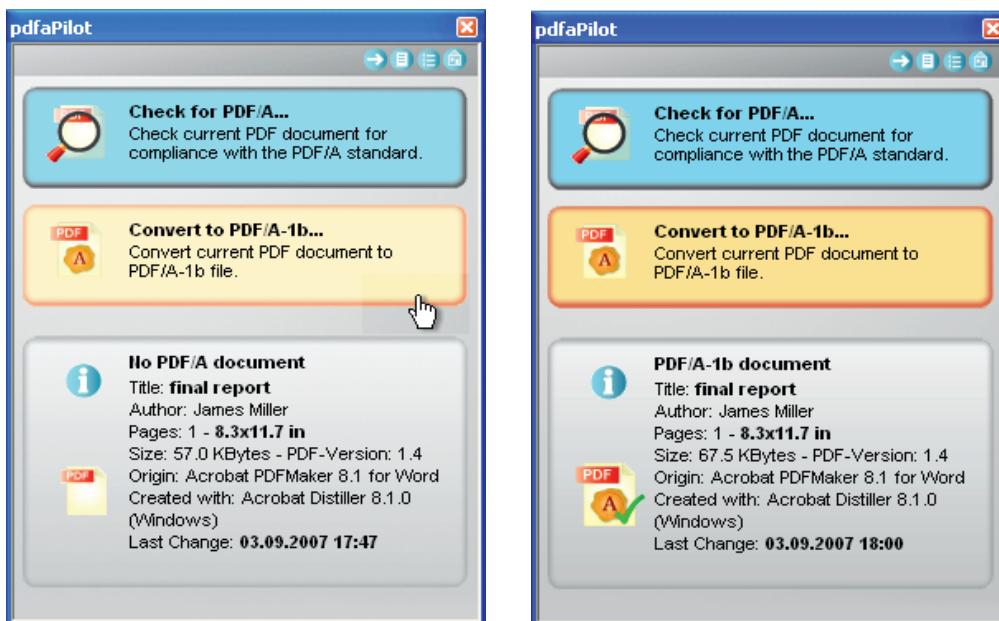
Direct selection of a profile

Experienced users can take advantage of a more direct way of selecting the required PDF/A test or conversion profile.

For example, they can choose to select one of the PDF/A profiles from the list in order to check a document for PDF/A suitability or, if possible, to immediately convert it to PDF/A-1a or PDF/A-1b using a specified output condition. The conversion profiles are all assigned one of the four most common output intents.

If the output intent required for a special workflow is not contained in the list, a new, modified PDF/A profile can be set up on the 'Edit Profile' screen.

The user selects the required profile for the verification or conversion from the list and clicks 'Execute'. Processing can also be started by double-clicking the corresponding profile name. ■



First and last steps: pdfaPilot starts the conversion process when the user clicks on the orange pushbutton. When the conversion has finished, the info field shows that the document is now PDF/A-compliant.

Converting PDF to PDF/A with pdfaPilot

Thanks to its largely self-explanatory user interface, callas software's pdfaPilot allows even unexperienced users with no prior knowledge to convert documents to PDF/A and verify them. This professional tool is a plug-in for Adobe Acrobat Standard and Professional Versions 6, 7, and 8. The conversion from existing PDF documents to PDF/A normally needs three steps and can be achieved in maximum of four:

- The PDF document to be converted is opened in Acrobat. pdfaPilot is called up from the tool icon or using the 'Plug-Ins' menu item.
- Clicking on the 'Convert to PDF/A-1b' pushbutton causes pdfaPilot to start the conversion process.
- If the conversion can be carried out without problems, a dialog box informs the user that the conversion was successful. If the tool found elements or settings for the PDF file that prevent it from being converted to a PDF/A-compliant document, it reports these elements instead. Users can open these error messages by clicking them. They then re-

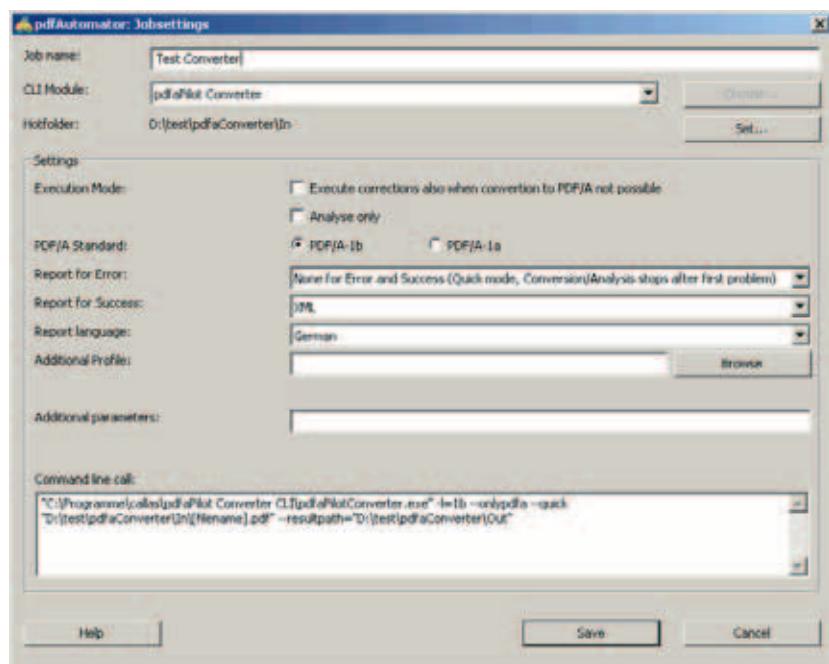
ceive tips on how to solve the problems encountered in order to be able to carry out a successful conversion to PDF/A next time.

High-volume processing with pdfaPilot CLI

The pdfaPilot CLI (Command Line Interface) is designed for high-volume PDF/A conversion and validation. This solution enables the server-based, automated generation of PDF/A files in companies or administrative departments.

For information on pdfaPilot, including a downloadable demo version, go to the following Internet address:
www.callassoftware.com

Automation: pdfaPilot is also available as a command-line (CLI) module. pdfaPilot Validator CLI is a pure validation tool and pdfaPilot Converter CLI can validate, correct, and convert files.



4.

Is this really a PDF/A file? PDF/A validation

A PDF/A document created with Adobe Acrobat can be easily recognized by the file name extension '_A1a' or '_A1b'. Other PDF/A generators use similar procedures. So why is an additional check needed when you receive a PDF/A file by e-mail or open a document from an archive?

The answer is simple: Because PDF/A files cannot be protected from further editing by measures including encryption or passwords. Doing so would contradict the PDF/A regulations, since PDF/A content must be available in its entirety without security measures.

This means that a PDF/A file that was once standard-compliant can lose that

compliance as a result of unintentional or deliberate changes without it being obvious that it is no longer compliant with the standard.

However, further investigation using tools such as Adobe Acrobat Preflight, cal-las software's pdfaPilot, or PDFlib 7 by PDFlib, all of which are specially designed for PDF/A validation, can safely and reliably uncover this kind of problem.

Of course, even deception cannot be ruled out – it is quite possible for users to manually add a file suffix such as '_A1b' to a PDF file before sending it even if the file in question has never actually been converted to PDF/A. This is why checks consti-



Paul Schubert – PixelQuelle.de

tute a prerequisite for the successful use of PDF/A. PDF/A data should be validated for standard compliance at two places in the process flow: When PDF/A files are received and before (external or internal) PDF/A documents are transferred to a digital archive (data storage drive, CD-ROM, or DVD-ROM).



Validation with Preflight

Acrobat 8 Professional's Preflight tool is not designed only for the creation of PDF/A



Calling up Preflight: This tool is called from the Acrobat menu (using the 'Advanced' menu item), by pressing Ctrl+Shift+X, or by clicking the tool icon.

files – it can also be used to test and validate PDF/A documents for their actual compliance with the standard.

The PDF/A icon at the bottom left of the Preflight window gives a quick overview of the PDF/A compliance of an open document. If a user opens a PDF/A file that has not yet been validated, the yellow question mark icon appears along with a message

that names the output intent contained in the PDF document and informs the user that the file has not yet been validated.

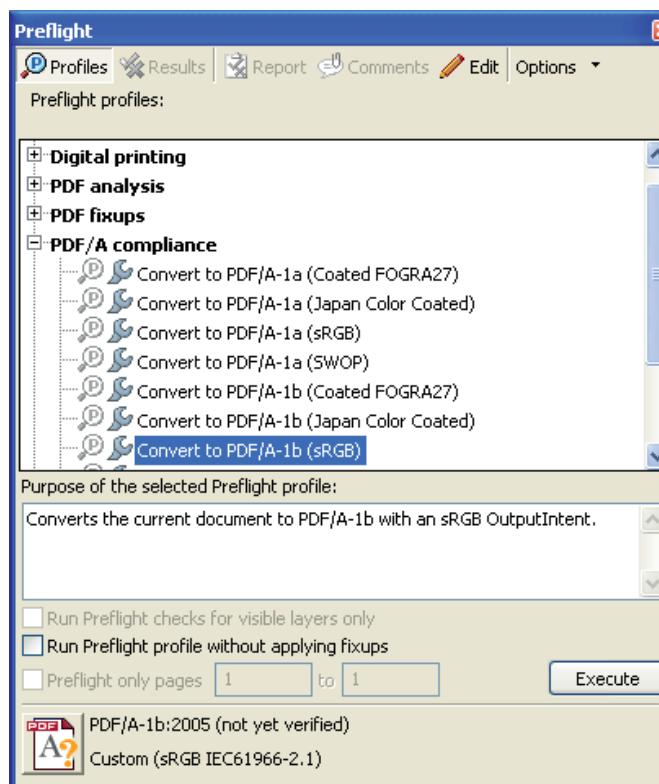
If the PDF/A icon does not appear in the Preflight window, the status display may be deactivated in the Preflight preferences.

Clicking the icon starts the Preflight PDF/A check. The tool works through a list of conditions that the PDF document must fulfill in order to comply with the PDF/A standard. More than one hundred specifications must be observed in order for a document to be declared standard-compliant.

If the check finds no deviations from the standard, the software indicates that the PDF/A file is standard-compliant (indicated by the green tickmark) and names the output intent.

PDF/A status: The status icon has three possible states: A file can be not yet validated, successfully validated, or have failed the validation.

Successful validation: Clicking on the PDF/A icon with the yellow question mark starts the validation process. The result (in this case – successful) appears after a few seconds. Everything's fine.

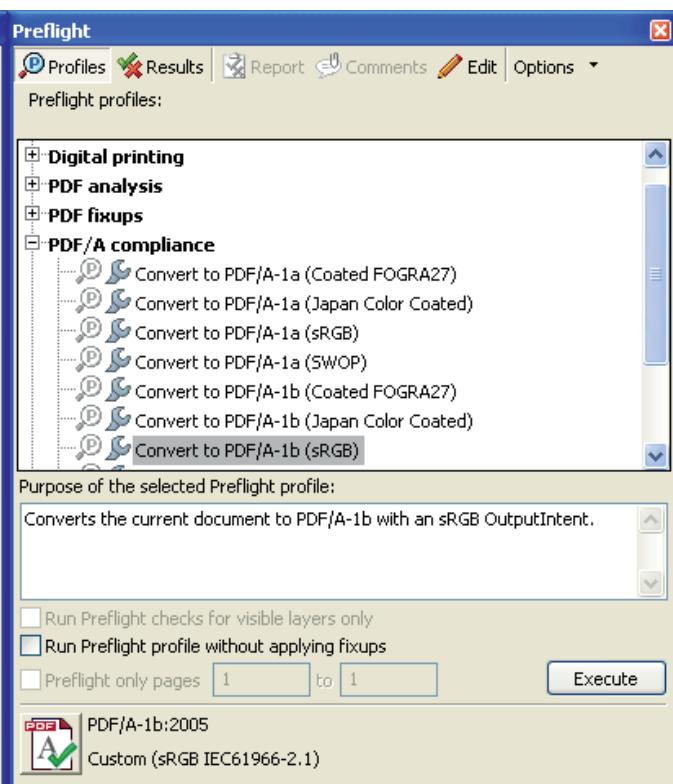


Purpose of the selected Preflight profile:
Converts the current document to PDF/A-1b with an sRGB OutputIntent.

Run Preflight checks for visible layers only
 Run Preflight profile without applying fixups
 Preflight only pages 1 to 1

Execute

PDF/A-1b:2005 (not yet verified)
Custom (sRGB IEC61966-2.1)



Purpose of the selected Preflight profile:
Converts the current document to PDF/A-1b with an sRGB OutputIntent.

Run Preflight checks for visible layers only
 Run Preflight profile without applying fixups
 Preflight only pages 1 to 1

Execute

PDF/A-1b:2005
Custom (sRGB IEC61966-2.1)

The Goal of PDF/A

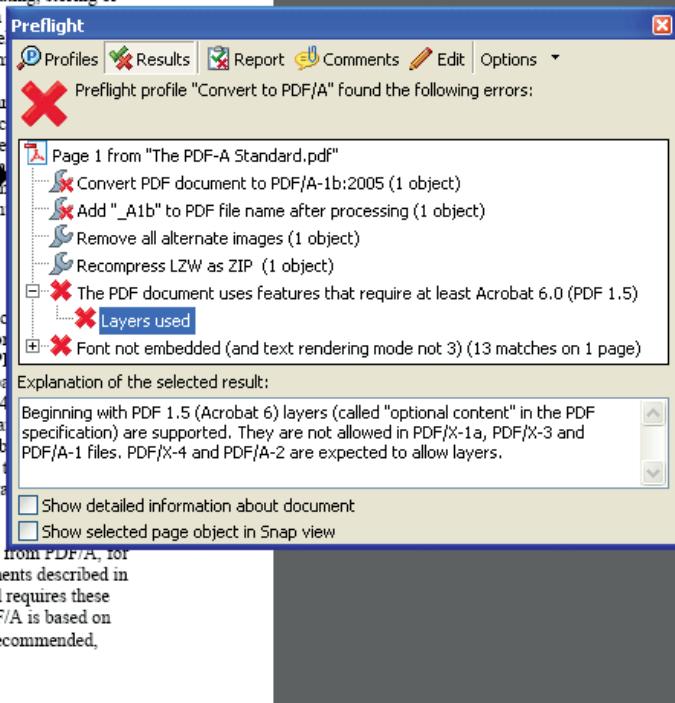
ISO 19005-1 defines “a file format based on PDF, known as PDF/A, which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems used for creating, storing or rendering the files.” (from ISO 19005-1). The Standard does not define an strategy or the goals of an archiving system. It identifies a “profile” for e documents that ensures the documents can be reproduced in years to come.

A key element to this reproducibility is the requirement for PDF/A documents to be self-contained. All of the information necessary for displaying the document in the same manner every time is embedded in the file. This includes all visible text, raster images, vector graphics, fonts, color information and much more. A document however is not permitted to be reliant on any information from indirect external sources, for example links to external images or fonts not embedded.

PDF vs PDF/A

PDF in its native form cannot guarantee long-term reproducibility and no “WYSIWYG” principle (what you see is what you get). Certain restrictions and amendments had to be incorporated into the Standard. To be accepted, PDF/A must be based on an existing version of the PDF Reference and not on anticipations in a future version. The ISO TC 171 chose the Adobe PDF Reference 1.4 as implemented in Acrobat 5, as the basis for the Standard. The ISO Standard PDF/A “shall adhere to all requirements of PDF Reference as modified by ISO 19005”. The Standard also identifies only differences with respect to the PDF Reference. In order to fully understand PDF/A, you have to also understand the PDF Reference 1.4.

Certain functionality allowed in PDF 1.4 has been specifically excluded from PDF/A, for example transparency and sound and movie actions. There are also elements described in the PDF Reference 1.4 that are not mandatory. PDF/A on the other hand requires these elements to be implemented, for example embedded fonts. In short, PDF/A is based on the PDF Reference 1.4, with specific features being either mandatory, recommended, restricted, or prohibited.



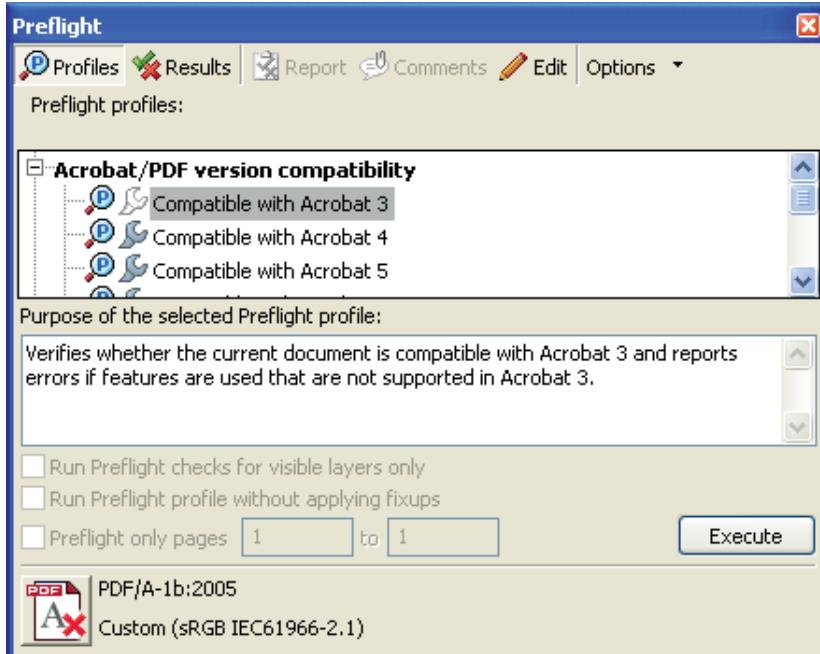
No valid PDF/A file: In this example, the Preflight validation process has found a problem: The insertion of a watermark after the creation of the file added a PDF layer to the file. PDF layers are not permitted in accordance with the PDF/A standard.

The PDF/A validation fails if the document being checked does not meet all of the specifications stipulated by the standard. If this is the case, the system informs the user that problems have occurred by means of a

red X. The Preflight results window contains a list of the problems encountered. Users can click the entries for more information on the various error messages. Preflight can also highlight the places where these problems were found (if the elements allow it to do so). The detailed information can also be viewed by double-clicking an entry in the list.

Because these error messages are not always self-explanatory, this publication contains an appendix that lists detailed background information on all possible errors in alphabetical order. Preflight also gives the user tips on how to repair errors that have occurred or how to avoid them next time around (see information starting on page 68).

Following a failed validation attempt, the PDF/A status is also indicated by a red X in the main Preflight window. ■



*Should be PDF/A-compliant – but isn't:
The Preflight main window uses a red 'X'
to indicate a failed PDF/A validation.*

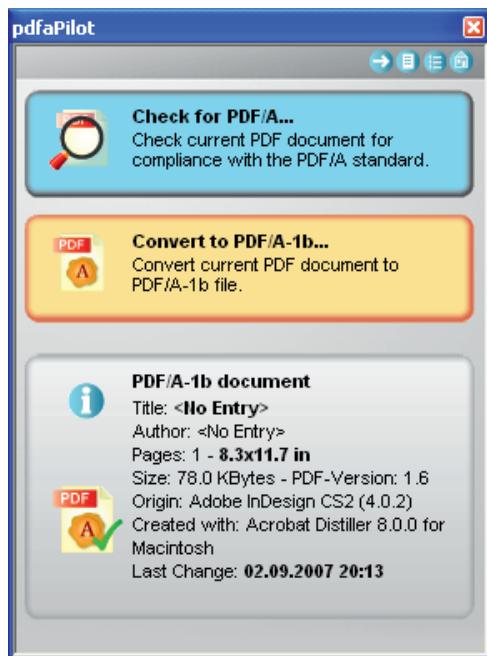
pdfaPilot PDF/A

Like the creation of PDF/A files, pdfaPilot can validate PDF/A files in just a few steps.

Clicking the 'Check for PDF/A' pushbutton causes the tool to examine the open file. If pdfaPilot does not find any problems, it reports a successful check by displaying an icon containing a green tickmark in the info area. It also gives the user further information on the PDF document, including the title, author, number of pages, page size, creation program, and program of origin.

If the validation fails, the system issues an error message informing the user of the problem and listing the ways in which the document in question deviates from the PDF/A standard.

In most cases, pdfaPilot can carry out a conversion to generate a PDF/A file that is suitable for long-term archiving even if problems occur during the validation process. To enable this, the developers of the tool integrated correction options in the tool that far exceed the functional scope of Acrobat Preflight. However, the application is no more difficult to use. If the problems are not corrected immediately in pdfaPilot,



Validation using pdfaPilot: The results show that the file is PDF/A-1b-compliant. In addition, the tool collects details on the existing file and presents them in an overview.



Problem found: Layers are not permitted in PDF/A-compliant files. Clicking on the orange conversion pushbutton eliminates this problem and generates a valid PDF/A file.

The lower pdfaPilot pushbutton changes depending on whether the PDF file validation resulted in serious, minor, or no compliance issues.

- If no problems exist, the PDF is declared to be a valid PDF/A file.
- If there are only minor problems, the document can be converted to a standard-compliant PDF/A file by simply clicking the pushbutton.
- Serious problems must be eliminated in the original file.

the software precisely explains the steps that need to be taken before the creation of the PDF in order to enable eventual conversion to PDF/A. ■

Explanation: Detailed information explains the context of the error messages and helps users to eliminate problems.

5.

Archive PDFs in everyday life: What issues might arise?

PDF/A requirements can change according to the environment in which the PDF/A files are used and the task to be done. One user might produce PDF/A files that only contain text and no illustrations, another might require signatures, and a third might need to create PDF documents that can be archived and also conform with accessibility requirements. The information below provides details on several usage possibilities and areas where PDF/A can be used.

Images

All images contained in PDF/A files must be clearly reproducible. This can only be ensured by integrating them into the files.

They may not be allowed to ‘go missing’ over the course of time, as can happen with other file formats that specify a link to an external storage location rather than integrating images into files. Most of us will, at some point, have called up a Web page only to find that the illustrations are missing and question marks or red crosses in frames are displayed instead. This cannot happen with PDF/A.

An image on a PDF/A page is also clearly reproducible because it exists once and only once. On rare occasions – and only in the prepress area – alternate images are used. These images contain a lower-resolution variant for the screen and a high-resolution variant for printing. PDF/A does not permit alternate images, partly



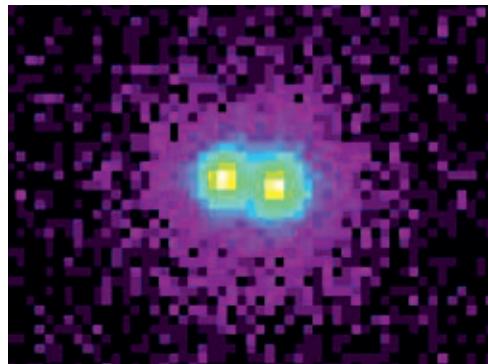
Markus Hein – PixelQuelle.de

because it cannot be guaranteed that the two variants have exactly the same content.

