



PDF for Archiving

Everything you need to know in
plain, simple English

by Dean Sappey

CONTENTS

Executive summary	3
What is PDF/A and how is it different from a non-PDF/A document?	4
How do I know if a document is PDF/A-compliant?	7
How do I create a PDF/A document?	8
Is there a way to make documents PDF/A ready, without actually converting them to PDF/A?	9
Should I implement these changes to existing PDF workflows?	9
Should I be converting every document to PDF/A now?	9
Summary	10
References	11

EXECUTIVE SUMMARY

Recently there have been several announcements by the US Federal Courts indicating that they are moving to adopt the PDF/A (Portable Document Format for Archiving) standard for electronic filing of case documents. In addition, government and industry bodies from various countries are mandating that electronic documents for archiving must be PDF/A compliant.

While most people understand the basic concept of the Portable Document Format (PDF) and why it is useful, it is not necessary to understand how the PDF document is created—nor should it.

IT departments and staff will need to familiarize themselves with the standard so that they can implement new workflows, and can support their users who will undoubtedly have questions.

However, the adoption of the PDF/A standard changes this “hands-off” approach to some degree. IT departments and staff will need to familiarize themselves with the standard so that they can implement new workflows, and can support their users who will undoubtedly have questions. The key then is to meet the PDF/A requirements in a way that gives users a general understanding without having to know all the technical “nuts and bolts” of the standard.

With that said, the aim of this whitepaper is to explain the reasoning behind the standard and how it is different from a “normal” PDF with as few acronyms and as little jargon and technical explanation as possible.

More specifically, this paper will answer the following frequently asked questions:

1. What is PDF/A, and how is it different from a non-PDF/A document?
2. How do I know if a document is PDF/A compliant?
3. How do I create a PDF/A document?
4. Is there a way to make documents PDF/A ready, without actually converting them to PDF/A?
5. Should I implement these changes to existing PDF workflows?
6. Should I be converting every document to PDF/A now?

The the aim of this whitepaper is to explain the reasoning behind the standard and how it is different from a “normal” PDF.

WHAT IS PDF/A AND HOW IS IT DIFFERENT FROM A NON-PDF/A DOCUMENT?

The Portable Document Format (PDF) was developed almost 20 years ago. It was intended as a way of allowing users to share documents—this wasn't possible previously. PDF allows users to share content with others irrespective of applications, fonts or system settings. The PDF can be read on screen or printed from any machine. The PDF document provides us with a number of other benefits:

Portability – PDFs are usually smaller than the original document, making it easier to email and to post on web sites.

PDF allows users to share content with others irrespective of applications, fonts or system settings. The PDF can be read on screen or printed from any machine.

Fidelity – a PDF document is an exact copy, image or replica of the original. This cannot be said of Microsoft Word documents. For example, sending a MS Word document to a colleague or external third party, who might have a different version of MS Word or different printer, can produce differences between the documents.

Reliability – you can open a PDF document on any system (Windows, Mac, Unix etc.) and it will display exactly the same on each machine as long as you have Adobe Reader.

Security – PDF documents can be secured in a number of ways. Users can apply an open password to the document.

The recipient needs to know the password in order to open the document. You can encrypt PDF documents to prevent users from editing, changing or manipulating the document. You can also apply digital signatures to the PDF to authenticate the document. A PDF application should also have redaction capability.

Like many software products in their early days, PDF had a clear and simple set of features and functionality. As the format has developed and been adopted over the years, PDFs have become much richer and much more complex. For example, a user can attach or embed non-PDF content such as multimedia, attachments (Word, Outlook, and Lotus Notes), links to other external documents and websites as well as JavaScript code, which performs special automated actions.

The end result is that PDFs may not be device or software independent. Adding native content to the PDF document requires the recipient in many cases to have the media player and software (Word, Outlook, and Lotus Notes) to view this content reliably. Often when a PDF is provided to a government agency or a Court, the PDF cannot be viewed as intended because the receiver doesn't have the appropriate software applications or versions installed.

Like many software products in their early days, PDF had a clear and simple set of features and functionality. As the format has developed and been adopted over the years, PDFs have become much richer and much more complex.

The International Organization for Standardization (ISO) agreed on a 'subset' of features that would be included in a PDF/A to ensure device and software independence.

