

# Certified PDF

## A Smart Digital Master

David van Driessche \*

Keywords: Standards, Workflow, PDF

**Abstract:** This paper introduces the Certified PDF technology designed to make quality control of PDF documents throughout a PDF workflow more manageable and reliable. Files can be created correctly from the start (rather than possibly fixed later), kept consistently correct throughout the workflow, unnecessary preflighting is avoided, liability issues around corrections are solved, and a full document history is available at any time. These benefits are accomplished through the use of Certified PDF documents; smart PDF documents which remember their own preflight status and editing history.

Certified PDF has currently been accepted as a compulsory national standard for digital document interchange in a large part of the publishing industry in Belgium, the Netherlands and France. In those countries it replaces traditional exchange of native application files, PostScript and plain PDF files and is set next to TIFF/IT files in terms of press-readiness and reliability.

### THE PLAYING FIELD

Wherever you currently are, take a second to look around you and realize for once how much of your environment is actually the result of some publishing workflow. From the colorful sheets you sleep on, to the wallpaper you are so used to in your living room. From the package of your favorite breakfast cereal to the ads on the bus you pass on your way to work. From the security badge you have to wear to the annoying html email about free university diplomas you receive...

---

\* Director of Software Development, Enfocus Software

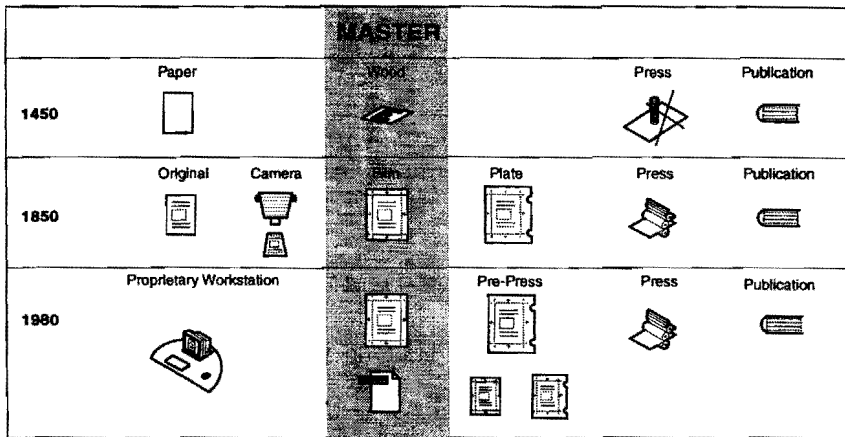
Not all of these are the result of a complex publishing workflow, but a surprisingly high number of the things we are so used to in our daily life actually are. They have been created by someone, revised, edited, sent off, received, pieced together, printed... the list of variations of such workflows is endless.

The challenge that the graphic arts industry has been struggling with since the beginning of times is how to control that process. How to make sure that the final product equals the vision of the original creator. How to get this design from the creator to the final result with a maximum of ease, reliability and flexibility.

### DIFFERENT APPROACHES

The first master used to convert a designed page to a printed piece consisted of pieces of lead, shaped as individual letters. It's a good thing for everyone in the publishing industry that we actually managed to find a lighter physical master over time... That lighter physical master, which is still used today in a lot of contemporary workflows, is film.

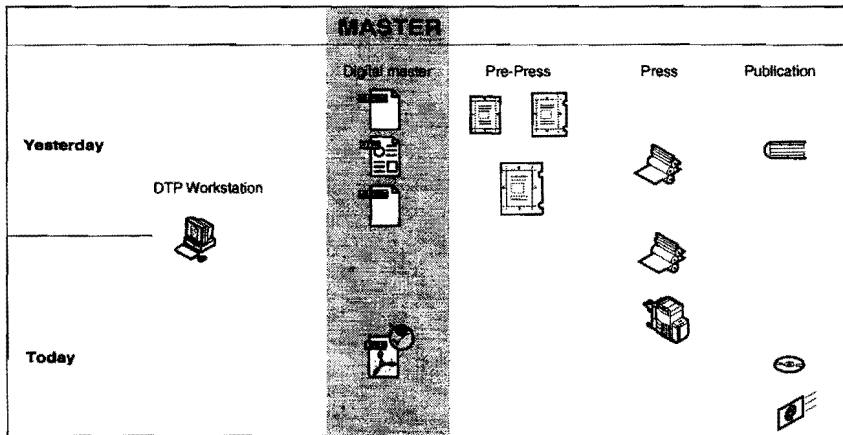
Using film to carry the design of a page from creator to publisher proved to be a reliable solution and something that worked for quite a while. It does however also carry substantial disadvantages. Moving pieces of film around is time-consuming and expensive. Storing them is bulky and the material decays over time. On top of that they offer very limited editing capabilities.