Resolution is not part of the PDF/A standard

Image resolution does not have a role to play when it comes to compliance with the PDF/A standard. This is because there is no single image resolution that is considered to be universally ‘correct’. For example, screenshots tend to have a resolution of of 72 or 96 ppi (pixels per inch). A common resolution for printing is 300 ppi, but it would not make sense to increase the resolution of a screenshot to the normal printing resolution because it would not convey any additional information to the user. If the worst comes to the worst, ill-considered increases in resolution can cause fuzzy edges. The maximum sensible image resolution for screenshots is 72 ppi.

Another area that often deals in low resolutions is astronomical photography. Some of the images transmitted to Earth by telescopes such as the Hubble Space Tele-



‘X-Ray Stars in M15’ – N. White & L. Angelini (LHEA), GSFC, CXO, NASA; www.nasa.gov

Low image resolution is no problem for PDF/A: Since there are subjects where only low resolution can be achieved, the PDF/A standard does not regulate image resolution.

scope are extremely grainy images of distant stars or galaxies. These low-resolution images are the best that can be achieved, and it is, of course, quite possible to create valid PDF/A documents from them.

Image resolution is not regulated by the PDF/A standard – it is left to the decision of the creator of each PDF/A file. Users must decide for themselves whether or not the image resolution used is the best resolution possible.

Color depths and grades

Black and white:

Line art image: 1 bit 2 grades



Continuous tone:

Grayscale: 8 bits 256 grades



Color/RGB: 24 bits 16.7 million grades

Color/CMYK: 32 bits 4.3 billion grades



Line art and black and white images only have two grades.

Half-tone images have different numbers of grades depending on the color model (grayscale, RGB, or CMYK). The compression options differ for black and white images and half-tone images.

Permitted and prohibited compression types

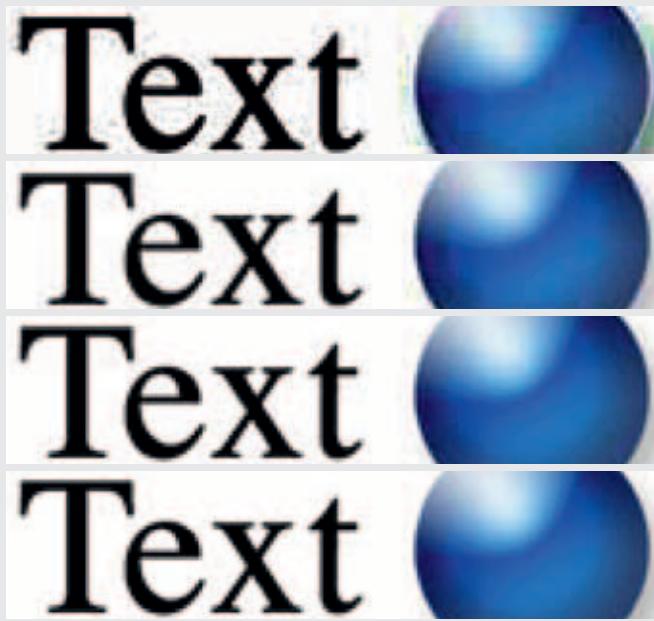
The choice of image compression type – the procedure used to minimize the quantity of image data – is not entirely down to the user. There are two types of pixel image: Half-tone images (grayscale and color images) and line art images that consist of only two colors. Line art images can be compressed for use in PDF/A using ‘CCITT Group 4’, a technology that is effective and prevents loss of data. Programs that use this compression type include Acrobat Distiller and Acrobat Professional.

The choice of compression types for half-tone images is greater, and not all types are PDF/A-compliant.

Of the compression types that prevent loss of data, LZW, a rather old compression type, is prohibited. It was decided to prohibit the use of this compression type because it was once protected by license. Since the more modern ZIP compression type is both permitted and prevents loss of data, it is the recommended compression method to be used for the compression of half-tone images without data loss. →

There are two basic types of compression: Lossless compression and techniques that can damage the quality of images to a lesser or greater extent (‘lossy’ compression).

JPEG compressions (magnified)

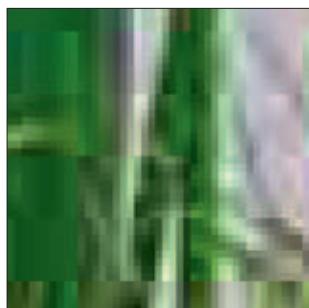


The JPEG compression rate affects the image quality. The compression rate is down to the decision of the user, and is not regulated by the PDF/A standard.

ZIP is not subject to license-related restrictions. LZW compression in a PDF can also be replaced by ZIP compression later on. The PDF Optimizer feature in Acrobat is designed for this purpose.

JPEG was the first procedure to achieve relatively high-quality results from comparatively small images, despite being subject to data loss. For this reason, the triumph of PDF in some situations would not have been possible without JPEG. JPEG enables different file sizes. The image quality can be set in steps from ‘minimum’ to ‘maximum’. If a high level of compression is chosen, block artifacts form. Depending on the nature of the image, they can be clearly visible on the screen. In the case of images with sharp edges (such as text), high levels of compression can be particularly awkward. However, just as for image resolution, the user decides on the level of JPEG compression; the PDF/A standard does not make any specifications.

However, the standard does prohibit the use of JPEG2000, a compression type that was developed by the same group as JPEG (the Joint Photographic Experts Group). JPEG2000 was introduced for PDF with



Blocks: Zoomed view of JPEG artifacts resulting from heavy compression.

JPEG minimum
285 KB
(US Letter page)

JPEG medium
325 KB
(US Letter page)

JPEG high
405 KB
(US Letter page)

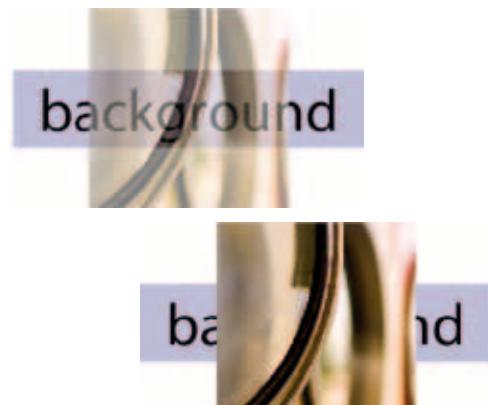
JPEG maximum
509 KB
(US Letter page)

Acrobat 6 (PDF 1.5). Because PDF/A-1a and -1b are based on PDF 1.4, JPEG2000 is prohibited simply because it was introduced too late. However, the version of the standard that is currently being compiled, PDF/A-2, will include JPEG2000. For the moment, if a PDF document contains images compressed using JPEG2000, the user can replace the prohibited compression type with JPEG or ZIP using the PDF Optimizer in order to achieve PDF/A compliance. This checklist lists specifications for images in PDF/A documents:

- All images must be an integral part of the PDF file in which they appear.
- Alternate images are not permitted.
- The user must decide on the image resolution and compression level (both factors influence the image quality).
- The compression types LZW and JPEG2000 are prohibited.
- For more information on colors (including colors in images) see the information starting on page 46.

Transparency

Transparent objects are not allowed in PDF/A-compliant documents. At the point when the PDF/A standard was adopted,



Transparency: The upper image above is transparent. The transparent object can be recognized easily, since the background is visible through the image. The lower image has an opaque foreground image.

Adobe had not yet completely formulated the algorithms for evaluating transparency in a completely clear manner. As a result, transparency is currently prohibited in the PDF/A standard. This will change in PDF/A-2.

Transparency can effect images, graphics, and text. Transparent objects are not 100% opaque – instead, their background shows through, as is the case for glass or thin parchment. Transparent objects cannot always be detected with the naked eye, especially since opacity can be as high as 99%.

Transparent elements do not only occur if they are explicitly created. Certain popular design functions such as drop shadows and soft edges can create ‘sneaky’ transparent objects. For example, many PowerPoint presentations contain transparent objects, even if they cannot be detected at a glance. If text or other elements are given drop shadows, transparent objects are created when they are converted to PDF.

Another widely used aid in Office environments is the function that allows text to be highlighted with a digital marker pen. This function is also available in Acrobat Professional. However, this also causes transparent objects to be created in PDF files. So how can such transparent



objects be avoided or removed later on? This is normally done by transparency flattening (transparency reduction). This procedure involves merging the transparent area and the background in a way that retains the appearance of the image. In addition to certain professional layout programs that can carry out transparency flattening in advance, this process can also be carried out during PDF optimization in Acrobat Professional. →

Risk of transparent objects: Drop shadows and soft object edges can cause the creation of transparent elements in a PDF file. This example is from a PowerPoint presentation.

The PDF/A-Standard

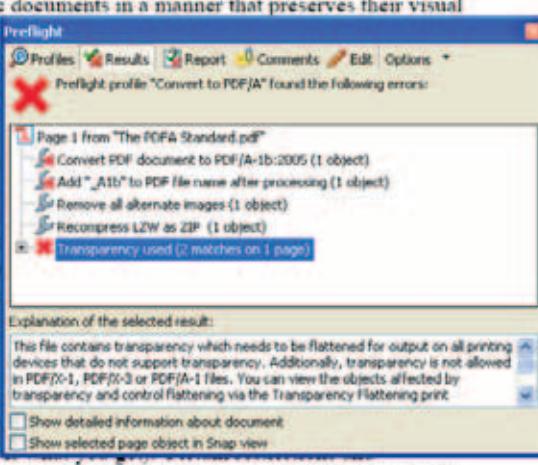
The Goal of PDF/A

ISO 19005-1 defines “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 rendering the files.” (from ISO 19005-1) strategy or the goals of an archiving system documents that ensures the document

A key element to this reproducibility is self-contained. All of the information in the same manner every time is embedded in the document. Text, raster images, vector graphics, form fields, and other data in the document however is not permitted to come from indirect external sources, for example being embedded.

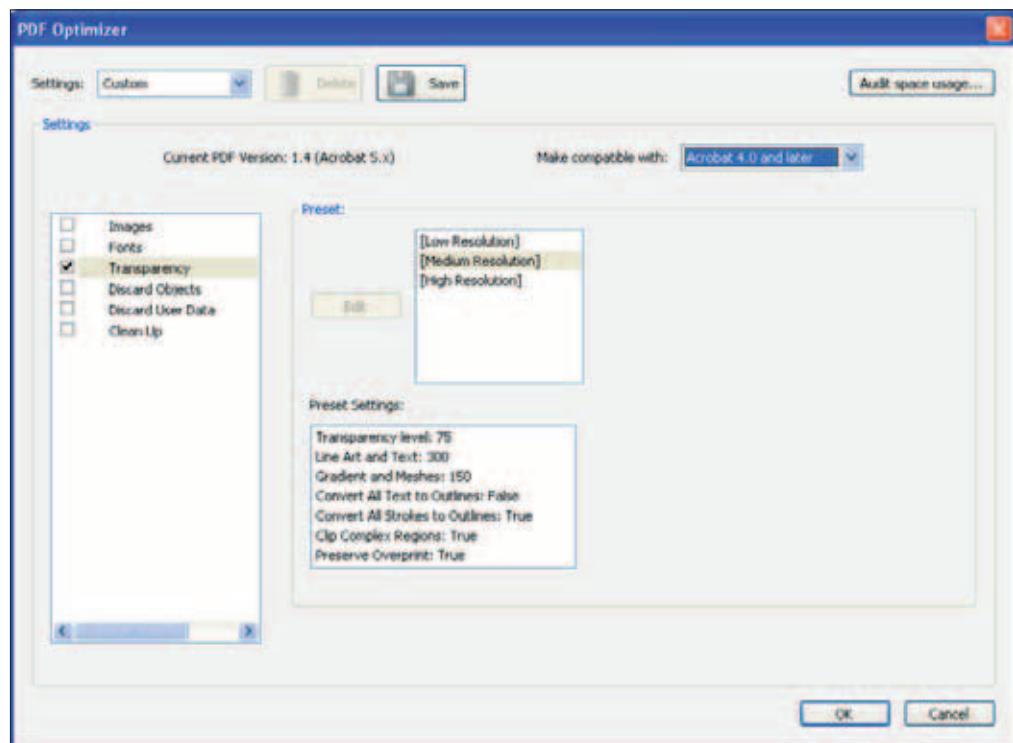
PDF vs PDF/A

PDF in its native form cannot guarantee the “WYSIWYG” principle (what you see is what you get). Amendments had to be incorporated into the Standard. To be accepted, PDF/A needed to



Transparency resulting from the use of a highlighter: Use of the Highlight Text Tool can cause the creation of transparent objects in a PDF document.

PDF Optimizer: This Acrobat function enables transparency flattening with different quality settings. To ensure PDF/A compliance, it is important to select a compatibility level no higher than Acrobat 5 (PDF 1.4).



When flattening transparency, the user can choose between different quality levels (from low resolution to high resolution), since this process generates new images out of overlapping graphic objects.

However, users must be careful when removing highlighted text. Instead of using transparency flattening, which would make the yellow highlighting opaque and hide the text, the Acrobat PDF Optimizer function ‘Discard all comments, forms and multimedia’ should be used. This function can be called from the ‘Discard User Data’ area.

of colors for text, image, and graphical elements.



Which color should it be? Without color management, the correct depiction of colors in company logos is a question of luck.

Color management

PDF/A uses color management to safely depict colors. Color management is based on the use of color profiles that are appended to image files, graphical documents, and PDF files to act as a kind of instruction manual.

The RGB color space is widespread in Office environments. sRGB ('Standard RGB') is now being used to enable colors to be displayed or printed as reliably as possible on different devices and printers. The sRGB profile is suitable for images, graphical elements, and text in Office documents. It was developed by Hewlett-Packard and Microsoft in 1996 to make printed pages as similar to those displayed on the screen as pos-

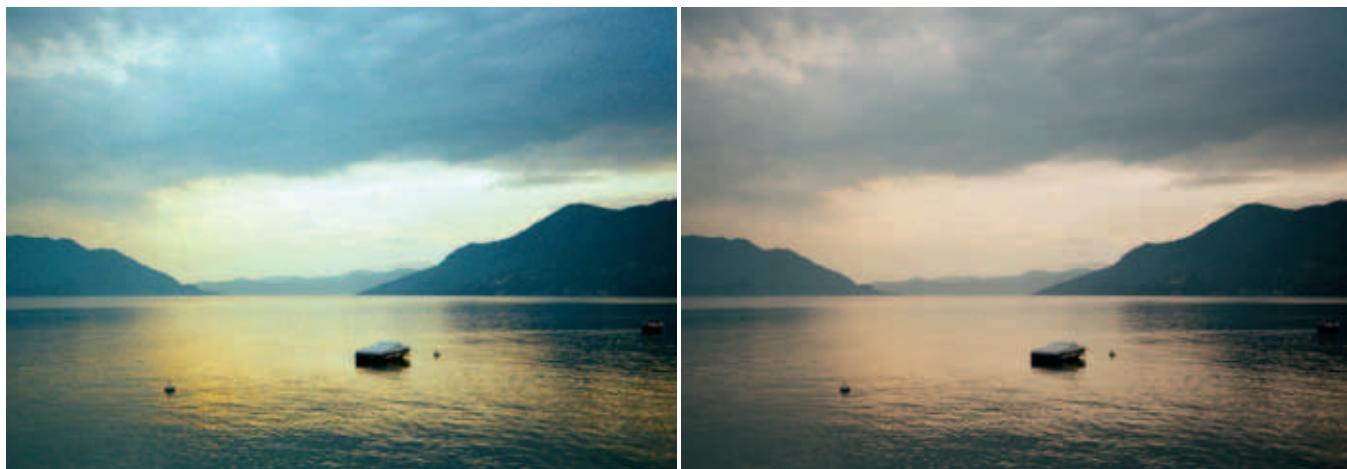
The PDF/A Standard
The Goal of PDF/A:
“A file format based on PDF, known as PDF/A, designed for representing electronic documents in a manner experienced over time, independent of the tools and systems in reading the file.” (ISO 19005-1). The Standard does not change the goals of an existing system. It identifies a “pdf” document that ensures the documents can be reproduced in a
Key element to this is the requirement for PDF/A-compliant: All of the information necessary for display

Hidden text: Transparency flattening
should not be used for highlighted text. It is better to use the ‘Discard all comments, forms and multimedia’ function, since the Highlight Text Tool is a comments tool.

Colors

The colors of illustrations and graphics in a document should always appear exactly the same – whether displayed on one’s own monitor, on a colleague’s monitor, or viewed as a printout. Nothing is more annoying than a company logo that, when used in a presentation or brochure, fails to depict the corporate identity because, for example, it appears orange rather than magenta.

Thanks to PDF/A, such problems are a thing of the past, since the PDF/A standard guarantees the reliable reproduction



sible. Common modern monitors and printers support sRGB color adjustment.

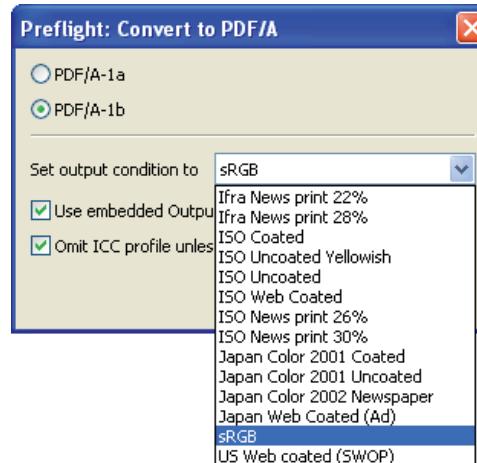
Adobe RGB is another widespread RGB profile. It was published by Adobe Systems in 1998. This profile is most useful to people who work with digital photographs, since cyan and green tones appear to be more natural with Adobe RGB than with sRGB. For documents always intended for four-color printing (production or digital printing), the ISO Coated color profile constitutes a good choice.

Output intent (output condition)

The profiles named above (and other profiles) can be passed to the conversion process along with each individual object placed in a document, but there is another, more practical procedure, that is applied to the entire PDF/A file. An output intent (the intended output condition) can be specified for the PDF/A conversion process. For example, if a PDF/A file is to be archived for the purpose of being displayed on a monitor later on, the sRGB profile, which is a standard part of PDF/A converters such as

Acrobat, Preflight, and pdfaPilot, is ideal. On the other hand, PDF/A files that are intended for printing can be given an ISO Coated profile.

The incorrect reproduction of colors can sometimes affect the message of an image: Was the evening spent at the lake depicted in these two photographs a warm evening or a cool one?



Output intent: What is the purpose of the PDF/A document? In this case, sRGB is the chosen output intent. Acrobat (Preflight) and other converters deliver a range of profiles. In addition, users have access to other profiles stored on their computers.

If the output condition of a PDF/A file changes at any point in the future, color conversion processes can be triggered. ■

Safe depiction of colors in PDF/A

- If device-dependent colors are used, an output intent must be specified.
- If there is already a source profile for all colors used, there is no need to specify an output intent.
- If an output intent is used, it must have one output profile only.
- Objects such as images and graphics can exist in different color spaces (RGB, CMYK, spot colors, grayscale and Lab).
- It is also possible to use a single device-dependent color space (with no ICC profile).
- Device-dependent CMYK and device-dependent RGB may not be used together. If device-dependent colors are used, there must be an output intent for the same color space (RGB, CMYK, or Gray). However, only one output intent color space may be used.

Global depiction of characters: PDF/A ensures that international texts are displayed properly, since the embedding of texts in documents means that all required characters are actually available within the document in which they are used.

International languages		International languages
العربية	日本語	□□□□□□□
Česky	Norsk (bokmål)	□□□
English	Polski	Norsk (bokmål)
Ελληνικά	Português	Polski
Español	Русский	Português
Français	Slovenčina	□□□□□□□
עברית	Slovenščina	□□□□□□□
Italiano	Türkçe	Slovenčina
Lëtzebuergesch	中文	Slovenščina
		Türkçe
		□□

Fonts

Because entire fonts can constitute large datasets, PDF/A permits the embedding of subsets. This means that only characters used in a particular PDF document are embedded. This limits the file size.

If a PDF file contains text that uses fonts – that is, text that has not been converted to paths/pixel-image text – there are a range of specifications for achieving PDF/A-compliance. The specifications for PDF/A-1a and PDF/A-1b are different. However, we shall first deal with the common specifications for both standards.

Embedding fonts

The following applies to both compliance levels – PDF/A-1a and -1b: All used fonts must be embedded into the PDF file in question. If this were not the case, text dis-

played on a computer that does not have the font used might not be displayed in its entirety. This is incompatible with the required visual reproducibility. The entire font does not need to be embedded; it is sufficient to embed only the characters that are used in the document. This is known as ‘embedding subsets’.

In the light of the international exchange of documents containing special characters that the recipient might well not have on his or her computer, the use of embedded fonts is a significant advantage. Modern operating systems increasingly provide Cyrillic, Asian, and Eastern-European fonts in order to enable the display of interna-



photocase.com/de

Eine Überweisung von 100 €

Missing characters: In this case, it is impossible to tell whether the transfer should be 100 €, £, or ¥. This cannot occur with PDF/A.

tional Internet pages, but there is no guarantee that the fonts delivered will be the fonts used by the creator of a particular PDF document.

Unlike in the early days of PDF, embedding fonts with the current program versions of Acrobat and many other professional tools for creating PDFs is not difficult. However, even today there are solutions – even at industry level – that do not fulfill the font handling specifications stipulated by the PDF/A standard.

Fonts must be uniquely encoded

Problems with character set encodings can cause individual font glyphs to be displayed incorrectly or not at all in documents including Word documents and e-mails. ‘Glyphs’ are the graphical depictions of characters.

What might happen if character set encoding is inconsistent? When the euro was introduced, problems were often experienced with the € sign. The characters ‘ä’, ‘ü’, and ‘ö’ often cause difficulties in international communication. The PDF/A standard now requires glyphs that are used in documents to be uniquely encoded in order to guarantee correct reproducibility.

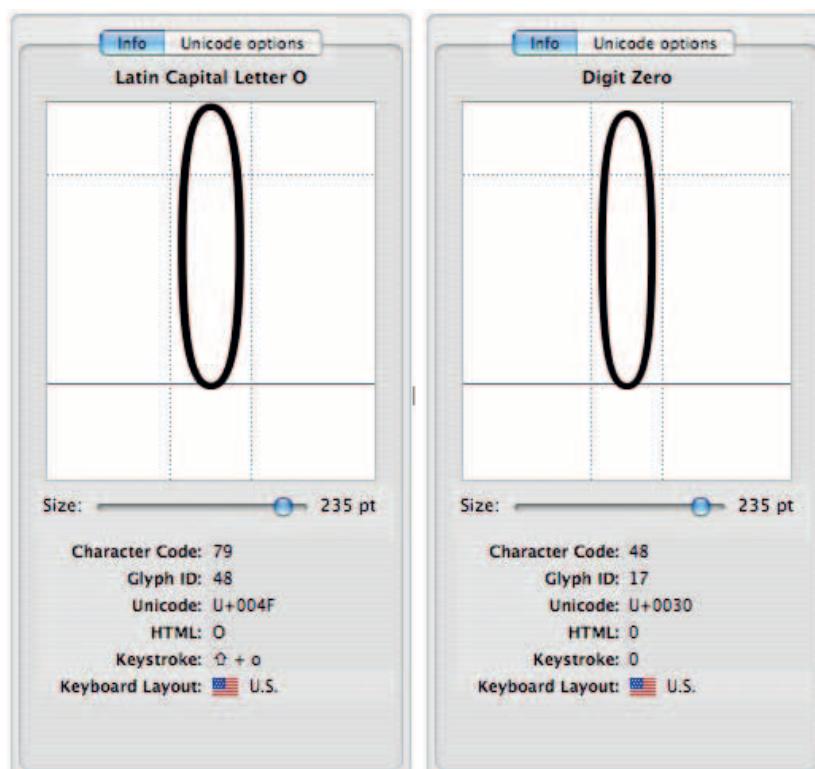
The quick brown fox jumps over the lazy dog
The qu c k b r n f x j u m p s o v e r t h e l a z y d g

Tracking information: The information on tracking has been lost in the case of the overlapping letters.