This is an issue particularly for long-term document archiving. Government, corporate and legal archivists need certainty when storing documents that they can be viewed and printed reliably long into the future. It is essential that the document can be viewed regardless of the operating system, installed software applications and fonts. As long as a PDF reader is available, PDF documents from the past should still be readable.

Due to the increasing complexity and in some cases loss of document fidelity and reliability, the International Organization for Standardization (ISO) agreed on a 'subset' of features that would be included in a PDF/A to ensure device and software independence. This subset of features or standard is referred to as PDF/A - ISO-19005-1:2005.

According to the specification, the following terms are recommended when referring to the ISO 19005-1:2005 specification when the full ISO name is not being used:

PDF/A – a synonym for the ISO 19005 family of standards

PDF/A-1 – a synonym for ISO 19005-1

PDF/A-1a – a synonym for ISO 19005-1 Level A conformance

PDF/A-1b – a synonym for ISO 19005-1 Level B conformance

Put simply, PDF/A takes a subset of features from the PDF format to ensure the fidelity of the PDF document, and that there is no reliance on any other software components or applications being installed (other than a 'Reader' application), which could prevent viewing or printing of the document in the future. Whilst the PDF specification continues to change each time a new version of Adobe Acrobat is released, the ISO decided on a base line version that included the important features of the PDF. It chose version 1.4 of the PDF specification (this was released with Adobe Acrobat version 5) and these documents can be opened by Adobe Reader 5 and higher.

PDF/A documents are like any other PDF—they have the same windows file extension (.PDF), same icon and can be opened by any PDF application. Think of PDF/A as a normal PDF document but with certain restrictions. Once a PDF document passes a validation test, it can be marked as PDF/A compliant.

There are two levels of PDF/A conformance—PDF/A-1a, and PDF/A-1b. The "a" designation refers to those required "Accessible" attributes for users with disabilities, and the "b" designation refers to "Basic." For the US Courts and most government agencies either format is generally acceptable, but you should check to confirm requirements as this can vary from country to country. Certainly, the PDF/A-1b conformance level is the easier to create and is the most widely supported by software developers globally.

There are also two levels of PDF/A conformance—PDF/A-1a, and PDF/A-1b. The "a" designation refers to those required "Accessible" attributes for users with disabilities, and the "b" designation refers to "Basic."

KEY PDF/A REQUIREMENTS

This table presents a summary of PDF/A requirements:

FEATURE	RULE	COMMENTS
PDF version	Based on 1.4 version PDF (Acrobat 5 or higher)	Technically can be higher than 1.4 format as long as it doesn't include features and functionality not part of the 1.4 spec. But for simplicity, save as 1.4 format
Security	Not allowed	No encryption of any kind, no passwords
Multimedia	Embedded audio and video not allowed	No movies or sound files
JavaScript	Not allowed	No embedded JavaScript such as automated functions in PDF forms
Embedded fonts	All fonts must be embedded	Includes the 'base' fonts such as Arial, Times. Invisible text (as created by OCR software) does not need to be embedded
Attachments	Not Allowed	No attachments
Transparency and Layers	Not allowed	Must be one layer and no "transparent" objects
PDF Portfolios	Not allowed	Acrobat Pro Portfolios (Binders) not allowed
pdfDocs Binders	Allowed	A pdfDocs Binder can be converted to a single, PDF/A-compliant document

Most other technical requirements of the PDF/A standard are covered by your software application and provider, so are not highlighted here in the interests of not getting too technical.

FEATURE	PDF/A-1A	PDF/A-1B
Standard	Recommended for creating accessible PDF/A-compliant documents	Recommended for most archiving purposes
Text-searchable	Must be text-searchable (Image PDFs should be OCR'd)	Optional requirement
Unicode	Without Unicode, text cannot be read out loud by screen reader software. A precise and complete text search can only be ensured using Unicode. If Unicode is not used, content analysis is not possible.	Optional requirement
Tagged (Accessible)	Must be tagged	Optional requirement

HOW DO I KNOW IF A DOCUMENT IS PDF/A COMPLIANT?

For a document to be PDF/A compliant it must pass a test confirming that only allowable PDF/A features are contained in the document. When a software product creates a PDF/A document you can assume that as long as the software is of professional quality, the document will be PDF/A compliant. You should use software applications from reputable developers like pdfDocs Desktop or Acrobat Professional to create PDF/A-compliant documents.

