

Long-Term PDF Storage

Donna L. Baker, ACE, Baker Communications

PDF/A is the international standard for long-term archival storage. Not every file needs archiving, nor is every file capable of archival formatting. In this article, rather than taking a tutorial approach, I've looked at several issues surrounding the use of PDF/A, features of the standard you may not be aware of, and where you can find information within your file and Acrobat 9.

Use the PDF/A standard for documents intended for long-term storage in PDF format. The Standard isn't involved with an organization's archiving system or strategy, but profiles a format for reproducible documents that become part of that archiving system or strategy.

A growing number of government and industry bodies such as libraries, newspapers, regulators and legal systems require the assurance that their electronically archived documents can be preserved over a long period of time, while still allowing for predictable document retrieval and rendering.

What the Standard Includes

The purpose of the PDF/A standard is maintaining longevity of your PDF files. How is this accomplished? Simply by excluding anything that can't be totally contained within the file. Disallowing external features means you can display the document the same way every time under any circumstance.

That's the Standard in a nutshell, but there's more. You can further define compliance based on the presence or absence of a document structure.

Pick a Compliance Level

You've seen the PDF/A standard listed as PDF/A-1a and PDF/A-1b, identifying two levels of compliance. Both levels include the requirements and exclusions listed previously.

In addition, PDF/A-1a requires a document structure. Tagging the file allows for repurposing and searching the document.

Basic Features

To comply with PDF/A, a document needs these features:

- All visual content, including text, raster and vector images within the document must be embedded
- Information derived from external sources is disallowed
- All fonts must be embedded and capable of universal rendering (and be legally embeddable)
- The colorspace defined in the document must be device-independent
- All document metadata must be standards-based

On the other hand, the document can't contain these features:

- Executable files and JavaScript
- Audio and video content
- Any form of encryption or document protection

Acrobat PDFMakers allow for three PFD/A selections (Figure 1). In the General tab of the Acrobat PDFMaker, select the “Create PDF/A-1a:2005 compliant file” check box. Once the option is chosen, the remaining items on the dialog box are disabled. To use one of the PDF/A-1b choices, click the Conversion Settings drop down arrow and select either the CMYK or RBG versions from the list. Click OK to close the dialog box and carry on with your file conversion.

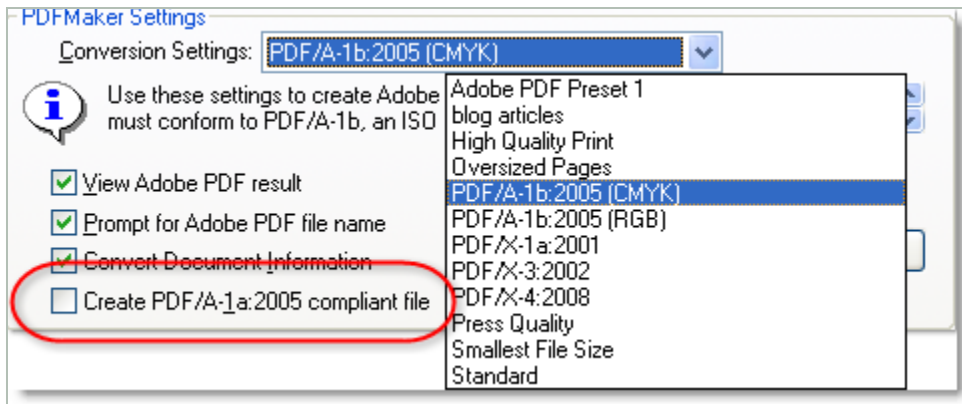


FIGURE 1. Choose the desired PDF/A version.

Implementation Issues

Just because a document calls itself “PDF/A Compliant” doesn’t mean it necessarily complies with the full standard. For example, a document can be PDF/A-compliant and not use PDF/A-compliant metadata. In addition, documents may claim to be compliant but in fact contain disallowed features. The only way to be sure is to test the files yourself.

Automatic Evaluation

Acrobat 9 manages the compliance evaluation issue for you. When you open a standards-compliant file, the Document Message Bar shows the notation shown in Figure 2, and the Standards navigation panel activates.

You’ll see a listing of both Conformance and OutputIntent. In the figure, you see the file was created using the device-independent sRGB color space. Click Verify Conformance to process the file. Upon completion, the Status shows as *verified* on the Standards panel.