Overlapping letters such as those that can occur when copying text are also eliminated by compliance with the PDF/A standard. The gobbledegook shown here is caused by missing tracking information. This problem cannot occur if PDF/A is used.

Unique characters with PDF/A-1a – thanks to Unicode

In addition to the points mentioned above, a further font requirement applies to



PDF/A-1a. All of the characters embedded in a PDF/A-1a file must be uniquely identifiable by means of their Unicode name. Unicode is an international standard that assigns a unique ID number to every char-

The letter 'O' or the digit '0'? This example shows that it is not always possible to distinguish between certain characters with the naked eye. This is where Unicode comes into play, since it defines a Unicode name for each individual character. (Illustration: Linotype FontExplorer X)



U+0061: All of these letter 'a's have the same Unicode numbers, regardless of the font

acter and symbol that exists worldwide (even for historic script). The Unicode Consortium and ISO work together on this project. Unicode encodes only abstract characters, not glyphs (the various graphical depictions of letters).

The use of Unicode encodings for PDF/A-1a brings the advantage of all character-based text being completely unique. This enables text to be searched precisely and reliably for content as well as allowing content to be reused. This is not completely guaranteed in the case of PDF/A-1b documents, although it should usually be the case.

Metadata

Many current file formats allow the storage of metadata. Metadata is data passed for a document over and above the actual working data. This might include technical information (for example, a digital camera saves additional information called EXIF data for each image file, including the exposure, aperture, and focus). In addition, users can add descriptions to files later on, such as keywords or copyright notices. IPTC metadata, primarily used by professional photographers, has been around for years.

Many programs allow metadata on a file to be viewed or even changed in the 'Properties' area. This normally includes core information such as the title of the document, the author, and the program used to create it.

Metadata provides useful information that can simplify the organization of large numbers of digital documents, either in database solutions or using search functions. Metadata is particularly important for archiving, since it can provide infor-

mation on a person or place depicted in an image, the author of a document, and any copyright restrictions.

PDF/A and metadata

Many issues relating to the use of metadata when creating valid PDF/A documents are left to the decision of the user. However, the following directives apply:

■ One metadata field is mandatory: The PDF/A identifier. This identifier is normally automatically written to the relevant field in its correct form by the PDF/A converter used to create the PDF/A document in question.

■ All metadata attached to a PDF file must exist in a certain form and must be encoded in an XMP-compliant manner. Although PDF/A only stipulates a single mandatory field, it makes sense to make the most of the possibilities of XMP to enable efficient archiving and powerful search and sorting functions.

What is XMP?

Metadata is another topic where standards are important. This type of data cannot be used effectively if every single user or user group develops their own system for creating and managing additional information. In the case of metadata systems that already exist in parallel, a reliable method for converting one to the other must be provided at the very least.

To promote the standardization of metadata systems, Adobe Systems is now using the Extensible Metadata Platform (XMP). XMP is a procedure that acts as a kind of bracket that pulls together established metadata formats such as IPTC and EXIF. Acrobat Professional and Adobe Reader are two of the applications that display XMP metadata; Acrobat Professional also allows it to be edited. Other manufacturers also use XMP – the technology is not limited to Adobe.

Viewing and editing PDF metadata in Acrobat

Metadata can be viewed at the following menu path: 'File' → 'Properties'. The 'De-

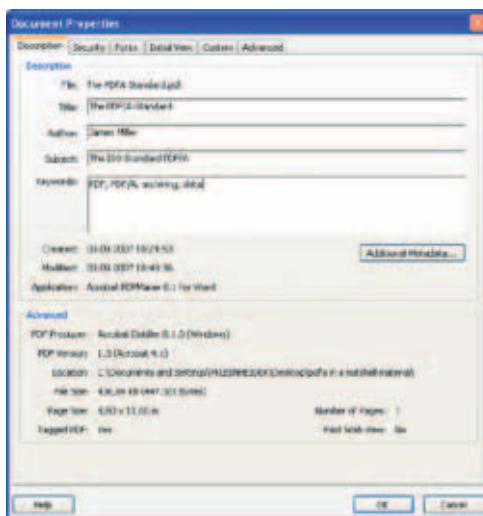


aoe

Analog metadata: Metadata also occurs in the analog world in the form of information such as imprints and mastheads for books and journals.

scription' tab contains fields that specify the title (which does not have to be the same as the file name), author, subject, and keywords (freely definable). The 'Title' field is normally filled with the file name of the original file. The other fields can contain metadata from the original file if the user gave them XMP-compliant data and as long as the PDF is not being created using the Distiller. Programs in the Adobe Creative Suite pass on XMP metadata to PDF documents that are created using the Export function. The extent to which metadata can be transferred from Word or Excel files to the corresponding PDFs depends on factors including the program version being used.

Clicking the 'Additional Metadata...' button displays a whole range of further categories including options for copyright information, personal processing notes, and



Document Properties: There are four basic metadata fields on the initial screen of this area: Title (this field is usually prefilled on the basis of the source document), Author, Subject, and Keywords. Note the 'Additional Metadata...' button. It calls the dialog box shown below.

additional descriptions. Other programs than Acrobat (such as Adobe Bridge) and products and solutions offered by other suppliers are recommended for the mass allocation of metadata in PDFs. ■

The PDFA Standard.pdf

Description

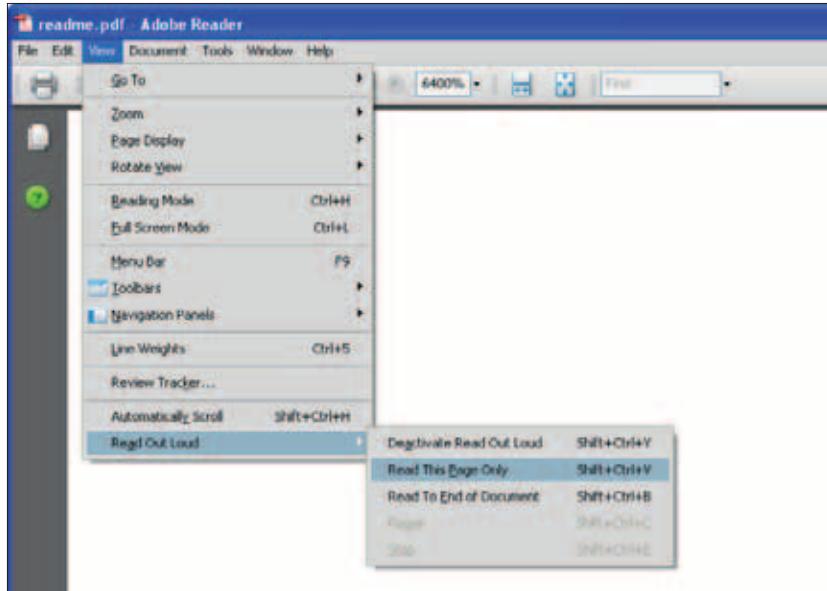
Description	Document Title: The PDF/A-Standard
	Author: James Miller
	Author Title:
	Description: The ISO Standard PDF/A
	Description Writer:
	Keywords: PDF; PDF/A; archiving; data
	<i>i</i> Commas can be used to separate keywords
	Copyright Status: Copyrighted
	Copyright Notice: (c) 2007
	Copyright Info URL: Go To URL...
Created: 03.09.2007 18:24:53 Modified: 03.09.2007 18:43:36 Application: Acrobat PDFMaker 8.1 for Word Format: application/pdf	

Powered By **xmp™**

Additional Metadata...: This section contains additional fields for XMP-based metadata. These fields can be filled in for the PDF but may already contain entries that were adopted from the metadata in the source document.

Accessibility

Accessibility refers to the concept of providing technical aids that make it easier for people with disabilities to participate fully



Accessibility feature: Adobe Reader can read PDF files out loud – including form fields. However, operating systems only deliver English ‘voices’ – for other languages, such as German, digital voices must be purchased at additional cost.

in everyday life. Examples of successful accessibility measures include wheelchair access to the metro and buttons with braille lettering in elevators. In today’s information society, it is important to ensure that

nobody is prevented from accessing public information simply because he or she has a physical disability. Access to the increasing amount of digital information must also be ensured for members of the public who are visually impaired or have restricted motor skills.

PDF has a range of useful functions for enabling the accessibility of content: The free Adobe Reader can read text out loud. Magnification and contrast options enable text to be read by readers who have impaired vision.

Structured: PDF/A-1a-compliant documents and accessible PDFs

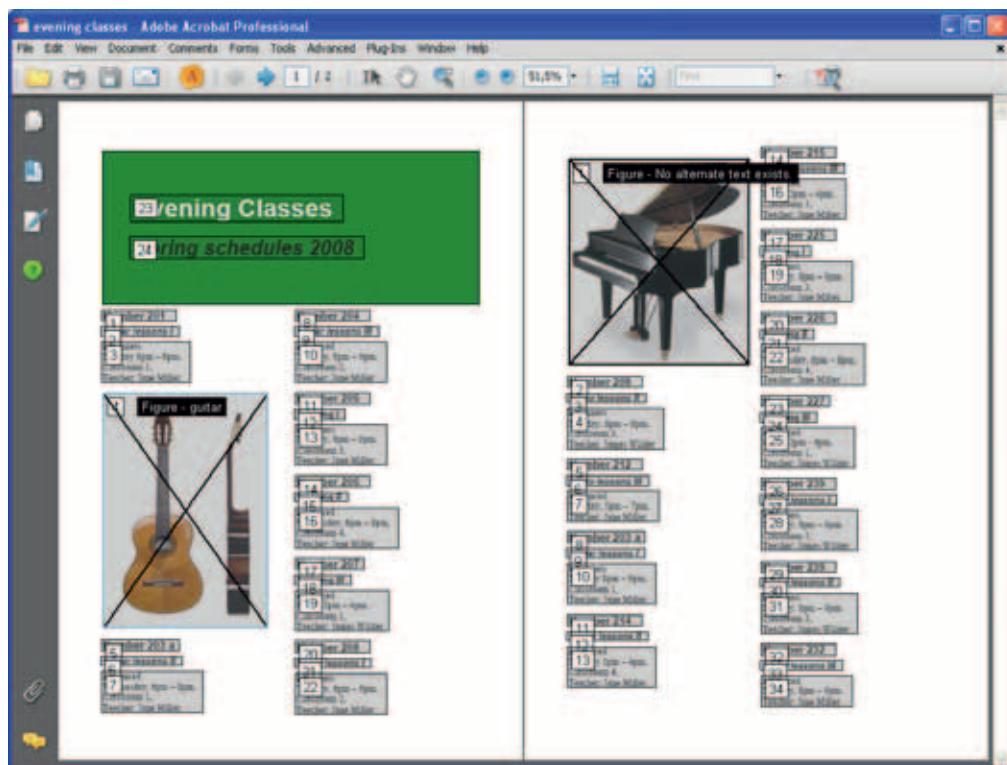
Accessible PDFs and PDF/A-1a-compliant documents have many things in common, and it is perfectly possible to create files that are both. Both PDF/A-1a-compliant documents and accessible PDFs require a structure with well-defined content.

This structure is realized by means of ‘tagged PDF’. The tags used give each PDF element additional information on content, position, and type.

Tags also define the order of content, which is particularly important in the case of pages with multi-column layouts. In



Jens Goetzke – PixelQuelle.de

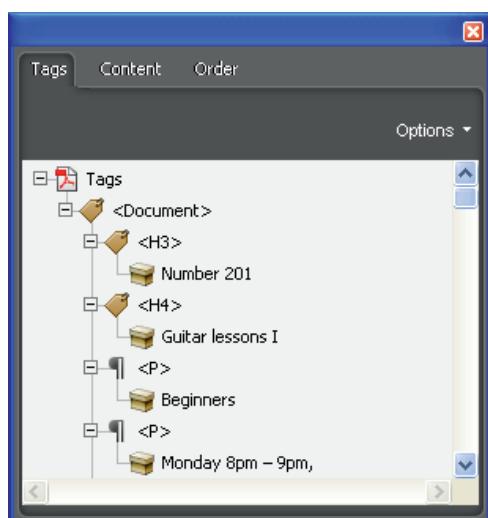


Tagged PDF in Acrobat: Structured PDFs
specify the exact sequence of their content.
In the case of multi-column layouts, it
might be impossible for software to auto-
matically determine the order of the con-
tent. The authors of documents must spec-
ify this sequence.

addition, tags can be used to distinguish between content and additional elements such as headers and footers or other background elements that do not directly belong to the content. Tags are also helpful for graphics and images on PDF pages. How do screen readers deal with images? If the creator has given the image 'alter-

nate text' that explains the subject, the user is told not only that there is a graphic at the relevant point in the text but also that the graphic displays a guitar, for example.

It is relatively easy to generate a PDF/A document from an accessible PDF and vice versa. Note that the conversion to PDF/A takes place at the very end of this process. Once a valid PDF/A file has been created, it cannot be changed – otherwise, it loses its compliance status.



Tags: Each element in a tagged PDF file has a tag that contains information on its type, position, and content. These tags are used to define the exact structure of the document.

Advantages of accessible PDFs

Thanks to tagged PDF, there are tangible advantages of accessible PDF files. Structured PDF files are easier to reuse than traditional documents. This means that reliable results can be obtained when converting formats (for example, PDF to HTML, TXT, or RTF).

In the case of PDF documents that are to be displayed on the small screens provided by mobile devices such as handhelds and cellular phones, the reflow function enables better readability. This text function can only be carried out without errors if tagged PDF is used. →



Reflow: PDFs can also be displayed comfortably on mobile devices following the successful 'reflowing' of content that is enabled by tagged PDF.

Accessible PDFs also permit safe full-text indexing and searching, since there are no ambiguities in the text flow.

Automatically meaningful structures?

Neither accessible PDF nor PDF/A-1a can enable a check of the tagged PDF to make sure that the structures of a document are meaningful or correct. Both types of check can only determine whether structural information exists in the specifications of the PDF file – not whether any structural information found makes sense.

For this reason, the standard stipulates that structural information may not be added automatically later on. It must be imported during the creation of the PDF or added manually afterwards.

The automatic creation of structures might be possible without causing problems for very simple PDF files. However, if a user uses an automated process to reconstruct a structure, he or she must make sure that the process is validated.

Creating an accessible PDF file from Word

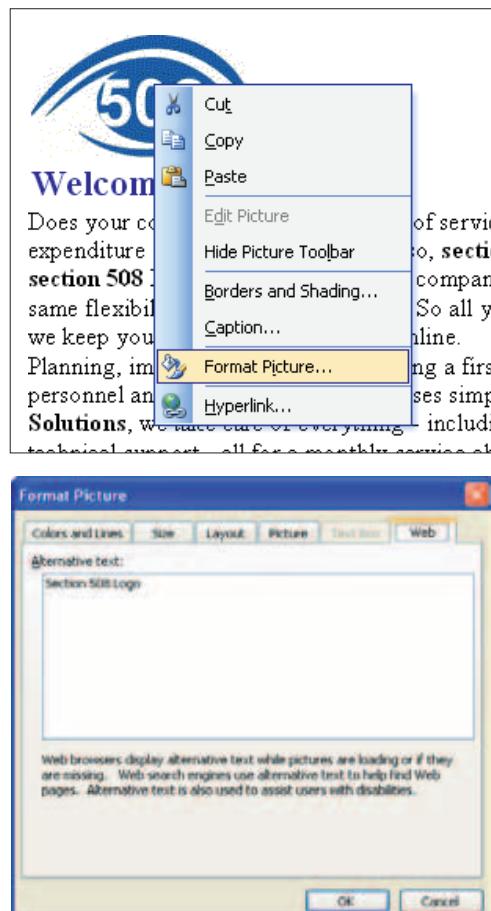
This example uses Word 2003 and the PDFMaker. The PDFMaker also accesses Acrobat 8 PDF settings. The following steps must be observed to obtain a successful result:

- To ensure accessibility, a language must be defined for the Word document. This setting is made under ‘Tools’ → ‘Languages’ in the menu. However, PDFMaker does not always transfer this information to the PDF file.

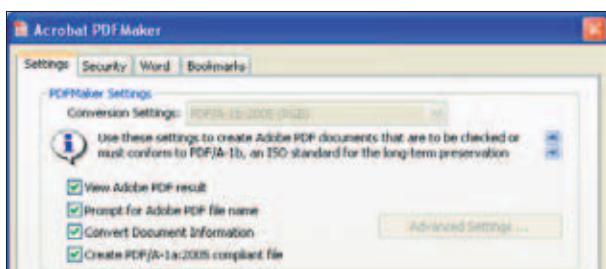
- The text in the Word document should be structured using styles (such as ‘Heading’, ‘Body Text’, and ‘List Bullet’).

- Multi-column layouts must be defined using the ‘Format’ → ‘Columns...’ function, and not using the tab key.

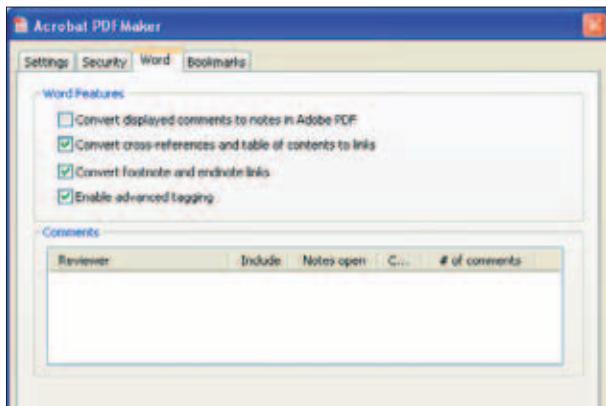
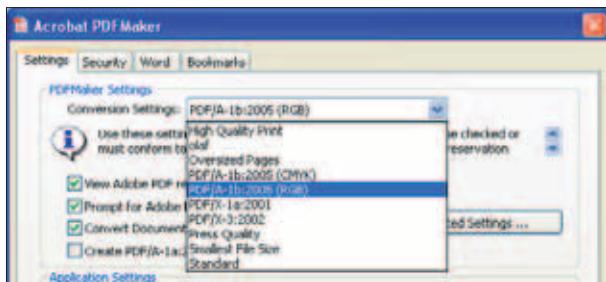
- The ‘Web’ tab reached by clicking the ‘Format Picture’ context-menu command must be used to give graphics and images alternate texts in Word. These image descriptions are transferred to the PDF document by PDFMaker during the creation of the PDF.



Formatting pictures in Word: Right-clicking opens the ‘Format Picture’ dialog. Alternative text can be entered on the ‘Web’ tab.



Accessibility and PDF/A: The user either selects the PDFMaker PDF/A-1a checkbox on the 'Settings' tab...



...or chooses one of the two PDF/A-1b variants for RGB and CMYK from the menu. In the case of PDF/A-1b, the 'Enable advanced tagging' checkbox must be selected in the 'Word' section to allow the generation of accessible PDF files.

■ The user then selects the 'Change Conversion Settings' command from the 'Adobe PDF' Word menu. On the 'Settings' tab, the user can activate the PDF/A-1a checkbox or select one of two PDF/A-1b settings from a pulldown menu. The 'Enable advanced tagging' option in the 'Word' area is active by default and ensures that PDF/A-1b files meet the tagged PDF requirements.

■ Choosing 'Convert to Adobe PDF' from the 'Adobe PDF' menu now causes the PDF-Maker to create a PDF/A-1a/ or PDF/A1b-compliant file that also meets the requirements for an accessible PDF file.

Adjustments in Acrobat Professional:

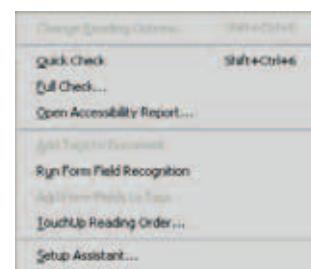
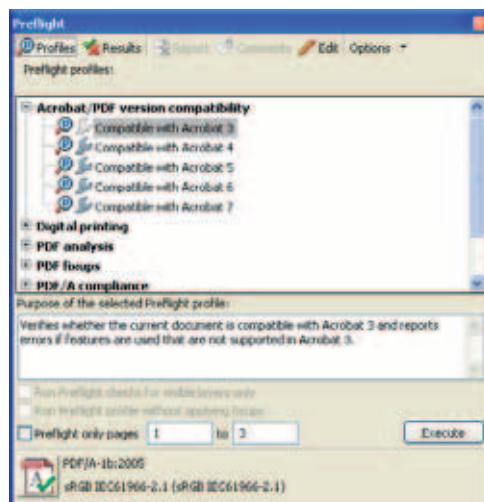
In some circumstances, certain additional steps may need to be taken in Acrobat Professional:

■ The file is modified in line with accessibility requirements. The user may need to set the language in the 'Advanced' area of the 'Document Properties' screen so that screenreader software functions correctly with the document.

■ The Acrobat 'Advanced' menu contains further accessibility options. The user should carry out the 'Full Check' function. If there are accessibility problems with the file being checked, Acrobat informs the user and proposes ways of repairing them.

■ The 'TouchUp Reading Order...' function can be used if repairs to the structure and alternate image text are required.

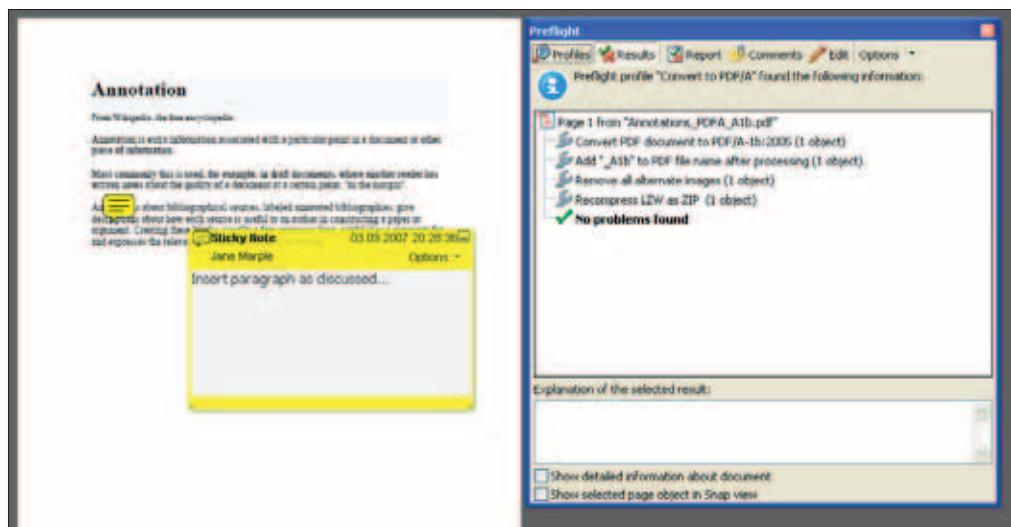
■ Once the PDF document has successfully passed the accessibility check, it can be validated in Acrobat's Preflight tool to make sure that it is suitable for conversion to PDF/A. It can then be converted. The PDF/A conversion/validation is always the very last step. ■



Accessibility options: This illustration shows the location of tools for accessible PDF in Acrobat Professional.

