
PDF Standards FAQ

Author **David van Driessche**

Executive Director

Co-chair Specification and Process
Control subcommittees

david.van.driessche@fourpees.com

Date January 14, 2022



Ghent
Workgroup

www.gwg.org

Table of Contents

Introduction	3
Is PDF itself a standard?.....	3
Why do we need other standards around PDF then?	3
What's the difference between PDF standards and PDF specifications?.....	4
Who invented PDF standards?	4
How many ISO-developed PDF standards are there?.....	4
What is PDF/X?	5
Can I still have problems when using PDF/X?	5
Why are there different versions of the PDF/X standard?.....	5
What should I know about PDF/X-1a?.....	6
What should I know about PDF/X-4?	6
What should I know about PDF/X-6?	6
Can a PDF be compliant to PDF/X-1a <i>and</i> PDF/X-4?	7
Can a PDF be converted from PDF/X-1a to PDF/X-4 (or vice-versa)?	7
What version of PDF/X should I be using?	7
What is the Ghent Workgroup (GWG)?	8
What's the difference between PDF/X and the GWG specifications?	8
What are PDF/X-plus specifications?.....	8
Should I use PDF/X or the GWG specifications?.....	8
What is PDF/A?	9
What are the different versions of PDF/A?	9
What are the different conformance levels of PDF/A?	9
What is PDF/UA?	10

Introduction

If you work with PDF documents, chances are that sooner or later you're going to come in contact with PDF standards. You might knowingly save a PDF conforming to this or that standard, but you might also use standards without even being aware of it. Why? Because many of the vendors who create software using PDF implement standards for you: saving a PDF file for a specific use from that software likely uses some standard under the hood.

It is, of course, always better to be aware of what you are doing. Using PDF standards correctly is important, and knowledge about the different standards and what they can and will not accomplish for you can make all the difference in getting the result you're looking for.

This document was written to clarify some of the language you might encounter, demystify some of the curious statements you might hear and provide pointers to help you navigate the wonderful world of PDF standards.

Is PDF itself a standard?

PDF (which stands for Portable Document Format) was invented by Adobe, and at that time it was not a standard. Meanwhile, PDF is officially owned by the International Standards Organization (ISO) and is a proper international standard. Its number is 32000 and there are (as of when this document was written) two versions: 32000-1 which is used almost everywhere and 32000-2 as an upcoming format.

So yes, PDF (today) officially is a standard.

Why do we need other standards around PDF then?

While PDF is standardized, that doesn't mean all of its features are appropriate for use in all situations. From the moment it was invented, PDF was a very broad format, and that has not changed over the years. The official description of the PDF standard from ISO is well over 1000 pages; many software and hardware implementations support only parts of the functionality that is possible.

The fundamental goal of all PDF based standards is to remove uncertainty. To accomplish this, all PDF standards follow essentially the same principle:

- They forbid the use of certain PDF capabilities or parts thereof,
- They impose the use of other capabilities, in strictly proscribed ways.

If you are mathematically inclined, this means that PDF standards are subsets of PDF, aimed to be used in specific, controlled, circumstances.

What's the difference between PDF standards and PDF specifications?

You're quite right to notice that both terms are used. For all practical purposes, there is no real difference. When trying to be strict (some would say pedantic), a standard implies it's been created by an accredited standards organization, while a specification is created by "some other organization".

Generally, we talk about ISO-developed PDF standards for example, but about PDF Specifications from the Ghent Workgroup. Realistically both try to increase standardization and reliability of PDF documents, though the ISO is more theoretical and industry-driven, while the Ghent Workgroup is more practical and customer-driven.

Feel free to mix both terms to your heart's desire and use what you find more logical.

Who invented PDF standards?

Historically PDF standards were first established as the result of the graphic arts industry adopting the (then quite new) PDF format, primarily for use in the exchange of advertisements and for commercial print jobs around the turn of the century. This resulted in two separate developments around 2001:

In the US, a number of organizations amongst which DDAP (Digital Distribution of Advertisements for Publishing) developed a PDF standard specifically for ads. This was later adopted by ISO and became the PDF/X-1a standard.

In Europe, a number of industry associations ran into the same problems around the use of PDF for advertisements and in commercial print and ended up forming the Ghent Workgroup (GWG). This resulted in the development of PDF specifications for those market segments.

How many ISO-developed PDF standards are there?

When this document was written, the following standards existed:

- PDF/X
- PDF/A
- PDF/UA
- PDF/E
- PDF/VT
- PDF/R
- PDF/VCR

That's a grand total of 7 ISO standards based on PDF. There are some other initiatives that could cause confusion because they are named similarly (such as PDF/H), but are not in fact (at this point) ISO standards. And in case you're wondering, the PDF/VCR standard *is* a real ISO standard (and no, there is no PDF/Betamax standard).