An easy way to determine if a document meets PDF/A compliance is to simply open the document in Adobe Reader.

On the Standards pane you will see a small icon, which when clicked will indicate the status of the document, ie whether it is compliant or not.

Another way of making this even more obvious to users working with PDF/A documents in Adobe Reader is to select Edit Menu > Preferences > Documents and to ensure the View documents in PDF/A mode is selected.

This will display the blue status bar at the top of the document (which can be hidden by clicking on the blue 'i' icon).

TAGGED DOCUMENTS

These are PDF documents that have been 'tagged' with information indicating the order in which paragraphs and columns of text should be read aloud to the visually-impaired.

Although software products can make a 'best' guess as to these tags, human intervention is required to make sure the order in which the text is read out is the right order – this means manually checking and updating each paragraph.

In addition, you must include a text description for any graphics in the document so the image can be explained to the visually-impaired. This is of course very time consuming, hence it is not required generally for PDF/A compliance and is a good reason to not consider creating PDF/A-1a documents as your output standard.

Compliance is Not Guaranteed

The fact that this status of compliance is displayed should not be taken as a guarantee that the document is compliant. The reason being that any further editing, merging or changing of the document is likely to invalidate PDF/A compliance. The only way to guarantee that a document is PDF/A compliant after editing or modifying is to test it again. This can only be accomplished if your software provides this functionality.

DocsCorp will be releasing its own 'Validate PDF/A' option (available as a free enhancement for existing users of pdfDocs Desktop 3) to test external PDF documents for validation. Adobe provides this capability in its Acrobat Professional 6 and higher software (not Reader or Standard) using its Pre-flight engine.

Should I validate every time?

For PDF files created by professional software you trust and know, there is no need to validate. PDF files created by an external party need to be validated.

WHAT ARE EMBEDDED FONTS?

Installed on your computer are various font files. These contain mathematical formulas that describe how a specific font should display in size and style. When a PDF is created, you can choose (best practice) to embed the font. By embedding the font or a sub-set of the font in the PDF, the text will display correctly when the document is opened. If the font is not embedded, the PDF viewer will substitute the font with a graphical equivalent.

So, why doesn't this just happen every time. Some PDF products do not embed any fonts, or at least don't do this automatically. This means if you receive a PDF from outside your organization, it may not look as the author intended. Some software products, like Acrobat do not embed in their default conversion setting what is generally referred to as the 'base' fonts, ie those fonts that come pre-installed in every computer or laptop (Arial, Times, etc). There are 14 of them. It is recommended you modify your Acrobat "job option" setting to automatically embed all fonts for all documents.

HOW DO I CREATE A PDF/A DOCUMENT?

Creating a PDF/A document is only possible with software applications designed to do so. Typically they have the ability to save or output as PDF/A compliant. If the conformance level is not specified, you can assume it is PDF/A-1b.

If you are using pdfDocs Desktop, you can simply set the Organizer output options for Compliance to PDF/A. For Acrobat Professional 8 and higher, simply select 'Save as PDF/A.'

However, and this is a big HOWEVER, you may find that when converting existing PDF documents to PDF/A, some parts of your document may now look a little different. Here's why:

Fonts - you may find that some of the fonts have changed. If the fonts in the original PDF were not 'embedded' and you do not have those fonts installed on your computer, they will be substituted during the conversion process. This could produce changes to the document as the font change may result in overlapping paragraphs and/or incorrect formatting.

How can you tell if this is going to happen? Simply open up your PDF in Adobe Reader, select File Menu > Properties > Fonts. Check that the word "embedded" displays against each font used in the document. If the fonts are embedded, there should be no problems.

Missing Bits - objects may be removed from the PDF. If your document contained videos, audio, or embedded attachments for example, they will be removed from your document without warning. So, beware!

In particular, if you have PDF forms that have automatic functions that calculate fields, perform automated routines

such as moving to another page, inserting pages, etc, all this functionality provided by JavaScript will be lost when you convert the PDF document to PDF/A.

IS THERE A WAY TO MAKE DOCUMENTS PDF/A READY, WITHOUT ACTUALLY CONVERTING THEM TO PDF/A?