For all of those reasons the focus shifted to finding a digital master as computers emerged in our world. While film is still used at a certain point in the workflow, the master used to transfer before that stage is a digital file.

Initially the systems to work with digital masters are expensive, high-end systems that require state-of-the-art computers and use proprietary digital masters. Most of these have one thing in common; their digital masters are raster based files such as copydot and TIFF/IT-P1 files.

The advantage of using a raster based digital master is that it is very, very reliable. The downside is that these masters are very big and thus difficult to transfer. In most cases they have to be transferred on physical media (CD-ROM for example) because they are too big to transfer digitally (even with all the advances we see today in network capacity and price it is often still too time-consuming to transfer these files over a network). And on top of that flexibility (both towards editing and repurposing) in a workflow based on raster based masters is very limited.



The publishing world changed forever with the advent of inexpensive 'personal' computers. These mainstream computers became part of the publishing workflow but did not have the processing power required to work with the huge raster based files needed in publishing.

A digital master that needs to work in such an environment *has* to be object based. Different objects on a page are stored as what they are; paragraphs of text using specific fonts, images, illustrations composed of lines in different colors... No longer are they converted to dots in the digital master. This reduces file size and makes the file usable on a less powerful computer.

As an added advantage using an object based file format increases flexibility. If information about paragraphs of text is retained, it is much easier to change that text at a later stage in the workflow. If high level information about an illustration and the shapes and colors used in it is remembered, it is much easier to change one of those colors later on...

There are downsides to this approach of course... Object based files are much more complex than raster based files and as such have a much higher chance to produce errors when they are ripped (converted from the original file to the dots that are actually printed) to film or direct to plate or press. And even though these workflows tend to be a lot more open than the traditional proprietary systems all sorts of compatibility and exchange issues still commonly appear...

Examples of this approach are workflows where native application files (such as Adobe InDesign or Quark XPress files) are transferred. Or workflows based on PostScript or PDF files.

## PDF AS A SOLUTION

Even though PDF is mentioned before together with native application file workflows or PostScript workflows, it actually stands apart in many ways.

It is truly an object based file format; text remains text, illustrations and images are stored as such. But going further than that, information such as fonts, color spaces and high-end prepress information (such as transfer curves and halftone definitions) is stored as separate objects too. All those objects are randomly accessible, which is quite different from a PostScript file where the file has to be read from beginning to end and information about page 25 for example is only available after reading the 24 preceding pages...

PDF is standardized by Adobe Systems and as everyone uses the same file definition much of the file interchange and version problems are avoided. The format was conceived from the very start as a cross-platform format and is thus excellently placed for use in a hybrid workflow.

It is a reliable format, converting a PDF file into dots for use on a press is a much easier task than for example converting the equivalent PostScript file into dots on the same press...

But with all that PDF is still a flexible solution that is suited for much more than a pure prepress workflow. In the very file format itself are provisions that allow it to be used perfectly for CD-Rom publishing, eBook publishing or Internet use, just to name a few alternatives... PDF lends itself perfectly for what is now known as cross-media workflows (usage of the same source material towards

different publishing workflows such as prepress and Internet) and allows easy repurposing of material. A perfect example of such “alternative” use of PDF is the IRS in the United States that makes all of its tax-forms electronically available as PDF files on their web site.

As a result of this, workflow systems have been migrating to be based on PDF rather than on a raster based format for the last couple of years. More than that, having a PDF based solution more and more became the equivalent of having a modern, flexible though reliable, future-oriented solution.

## NO ROSES WITHOUT THORNS

While it is undoubtedly true that a PDF based solution offers many advantages and that the publishing industry has acknowledged that PDF is the way to go, it is still important to realize that a system based on PDF leaves much of the fundamental problems in a workflow unanswered.