Last step: Finally, the PDF file is converted to PDF/A or tested to see whether it conforms to the standard.

Sticky Note in a PDF/A file: As a rule, comments are permitted in PDF/A-compliant files. However, multimedia comment types such as audio comments are prohibited. Users must be especially careful with colors and transparency when using graphical annotations.



Interactive PDF files

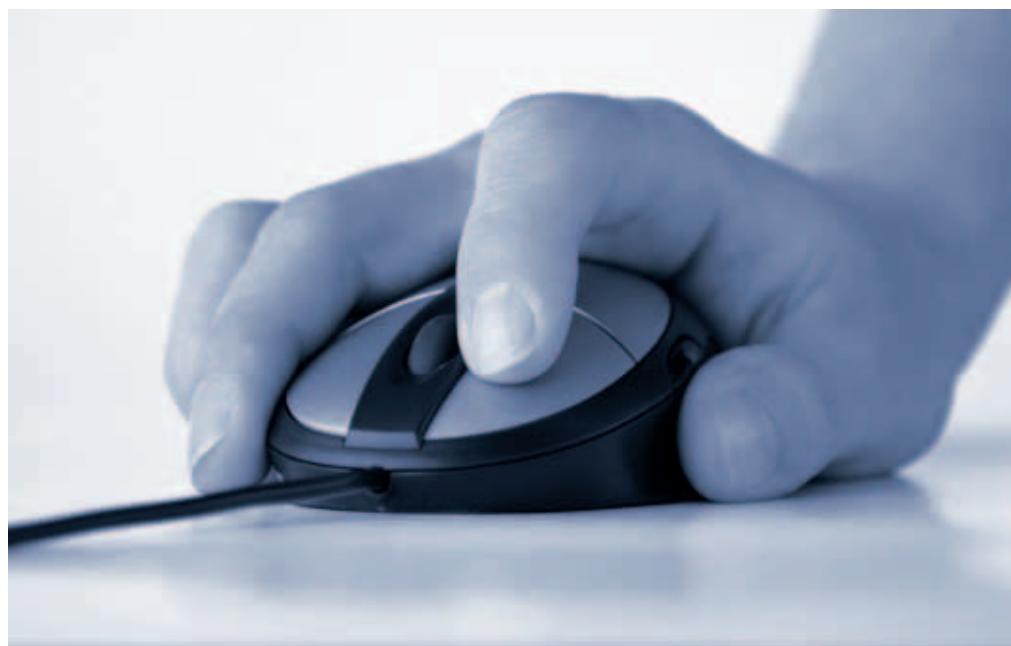
Interactive elements significantly enhance the functional scope and usage possibilities of PDF documents. Interactive elements create connections – whether in the form of navigation between documents or interaction between companies and clients or public authorities and citizens.

PDF documents can be given hyperlinks, comments, and form elements. But to what extent are these interactive functions compatible with PDF/A? The PDF/A standard stipulates that files may only

contain future-proof elements that do not impede clarity.

Comments and annotations

PDF/A aims to make all of the content in PDF files reproducible and permanently accessible. This includes comments. They may not be hidden or set as 'non-printing'. However, if a user wants to give a PDF file permanent comments – that is, he or she wishes to retain the comments in the PDF/A file – this is technically possible. It is not difficult to define a comment as a note in Acrobat and generate a valid PDF/A file from the document in question.



PixelQuelle.de

The screenshot shows the Acrobat Preflight interface. At the top, it says 'Preflight' and 'Preflight profile "Convert to PDF/A" found the following errors:'. Below this is a list of errors:

- Page 1 from "Annotation.pdf"
 - Convert PDF document to PDF/A-1b:2005 (1 object)
 - Add ".A1b" to PDF file name after processing (1 object)
 - Remove all alternate images (1 object)
 - Recompress LZW as ZIP (1 object)
 - Transparency used (1 match on 1 page)**
 - Summary
 - Page 1: Form XObject (05.791/636.563/140.274/653.317)

Below the errors, there is an 'Explanation of the selected result:' section with a detailed description of transparency issues. At the bottom are two checkboxes: 'Show detailed information about document' and 'Show selected page object in Snap view'.

Annotation

From Wikipedia, the free encyclopedia

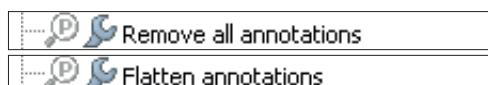
Annotation is extra information associated with a piece of information.

Most commonly this is used, for example, written notes about the quality of a document.

Annotations about bibliographical sources, descriptions about how each source is used in an argument. Creating these blurbs, usually expresses the relevance of each source.

Since note icons and input masks work with RGB, the PDF/A file in question must have an RGB output intent such as 'sRGB'.

There are also comment types that are not permitted. It is easy to understand why text edit comments are prohibited. If such annotations exist, it is to be assumed that a text correction that should have been made has actually been overlooked. Care should also be taken with comments that use transparency to mark a document. This includes the Highlight Text Tool and the stamps delivered with Acrobat, e.g. 'Approved'.



Removing or flattening annotations: Preflight corrections can remove or flatten comments. In the latter case, the annotations are still visible but they lose their typical comment features.

What can be done if these types of comment are important and need to be transferred to the PDF/A file being created? The Acrobat Preflight tool provides a solution. If you carry out corrections to flatten comments and transparencies before carrying out a PDF/A conversion, the visual nature of the comments is retained but the comment functionality is completely lost. For example, following flattening in the Preflight tool stamp notes can no longer be opened by double-clicking on them.

*Not all comments are suitable for PDF/A:
Unsuitable annotations include text edit annotations and annotations with transparencies.*

Hyperlinks are comments

It might be surprising, but from a technical point of view hyperlinks are also comments. They may not be retained in their original form if PDF/A-compliance is to be achieved – instead, they must be flattened. If a user attempts to convert a PDF file that contains links into a PDF/A file, the system issues two error messages per hyperlink: 'Annotation has no Flags entry' and 'Annotation not set to print'.

The Preflight correction profiles 'Remove all annotations' and 'Flatten comments' can be useful here. In this case, the result of both procedures is the same. Once the links have been discarded, Preflight can usually convert the PDF file into a PDF/A file without any difficulty.

Hyperlinks are comments: The illustration below shows that this PDF/A conversion cannot be carried out in Preflight because of links in the document.

The screenshot shows the Acrobat Preflight interface. At the top, it says 'Preflight' and 'Preflight profile "Convert to PDF/A" found the following errors:'. Below this is a list of errors:

- Page 1 from "Hyperlink.pdf"
 - Convert PDF document to PDF/A-1b:2005 (1 object)
 - Add ".A1b" to PDF file name after processing (1 object)
 - Remove all alternate images (1 object)
 - Recompress LZW as ZIP (1 object)
 - Annotation has no Page entry (1 match on 1 page)**
 - Annotation not set to print (1 match on 1 page)

Below the errors, there is an 'Explanation of the selected result:' section with a detailed description of annotation issues. At the bottom are two checkboxes: 'Show detailed information about document' and 'Show selected page object in Snap view'.

Hyperlink

From Wikipedia, the free encyclopedia

A **hyperlink** (often referred to as a **link**) is a reference to another section or document, or a section of another document, that the user can click on to go to the referenced document or section. A **hypertext link** is a **hyperlink** that is part of the **hypertext** system, but are not limited to **electronic media**.

Hyperlink formats

There are a number of ways to distinguish the various forms of hyperlinks.

If forms use interactive elements, it is not usually possible to depict them in a 1:1 manner in PDF/A. PDF/A-compliance is easiest to achieve for simple PDF forms with no calculation or validation functions, reports, and so on.

Forms

The PDF/A standard does not prohibit forms as such, but there are form field types that work with actions, and actions can prevent PDF files from being suitable for long-term archiving. Problems are to be expected in the following cases:

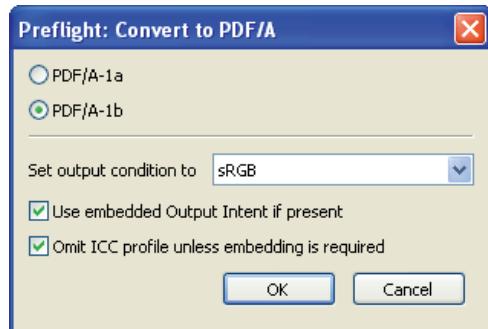
- Actions of the type ‘Submit a Form’, ‘Import Form Data’, and ‘Reset a Form’ are prohibited. This is understandable, since these actions change document content.
- Additional actions are not permitted because they too can change content.
- JavaScript actions are prohibited because they endanger the reproducibility of the actual state of a file.

An attempt to convert a form PDF to a PDF/A-1b-compliant document might give the result shown in the illustration below. Such a result is caused by critical events including ‘Reset Form’ and the ‘Send’ button.

As is shown by the Preflight result in the illustration, problems with non-embedded fonts have also occurred in this example. This is not so easy to solve at a later stage

(for more information, see below). In addition, the tool sometimes reports ‘DeviceRGB’ colors (device-dependent colors) that result from the colored background of form fields. (Acrobat Professional only provides device-dependent RGB for setting up form fields and buttons.) What can be done?

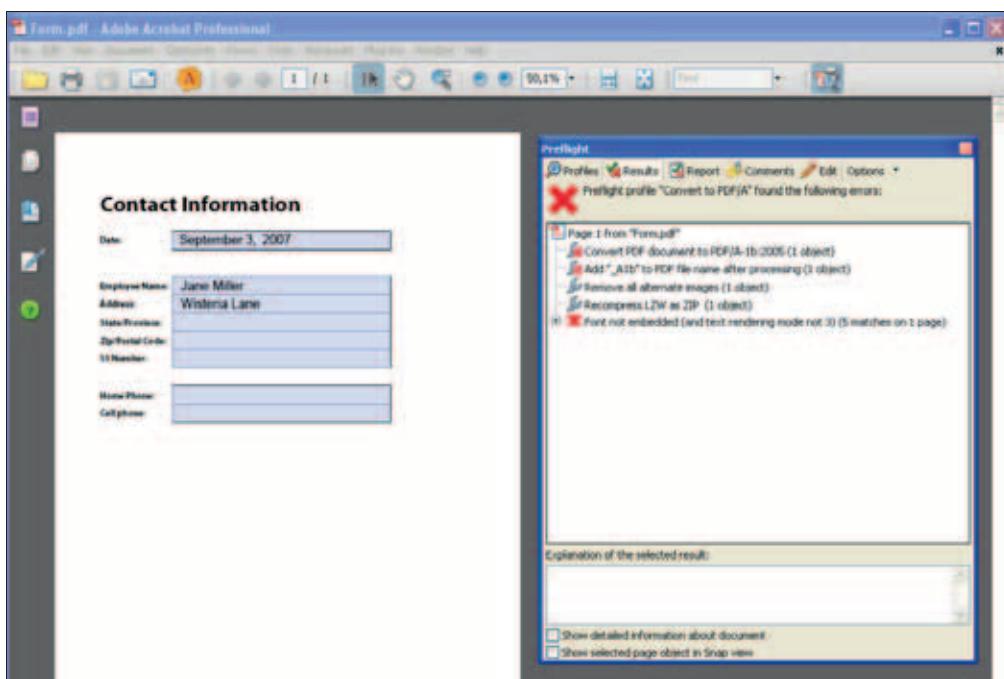
- First, actions and JavaScript must be removed from the source form. This causes certain restrictions in functionality.



Device-dependent: DeviceRGB requires ‘sRGB’ as the output intent.

- If the device-dependent RGB colors are not to prevent compliance with the PDF/A standard, the user has to select ‘sRGB’ as the output intent for the conversion. The trimmed down, pre-treated form PDF can then be converted to PDF/A-1b. However,

Forms and PDF/A: It is not form fields themselves that cause problems when converting files to PDF/A – it is certain actions and JavaScript used in form fields. Problems can also occur with fonts and colors.



Preflight: The conversion of the form to PDF/A cannot be carried out because the tool is unable to embed the required fonts in the file.

an alternative solution must be found for non-embedded fonts in form fields.

Embedding fonts for PDF/A forms

Many current tools do not enable the embedding of fonts in PDF form fields. However, these fonts must be contained within the PDF file in order to achieve PDF/A-compliance.

The Acrobat Preflight tool cannot embed fonts in form fields. Following a failed attempt to convert a document containing them into a PDF/A document, the tool is-

sues the following error message: 'Font not embedded'.

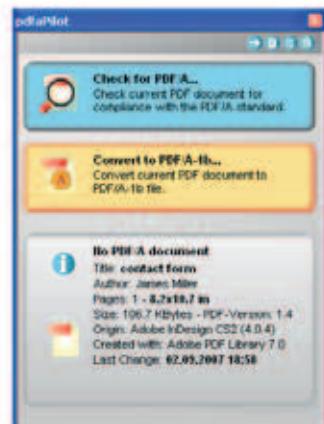
There is now a tool that can carry out this task – the Acrobat pdfaPilot plug-in from cal-las software. Among many other correction functions, it allows form PDF files to be converted into PDF/A-compliant documents. For the process to work, all of the fonts required for the PDF document being converted must be available and accessible on the computer. In addition to the function for embedding fonts, pdfaPilot also solves many of the common color problems that occur in forms. ■



Access to fonts: All used fonts must be available on the computer being used for the conversion in order for font embedding to be successful.

The screenshot shows a 'Contact Information' form with fields for Date (September 3, 2007), Employee Name (Jane Miller), Address (Western Lane), State/Province, Zip/Postal Code, SS Number, Home Phone, and Cell phone.

PDF/A form: pdfaPilot can create valid PDF/A forms with embedded fonts.





Cahloc - PixelQuelle.de

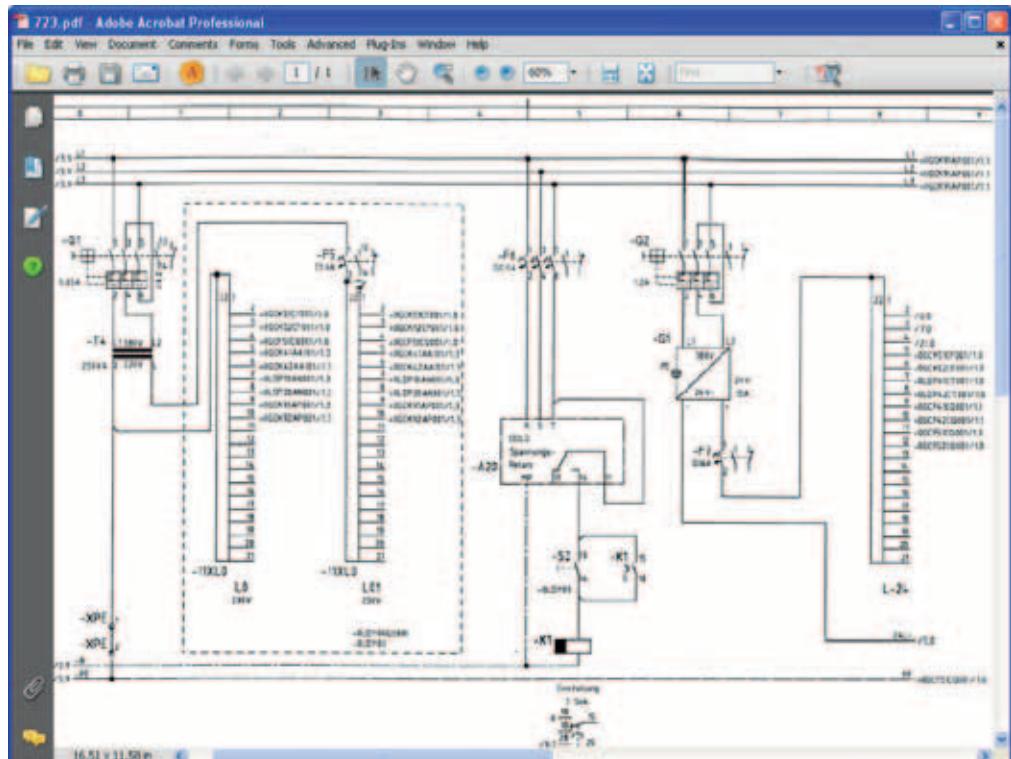
PDF/A for design drawings

Drawings such as CAD plans and maps are ideal candidates for long-term archiving as PDF/A files. Drawings from the fields of architecture and statics must often be kept for long periods of time. Technical blueprints frequently require long-term storage, too.

PDF 1.4 can handle the large page formats with lengths of several meters that often occur in technical drawings. PDF 1.4,

the PDF specification on which PDF/A-1a and -1b are based, supports a maximum page size of 200 x 200 inches, which corresponds to 5.08 x 5.08 meters. As of PDF 1.7, virtual page sizes of up to 381 kilometers in length are permitted, but the PDF/A standard only permits the maximum page size for PDF 1.4.

Design drawings can be output to PDF or even to PDF/A directly from current CAD programs – the PDFMaker is often



used for the PDF conversion process. However, the PDF/A conversion may take place in Adobe or in another PDF/A converter. Older plans are often line scans in formats such as TIFF G4. Such plans can be converted to PDF and then to PDF/A using Acrobat Professional (or other PDF conversion solutions). It is often possible to use the text recognition function to give drawings searchable text during this process.

No 3D models in PDF/A

Designs created in 2D can be archived as PDF/A without any problems. This is not the case for 3D models. Three-dimensional designs have only been supported since Acrobat 7 (PDF 1.6). They are therefore not permitted in PDF/A-compliant files.

Electronic signatures

Our everyday life is now digital. Within the space of a few years, e-commerce has become much more widespread and business agreements are now often made online using e-mail. Digital communication between public authorities and citizens is no longer a thing of the future – just think of electronic tax return systems such as EFTPS.

Because parties taking part in such transactions are not in the presence of each other or witnesses, it is more important than ever to ensure that digital documents can be reliably checked for authenticity. Electronic signatures enable a completely digital flow of communication and transactions of a contractual nature.

Proving authenticity by means of a mark or signature dates back nearly as far as the first written evidence of mankind. Even the Mesopotamians signed their records with a seal or stamp. The practice of signing documents with a stamp instead of by hand – which is still used in China and Japan today – has a history that dates back over millennia. Magnificent wax seals are known to have been used during the Middle Ages. Placing one's own signature at the bottom of a contract is a relatively new procedure, just as general literacy is a relatively new achievement for our culture.

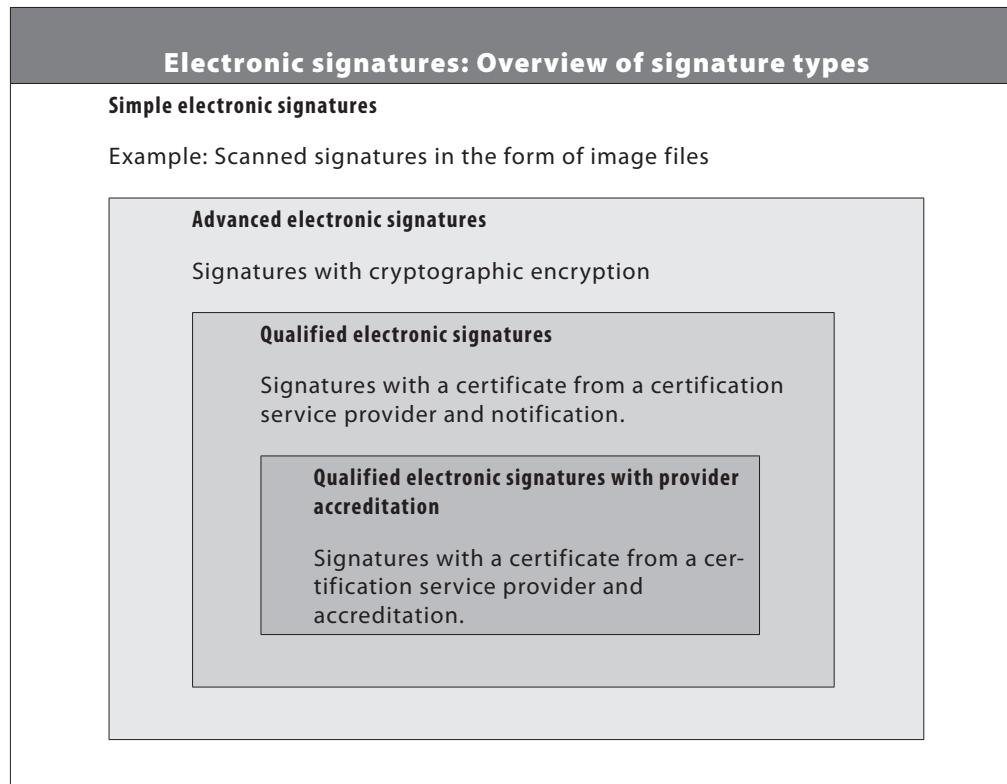
But now we are faced with another problem – how can we make digital files into legal documents?

The simplest way of electronically signing a file is to place a scanned signature on a page of the document in the form of an image file. This procedure can be legally recognized, as it is in the United States. →

The terms 'digital signature' and 'electronic signature' are often used interchangeably. However, the term 'digital signature' refers to a cryptographic, technical process, whereas „electronic signature“ is a legal term.



photocase.com/de



However, it is obvious that this solution provides no security against signature fraud. As a result, the development of powerful and reliable systems for digital signatures is extremely important.

Identity, integrity, and time stamps

Nowadays, a great number of agreements and contracts are concluded digitally. To a great extent, the Internet has now replaced communication channels such as couriers and postal services. In today's world, users carry out transactions including quoting, ordering, and invoicing in e-business or motions and rulings in e-government using only digital means.

Certain basic prerequisites must be fulfilled for these kinds of transactions. Recipients of data must be able to easily determine whether the person who sent the files really is the person specified. In addition, they must be able to ascertain that the content received has not been changed or falsified. These requirements therefore relate to identity (determination of the writer) and integrity (determination of intact content).

Advanced electronic signatures ensure that these requirements are met. They allow recipients to use cryptographic technology to detect any changes to content.

Recipients can also use a cryptographic key to definitively identify the author of a signed message.

Time is often another criteria for ensuring legally valid transactions or agreements. Time stamps that specify the date and time of a content version are often used for this task. Electronic signatures and encryption have different purposes. Whereas electronic signatures ensure that the content, involved party, and time of digital transactions can be clearly identified and not changed, encryption protects confidential data from being viewed by unauthorized parties by, for example, only permitting documents to be opened with a password.