What is PDF/X?

PDF/X is an ISO standard (ISO 15930) that governs the use of PDF in the graphic arts market segment. As all PDF-based standards, it defines a “safe” subset of PDF, a definition of which features from the very rich PDF format can safely be used without causing uncertainty or confusion.

The “X” comes from “Blind eXchange”. It’s helpful to know this as the concept of blind exchange is at the heart of the standard. The thought behind PDF/X is that, if I am a creative person and you are a printer, I should be able to create a PDF/X file, send it to you without further explanation and you should be able to print it exactly as it was intended by me.

When exchanging regular PDF files, this proved to be very difficult as PDF was (is) a very broad format. The PDF/X standard forbids all PDF features (or parts of features) that would make this concept of blind exchange impossible and imposes the use of other features to increase reliability.

Can I still have problems when using PDF/X?

Yes, you can. There are two main reasons you should be aware of.

First of all, PDF/X reduces the complexity of PDF by limiting which PDF features are allowed to be used. That doesn’t take away the fact that even PDF/X is still technically complex. As no software is absolutely perfect, and dealing with complex file formats is difficult, there can – at times – be things that go wrong.

But perhaps more important and certainly more fundamental: PDF/X is an ISO standard for the complete graphic arts industry. And if you think about it, the graphic arts industry is amazingly broad: the technical requirements to print a business card for example, are very different from the requirements to print a billboard. As a result, PDF/X has to remain vague about some aspects. You won’t find ink limits in the PDF/X standard, nor minimum point sizes for text, nor minimum and maximum resolutions for images...

Doesn’t that mean PDF/X is essentially useless? Nothing could be further from the truth. While it’s not perfect, it’s definitely much better than using plain PDF. Think of it as a sliding scale; if you have to exchange native application files, that’s not so good. If you can use PDF instead, that’s much better already. If you can exchange PDF/X, that’s much better again.

Why are there different versions of the PDF/X standard?

Life would be simpler if there was just PDF/X, and that was it. Unfortunately, that’s not how technology works... instead the software and the hardware we use evolve over time, as do business practices and methods. The PDF/X standard was originally invented in 2001 and different flavors of the standard were created for different (good and less good) reasons since then. By the way, someone with knowledge of the process around ISO standards might refer to different “parts” of the PDF/X standard, rather than flavors or versions.

Are all of those flavors equally important? While it’s hard to come up with a truly objective answer to this, you would do well to remember these flavors:

- PDF/X-1a
- PDF/X-4
- PDF/X-6

What's with the gaps in numbering? Let's keep it at the fact that history is a harsh teacher and some things sound like a better idea than they really are in hindsight. Also, while these are the most important flavors, it's not impossible you would run into another flavor in specific workflows.

What should I know about PDF/X-1a?

PDF/X-1a is the flavor of PDF/X that was used first by large groups of people and it is still the most used PDF/X flavor throughout the world, even though it has not aged all that gracefully in some areas. Invented in 2001, PDF/X restricts PDF documents to a very limited subset of PDF features. Most notably, a PDF/X-1a document:

- Can only contain gray, CMYK and spot colors. All other color spaces are disallowed.
- Can feature only one ICC profile, which has to be a CMYK ICC profile, in its output intent.
- Cannot use live transparency, which results in PDF/X-1a files having to be created fully without transparency effects or having to be transparency flattened.
- Cannot use a host of other modern PDF features such as Optional Content Groups (aka layers), modern types of image compression, 16-bit images, OpenType fonts...

A short, and subjective, summary could be: PDF/X-1a is a very reliable format, but due to its limitations it can at times cause as much problems as it resolves in modern workflows. Many experts advise using a more modern PDF/X flavor instead.

What should I know about PDF/X-4?

PDF/X-4 is more modern as a standard than PDF/X-1a: the first version of it was released in 2008. It also supports many more modern PDF features than the earlier standard. So in this flavor of PDF/X:

- Live transparency *is* supported (PDF files do not have to be transparency flattened).
- Optional content groups (layers) *are* supported.
- More modern image compression, 16-bit images, OpenType fonts... *are* supported.

The biggest change however is support for color-managed content. Where in PDF/X-1a all content essentially has to be CMYK (gray can be seen as the black separation in CMYK, and spot colors need to have a CMYK alternate color space), in PDF/X-4 you have the possibility to have other color spaces, as long as they are color-managed. This means that you may have Lab images in a PDF/X-4 file, but just as well RGB content (as long as that content has an RGB ICC profile attached to it so it can be properly color managed).