- From its very roots the PDF file format was conceived as having a much wider focus than just prepress. While this makes the file format ideal to be used in modern cross-media workflows, it also makes for a lot of possibility for error when using PDF files coming from various sources. Creating a PDF that is ready for a prepress workflow is not a trivial task.

*PDF is too flexible to be a true Digital Master.*

- Due to the flexibility of PDF, there is a need to *preflight* a file before committing it to film. Preflighting is performing quality control; making sure the file conforms to the set of specifications the printer knows his files need to have if he is to output them with good results.

In those modern workflows where physical film is avoided altogether, because of computer to plate or computer to press technologies, that quality check becomes even more important. In these workflows the printer no longer has the comfort of examining a piece of film with a magnifying glass before he commits to making an expensive plate or – even worse – doing a press run...

And preflight is not only needed before printing the PDF file, it becomes more and more important to move preflight upstream to earlier stages in the workflow. There are two trends in the publishing industry that make this necessary.

First of all deadlines become shorter and shorter and it becomes crucial not to waste time by transferring incomplete or incorrect PDF files.

And secondly, document creators and designers are increasingly interested in the quality of the final printed piece. They now have desktop tools to create good quality files and they want to make sure the PDF file they hand over is good.

*Workflows need constant quality control.*

- Even if the tools for good PDF creation and preflight are available, in most workflows one still relies on people to do the actual quality control at different stages in the workflow. One trusts that everyone will do “the right thing”. In a complex workflow with a flexible file format as PDF that turns out not to be sufficient; it is dangerous to trust that everyone in a workflow will consistently do the right thing without ever making a mistake.

*People in a workflow are not always reliable.*

- Using PDF files as Digital Master introduces what is seen as a huge benefit in many workflows; the possibility to correct problems until the last minute. Even though that is indeed a big advantage of a PDF solution it also introduces a liability problem. What if many thousands of copies of an ad are printed with a mistake in them? Who can prove where that error is introduced in the workflow if you use a flexible file format as PDF?

*Being able to edit a PDF introduces liability concerns.*

## PDF/X AS A SOLUTION?

In an effort to take away the extraneous flexibility of regular PDF documents, the Committee for Graphic Arts Technical Standards (CGATS) developed the PDF/X standards. A family of file specifications based on PDF and restricting the feature set one is allowed to use. The restrictions are targeted specifically to the use of PDF/X in prepress for the exchange of digital advertisements in different workflows.

Two of those standards, PDF/X-1:2001 and PDF/X-1a:2001 have been proposed and accepted as official ISO standards, the other members of the PDF/X family are well on their way to become ISO standards.

Introducing PDF/X in a workflow is aimed specifically at taking away the flexibility concern. Limiting the features that one is allowed to use in a PDF file also limits reliability concerns; less features means less chance for problems. And that is undoubtedly a step in the right direction.

But even if one looks only at the flexibility problem (which is only one of the fundamental workflow problems we identified), PDF/X is not a complete solution. The reason for this is that PDF/X is a general standard for one part of the industry and as such it is both too wide and too narrow in scope.

It is too wide because it attempts to target everyone involved in transferring digital ads. Specific workflows or market segments have stricter rules which are not covered by PDF/X; a magazine publisher for example will have higher expectations in terms of image quality for ads than a newspaper publisher. Even though both are helped by the PDF/X standard because it at least gives them a higher chance of success, they will need quality checks that go beyond PDF/X.

And at the same time it is too narrow because it focuses only on digital ad delivery, which leaves a lot of the traditional prepress industry and even more of the cross-media publishing industry unaccounted for. These industries generally need more or other limitations to the feature set used than what is defined in PDF/X...

The conclusion from all this is twofold:

- 1) If you can use PDF/X in your workflow it is a very good idea to do so,
- 2) PDF/X is not *the* general workflow solution; it offers only a partial solution to just one of the many fundamental workflow problems.

## INTRODUCING CERTIFIED PDF

To overcome not just the flexibility problem but all fundamental workflow problems described before, Enfocus Software developed a technology called Certified PDF.