There are some simple steps you can implement so that any PDF document you create can be easily and reliability converted to PDF/A.

There are some simple steps you can implement so that any PDF document you create can be easily and reliably converted to PDF/A.

Set the document output format to 1.4 version—this limits the objects you can save in your PDF to those most closely matching the PDF/A specification so that you don't have a concern that items stored in your PDF will be removed when converted to PDF/A.

Embed all fonts - if your software does not provide this automatically for all fonts (including the base fonts), change the settings so that this does occur. If you are using pdfDocs Desktop, this setting is on by default. If using Acrobat you will need to change the default standard conversion setting.

SHOULD I IMPLEMENT THESE CHANGES TO EXISTING PDF WORKFLOWS?

A dilemma facing law firms in particular in the United States is when to start preparing documents to be e-filed with the Courts in the PDF/A standard. Should they wait until it is made mandatory by a particular Court, or should they start submitting it now in the PDF/A standard?

From the information already provided in this document you should now be clear that a PDF/A document is actually just a normal PDF document, but with feature restrictions. Therefore, you can start e-filing all your documents in PDF/A now as these will be valid. The Courts in the United States are accepting case documents in both PDF and PDF/A.

The advantage of setting up a PDF/A workflow now for e-filing is that you will have had time to tweak and fine-tune any new procedures so that you are ready whenever filing case documents in PDF/A with the Courts becomes mandatory.

Therefore, you can start e-filing all your documents in PDF/A now as these will be valid.

SHOULD I BE CONVERTING EVERY DOCUMENT TO PDF/A NOW?

There are two possible answers to this question. If PDF/A is required for archiving and compliance, then this is an easy answer—of course they must be PDF/A compliant.

If you are using a Document Management System and are storing documents for short-term use, (not for archiving), then PDF/A is not for you. Just convert the documents to PDF/A when you need to file them with a Court or provide to a government agency. Why is this? – mainly because you will lose some PDF functionality, particularly if you are storing PDF forms in your Document Management System.

SUMMARY

The PDF/A standard is gaining momentum and is being widely adopted by government and industry bodies globally. 2011 will be the year that law firms in the United States must file case documents electronically with the Courts in PDF/A.

PDF/A is simply a subset of the PDF 1.4 specification (Acrobat 5), which excludes certain elements from the document as well as requiring other elements such as fonts, color profile and metadata to be embedded in the PDF to ensure that the document can be viewed and printed in the distant future and will remain true to the original.

Even if you are not required to create PDF/A currently, it is a good idea to at least make them PDF/A ready. This will ensure that you will be able to convert them to PDF/A reliably when required. Equally, it is good practice to create this type of PDF document as it results in smaller files, ensures document fidelity and can be reliably viewed by your recipients.

So that's it – now you know as much as you really need to know about PDF/A documents without being too technical.

REFERENCES

pdfA Competence Center: www.pdfa.org

DocsCorp: www.docscorp.com

Adobe: www.adobe.com

Wikipedia: <http://en.wikipedia.org/wiki/PDF/A>

ABOUT DEAN SAPPEY

B. App. Sc (Comp Sc)

Dean Sappey is President and co-founder of DocsCorp. After selling his first legal software business in 2000 to Solution 6 (ASX), he co-founded DocsCorp to develop software products for document management professionals. Dean combines his technical background (having a degree in computer science) and PDF technology in particular with extensive experience in building and delivering software solutions that enhance productivity in law firms, accounting firms, government, and corporations.

ABOUT DOCSCORP

DocsCorp provides a range of software solutions to document management professionals that enable them to share critical information in a format that is safe, secure, complies with industry standards and is universally readable. DocsCorp software integrates with Document Management, CRM and Practice Management systems to enhance business efficiency, productivity and workflow. DocsCorp has more than 3,500 clients worldwide, including professional service firms, Government departments and corporate enterprises. DocsCorp has offices located throughout the world, with teams of software professionals based in Australia, the United States and the United Kingdom. DocsCorp is a member of the PDF/A Competence Center. www.pdfa.org

SYDNEY
MELBOURNE
LONDON
NEW YORK
WASHINGTON
CHICAGO
PITTSBURGH
PORTLAND

info@docscorp.com
www.docscorp.com