These changes make PDF/X-4 more versatile and remove a number of potential issues (mostly related to transparency flattening) that inflict PDF/X-1a files. They are however more complex than PDF/X-1a files (or at least have the potential to be).

What should I know about PDF/X-6?

PDF/X-6 is a much more modern flavor of PDF/X that is intended to replace PDF/X-4 (and PDF/X-1a) in due time. Late 2020 this was not yet an official ISO standard (it was still under development). It was expected to be published by ISO soon though.

The major difference in PDF/X-6 is that it is built on top of a new version of the PDF standard: ISO 32000-2, and this allows PDF/X-6 to use a number of major new features from this new PDF version:

- The use of Black Point Compensation in PDF documents.
- Page level output intents.
- The ability to store better ink information (mixing hints, spectral data) in a PDF document.
- DPart information to attach metadata to page or ranges of pages in a PDF document.

As this is at the time of writing a “future” standard, which has no to little support from existing software and hardware, the expectation is that it will take a while before PDF/X-6 becomes an important player in PDF/X land.

Can a PDF be compliant to PDF/X-1a and PDF/X-4?

The short answer is no. That is not possible. The longer answer is that it’s *almost* possible.

You can create a PDF that adheres to almost all of the rules for both PDF/X-1a and PDF/X-4. But... PDF/X contains a rule that declares that the PDF must contain an identification of the PDF/X version it claims to be. And here, it isn’t possible to have a file that claims to be either of those two versions, it must choose exactly one PDF/X version. So while the content of the file might perfectly well be in correspondence with both PDF/X versions, technically it’ll have to choose one.

Can a PDF be converted from PDF/X-1a to PDF/X-4 (or vice-versa)?

Given that PDF/X-1a is a much more limited standard than PDF/X-4, converting from PDF/X-1a to PDF/X-4 is typically very easy. In many cases it likely only involves changing the metadata that contains the PDF/X version identification.

Going from PDF/X-4 to PDF/X-1a should typically be possible as well, as PDF/X-4 files are (by design) well-behaved PDF files. However, in this case there might be a lot more work, because PDF/X-4 is so much broader than PDF/X-1a. And some of this conversion work is not trivial; your PDF document would have to be transparency flattened for example, and there might be color conversions from non-CMYK content. Possible, not trivial.

What version of PDF/X should I be using?

This is a question on which opinionated answers are the only possibility. Still, there seems to be growing consensus that it is better to use PDF/X-4 if there are no good reasons why you “are stuck with” PDF/X-1a.

From a theoretical point of view, PDF/X-4 is clearly better. From a practical point of view, one also has to consider the different parties involved in a print workflow, and their equipment both on the hardware and software side. Depending on the workflow, there might also be business issues (such as who takes responsibility for what) that need to be taken into account.

What is the Ghent Workgroup (GWG)?

The Ghent Workgroup (GWG) is a standardization group formed by software and hardware vendors, end-user associations, industry experts and consultants, printers and publishers and educational experts. They are primarily from Europe and North America and have worked on best practices for graphic arts workflows since 2002.

The group has two main focus points:

- The creation of standards, application settings, test suites... Deliverables that make the life easier for people working and breathing graphic arts.
- The dissemination of all of the above materials and the knowledge surrounding them through white papers, webinars, video recordings, seminars...

What's the difference between PDF/X and the GWG specifications?

ISO creates a standard for the entire graphic arts industry with PDF/X; the GWG builds on that PDF/X standard to create their own specifications. That means the main difference between the two is that PDF/X by design is a lot more vague (as it addresses the whole graphic arts market segment), while the Ghent Workgroup creates specifications based on PDF/X for specific market segments inside the graphic arts market.

What are PDF/X-plus specifications?

The term is used to refer to the Ghent Workgroup specifications. Vicki Blake, then Executive Director of the Ghent Workgroup, coined the term to refer to the practice where the Ghent Workgroup specifications *add* additional restrictions to the PDF/X standard. In other words, the Ghent Workgroup specifications contain all restrictions from PDF/X *plus* additional requirements.

Should I use PDF/X or the GWG specifications?

To be completely correct, preferably neither... But hold on, let me explain that!

Preferably, if at all possible, you should talk to the next person in the workflow. If you're creating content that needs to then get printed, you should, if possible, talk to the printer you're working with. Find out what standards they use. How *they* would like to receive PDF documents from you, what preflight profiles they are working with. That is your number one choice.