Security levels

There are various well-used procedures for signing digital documents. They differ in their scope and the security level that they provide for users in cases of uncertainty or contention.

Simple electronic signatures

Simple electronic signatures include image files of scanned signatures. They only provide low levels of authentication.

Advanced electronic signatures

These signatures are subject to more stringent requirements: They must enable the detection of content manipulation and it must be possible to determine the authenticity of the signatory using an electronic certificate. Advanced electronic signatures provide less authenticity in a court of law than qualified electronic signatures.

Qualified electronic signatures

These signatures provide the highest level of security. A qualified certificate is used to assign each electronic signature to its originator. Qualified certificates are signed by a certification service provider (CSP/CA). CAs are required to comply with the requirements of signature legislation, which stipulates that they must operate their certification services in a protected environment (trust center).

Common Criteria

'Common Criteria' refers to an international standard that defines the criteria for the evaluation and certification of the security of computer systems in relation to data integrity and protection. The Common Criteria standard avoids the need for components and systems to be certified more than once in different countries. Digital signature solutions can also be certified in accordance with the Common Criteria standard.

Digital signatures in PDF with Acrobat

Digital signatures in PDF documents have been supported since Acrobat 4.0 and Adobe Reader 5.1. Whereas older Reader versions only enable digital signatures to be viewed and checked, users of Adobe Reader 8 can actually sign documents as long as the Acrobat function for doing so has been activated.

With Acrobat, signatures are embedded into the PDF documents they authenticate. The maximum security level supported is

that provided by 'advanced electronic signatures'. Users who wish to use qualified electronic signatures must use specialist software such as Openlimit PDF Sign for Adobe. For more information, see www.openlimit.com on the Internet.

PDF/A and digital signatures

It is not uncommon for users to be faced with a dilemma regarding the correct sequence of steps when using digital signatures in PDF/A documents: The creation of the PDF/A version and the process of digitally signing a document both take place at the end of the work process. A PDF is saved as a PDF/A-compliant document once it has the precise form in which it is to be archived – this means that it can no longer be changed once it has been converted to PDF/A. However, a digital signature certifies the authenticity of a PDF document at a specified moment in time – thereafter, no changes are permitted. However, in practice signed PDF/A files do not pose a problem. It is entirely possible to give a PDF/A file a digital signature without causing the document to lose its PDF/A-compliant validity. The PDF/A file is generated first and it is then given a digital signature. This is because a signature does not actually constitute a document change – it merely indicates that the document has exactly the same form at the point when it was signed. This means that a digital signature does not damage or impede the compliance of a PDF/A document. →



Strictly speaking, even the addition of a digital signature constitutes a change to a PDF document. However, this is one exception where the change does not affect the PDF/A status or content of the document in question. Digitally signing a PDF/A file is therefore both permitted and sensible.

Adobe Reader 8 enables the digital signing of documents as long as the function has been activated in Acrobat.

Enable Usage Rights in Adobe Reader

The following features will become available for this document when opened in the free Adobe Reader.

- Save form data (for a fillable PDF form only)
- Commenting and drawing mark-up tools
- Sign an existing signature field
- Digitally sign the document anywhere on the page (only supported in Adobe Reader 8.0)

Note: Once Reader Enabled, certain functions, such as editing document content or inserting and deleting pages, will be restricted.

Save Now

Cancel

In fact, it underwrites compliance as long as the signature itself meets the requirements of the PDF/A standard.

Challenges in practice

There are certain fields of application where digital signatures come up against brick walls for technical reasons or where users have to plan the flow of individual process steps in detail.

Multiple signatories

In business and politics, documents often have to be signed in multiple. This might be the case if all members of a board of directors have to sign a ruling or if several secretaries of state need to ratify a resolution. This causes a technical problem for digital signatures: Each time a signature is added to a PDF document, the validity of the previous digital signature is nullified, since the addition of the new signature constitutes a change to the PDF document. This is main-

ly a problem that is specific to PDF because signatures are embedded; other external types of signature are not subject to this issue due to the different procedures used to apply them to documents.

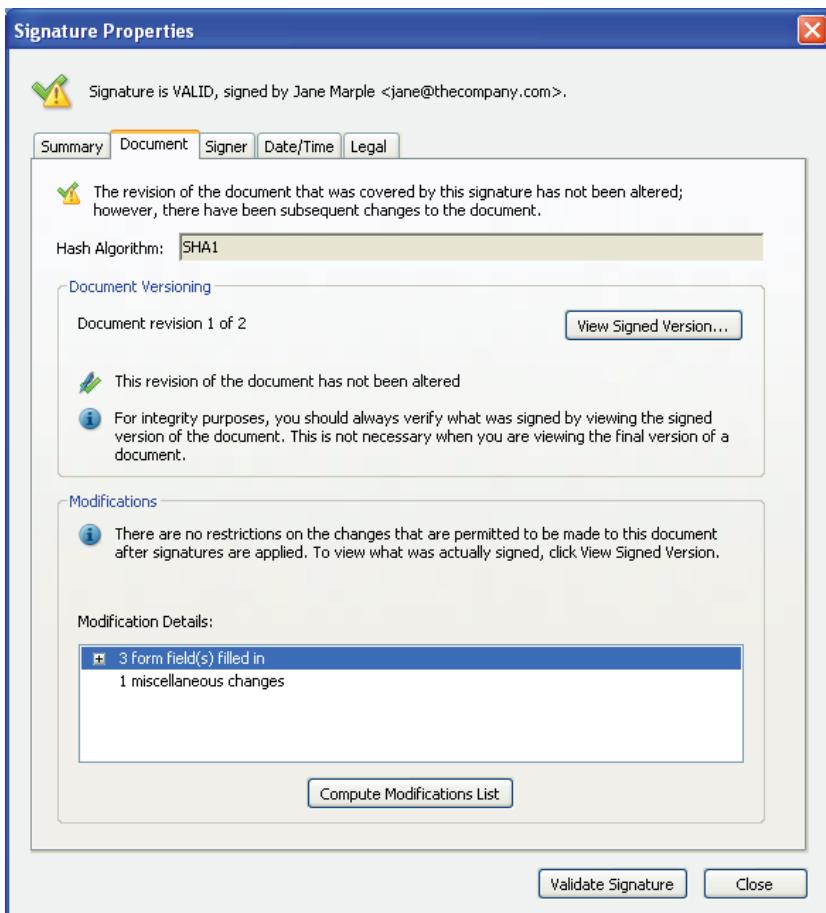
Signing documents again

For security reasons, digital signatures are only afforded restricted validity periods. Since we can assume that computer performance will continue to rise constantly in years to come, security codes that are extremely difficult if not impossible to crack today might be eradicated in a few hours in the future by simple trial and error. For this reason, documents have to be signed again after a certain period of time to keep the validity of the signature up-to-date. This makes the previous digital signature invalid since, as explained above, each newly added digital signature invalidates any previous signature.

Archiving signed PDFs as PDF/A files

Users who receive signed documents that are not yet PDF/A-compliant face practical problems if they wish to archive them as PDF/A documents. The only solution is to have such documents signed again following conversion to PDF/A.

Another complication arises when a PDF form (such as a contract) that needs to be signed in order to be valid then needs to be stored in an archive as a PDF/A file. This is another case where the document in question needs to be converted to PDF/A before it is signed. This means that JavaScript functions in the form need to be removed and fonts used in filled text fields need to be embedded before digital signing. ■



Tracking changes in Acrobat: To find out when and what changes have been made to a PDF document following the addition of a signature, users can go to the 'Signature Properties' area. This dialog box is opened by clicking on a signature. The 'View Signed Version...' option in the 'Signature Properties' dialog box can be used to display saved versions of a PDF file.

The outlook: PDF/A in the future

6.

PDFs are extremely practical and it is difficult to argue against their usefulness for many application areas. PDF as a format has 'grown up' over a period of 14 years, and today the format itself and the software required to use it take various mature forms. In addition, the adoption of the PDF/A standard has made PDF a highly reliable format, both for today and for the future. Does this make the issue of technical formats and procedures for the long-term, secure archiving of digital documents a closed topic? What has PDF/A achieved and what remains to be done?

Enhancements in PDF/A-2

The PDF/A standard constitutes an extremely solid base, at least regarding the unambiguous and reliable visual reproduction of content. Without a doubt, the PDF/A standard will be developed further

to incorporate technical fine-tuning of the PDF format in a manner that allows archiving.

The second part of the PDF/A standard – PDF/A-2 – is planned for 2009. It is important to note that the second part will not invalidate PDF/A-1; PDF/A-1-compliant documents will still be valid and reliable archive files. It will not be necessary to migrate existing PDF/A-1 archives to PDF/A-2 once the new PDF/A standard is published. Doing so would benefit nobody. However, in some cases it might make sense to archive new archive documents as PDF/A-2 files. For example, PDF/A-2 will support the JPEG2000 image compression format. If files contain image data in JPEG2000 format, it is clearly sensible to archive that data in JPEG2000, thereby avoiding the need to carry out a recompression into JPEG (which can cause albeit low



Wichert – PixelQuelle.de



Claudia Hautumm – PixelQuelle.de

levels of data loss) or ZIP (which increases the amount of memory required to save files).

Looking towards PDF/A-3

A third part to the standard is already being discussed – PDF/A-3. This part should deal with ‘dynamic’ PDF documents. PDF/A-1 deals exclusively with PDF documents whose content and depiction does not change and may not be modified (as is the case with paper documents). In the case of PDF files that contain audio or video data, self-playing animated presentations, ‘walkable’ 3D models, or complex form logic with database connections, it is only possible to preserve a snapshot of a certain display point or specific content form when printing them or archiving them as PDF/A-1. This is hardly an ideal solution. It is certain to take several years to complete and adopt the PDF/A-3 standard, since depicting dynamic content is far more difficult than capturing static, two-dimensional visual content.

PDF/A-1 developments

In any case, there are bound to be new developments for PDF/A-1 itself – not for the PDF/A-1 standard, but for related issues.

Digital signatures

One closely monitored issue is the interaction of PDF/A-1 and digital signatures. The

PDF/A standard intentionally permits digital signatures, but just as deliberately refrains from stipulating actual implementation methods. One important reason why there is not yet an ISO standard on digitally signing PDF/A documents is that requirements and legislation for digital signatures differ from country to country. Despite the prevalence of economic processes that are generally globalized, this area is subject to an extremely high degree of disparity and incertitude. In addition, digital signature technology is still a long way from being mature, and is not as wide-spread or easy to use as PDF technology, which is accessible to all and sundry.

Full-text searching

Another important aspect is full-text searching. This function usually works so well with traditional PDFs that we take its permanent availability for granted. However, there are always a couple of actual hits that are, in fact, missed – even if we do not notice it, precisely because the hits in question are not found by the text search. The failure of a full-text search to find certain hits can result from something as basic as a typing error (for example, if the name ‘Smith’ is typed as ‘Smiht’). However, perfectly legitimate differences in spelling or punctuation can also cause a hit to be missed: For example, the number one thousand point zero is written

as 1,000.0 in the US, 1.000,0 in Germany, and 1'000,0 in Switzerland. There is also a great variety of ways to write down telephone numbers – various countries use spaces, brackets, or hyphens to improve readability or conform with different format rules. The PDF/A standard requires all characters to have unique Unicode names for the PDF/A-1a compliance level. However, it is not possible for the PDF/A standard to ensure that a unique Unicode character-to-code assignment is correct. Only a human operator can decide whether or not an X is being passed off as a U.

Structured content

There is also room for improvement in relation to the structure of content in PDF files: Countless documents that require archiving do not only contain structured content (a reading order and important specifications such as title, image caption, or sequential body text) but also include specific, uniquely identifiable data specifications. For example, telephone bills always contain fields such as 'Customer Number' and 'Invoice Number', and they always state the amount owed. Tagged PDF (which is also specified in PDF/A-1a for the structure of content) already helps a great deal in this area. However, it would be even more useful if this kind of data specification could be determined and read directly and uniquely, as data records are read from a database. Format-related ambiguities also need to be eliminated. In fact, the current state of technology enables these tasks to be accomplished, but a standard that makes documents and software interoperable is still required.

PDF/A in one hundred years time

The aspects mentioned above will probably be implemented during the next five or ten years. But what will the world of PDF/A be like in fifty or even one hundred years time? For example, what is the probability of a person interested in the beginnings of PDF/A being able to find and read a printout (or microfilm or TIFF version) or a PDF/A-1 version of this publication in the year 2107? This can partly be answered by a harsh truth: Money makes the world go round. As long

as PDF/A is widely used, its popularity will create a market where providers of solutions and services can make a profit. There is no need to worry about the future of the format unless this market shrinks to a critical size. Unwanted stumbling blocks in the form of patents and other industrial property rights are increasingly less likely than for many other formats, even if they cannot be completely ruled out in the society in which we live. Consider, for example, Unisys and its LZW patent that has only recently expired, Forgent and its JPEG patent claims, or Microsoft, which had to pay over one and a half billion US dollars to Alcatel-Lucent in a dispute over the widely used MP3 format. In any case, one thing's for sure: In fifty or one hundred years time, all of these patents will have expired.

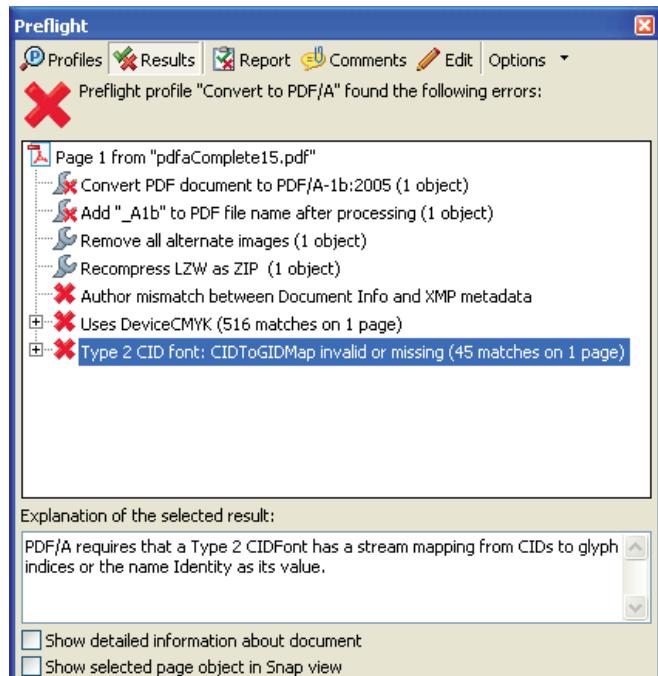
However, one question remains: Have we already reached the critical PDF/A target required to ensure an enduring market, and, if not, when will the target be reached? As far as the rollout and practical implementation of PDF/A are concerned, we are still only beginning. However, in terms of the advantages offered by PDF/A, there cannot be any doubt that, by 2010, PDF/A will be so widely used that this critical target is certain to be met. This is why: No format other than PDF and/or PDF/A is so ideally suited, practical, and widespread when it comes to archiving the rapidly increasing number of digital documents. Moreover, no other format has been adopted as an ISO standard, which makes manufacturers far more likely to use it.

As mentioned previously, the PDF/A landscape will continue to develop in various ways in order to achieve technical progress and meet application-specific requirements. However, it is important to note that this will not result in a need to revise the basic principles of the standard: The ISO PDF/A-1 standard provides a very solid basis for the field – even in the light of planned additional parts to the PDF/A norm. It describes a strong foundation that is not subject to noteworthy change in the long term. This fact is sure to facilitate the strategic and economic justification of investment when implementing PDF/A-based archiving processes. ■

What the error messages mean

Preflight results and troubleshooting for PDF/A

During PDF/A conversion or validation in Acrobat 8 Professional, the Preflight tool informs the user of problems that prevent compliance with the standard. Although the user receives a short explanation of each error in the info window, not all descriptions are easy to understand. For this reason, this alphabetical list of all PDF/A error messages that might occur is intended to help provide an overview of why the tool has declared a document to be non-PDF/A-compliant. In addition, each error message is followed by a description of measures that users can carry out in advance or later on in order to enable the production of a valid PDF/A file.



Puzzling error messages: Not all explanations that the Preflight tool provides for PDF/A problems can be understood without further explanation. This list intends to shed some light on error messages users might see.

■ 3D annotation used: Three-dimensional comments can contain complex 3D models that generally originate from CAD programs. They can be embedded as comments in PDF files and rotated interactively on a monitor, for example. Interactive rotation is obviously not possible when a document is printed out. Because the PDF/A stipulates that a PDF must look exactly the same (to the greatest extent possible) regardless of whether it is displayed on a monitor or printed out, PDF/A files may not

contain 3D comments. *Remedy: The Acrobat 8 Preflight module can be used to discard 3D comments.*

■ Additional actions (AA) used: PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The Acrobat PDF Optimizer can be used to remove these actions.*

■ Alternate image present: An image in a PDF file can contain 'alternate interpretations'. This allows monitor visualization to be accelerated if, for example, high-resolution images also have a low-resolution alternate interpretation. However, the visualization of the PDF is then no longer uniform but varies depending on the output device used. Because PDF/A demands the unambiguous visualization of all page objects, alternate interpretations of images are not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard all alternate images' in the 'Discard Objects' area.*

■ Annotation has C entry but no OutputIntent present: In this PDF file, at least one comment uses a DeviceRGB-defined color for drawing the outline of a comment field. PDF/A does not differentiate between page objects and comments for colors, so the requirements for text, graphics, and images also apply here. *Remedy: Such files can normally be converted to PDF/A using the 'sRGB' output intent.*

■ Annotation has CA value that is not 1.0: A comment symbol in this PDF file is set to 'transparent' so that the background shows through the symbol when it is displayed on a monitor or printed out. The PDF/A standard stipulates that all features used in a PDF file must be displayed in a single unique way on a monitor or in a printout. Because this cannot be ensured in the case of transparent objects and their backgrounds, transparency is not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer enables comments to be discarded.*

■ Annotation has IC entry but no OutputIntent present: In this PDF file, at least one comment uses a DeviceRGB-defined color for drawing the background of a comment field.

PDF/A does not differentiate between page objects and comments for colors, so the requirements for text, graphics, and images also apply here. *Remedy: Such files can normally be converted to PDF/A using the 'sRGB' output intent.*

■ Annotation has no Flags entry: A comment element in a PDF file must contain certain additional information that determines its appearance when displayed on a monitor or printed out. This information is missing for the comments in this PDF file. This means that it is unclear whether/how the comment will be rendered when the document is displayed or printed out. The PDF/A standard stipulates that all information required for visualizing a comment must be contained within the PDF file in which the comment appears. *Remedy: These comments are normally invisible components of hyperlinks. They can be removed using the Acrobat PDF Optimizer ('Discard all comments, forms and multimedia').*

■ Annotation Hidden flag set: Comments can be set to 'Hidden' to prevent them from being displayed on the monitor. The 'Hidden' flag is used to do this. Because the PDF/A standard stipulates that the visualization of a document must be ensured and because it is impossible to guarantee that the 'Hidden' flag will be correctly evaluated, PDF/A documents may not use this flag for comments. *Remedy: These comments can be discarded using the Acrobat PDF Optimizer ('Discard all comments, forms and multimedia').*

■ Annotation Invisible flag set: Comments in PDF files can be set to 'invisible' to prevent them from being displayed on a monitor. The 'Invisible' flag is used to do this. Because the PDF/A standard stipulates that the visualization of a document must be ensured and because it is impossible to guarantee that the 'Invisible' flag will be correctly evaluated, PDF/A documents may not use this flag for comments. *Remedy: These comments can be discarded using the Acrobat PDF Optimizer ('Discard all comments, forms and multimedia').*

■ Annotation not set to print: Comments in PDF files can be defined as non-printing to prevent them being printed out. Because the PDF/A standard stipulates that the visualization of a document must be ensured and that a document must be identical whether displayed on a monitor or output using a printer, comments in a PDF/A document must not be defined as not to be printed. *Remedy: Current PDF-to-PDF/A converters correct this error during the conversion process.*

■ Annotation NoView flag set: Comments in PDF files can be set to 'NoView' to prevent them from being displayed on a monitor. Because the PDF/A standard stipulates that the visualization of a document must be ensured and that a docu-

ment must be identical whether displayed on a monitor or output using a printer, comments in a PDF/A document must not be defined as not to be displayed. *Remedy: These comments can be discarded using the Acrobat PDF Optimizer ('Discard all comments, forms and multimedia').*

■ Annotation's AP (appearance) contains only N entry is not true: Comments in PDF files can contain differing visualization methods that are used, for example, depending on whether the mouse cursor is moved over a comment symbol or the comment symbol is clicked. These effects are, of course, not possible when a document is output on a printer. Because the PDF/A standard stipulates that the visualization of a document must be guaranteed and that it must appear identical when output on a printer and when displayed on a monitor, comments in a PDF/A document may not have different visualization variants for mouse effects. *Remedy: The Acrobat Professional PDF Optimizer has an option called 'Discard all comments, forms and multimedia' in the 'Discard User Data' area. This option corrects this error.*

■ Author mismatch between Document Info and XMP metadata: In this PDF document, the data on the author in the XMP area does not match the data in the general document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: New PDF/A conversion.*

■ Belongs to transparency group: A group of page objects is defined as 'transparent'. The PDF/A standard stipulates that all features used in a PDF file must be displayed in a single unique way on a monitor or in a printout. Because this cannot be ensured in the case of transparent objects and their backgrounds, transparency is not permitted in PDF/A files. *Remedy: Adobe Acrobat Professional (Version 6, 7, or 8) includes a flattener module that can be used to remove transparencies.*

■ Bits per color component > 8: Images with a color depth other than 8 bits are used in this PDF. Color depths that are not 8 bits are not reliably supported by all visualization devices (monitors and printers). In addition to this, such fine nuances cannot be visualized technically on most devices in a way that ensures that differing color depths do not lead to differences in color or brightness when visualized. For this reason, only 8 bit images are permitted in PDF/A files. *Remedy: The PDF must be regenerated using images that have an 8 bit color depth. Acrobat 8's Preflight module also has a correction option that reduces the color depth of images from 16 bits to 8 bits.*

■ **CharSet missing or incomplete for Type 1 font:** A font is not fully embedded and contains no list of embedded symbols (CharSet). If a font in Type 1 format is not fully embedded, it must contain a list of the embedded characters to enable conversion to PDF/A. The list must include all characters used in this font in the PDF file. In this case, a font is not fully embedded in this PDF file and its list of embedded symbols is missing or is incomplete. *Remedy: In order to resolve this problem, the PDF file must be created again using a different font or with the same font but in its complete form. Alternatively, the incomplete font may be used, but only with the relevant CharSet.*

■ **CIDset in subset font is incomplete:** A font is not fully embedded and contains no list of embedded symbols (CharSet). If a font in CID 1 format is not fully embedded, it must contain a list of the embedded characters to enable conversion to PDF/A. The list must include all characters used in this font in the PDF file. In this case, a font is not fully embedded in the PDF file and its list of embedded characters is incomplete. *Remedy: In order to resolve this problem, the PDF file must be created again using a different font or with the same font but in its complete form. Alternatively, the incomplete font may be used, but only with the relevant CharSet.*