A Certified PDF workflow is a workflow in which the files that are transferred are Certified PDF files. These Certified PDF files are regular PDF files in which extra information is stored which allows everyone in the workflow to know:

- what the preflight rules are for this particular document
- whether the document adheres to these rules
- how previous versions of the document looked
- who edited the document, and when
- what exactly was edited in the document

In a modern and open workflow it is important that these Certified PDF files can be used by anyone in (or outside) the workflow as regular PDF files; the extra information stored in the files is invisible for anyone who is not aware of it. That means that a Certified PDF can be viewed using Adobe Acrobat, edited,

transferred, printed... as any other PDF file. That also ensures that Certified PDF is perfectly compatible with other PDF oriented standards, such as PDF/X and – to give just one other example – PDF based job-ticketing (PJTF or JDF).

Perhaps even more important than storing the Certified PDF information unobtrusively, is making sure it is tamper-proof... The Certified PDF information is recorded in such a way that workflow tools know a file has been changed, even if that happens outside of the Certified PDF workflow or with a tool that does not know about Certified PDF at all.

## ENSURING QUALITY THROUGHOUT THE WORKFLOW

The number one concern throughout a modern workflow is quality control, from the moment the creator sends off his design to the moment it is published. This is accomplished in a Certified PDF workflow by embedding a *PDF Profile* inside the Certified PDF file.

A PDF Profile is the collection of preflight rules this particular file has to conform to. Examples of preflight rules are the minimum or maximum resolution for images, the color spaces that can be used, whether fonts should be embedded or not and which fonts can be used...

When a creator prepares a Certified PDF file, he picks the appropriate PDF Profile for that file. This PDF Profile can come from a variety of sources; from a standards organization which has developed standard PDF Profiles for a specific workflow, from the Enfocus web site which contains a library with more generally suitable profiles (such as a PDF/X-1a:2001 profile for example) or – ideally – from the publisher who will publish that particular Certified PDF file.

Compare this concept of a PDF Profile with that of a standard as PDF/X for a second. In essence both do the same thing, make sure that certain features that are allowed in regular PDF files are not used in a specific workflow. The difference between the two is that a PDF Profile can check for much more problems than what PDF/X checks for. And at the same time a PDF Profile can be customized for a specific workflow, where PDF/X is linked to 'just' digital ad transfer.

Once the PDF Profile is selected, it is embedded in the Certified PDF file. At each point further down in the workflow, the embedded PDF Profile is then compared to the PDF Profiles available on the system the Certified PDF file is opened on. The reason for this is to ensure that nor the creator nor anyone else in the workflow tried to tamper with the PDF Profile or is using an older version of the PDF Profile.



Each time a Certified PDF file is edited it is not just saved but preflighted first. The reason for this is that any change to the file – no matter how innocent it looks – can invalidate it.

To give just one example, assume a creator creates a perfect PDF/X-1a file and saves it. At the last minute, just before sending it off, he decides to add a note that asks his publisher to make sure the correct spot color is used for his logo... That small change invalidates the file as a PDF/X-1a file as the PDF/X-1a standard does not allow the use of annotations on the printable surface of a document.

By providing tools that enforce – or at least promote – preflighting of a Certified PDF file before it is saved, one overcomes human error in a workflow. The importance of embedding the PDF Profile for a Certified PDF file inside the file becomes clearer here as well. Not only does it make sure one can check whether the correct PDF Profile was used, but it is now also possible to preflight the file whenever necessary without having to know which PDF Profile to use.

But let's assume the worst and assume that one party in a workflow neglects to follow the prompts to preflight the document when he or she should... Even in that scenario real problems are avoided as a Certified PDF file not only has its PDF Profile embedded, but also stores its own preflighting status. That means that at any point in the workflow, just opening the Certified PDF file, without doing any quality control makes it clear whether the Certified PDF file:

- has a valid and correct PDF Profile embedded
- has been preflighted with that PDF Profile and has not been edited since that preflight
- has been preflighted without successfully (i.e. without errors during preflight)

A creator who is trying to deliver good files to the next person in the workflow also gets an easier job. All he is required to do is pick the correct PDF Profile and he instantly gets a green or red light. A green light indicates a good quality file that can be sent off, a red light indicates a file that has problems which need to be corrected first.

A receiver has an equally easy job; as a Certified PDF file stores its own status a new preflight is only necessary in case the creator would have used the wrong PDF Profile. In all other cases it is immediately obvious whether a file is correct or not.