If that is not a possibility, look at the GWG specifications and select one that is suitable for the type of work you are creating. The GWG has application settings and preflight profiles you can use for popular software applications. Use those (for free, from their site: www.gwg.org).

If for whatever reason you cannot use GWG specifications, try to use PDF/X. If that too is not possible (you'll need a very good reason at this point for that to be true), use plain PDF.

Your risk on problems goes up with every step essentially. Ghent Workgroup compliant files are more reliable than PDF/X, which is more reliable than regular PDF. Following the standards provided by your printer is better than everything else.

What is PDF/A?

PDF/A is an ISO standard (ISO 19005) that governs the use of PDF for long-term archival. The stated goal for the standard is to provide guidelines for the creation of PDF documents so that they can be archived and are still usable after a long period of time (50 years up to infinity). What "are still usable" means, depends on the flavor of the PDF/A standard used. PDF/A is a bit of a complex standard in the sense that it comes in different versions, where each version has different "conformance levels".

What are the different versions of PDF/A?

At the time of writing, there are three different PDF/A versions, with the following characteristics:

- PDF/A-1 is the oldest version of the standard, released in 2005, and built on top of a quite old version of PDF. As a result it doesn't support many newer PDF features common today.
- PDF/A-2 was released in 2011 and came with two major changes: it's built on a newer PDF version and so supports many more modern PDF features, and it added the ability to embed files into a PDF/A document (so the PDF/A document can be used as an archive). Any embedded files have to be PDF/A compliant themselves though.
- PDF/A-3 was released in 2012 and is nearly identical to PDF/A-2. However, the files that can be embedded in a PDF/A-3 do not themselves have to be compliant to the PDF/A standard. This opens the way for PDF/A files containing an XML payload inside for example.

What are the different conformance levels of PDF/A?

That differs depending on the version of PDF/A you are using, but there are in total three different conformance levels:

- **"b" or "basic"**, as in PDF/A-1b, PDF/A-2b and PDF/A-3b
The lowest conformance level is only interested in visual representation. If I create a compliant file today, the goal is that I will still be able to reproduce it visually (for example print it), say 50 years from now. This is the easiest conformance level to obtain and 99.99% of input files can be converted to it automatically.
- **"a" or "advanced" / "accessible"**, as in PDF/A-1a, PDF/A-2a and PDF/A-3a
The highest conformance level is interested both in visual representation and in meaning. So the goal is not only that a file will be able to be reproduced visually, but also that the meaning of the content in the file will still be there. I'll be able to search it, extract text from it, understand what the graphical elements (such as images) actually are... While having files with this conformance level is very interesting, it is often very hard to create correct files. Images have to be tagged, file structure has to be present and those things are very difficult to obtain automatically if the input file isn't already quite good.
- **"u" or "Unicode"**, as in PDF/A-2u and PDF/A-3u
Because of the difficulty in obtaining the "a" conformance level, an intermediate level was invented. At this conformance level text still has to be in the PDF file in the correct way so that the file can later be searched or that text can be copied out but tagging of file structure and images isn't required. This conformance level is moderately easy to obtain if reasonably proper input files are used.

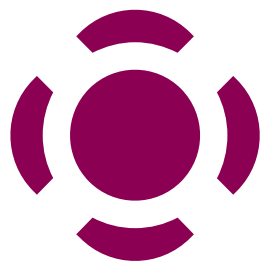
What is PDF/UA?

PDF/UA is an ISO standard (ISO 14289) that governs the use of PDF for accessibility purposes; the "UA" suffix comes from "Universal Accessibility". The goal of the standard was to ensure PDF documents can easily and correctly be used by people with disabilities who use assistive devices and software. Some of the most important requirements for a PDF/UA file are:

- All text in the file must be properly encoded and it must be tagged with a proper language tag. The purpose of this is for assistive technology to be able to read the text aloud using the proper language / voice.
- All content in the file must be properly structured / tagged. Again this is so assistive technology knows the proper reading-order and can use logical file structure (headings, paragraphs, tables...) to guide a user through the document quickly and reliably.
- For the same reasons, all non-textual elements must either be clearly marked as an artefact (decoration or similar, not important for the understanding of the document) or they have to be identified and tagged with alternative text (an explanation of the graphical element for those who cannot see it).

The difficulty of the PDF/UA standard is that it is very difficult for software to create automatically and even to verify for compliance automatically.

This is easily illustrated with a few examples. It is easy to check whether text is tagged with a language, but it is incredibly hard for a computer to make sure that language is the correct one. It is easy to check whether an image has an alternative text associated with it; but very difficult to automatically check whether that alternative text makes sense for that particular image.



Ghent
Workgroup