■ **CIDset in subset font missing:** A font is not fully embedded and contains no list of embedded symbols (CharSet). If a font in CID 1 format is not fully embedded, it must contain a list of the embedded characters to enable conversion to PDF/A. In this case, a font is not fully embedded in the PDF file but the list of embedded characters is missing. *Remedy: In order to resolve this problem, the PDF file must be created again using a different font or with the same font but in its complete form. Alternatively, the incomplete font may be used, but only with the relevant CharSet.*

■ **CIDSystemInfo and CMap dict not compatible:** A font is not fully embedded and contains no list of embedded symbols (CharSet). The characters stored in a font are allocated number codes in accordance with an allocation table. These number codes are used to display the characters in the PDF that uses them. No allocation table has been specified for a font in this PDF. *Remedy: In order to resolve this problem, the PDF file must be created again using a different font or with the same font but in its complete form. Alternatively, the incomplete font may be used, but only with the relevant CharSet.*

■ **CMap not embedded for custom CMap:** A font is not fully embedded and contains no list of embedded symbols (CharSet). This font has no clear information regarding the assignment of characters to letters (the CMap is missing). The

letters and other characters used in PDF texts require ‘fonts’ that determine their exact appearance when visualized. The characters stored in a font are allocated number codes in accordance with an allocation table. These number codes are used to display the characters in the PDF that uses them. These allocation tables are made up differently depending upon the font format (PostScript Type 1, Type 3, or TrueType) and are known as ‘encodings’. MacRoman (Macintosh) and WinAnsi (Windows) are standard encodings. ‘CID fonts’ can use encodings that deviate from these standards. The PDF/A standard stipulates that a font that uses its own encoding must get the encoding in question from a corresponding table (CMap). This PDF does not use standard encoding and does not contain an encoding table (CMap). This PDF can therefore not be converted to PDF/A. *Remedy: In order to resolve this problem, the PDF file must be created again using a different font or with the same font but in its complete form. Alternatively, the incomplete font may be used, but only with the relevant CharSet.*

■ **CMYK used but PDF/A OutputIntent not CMYK:** Device-dependent color (DeviceCMYK), but no CMYK output intent. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot also solves this problem.*

■ **CMYK used for alt. color but PDF/A OutputIntent not CMYK:** A spot color has been defined in DeviceCMYK but the output intent is not defined for CMYK. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot also solves this problem.*

■ **Compressed object streams used:** Since PDF 1.5, which Adobe introduced with Acrobat 6, some objects in PDF files can be compressed as object streams. This technique is used in this PDF. The PDF/A standard only permits objects that are compatible with PDF 1.4. Cross-object compression is therefore not permitted in PDF/A files. *Remedy: Use the Acrobat PDF Optimizer to save the file as a PDF 1.4 file.*

■ **Contains action of type ImportData:** Active content that imports data from an external file. PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The PDF must be redesigned and generated again so that all content is present within the file itself.*

■ **Contains action of type Launch:** Active content that triggers another application. PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. This PDF file contains an action for launching another application. The PDF/A standard stipulate that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The PDF must be redesigned and generated again so that all content is present within the file itself.*

■ **Contains action of type Movie:** Active content that shows a movie in another window. PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. This PDF file contains an action for playing a movie. The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard all comments, forms and multimedia' that can be used to remove movies.*

■ **Contains action of type ResetForm:** Active content that influences the content of form fields. PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. This PDF file contains an action for emptying form fields (ResetForm). The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard all comments, forms and multimedia' that can be used to remove movies.*

bat PDF Optimizer contains an option called 'Discard all form submission, import and reset actions' that corrects this problem.

■ **Contains action of type Sound:** Contains audio data (sound). PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. This PDF file contains an action for playing sound. The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard all comments, forms and multimedia' that can be used to remove audio data.*

■ **Creation Date mismatch between Document Info and XMP metadata:** The 'CreateDate' entry in the XMP document information deviates from the 'Created' entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: A PDF file can contain descriptive document information including the author, creation date, title, and other details. This information can be opened using the 'File' menu and changed in the general Document Properties dialog. Otherwise, the PDF/A file can be recreated from scratch.*

■ **Creator mismatch between Document Info and XMP metadata:** The 'PDF creator' entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: New PDF/A conversion.*

■ **Custom annotation used:** A comment in the PDF document does not use a standard PDF comment type. The PDF specification allows PDFs to contain comments in custom formats. This function can be used in specialized applications to position special, additional elements in PDF files. These comments much have the comment type 'Custom'. Use is seldom made of these options. Because custom comments can only be visualized on specialized output devices, they may not be used in PDF/A files. *Remedy: These comments must be discarded.*

■ **Destination profiles in OutputIntents differ:** There are multiple output intents with different profiles. Because the PDF/A standard stipulates that colors must appear the same

(as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. The output intent must describe the color properties of the output device using an ICC profile. To ensure unambiguity, a PDF/A file may only have different output intents if all of the output intents used have the same ICC profile. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot can also carry out this task.*

■ Device process color used but no PDF/A OutputIntent: Device-dependent color exists, but no output intent. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot can also carry out this task.*

■ Device process color used in alt. color space but no PDF/A OutputIntent: A spot color has been defined as a device color, but no output intent exists. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot can also carry out this task.*

■ Document contains JavaScripts: Active content in the form of JavaScript changes the visualization of pages. PDF

files can dynamically alter their content during visualization. This active content can be contained in PDF files in the form of JavaScript, for example. This is the case in this PDF. JavaScript is often used in connection with active form elements (for example, buttons). The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer contains the 'Discard all JavaScript actions' option in the 'Discard Objects' area. This option can be used to correct the error.*

■ Document is damaged and needs repair: The PDF file is incorrectly formatted. Every PDF/A file must be a basically correct PDF file. The file that is currently open does not conform with the PDF specification and can therefore not be converted to PDF/A. *Remedy: The problem can possibly be resolved by opening the file in Adobe Acrobat and saving it once again using the 'Save As' option. Otherwise, it might be possible to clean it up using the 'Save Optimized As' option.*

■ Document is encrypted. The document is encrypted and cannot be analyzed. PDF files can be encrypted in order to password-protect certain functions. This means that a PDF can be displayed on the monitor without restrictions but a password is required in order to print or modify it. Encryption is not permitted in a PDF/A file, as its visualization would then be dependent upon information stored externally (i.e. a password). *Remedy: The PDF file cannot be converted to PDF/A in this form. If the required password is known, the PDF file's password protection can be removed in Adobe Acrobat and the file can then be saved.*

■ Embedded PostScript operator: The document uses PostScript code for the page description. PostScript code can also be used in PDF files. This option was primarily used at the beginning of the PDF format era by programs that did not offer full PDF support. However, there are very few programs that can visualize this PostScript code on a monitor. PostScript code is used in this PDF document to describe the page objects. Because the PDF/A standard stipulates that all components of a PDF file must be reliably visualized, the use of PostScript code is not permitted in PDF/A files. *Remedy: This error occurs very rarely. Current PDF-to-PDF/A converters solve this problem during the conversion process by removing the PostScript entries.*

■ EmbeddedFiles entry in Names dictionary: The document contains an embedded file. In PDF files, other files can be embedded as an 'attachment' in a similar manner as with an e-mail. The corresponding program is required to view these files (for example, Microsoft Word if a Word file is em-

bedded). Because the PDF/A standard stipulates that it must be possible to visualize all components of a PDF file without the aid of other software, file attachments are not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard file attachments' in the 'Discard Objects' area. This option corrects the error.*

■ Encoding entry prohibited for symbolic TrueType font:

This symbol font contains an allocation table for 'normal' fonts. The PDF/A standard stipulates that a TrueType font that is also a symbol font may not use an entry for this type of standard encoding, since standard encoding only defines 'normal' characters and not the special characters contained in symbol fonts. This PDF can therefore not be converted to PDF/A. *Remedy: In order to resolve this problem, the PDF file must be created anew, using a different font.*

■ File header not compliant with PDF/A: The PDF file header (PDF version entry or binary digit string) is not compliant. The PDF/A standard stipulates that a PDF file must comply with the general file header regulations in the PDF specification (1.6). *Remedy: The 'Save As' command in Acrobat can be used to solve this problem.*

■ File size is above 2GB: The file is too large (the maximum permitted size is 2GB). Extremely large files may lead to rendering problems when the files in question are printed out or displayed on a monitor. The maximum file size is therefore limited to 2 GB in the PDF/A standard. *Remedy: No repair possible. It may be possible to recreate the PDF file with a more effective compression type.*

■ Font not embedded (and text rendering mode not 3): Text uses a non-embedded text. The letters and other characters used in PDF texts require 'fonts' that determine their exact appearance when visualized. A text in this PDF uses a font that is not embedded into the PDF. It is therefore only possible to visualize this PDF correctly if this font is installed on the computer or printer being used. Because PDF/A stipulates that a PDF may not require external dependencies in order to be visualized, PDF/A files must not contain any fonts that are not embedded. The only exception is text that is not displayed but is merely used for the full-text search instead (text rendering mode = 3). *Remedy: The PDF must be regenerated with all used fonts embedded.*

■ Form field does not have appearance dict: Form field is 'invisible': A PDF file can contain form fields. These form fields must contain additional information to ensure that they can be visualized. This PDF contains form fields that do not contain the required additional information. Because the PDF/A standard stipulates that the visualization of a docu-

ment must be guaranteed and that it must be identical when output on a printer and when displayed on a monitor, a PDF/A document must only contain form fields with a visual representation. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard all comments, forms and multimedia' that can be used to remove form fields.*

■ Form field's AP (appearance) contains only N entry is not true:

Form fields in PDF files can contain differing visualization methods that are used, for example, depending on whether the mouse cursor is moved over a form field or the form field is clicked. These effects are, of course, not possible when a document is output on a printer. Because the PDF/A standard stipulates that the visualization of a document must be guaranteed and that it must appear identical when output on a printer and when displayed on a monitor, form fields in a PDF/A document may not have different visualization variants for mouse effects. *Remedy: The Acrobat Professional PDF Optimizer has an option called 'Discard all comments, forms and multimedia' in the 'Discard User Data' area. This option corrects this error.*

■ Glyphs missing in embedded font: A font does not contain all of the characters required. The letters and other characters used in PDF texts require 'fonts' that determine their exact appearance when visualized. This PDF contains an embedded font, in which however not all symbols that are used in texts using this font, are described. This means that there is no visual representation for the characters that are missing. The PDF/A standard stipulates that all fonts used must be embedded and that a visual representation for all used characters must exist. This PDF can therefore not be converted to PDF/A. *Remedy: To resolve this problem, the PDF file must be created anew.*

■ ICC profile version 4 or newer: This file uses an ICC profile for color definition that has a newer version than is permitted by PDF/A. The file can therefore not be converted to PDF/A. The ICC profile may also be defective. *Remedy: Note that it is very unusual for ICC profiles that are more recent than Version 3 to be used. Tools such as the Acrobat Preflight module can be used to find out which component uses the profile in question. This enables the error to be eliminated. The object in question must be recreated and the PDF file must then be regenerated.*

■ ID in file trailer missing or incomplete: No file ID entry available. Every PDF file should contain an internal ID that gives it a certain uniqueness and is altered each time the document is changed. The PDF/A standard requires the presence of this ID. *Remedy: The 'Save As' command in Acrobat can be used to solve this problem.*

■ **Image has OPI information:** OPI (Open Prepress Interface) is a procedure used in prepress. It involves the replacement of images with alternate images when printing via an OPI server. Since PDF files that carry OPI information can give different results depending on the output chosen (display on a monitor, printing on a desktop printer, or printing via an OPI server), OPI comments are not permitted in PDF/A files. *Remedy: OPI comments can be removed using the Acrobat PDF Optimizer.*

■ **Inadequate namespace URI for PDF/A entry:** The PDF/A entry in the document information is incorrectly formatted. A PDF/A file must have a corresponding entry in its document information. This entry can also be displayed with Adobe Acrobat. To display it, the user must choose 'Properties' from the 'File' menu in Adobe Acrobat and then click the 'Additional Metadata...' button. The entry appears in the 'Advanced' section under <http://www.aiim.org/pdfa/ns/id/>. The PDF entry is available in this document but it is incorrectly formatted. *Remedy: New PDF/A conversion.*

■ **Incorrect PDF/A version number (must be 1):** The PDF/A entry in the document information is incorrectly formatted (version number is not '1'). A PDF/A file must have a corresponding entry in its document information. This entry can also be displayed with Adobe Acrobat. To display it, the user must choose 'Properties' from the 'File' menu in Adobe Acrobat and then click the 'Additional Metadata...' button. The entry appears in the 'Advanced' section under <http://www.aiim.org/pdfa/ns/id/>. This group must contain a 'pdfaid:part:"1" entry. This entry specifies the version of the PDF/A standard. In this PDF file, the entry has a value that is not equal to '1'. *Remedy: New PDF/A conversion.*

■ **Incorrect PDF/A-1a conformance level (must be "A"):** This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). The PDF/A entry does not have the compliance level PDF/A-1a. A PDF/A file must have a corresponding entry in its document information. This entry can also be displayed with Adobe Acrobat. To display it, the user must choose 'Properties' from the 'File' menu in Adobe Acrobat and then click the 'Additional Metadata...' button. The entry appears in the 'Advanced' section under <http://www.aiim.org/pdfa/ns/id/>. This group must contain a 'pdfaid:conformance:' entry that specifies 'A' and declares that the file must comply with PDF/A-1a (and not 'only' with PDF/A-1b). PDF/A-1a files are always compliant with PDF/A-1b. *Remedy: Preflight can correct this entry if converting the file to PDF/A-1a.*

■ **Incorrect PDF/A-1b conformance level (must be "B"):** The PDF/A entry does not have the compliance level PDF/A-1b. A PDF/A file must have a corresponding entry in its docu-

ment information. This entry can also be displayed with Adobe Acrobat. To display it, the user must choose 'Properties' from the 'File' menu in Adobe Acrobat and then click the 'Additional Metadata...' button. The entry appears in the 'Advanced' section under <http://www.aiim.org/pdfa/ns/id/>. This group must contain a 'pdfaid:conformance:' entry that specifies 'A' or 'B' and declares that the file must comply with either PDF/A-1a or PDF/A-1b. PDF/A-1a files are always compliant with PDF/A-1b. This PDF contains a conformity entry, but it is not B or A. *Remedy: Preflight can correct this entry.*

■ **Interpolate key for image not false:** An image has an 'interpolation key' that is not supported by PDF/A viewers. The rendering or printout of an image is based upon the resolution of the output device. This means that it depends on the vertical or horizontal resolution if a monitor is being used and on the thickness of the lines with which the printing drum can be 'imaged' if a laser printer is being used. If the image resolution is significantly less than the resolution of the output device, additional pixels must be added. This process is known as interpolation. Interpolation is normally carried out in accordance with a standard procedure. However, an image in a PDF file can contain a key stating that a particular interpolation procedure must be used. Nevertheless, this option is rarely used these days and the key is ignored by most output devices. Because PDF/A files must appear the same regardless of the output device used, images in PDF files may not have interpolation keys. *Remedy: PDF-to-PDF/A tools such as Preflight or pdfaPilot have correction functions that make files containing interpolation keys standard-compliant.*

■ **Invalid rendering intent:** Only the following standard rendering intents are permitted in PDF/A files: Relative Colorimetric, Absolute Colorimetric, Perceptual, and Saturation. It is very unusual for a different rendering intent to be specified in a PDF. However, this PDF uses a different rendering intent. *Remedy: This is a very unusual error. It can be corrected using pdfaPilot.*

■ **Invalid WMode:** The stream direction is entered incorrectly in this font. The letters and other characters used in PDF texts require 'fonts' that determine their exact appearance when visualized. In addition to the appearance of the characters, a font must contain information regarding the font 'stream direction', since the characters of a font may not always be strung together horizontally from left to right as is the case with Latin fonts. For example, some Far Eastern fonts characters are strung together in a vertical direction (from top to bottom). This PDF uses a font with incorrect stream direction information. It is therefore impossible to ensure that the PDF will always be visualized in exactly the

same way regardless of the output device. This PDF cannot be converted to PDF/A. *Remedy: This problem can be avoided by using a different font for the text in question.*

JPEG2000 compression used: An image in this document is compressed in JPEG2000. Images placed in a PDF are usually compressed to keep the size of the PDF file to a minimum. Various compression methods such as ZIP, LZW, and JPEG can be used to compress image data. Images compressed using JPEG2000 can be decompressed in stages during visualization, meaning that an image can be displayed even if it is not fully decompressed. However, this process is only supported in more recent PDF versions and must not be used in PDF/A files. The PDF/A standard does not support objects that were not permitted in the PDF 1.4 specification that was published by Adobe with Acrobat 5. *Remedy: The Acrobat PDF Optimizer can be used to apply JPEG or ZIP compression (without downsampling) to all images. Alternatively, the file can be saved as a PDF 1.4 file.*

Keyword mismatch between Document Info and XMP metadata: The ‘Keywords’ entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: New PDF/A conversion.*

Last Modification Date mismatch between Document Info and XMP Metadata: The ‘ModifyDate’ entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: The ‘Save As’ command in Acrobat can be used to solve this problem.*

Layers used: The file contains layers that can be used to switch the visibility of objects on and off. Layers can be used in PDF files to define that certain page content should only be visualized under certain circumstances. Whether or not an object placed on a layer is visible depends on whether the viewer has set the layer in question to ‘visible’ or ‘invisible’. (It is also possible for visibility of a layer to be linked to other factors such as the zoom level with which a PDF is viewed – in this case, very small details in a drawing might only be visible when using a large zoom value.) Because PDF/A stipulates that the visual appearance of a PDF must always be exactly the same, layers cannot be used in PDF/A files. *Remedy: The Acrobat PDF Optimizer ‘Discard hidden layer content*

and flatten visible layers’ option or the Preflight ‘Merge Layers’ option can be used to correct this error.

LZW compression used: Objects used in a PDF are often compressed to keep the size of the PDF file to a minimum. Various compression methods are permitted for doing this, including ZIP, LZW, and JPEG (for images). This PDF file uses LZW compression. LZW compression is a lossless compression method that is patented. It can quite easily be replaced by ZIP compression, which is also lossless and uses a similar algorithm but is not patent-protected. The PDF/A standard only permits objects that can be visualized without restriction, even in the future. This also includes legal restrictions that might exist because of the LZW patent. For this reason, LZW compression is not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer contains an option that can be used to apply JPEG or ZIP compression (without downsampling) to all images in the ‘Images’ section.*

Marked entry in MarkInfo missing: This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). The document does not contain any information on its structure (in the document catalog). The stricter PDF/A-1a standard stipulates that a PDF file must contain structural information. *Remedy: To resolve this problem, the PDF file must be given the relevant structural information. This information can be added when the PDF is generated. Some PDF export modules have a ‘Tagging’ option or an option with a similar name. This enables structural information to be transferred into a PDF. It is also possible to add structural information later on in Adobe Acrobat Professional. Alternatively, it might be possible to convert the file to PDF/A-1b instead.*

Marked entry in MarkInfo not boolean: This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). The document contains no correctly formatted information on its structure. The stricter PDF/A-1a standard stipulates that a PDF file must contain structural information. *Remedy: To resolve this problem, the PDF file must be given the relevant structural information. This information can be added when the PDF is generated. Some PDF export modules have a ‘Tagging’ option or an option with a similar name. This enables structural information to be transferred into a PDF. It is also possible to add structural information later on in Adobe Acrobat Professional. Alternatively, it might be possible to convert the file to PDF/A-1b instead.*

Marked entry in MarkInfo not set to true: This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). The entry for structural information is defined

as ‘not available’ in the document. The stricter PDF/A-1a standard stipulates that a PDF file must contain structural information. *Remedy: To resolve this problem, the PDF file must be given the relevant structural information. This information can be added when the PDF is generated. Some PDF export modules have a ‘Tagging’ option or an option with a similar name. This enables structural information to be transferred into a PDF. It is also possible to add structural information later on in Adobe Acrobat Professional. Alternatively, it might be possible to convert the file to PDF/A-1b instead.*

■ **MarkInfo missing:** This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). The document does not contain any information on its structure (in the structure info directory). The stricter PDF/A-1a standard stipulates that a PDF file must contain structural information. *Remedy: To resolve this problem, the PDF file must be given the relevant structural information. This information can be added when the PDF is generated. Some PDF export modules have a ‘Tagging’ option or an option with a similar name. This enables structural information to be transferred into a PDF. It is also possible to add structural information later on in Adobe Acrobat Professional. Alternatively, it might be possible to convert the file to PDF/A-1b instead.*

■ **Max. nesting level of graphic states exceeded:** The PDF file contains very deeply nested page objects that can cause problems when it is printed out. Every PDF/A file must be a basically correct PDF file. The file currently open violates the restrictions of the PDF specification, which limits the degree of nesting for page objects. It is therefore not compliant with the PDF specification and cannot be converted to PDF/A. *Remedy: The problem can possibly be resolved by opening the file in Adobe Acrobat and saving it once again using the ‘Save As’ option. Otherwise, it might be possible to clean it up using the ‘Save Optimized As’ option.*

■ **Max. number of colorants for DeviceN exceeded:** An object uses too many color channels in a DeviceN object. DeviceN is a multi-channel color space in which spot colors can also be used, for example. DeviceN objects can be used in PDF/A files but the number of channels is restricted to a maximum of 8. *Remedy: It is extremely unusual for a DeviceN color space to use more than 8 channels. Tools such as the Acrobat Preflight module can be used to find out which component uses the color space in question. The error can then be corrected. The object in question must be recreated and the PDF file must then be regenerated.*

■ **Metadata does not conform to XMP:** The PDF/A standard stipulates that document information must exist in the XMP area. This document has XMP metadata but it is incor-

rectly formatted. *Remedy: The file must be either converted to PDF/A again or the PDF document must be regenerated.*

■ **Metadata entry missing:** No document information for the PDF/A entry. The PDF/A standard stipulates that document information must exist in the XMP area. There is no XMP document information in this PDF file. *Remedy: The ‘Save As’ command in Acrobat can be used to solve this problem. XMP metadata is created during the generation of PDF/A.*

■ **Metadata not embedded as plain text:** The PDF/A standard stipulates that document information must exist in the XMP area and must not be compressed. However, the XMP metadata in this PDF is compressed. *Remedy: The file must be either converted to PDF/A again or the PDF document must be regenerated.*

■ **More than one encoding in symbolic TrueType font’s cmap:** A symbol font has more than one allocation table, which means that characters cannot be uniquely identified. The PDF/A standard stipulates that TrueType font that is also a symbol font may only contain a single encoding entry. Otherwise, unique allocation is impossible. *Remedy: To resolve this problem, the PDF file must be created anew.*