The procedure of associating a PDF Profile with a specific Certified PDF file and preflighting the Certified PDF file whenever needed (and only then!)

eliminates the quality concerns that exist in a regular PDF workflow. It also reduces the human factor to just one decision; the creator needs to ensure he or she selects the correct PDF Profile when preparing the Certified PDF file. But if they do not, that is spotted as soon as the file is transferred to the next stage in the workflow...

## REMOVING LIABILITY CAUSED BY EDITING

As mentioned before one of the big advantages of using an object oriented file format such as PDF is the ability to do editing after creation of the PDF file. Remark that we are not talking about *creative* editing; that should be done in the original application (such as Quark XPress, Adobe InDesign, Adobe Illustrator...). There are however plenty of reasons to do *corrective* editing on the PDF file rather than go back to the original application file. In many cases the original application files are not available at that point in the workflow, or it would take too much time to get them. And in a lot of workflows it is simply more time-consuming and risky to go back to the original application and then re-create the PDF document. Those arguments get more important as the pressure on deadlines for publishers increases...

Whatever the reasons are, it is a fact that refusing to edit the PDF file is safer. As soon as one edits the file, one also takes responsibility of it. If something else goes wrong with that file, it will be very hard to prove who is at fault.

Certified PDF tackles this problem in two ways. First of all, it allows a user to see what different versions of a Certified PDF file were saved throughout the workflow. And secondly, it allows one to see in detail everything that was changed in each of those versions. Let's see that in a bit more detail...

## TRACKING FILE HISTORY

There are two ways to save a regular PDF file after it has been changed. The first way is to save a completely new copy – this removes information that is no longer used and yields the smallest possible file. The second way is to use what is called *incremental save*, which only saves the changed objects and appends them to the end of the original file.

The disadvantage of incremental save is that it increases the size of the file. The big advantage is that it allows a user to take steps back and see how the file looked the first time it was saved, and the second time, and... Incremental save is a feature that is native to regular PDF files; though Adobe Acrobat uses it only when digital signatures are used.

In a Certified PDF workflow, opening a PDF file, changing it and then saving those changes is called a *session*. Certified PDF files always use incremental save which means that a Certified PDF file always consists of the original file with one or more sessions appended to the end.

This allows anyone in a Certified PDF workflow to know how a file looked at any point in its lifetime. It offers the possibility to take a step back and save the original or any later version of a Certified PDF file. But it also allows comparing two different versions of a Certified PDF file to see what exactly did change between those versions.

Together with each session that is saved in a Certified PDF file, identification information for the person who saved that session and for the computer used to do the work is saved in the file. This allows pinpointing exactly where a certain change was introduced and who was responsible for it.

### TRACING INDIVIDUAL EDITS

Taking this concept one step further, a Certified PDF file does not stop at tracking sessions in a document that is pushed through a Certified PDF workflow. Each time a user edits the file (for example converts a color from RGB to CMYK, does text-editing, changes the position of an object) a description of that particular edit is added to the *edit log* for that session.

At any time the edit log for a session (or for the complete lifetime of the document) can be consulted to know exactly what changes were done, where they were done and who did them with what products. Edit log entries contain human readable descriptions of the changes, so that even changes that don't have a visual effect on screen (changing halftone information to name just one) can be traced without problem. Together with the human readable description Certified PDF also stores the areas on every page where a change had effect; doing this allows tools working with Certified PDF not only to show each change but also to highlight the corresponding areas on a page.

### REAL WORLD TECHNOLOGY

Certified PDF technology is more than a conceptual idea of the sort of technology that should be available in a workflow. About a year ago, the Belgian advertisement industry committee, Medibel+, decided to standardize all exchange of digital ads on Certified PDF files. From then on ads sent to newspaper or magazine publishers had to be not just plain PDF, but Certified PDF.

To accomplish this one PDF Profile specific for magazine advertisements and one specific for newspaper advertisements were created. That effort was then combined with a substantial communication effort to all members of the industry to make sure everyone had the correct tools and knowledge to deliver correct Certified PDF files. Implementing this caused the error rate of received files to drop dramatically. The failure rate in received PDF files dropped from 40% to under 1% over a couple of months.

The example set by Medibel+ was followed quickly by the organization uniting the commercial print sector in Belgium and over the last couple of months by their Dutch and French sister organizations.