■ **Named action with a value other than standard page navigation used:** PDF files can dynamically alter their content during visualization. Actions can be contained in the PDF file for this purpose. The PDF/A standard stipulates that the visualization of a document must be guaranteed and always the same. For this reason, active content is not permitted in PDF/A files. The only exception is elements for page navigation. *Remedy: This problem can be solved using the Acrobat PDF Optimizer or pdfaPilot.*

■ **NeedAppearances flag present but not set to false:** Form fields can be filled with variable content. These initially empty fields can have an entry that defines that they must be filled with variable content (either through user input or input that is determined dynamically such as the system environment or time). Because the PDF/A standard stipulates that the visualization of a document must always be identical whether it is displayed on a monitor or output on a printer, form fields in PDF/A documents must not contain this entry. *Remedy: As long as the content in question is not adversely affected, the ‘Flatten form fields’ PDF Optimizer function can be used to correct this error.*

■ **Number of PDF/A-1 OutputIntent entries > 1:** There are multiple output intents for PDF/A. This is not compliant with the PDF/A standard. Output intents are used to define the colors used in a PDF in accordance with a specific output

procedure (for example, printing or display on a monitor). This uniquely defines color specifications. Consequently, only a single PDF/A output intent may be specified for a PDF/A file. *Remedy: Acrobat 8 Preflight has a correction option that removes all output intents from a document. The user has to create a new profile and choose the option 'Remove Output Intent' from the list of predefined corrections. Since the correction removes all output intents, the document must then be converted to PDF/A again.*

■ OutputConditionIdentifier missing or empty in PDF/A OutputIntent: The output intent is incomplete. The output condition identifier is missing. *Remedy: New PDF/A conversion.*

■ Page description contains invalid operator: The PDF file uses invalid commands for the page description. Every PDF/A file must be a basically correct PDF file. The file that is currently open uses a command in its page description that is not defined in the PDF specification. This file is not a valid PDF file. It is therefore not possible to convert it to PDF/A. *Remedy: The problem can possibly be resolved by opening the file in Adobe Acrobat and saving it once again using the 'Save As' option. Otherwise, it might be possible to clean it up using the 'Save Optimized As' option. If the problem still exists, the PDF file must be regenerated.*

■ PDF contains data after end of file marker: Every PDF file should have an end of file marker. No further data should follow this marker. In this PDF, there is data after the end of file marker. *Remedy: The 'Save As' command in Acrobat can be used to solve this problem.*

■ PDF contains EF (embedded file) entry: The document contains an entry for an embedded file. In PDF files, other files can be embedded as an attachment in a similar manner as with an e-mail. The corresponding program is required to view these files (for example, Microsoft Word if a Word file is embedded). Because the PDF/A standard stipulates that it must be possible to visualize all components of a PDF file without the aid of other software, file attachments are not permitted in PDF/A files. *Remedy: The Acrobat PDF Optimizer contains an option called 'Discard file attachments' in the 'Discard Objects' area. This option corrects the error.*

■ PDF/A Destination profile version 4 or newer: The file has an output intent, but the ICC profile used is not compatible with PDF/A. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output in-

tent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: New PDF/A conversion.*

■ PDF/A entry missing: No PDF/A entry in the document information. A PDF/A file must have a corresponding entry in its document information. This entry can also be displayed with Adobe Acrobat. To display it, the user must choose 'Properties' from the 'File' menu in Adobe Acrobat and then click the 'Additional Metadata...' button. The entry appears in the 'Advanced' section under <http://www.aiim.org/pdfa/ns/id/>. There must be a 'pdfaid:part: 1' entry for the PDF/A version (only version 1 at present) and a 'pdfaid:conformance:' entry for the conformity level (PDF/A-1a or PDF/A-1b). The conformity level must be 'A' or 'B'. PDF/A-1a files are always compliant with PDF/A-1b. *Remedy: The file can be converted to PDF/A again.*

■ PDF/A OutputIntent has no destination profile: Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist, along with its destination profile. *Remedy: Preflight contains a correction option that converts the alternate visualization to CMYK (SWOP). This correction must be duplicated and an RGB color space such as sRGB must be used as the target. The correction can then be assigned to a profile. The alternate visualization of the spot color can then be modified. pdfaPilot can also carry out this task.*

■ Producer mismatch between Document Info and XMP metadata: The PDF 'Producer' entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: New PDF/A conversion.*

■ Prohibited annotation type: A PDF can contain different types of comment. Some of these comment types are intended for multimedia content: Sound and movie comment types. These types of comment cannot be reproduced by printers. The FileAttachment comment type allows files in other formats to be embedded into a PDF. Only specialized visualization systems are able to render these file attachments. These types of comment are not permitted in PDF/A files as they cannot be visualized on all output devices. In ad-

dition, all types of comment that were not specified in PDF 1.4 (from Adobe Acrobat 5 onwards) are not permitted in PDF/A since PDF/A is based on PDF 1.4. *Remedy: The Acrobat PDF Optimizer provides an option for removing file attachments in the 'Discard User Data' area.*

■ **RGB used but PDF/A OutputIntent not RGB:** Device-dependent color (DeviceRGB) is used but no RGB output intent exists. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: New PDF/A conversion.*

■ **RGB used for alt. color but PDF/A OutputIntent not RGB:** A spot color has been defined in DeviceRGB but the output intent is not defined for RGB. Because the PDF/A standard stipulates that colors must appear the same (as far as is technically possible) regardless of the output device, either a PDF/A document must only contain device-neutral colors or the color properties of the output device must be defined using an output intent profile. If a document contains DeviceRGB or DeviceCMYK colors, an output intent of the same type must therefore exist. *Remedy: New PDF/A conversion.*

■ **Scaling factor used:** Page contains a zoom factor or downsizing factor. This PDF defines a change of image scale. This particular characteristic was first introduced with Adobe PDF 1.6 (as of Acrobat 7). The PDF/A standard only permits objects that are compatible with PDF 1.4. A change of image scale is therefore not permitted in PDF/A files. *Remedy: This problem can be corrected using the Acrobat PDF Optimizer by selecting PDF 1.5 compatibility (Acrobat 6).*

■ **SMask entry present with a value other than None:** A partially transparent mask is used in this PDF file. Masks hide background objects. They can however be set to 'transparent' in PDF files so that objects positioned behind them still remain partially visible. You can set a percentage value on a scale of 0% to 100% to define the extent to which the background of a 'transparent' object should be visible. The color values of the foreground mask and background object must be offset against one another for the reproduction of such constructions when displaying files on a monitor or printing them out. However, this method of blending is not clearly defined. The PDF/A standard stipulates that all features used in a PDF file must be displayed in a single unique way on a monitor or in a printout. Because this cannot be ensured in the case of transparent objects and their backgrounds, transparency is not permitted in

PDF/A files. *Remedy: Adobe Acrobat Professional (Version 6, 7, or 8) includes a flattener module that can be used to remove transparencies.*

■ **Stream object contains F entry:** To be visualized in its entirety, this PDF requires additional files. Because the PDF/A standard stipulates that a PDF must be complete and must not require any other information for its visualization, this PDF cannot be converted to PDF/A. *Remedy: It is unusual for external files to be required for the visualization process. In order to trace the cause, a file that produces this error must be checked – with Adobe Preflight, for example – to see which objects are to blame.*

■ **Stream object contains FDecodeParams entry:** External files are required for the visualization of the file in question (FDecodeParams entry). To be visualized in its entirety, this PDF requires additional files. Because the PDF/A standard stipulates that a PDF must be complete and must not require any other information for its visualization, this PDF cannot be converted to PDF/A. *Remedy: It is unusual for external files to be required for the visualization process. In order to trace the cause, a file that produces this error must be checked – with Adobe Preflight, for example – to see which objects are to blame.*

■ **Stream object contains FFilter entry:** External files are required for the visualization of the file in question (FFilter entry). To be visualized in its entirety, this PDF requires additional files. Because the PDF/A standard stipulates that a PDF must be complete and must not require any other information for its visualization, this PDF cannot be converted to PDF/A. *Remedy: It is unusual for external files to be required for the visualization process. In order to trace the cause, a file that produces this error must be checked – with Adobe Preflight, for example – to see which objects are to blame.*

■ **Stream size is above 2 GB:** A data stream in this file is too large (the maximum permitted size is 2 GB). Extremely large files may lead to rendering problems when the files in question are printed out or displayed on a monitor. The maximum file size and also the size of internal data objects in PDF/A files is therefore limited to 2 GB. *Remedy: No repair possible. It may be possible to recreate the PDF file with a more effective compression type.*

■ **Subject mismatch between Document Info and XMP metadata:** The 'Subject' entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: New PDF/A conversion.*

■ **Text cannot be mapped to Unicode:** This message only appears when validating PDF/A-1a (and not when validating PDF/A-1b). This PDF file contains characters that could not be allocated to a Unicode ID since the relevant information is not available in the PDF file. *Remedy: In order to resolve this problem, the PDF file must be created anew, using a different font. Alternatively, it might be possible to convert the file to PDF/A-1b instead.*

■ **Title mismatch between Document Info and XMP metadata:** The ‘Title’ entry in the XMP document information deviates from the corresponding entry in the document properties. The PDF/A standard stipulates that document information must exist in the XMP area. If this data is also contained in the document properties, it must be identical to the entries in the XMP area. *Remedy: This information can be opened using the ‘File’ menu and changed in the general Document Properties dialog.*

■ **TR2 entry used with value other than Default:** Underlying gradation curve. The color values of a page object (text, images, graphics, and so on) can be changed with the aid of gradation curves (or ‘transfer curves’). On the basis of a ‘transfer curve’, a new value is determined for every color value when a document is displayed on a monitor or printed out. This value is then displayed or printed in place of the ‘original value’. Because not all devices can handle transfer curves, they are not permitted in PDF/A files. *Remedy: The Preflight ‘Apply transfer curves’ function corrects this error.*

■ **Transparency used:** Objects in this PDF file are defined as ‘transparent’. The PDF/A standard stipulates that all features used in a PDF file must be displayed in a single unique way on a monitor or in a printout. Because this cannot be ensured in the case of transparent objects and their backgrounds, transparency is not permitted in PDF/A files. *Remedy: Adobe Acrobat Professional (Version 6, 7, or 8) includes a flattener module that can be used to remove transparencies.*

■ **Type 2 CID font: CIDToGIDMap invalid or missing:** Not all characters (glyphs) can be allocated to this font (CIDToGIDMap is missing or incorrect). A font in this PDF does not have a complete allocation table for allocating character codes to character representations in the font. In PDF/A files, the reliable allocation of codes to character representations must be ensured. *Remedy: In order to resolve this problem, the PDF file must be created anew, using a different font.*

■ **Uses NChannel color:** An object in this file uses the NChannel color space, which is not allowed in PDF/A. The document defines colors in the ‘NChannel’ color space, in-

roduced with Acrobat 7 (PDF 1.6). NChannel is an extension of the DeviceN color space, permitted since PDF 1.3. Both color spaces allow the specification of color values in multi-channel color spaces in which spot colors, for example, may also be used. The PDF/A standard does not support objects that were not permitted in the PDF 1.4 specification that was published by Adobe with Acrobat 5. *Remedy: Use the Acrobat PDF Optimizer to save the PDF file as a PDF 1.4 version.*

■ **Uses OpenType font:** OpenType fonts may not be used in PDF/A files. There are various common font formats (PostScript Type1, Type3 or TrueType). They can be embedded in all PDF versions. As of PDF 1.5 (Acrobat 6), OpenType format fonts can also be embedded. This font format is used by one of the fonts embedded in this PDF file. The PDF/A standard only permits the use of objects that are compatible with PDF 1.4. Fonts in OpenType format must therefore not be used or embedded. *Remedy: In order to resolve this problem, the PDF file must be created anew, using a different font.*

■ **Width information for glyphs incomplete:** Width information is missing for some characters in a font used in this document. The letters and other characters used in PDF texts require ‘fonts’ that determine their exact appearance when visualized. The characters used in the text are then depicted and arranged in accordance with the representation stored in the font. The precise position of each character depends upon the tracking of the previous character. The PDF/A standard stipulates that width information must be available for every single character used in a document. *Remedy: To resolve this problem, the PDF file must be created anew.*

■ **Width information for glyphs is inconsistent:** Deviating specifications exist for character width. The letters and other characters used in PDF texts require ‘fonts’ that determine their exact appearance when visualized. The characters used in the text are then depicted and arranged in accordance with the representation stored in the font. The precise position of each character depends upon the width of the preceding symbol. The width specification for any one character is defined both in the font that is embedded in the PDF and in the PDF itself. The PDF/A standard stipulates that the width specifications in the embedded font and in the PDF file must be identical. *Remedy: To resolve this problem, the PDF file must be created anew.*

■ **Wrong encoding for non-symbolic TrueType font:** This font does not use standard character to symbol allocation (MacRoman or WinAnsi). This PDF can therefore not be converted to PDF/A. *Remedy: In order to resolve this problem, the PDF file must be created anew, using a different font.*

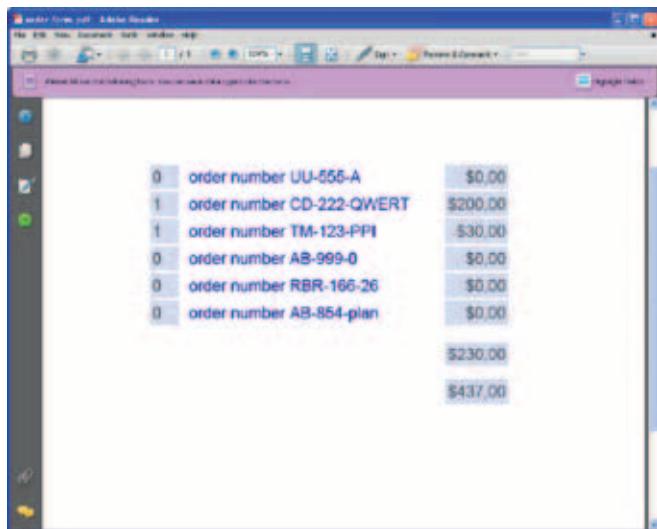
Glossary

Explanation of terms relating to PDF/A

■ **Accessibility:** In the digital world, accessibility aims to ensure that users with impaired vision, restricted neuro-muscular skills, and other disabilities can also take part in the exchange of information. Web pages and other files must be designed in a way that provides a clear flow structure to help screenreaders reproduce their content correctly. PDF files can be both accessible and PDF/A-compliant at the same time. Accessibility is increasingly regulated by legislation in the US and Europe.

■ **Adobe Acrobat:** Program for creating and processing PDF files. Version 1 was introduced by →*Adobe Systems* in 1993. The current version is Version 8. Acrobat Standard, Professional, and Elements offer different features but all belong to the Acrobat family. Adobe Professional includes Adobe →*Distiller*, which can create PDF documents from →*PostScript* and EPS data.

■ **Adobe Reader:** Adobe Reader (previously called Acrobat Reader) is Adobe's free →*PDF viewer*. This program runs on various computer and mobile-device platforms and has been downloaded from the company's sites millions of times. The free distribution of this program has contributed to the success of the PDF format. Adobe Reader 8 also allows form files to be saved if the creator of the documents has enabled the function in Acrobat Professional.



Version 8 of the free Adobe Reader can also save form files if the files permit it.

■ **Adobe Systems:** US software company founded in 1982 by John Warnock and Charles Geschke. Warnock and Geschke developed the →*PostScript* format for printing files. The name 'Adobe' refers to a type of clay or the brick made from it, and a river called Adobe Creek flows near the company's headquarters. Their well-known products include Photoshop, Illustrator, InDesign, and Acrobat. PDF was developed by Adobe.

■ **CCITT Group 4:** The 'Comité Consultatif International Téléphonique et Télégraphique' (International Telephone and Telegraph Consultative Committee) developed this lossless compression procedure for black and white images (line art) for use when sending faxes.

■ **CMYK:** This abbreviation stands for Cyan, Magenta, Yellow, and Key (=black). Different sized dots in these four colors can be distributed in various ways to realistically depict most color images as well as graphics and text. However, fluorescent colors and other shades cannot be displayed well using CMYK. Spot colors are used to display these shades.

■ **Color management:** This technology aims to enable the uniform display of colors regardless of whether they appear on a monitor or in proofs, newspaper printouts, or art printouts. Color profiles (usually →*ICC profiles*) are very important for color management – they enable the device-independent display of colors. Color management encompasses all production stages from digitalization using a scanner or digital camera to editing and displaying the result on a monitor or printing it out.

■ **Comments:** Also called annotations. Text comments enable sophisticated correction workflows with PDF files such as those required in editorial departments. →*Adobe Acrobat* enables the exchange of comments with recipients who only have the free →*Adobe Reader* (current version) at their disposal. PDF/A supports text annotations but prohibits comment types such as Sound and Movie.

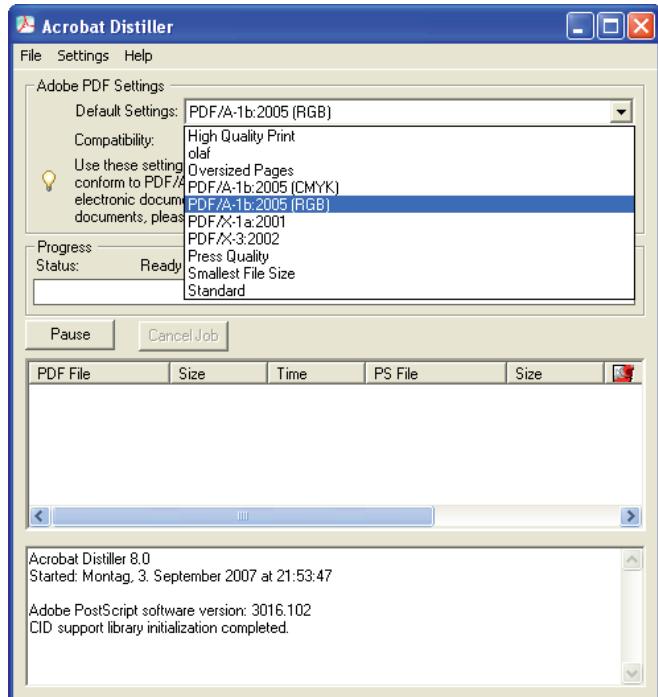
■ **Compression:** Technical procedure for reducing file size. There are 'lossy' compression types such as →*JPEG* and 'lossless' procedures such as →*ZIP*. PDF can use compression on page description objects, such as embedded

images compressed in JPEG. However, other elements of a PDF file that are not components of the page description can also be compressed. As of PDF 1.6, these elements can even be merged together into a compressed object (cross-object compression).

■ **Conversion:** The term ‘conversion’ refers to changing a file from one file format to another.

■ **Digital signature:** Electronic signatures are important in many fields of business and administration. They are used to identify the originator of a document as well as the read and usage authorizations of its recipient. Digital signatures must be suitable for encryption. They must also be impossible to falsify. They can be managed and used in PDF documents using programs such as Acrobat and →*Adobe Reader*. The use of PDF/A with digital signatures requires a precisely planned process flow.

■ **Distiller:** The Distiller is an auxiliary program for creating PDF files from the →*PostScript* print data format. The program has been delivered with →*Adobe Acrobat* since Version 1. The Distiller enables process flows to be automated by setting up watched folders.



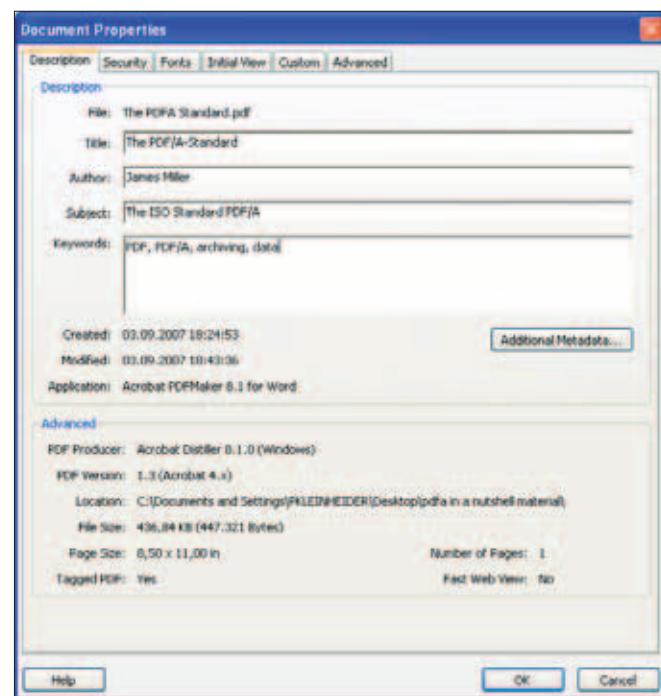
The Distiller enables the creation of PDF/A-1b documents. It is not possible to convert files to PDF/A-1a because the structures required for compliance with the stricter compliance level cannot be adopted or generated.

■ **DMS:** This abbreviation stands for Document Management System. It encompasses the management of docu-

ments that used to be printed out on paper in electronic systems. DMS is an important component of electronic document archiving systems.

■ **Document scanner:** These special devices are for creating large document sets in the shortest amount of time possible. Document scanners enable entire batches of documents to be digitalized (both the front and back of pages). Scanned material is increasingly stored as PDF files. If the files in question are to be archived, it makes sense to save them as PDF/A files.

■ **Document properties:** The document properties (document info) for PDF files contain four entries: Title, Author, Subject, and Keywords. These entries constitute basic →*metadata* information. In →*Adobe Acrobat* and →*Adobe Reader*, document info can be displayed by pressing Ctrl+D.



Document Properties in Acrobat 8 Professional. This area displays information such as the title and author of a document.

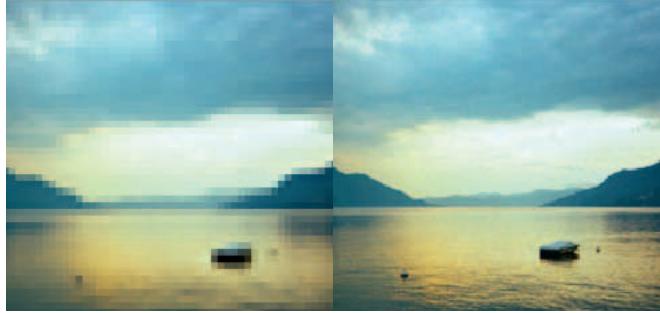
■ **Font:** Character set. A font contains all the letters in an alphabet as well as digits and sometimes graphical symbols.

■ **Glyph:** A glyph is a graphical representation of a character. A character is an abstract concept of a letter or symbol. A glyph is the actual graphic used to represent it.

■ **ICC profile:** ICC profiles are important features of →*color management*. An ICC profile is a data record that

describes the color space a device (a monitor, printer, scanner, or similar device) uses to specify or reproduce colors. ICC stands for International Color Consortium, a group that consists of manufacturers of graphics, image editing, and layout programs.

■ **Image resolution:** A digital image consists of pixels (image elements). The number of pixels per inch determines the quality of an image. Because they carry more information, high-resolution images have larger file sizes. The usual screen resolution is 72 ppi (pixels per inch – 1 inch is 2.54 centimeters). 300 ppi is often used for printing.



The left-hand image above is rendered in 72 ppi; the right-hand image has a resolution of 300 ppi. Both images have been significantly magnified.

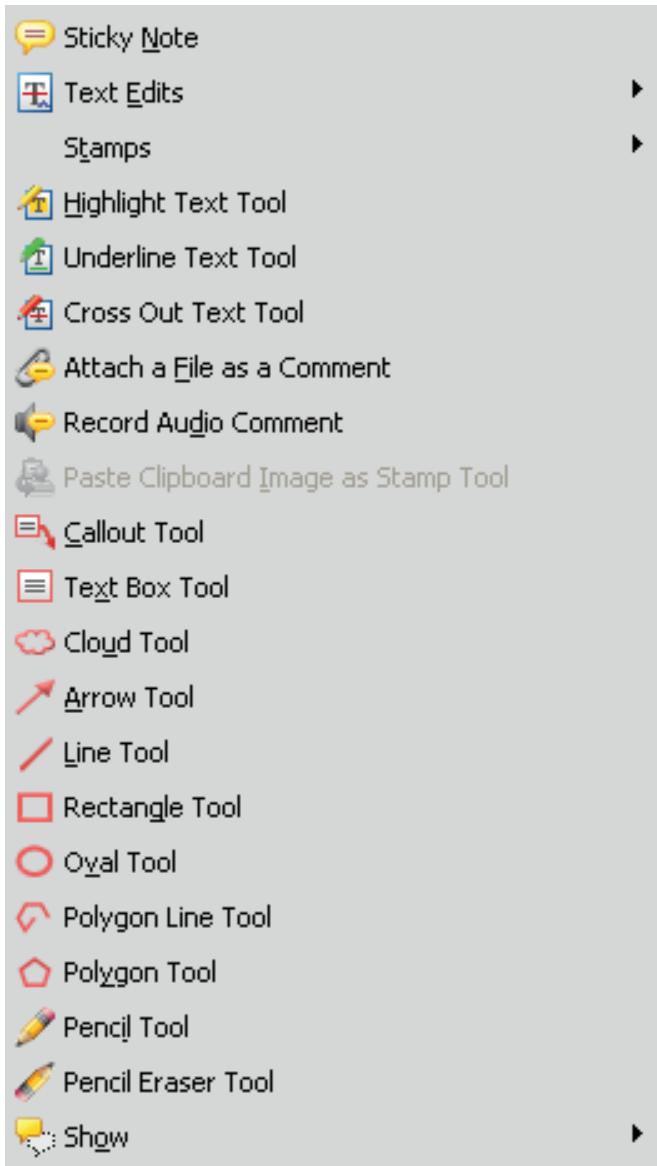
■ **ISO:** The International Organization for Standardization. This organization compiles international standards including the PDF/A standard (ISO 19005-1:2005). It was founded in 1947 in Geneva. More than 150 countries are now represented within ISO. The standards of the group are formulated in committees and sub-committees and are printed and published digitally once complete.

■ **JBIG:** JBIG is a standard for the lossless compression of digital images. Its name is taken from the first letter of each word in the name of its creators, the Joint Bi-level Image Experts Group. JBIG was specially developed for black and white images such as faxes.

■ **JPEG:** This abbreviation stands for Joint Photographic Experts Group, the group that developed this image compression procedure. When it was initially developed, this procedure was the first to enable the high image resolutions required for printing while retaining relatively low data volumes. However, it is a 'lossy' compression procedure and is not suitable for black and white images. It offers low, medium, high, and maximum quality levels that users can select when they save a document. The lower the quality, the smaller the file. PDF/A supports JPEG compression. This compression

procedure was the predecessor of →JBIG (bi-level images, black and white files) and →JPEG2000 (improved compression).

■ **JPEG2000:** An image compression standard that, like →JPEG, was developed by the Joint Photographic Experts Group. JPEG2000 supports both lossless and lossy compression. This image file format can include a range of metadata that facilitates file management and makes it easier to find images on the Internet. JPEG2000 is not permitted by PDF/A-1a and -1b, but PDF/A-2 will support it.



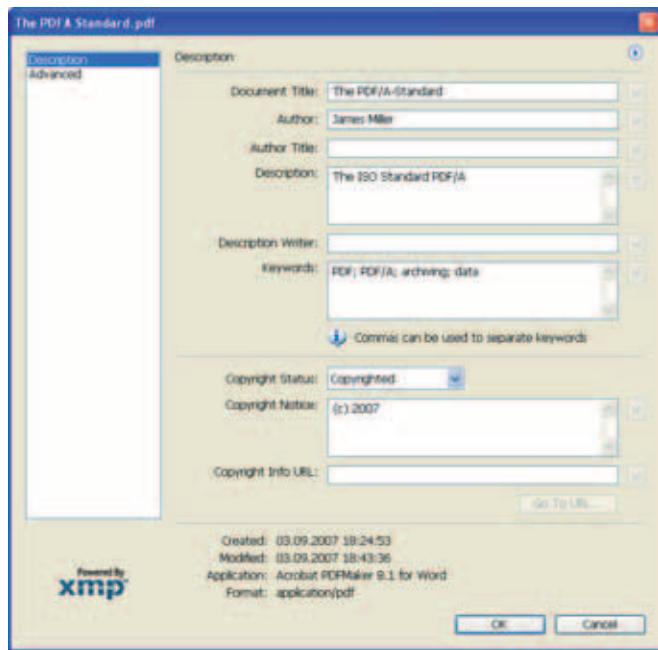
Acrobat offers a range of different comment types, not all of which are supported by PDF/A. The illustration above displays comment types in Acrobat.

■ **Library:** Program library. Unlike programs, a program library does not constitute an independent executable unit.

Instead, it is an auxiliary module that contains elements required to make programs available.

■ **LZW:** An older, lossless image compression procedure from the 1970s/1980s. It is named after its creators – Abraham Lempel, Jacob Ziv, and Terry A. Welch. This procedure is not supported by PDF/A because it was subject to licensing restrictions for a considerable period of time.

■ **Metadata:** A digital document can have additional data on its properties. This information is called metadata. Metadata provides information on attributes such as the author of a file and the title of a document. It enables documents to be categorized by keywords and supports the addition of copyright information. Adobe products use modern →*XMP* metadata.



Metadata in a PDF file, displayed in Acrobat Professional.

■ **OCR:** This abbreviation stands for Optical Character Recognition. It is an optical text recognition procedure. OCR allows the pixel-based data of scanned documents to be assigned searchable text later on.

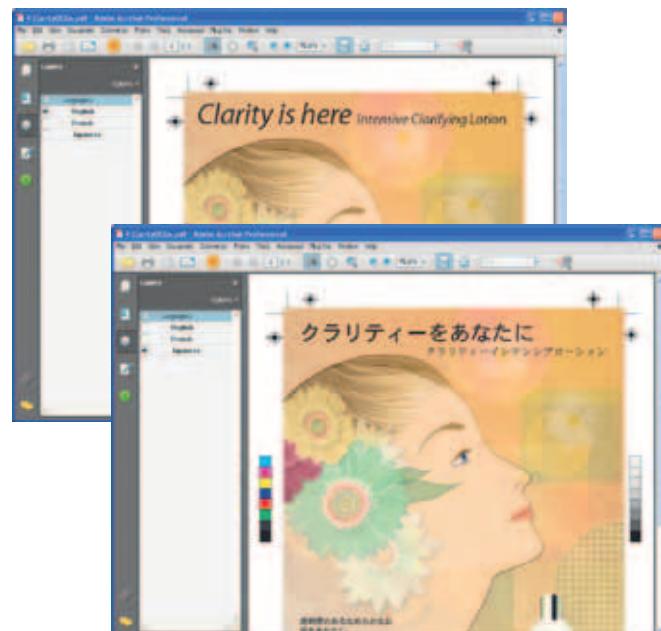
■ **Output intent:** An output condition or output intent is a component of color management. The output intent can be used to specify the intended output method for a (PDF) file. This is normally realized using an ICC output profile. For example, whereas a PDF file intended for offset printing might user the 'ISO Coated' output intent, a file for display on a monitor is better suited to →*sRGB*. The assign-

ment of an output intent can modify colors in line with the requirements of a different output device. For example, since Version 6, Adobe Acrobat has displayed a PDF with an output intent for offset printing differently to how it would display the same PDF with an output intent for newspaper printing.

■ **PDF:** This abbreviation stands for 'Portable Document Format'. It is a platform-independent, open file format that has been developed by →*Adobe Systems* since 1993. Like a container, a PDF document can contain diverse elements: Images, text, sound, movies, 3D objects, form elements, and many more. The functional scope of PDF is constantly being enhanced. The current version is PDF specification 1.7, which was introduced with →*Adobe Acrobat* 8.

■ **PDF viewer:** Program for displaying PDF documents. In addition to the →*Adobe Reader*, such programs include the 'Preview' program that belongs to the current version of Apple's operating system. There are both free PDF viewers for various platforms and viewers that must be purchased.

■ **PDF layer:** PDF files can contain layers. A better term is 'Optional Content Group'/OCG, since PDF layers are not layers like those in Photoshop – instead, they enable content to be made visible or invisible depending on the context defined for a file. PDF layers/OCGs are not permitted in PDF/A files because they enable different representations of the same file.



Layers in PDF: PDF layers (OCGs) can be used to create language layers in PDF files, for example.

■ **PDF version:** PDF is constantly being developed. With each new Acrobat version, Adobe publishes a new PDF specification. The document containing the specification is called a 'PDF Reference'. PDF 1.7 has been available since the rollout of Acrobat 8 (tip: to determine the corresponding Acrobat version, add one to the PDF version number – for example, PDF 1.3 belongs to Acrobat 4).

■ **PDF/A:** Standard developed by →ISO, the International Organization for Standardization, especially for the long-term archiving of PDF files. The PDF/A-1 standard was adopted under the name ISO 19005-1:2005 in 2006. This first version used the PDF 1.4 specification to define the elements that are permitted in PDF/A files. PDF components that were only introduced in later versions of the PDF specification are therefore prohibited in PDF/A files. Such components must be modified or removed. The PDF/A-2 standard, which is already being compiled, will be based on a more recent PDF specification.

■ **PDF/X:** Standard developed by →ISO for the prepress field. PDF/X is defined in ISO standards 15929 and 15930. This standard enables the reliable reproduction of PDF print files without the need for protracted talks beforehand. The PDF/X-1a and PDF/X-3 are already prevalent. PDF/X-1 is intended only for use with →CMYK (and possibly spot colors), whereas PDF/X-3 also permits profiled →RGB.

■ **PDFMaker:** The Adobe PDFMaker is a macro installed with →Adobe Acrobat. Among other things, it enables the creation of PDF files from Word. The PDFMaker works with the Acrobat →Distiller to create PDF files, meaning that it has access to all PDF settings for the Adobe program.

■ **Plug-in:** A plug-in is an additional module for a main program. These additional modules are often marketed by third-party suppliers. →Adobe Acrobat can be enhanced by many additional functions with plug-ins.

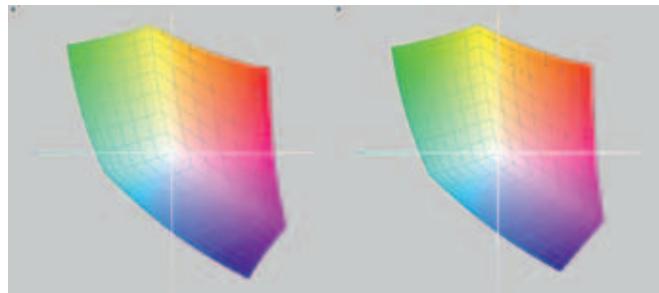
■ **ppi:** Unit of measure for image resolution. The abbreviation stands for 'pixels per inch'. One inch is equal to 2.54 centimeters.

■ **PostScript:** PostScript is a page description language that has been developed by the US company →Adobe Systems since 1984. It converts pages into PostScript format in order to output them on different output devices in any size and resolution without loss. The functional scope of the format has been enhanced twice, and the current version is PostScript Level 3 (available since 1998).

■ **Preflight:** This plug-in, which is delivered with Acrobat, is a tool for checking PDF files. It is developed by the Berlin-based company callas software. As of Acrobat 8, Preflight can carry out corrections as well as checking PDFs. Acrobat also uses Preflight to carry out all of its PDF/A validations and conversions. In addition to using the validation and verification profiles delivered with Preflight, users can also create their own profiles.

■ **RGB:** This color space consists of the primary colors red, green, and blue and is used for displaying documents on color monitors. The additive color model has 255 grades for these three basic colors. White is created if all three components have the value 255; black is formed if they all have the value 0.

■ **sRGB:** sRGB (standard RGB) color space. It was mutually developed in 1996 by Hewlett-Packard and Microsoft.

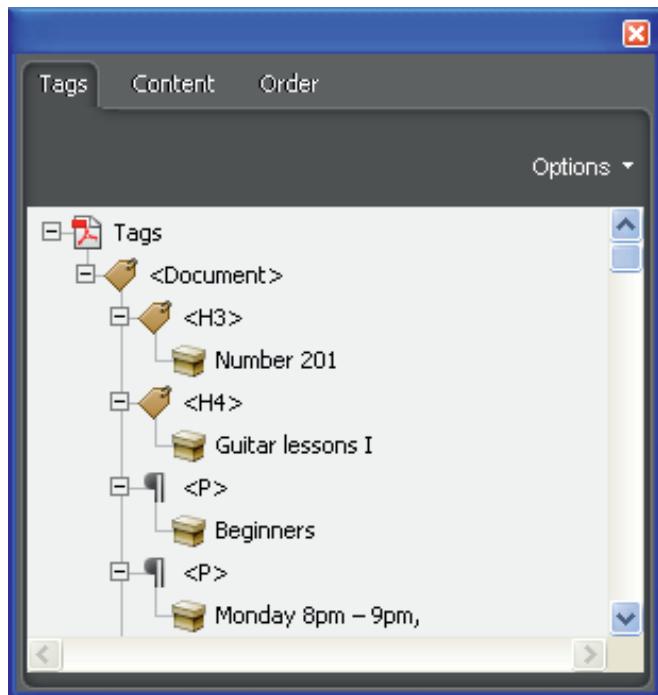


The image on the left shows a schematic depiction of the sRGB color space; the image on the right displays the RGB color space.

■ **Tagged PDF:** Structured PDF. The content structure of a PDF/A file must be specified using tagged PDF. Tagged PDF is also a prerequisite for accessible PDF files. The corresponding structures can be created in the source document (for example, in InDesign) or can be added later on in the PDF file. Unlike PDF/A-1b files, PDF/A-1a files must contain structural information (tagging).

■ **Tags:** Tags help to produce →tagged PDF. For example, an image can be given the tag 'Figure' and an alternate image description can be added for accessibility reasons.

■ **TIFF:** The file format TIFF (Tagged Image File Format) was developed by Aldus (taken over by Adobe in 1984) and Microsoft for scanned raster graphics for color separation. TIFF can contain layers. →JPEG or →ZIP compression can be used to reduce the file size. The variant for black and white images, →TIFF G4, has been an important format for archiving scanned documents for a long period of time.



Tags in Acrobat Professional: All elements of a tagged PDF file are given marks that clearly assign them to a content type and style. Tags also control their sequence.

■ **TIFF G4:** TIFF G4 is a monochrome type of TIFF that is compressed using the →CCITT Group 4 procedure. This type of TIFF file combines high readability of text documents with relatively small document size. This is precisely what is required for archiving purposes.

■ **Transparency:** Transparent objects can occur in PDF files. If the opacity of an element is less than 100 percent, the background is visible through the element in question. Transparencies have been supported in PDF documents since PDF 1.4 (Acrobat 5). Transparencies are not permitted by the PDF/A standard.

■ **Unicode:** International industry standard promoted by the Unicode Consortium since 1991. It aims to define a digital code for every single character in all known writing systems/character systems. It has different formats, but UTF-8 (Unicode Transformation Format) is the most common both on the Internet and in common operating systems.

■ **Validation:** From the Latin ‘validus’ ('strong'). A validation is the checking of an hypothesis. It concludes in verification ('true'), falsification ('not true'), or in no result. In the context of PDF/A, validation means checking a file that claims to be PDF/A-compliant to see whether this is actually the case.



Not yet validated: The Acrobat Preflight tool can be used to check the validity of PDF/A documents.

■ **XML:** Extensible Markup Language (XML) can be used to structure document content.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<edition>
  <title>Cities</title>
  <entry>
    <keyword>Phoenix</keyword>
    <text>Phoenix is the capital of Arizona...</text>
  </entry>
  <entry>
    <keyword>Berlin</keyword>
    <text>Berlin is the capital of Germany...</text>
  </entry>
</edition>
```

XML can be used to store structured information in a kind of tree structure.

■ **XMP:** This abbreviation stands for Extensible Metadata Platform. XMP is used to integrate →metadata into Adobe programs in a uniform manner.



XMP – developed by Adobe Systems – is used to integrate metadata into PDF/A.

■ **XPS:** This abbreviation stands for XML Paper Specification, a Microsoft document format.

■ **ZIP:** The ZIP format is an open format for compressing files. ZIP compression is ‘lossless’ and works well for images that contain large areas in one color or in repeating patterns. The use of ZIP compression is permitted for PDF/A documents.

About: The PDF/A Competence Center

Association for Digital Document Standards – ADDS

The PDF/A Competence Center is an initiative of the Association for Digital Document Standards (ADDS) e.V., founded in September 2006. A particularly important aim of the association is to promote the exchange of information and experience in the area of long-term archiving in accordance with ISO 19005 (PDF/A).



The new ISO standard for long-term archiving, PDF/A, is generating considerable interest in the market. In order to encourage the high demand for information and exchange of ideas concerning PDF/A, callas software GmbH, Compart Systemhaus GmbH, LuraTech Europe GmbH, PDF Tools AG and PDFlib GmbH have founded the PDF/A Competence Center.

The executive chairman is Thomas Zellmann, a managing partner of LuraTech. Dr. Hans Baerfuss, CEO of PDF Tools AG, Switzerland, is the executive vice-chairman.

The association is geared towards developers of PDF solutions, companies that work with PDF/A in the area of DMS/ECM, interested individuals, and also users who want to implement PDF/A in their organizations. Although the months directly after the founding saw new members predominantly from German speaking regions, the executive committee has expanded their activities internationally beginning in 2007.

Interested parties can thus benefit from the combined knowledge of competent PDF/A suppliers. The newly founded association offers numerous services including

conducting events, working on further standardizations and serving as a central competent point of contact for answering all questions about PDF/A.

Work on the ISO Standard

Several members of the PDF/A Competence Center are technically oriented and actively participate in the further development of the PDF/A standard as members of the responsible ISO committee (ISO TC 171 – Document management applications).

Member companies test each others products for compliance with the ISO standard and compatibility in order to guarantee a high level of quality. It is planned to also offer test suites and compliance checks for products from other suppliers. This happens in the context of the Technical Working Group (TWG).

Events around the PDF/A Standard

In order to meet the high informational needs around PDF/A in the market, the PDF/A Competence Center organizes seminars and events in different locations on a regular basis.



For details about current activities, please check the Events page at pdfa.org on the Internet.

The Enterprise Content Management Association

AIIM is the international authority on Enterprise Content Management (ECM). ECM is the technologies used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and technologies provide solutions to help users with the four C's of business: Continuity, Collaboration, Compliance, and Costs.



For over 60 years, AIIM has been the leading non-profit organization focused on helping users to understand the challenges associated with managing documents, content, records, and business processes. Today, AIIM is international in scope, independent, implementation-focused, and, as the representative of the entire ECM industry – including users, suppliers, and the channel – acts as the industry's intermediary.

As a neutral and unbiased source of information, AIIM serves the needs of its members and the industry by providing educational opportunities, professional development, reference and knowledge resources, networking events, and industry advocacy. Information about AIIM can be found at www.aiim.org.

AIIM provides:

Market Education: AIIM provides unbiased information through its ECM Solutions Seminar (held throughout the U.S. and Canada); the Managing Information and Documents Road Show (held throughout the UK); InfoIreland (held in Dublin); AIIM Webinars; AIIM E-DOC Magazine

and M-iD (Managing Information and Documents Magazine) – the leading industry print publications in North America and the UK; and our online Solution Centers for financial services, healthcare, and state & local government.

Professional Development: AIIM's industry education road map offers business and government professionals a variety of training opportunities. Our ECM & ERM Certificate Programs provide instruction on the Why?, What?, and How? of Enterprise Content Management and Electronic Records Management via Web-based and/or classroom courses.

Peer Networking: Through chapters, networking groups, programs, partnerships, and the Web, AIIM creates opportunities that allow, users, suppliers, consultants, and the channel to engage and connect with one another.

Industry Advocacy: As an ANSI (American National Standards Institute) accredited standards development organization, AIIM acts as the voice of the ECM industry in key standards organizations, with the media, and with government decision makers. Our Industry Watch research reports provide intelligent information about user trends and perceptions. ■



PDF/A in a Nutshell – Long Term Archiving with PDF

PDF/A is *the* PDF for long-term archiving. PDF/A – which was adopted at the end of 2005 – is the first file format which, since it is an ISO standard, guarantees that documents created today will also be able to be opened and used in the future. ‘PDF/A in a Nutshell’ allows the user to take a look behind the scenes of the standard and provides practical instructions on generating PDF/A that conforms with the stipulations of the standard in his or her working environment. This book also serves as a comprehensive introduction to a subject matter that is still very new as well as providing practical examples for different software tools and industry solutions able to generate and work with PDF/A.

Extracts from the content of the book:

- Why PDF/A?
- The PDF/A-1a and PDF/A-1b conformity levels
- PDF/A with Acrobat 8 Professional
- Archive PDFs from Microsoft Office 2003 and 2007
- Scanning documents to create PDF/A and applying text recognition
- High-volume PDF/A creation
- Validating PDF/A
- Accessible PDF/A documents
- Future-proof contracts
- Forms in PDF/A
- Fonts and images in PDF/A
- Reliable colors on monitors and when printing

The authors:



Olaf Drümmer:

Olaf Drümmer is the co-author of ‘Postscript- und PDF-Bibel’, and has played a crucial role in the standardization of PDF/X (since 1999) and PDF/A (since 2002). He is a member of several international institutions and associations: DIN, ECI, Ghent PDF Workgroup, PDF/A Competence Center, and PDF/X-ready.

Olaf Drümmer is CEO of callas software GmbH. callas software develops the Preflight functions integrated in Acrobat since Version 6 (2003).



Alexandra Oettler:

Alexandra Oettler is a technical writer who has worked as a freelance journalist specializing in software for many years. She regularly has articles on DTP software in practice published in specialist prepress journals. She also writes user manuals for PDF and prepress programs and teaches software training courses. As the editor in chief of the pdfnews.de Web site, she provided German-speaking readers with daily information on new products, schedules, and tips between 2001 and 2004.



Dietrich von Seggern:

After completing his university studies in print technology, Dietrich von Seggern worked as a prepress manager. He worked on research projects related to the transmission of digital print data. Later, he became the manager of the digital advertisement transmission department at the marketing organization of the German newspaper publishers (ZMG). He has been working as the head of Product Management at callas software GmbH in Berlin for several years now.