Tagging Features

Both PDF/A levels are designed to make a document display and print correctly. In addition, PDF/A-1a includes metadata requirements. First, text needs enough information to allow Unicode mapping. Second, document content requires tagging to identify the semantic relationship between objects.

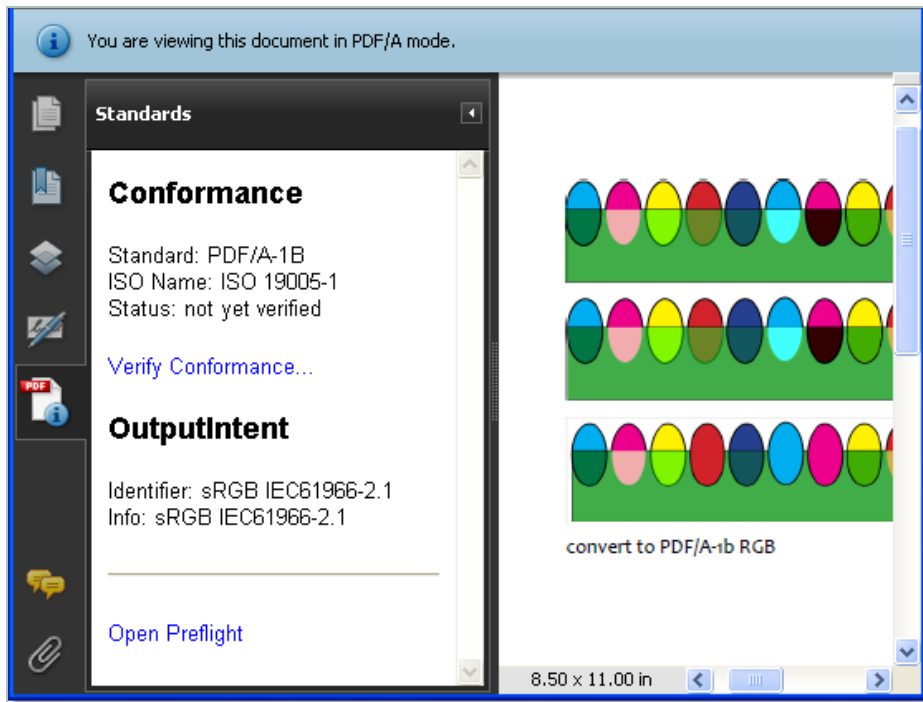


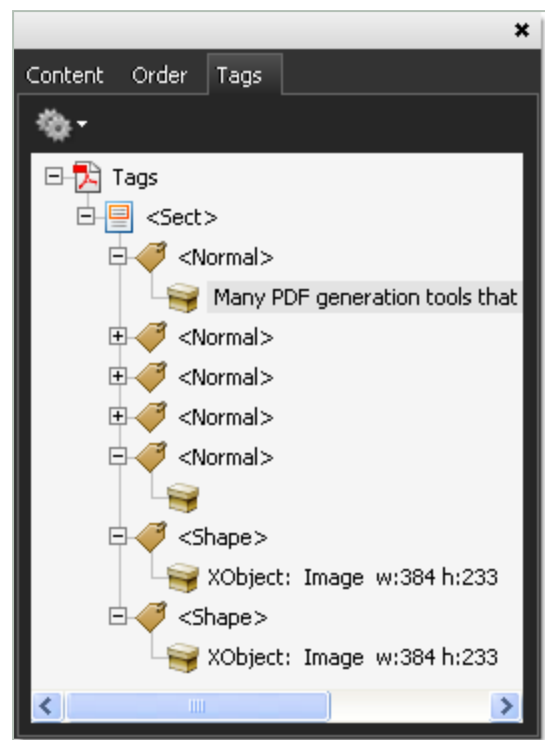
FIGURE 2. Test the file for compliance.

Fortunately, Acrobat again provides for maintaining standard compliance. For example, unlike other basic documents, once a file is defined as PDF/A compliant, you can't access or change the file's internal structures.

Here's an example. In a PDF/A1a-compliant document, I can access the Tags panel. However, I can't make any changes to the tags, nor can I change an existing tag to an artifact as the commands are disabled (Figure 3).

FIGURE 3. Tags are visible, but can't be modified.

As you see in Figure 4, I can look at the Class Map (style dictionary that stores attributes associated with the elements) in Read-only mode. Ditto for the Role Map, the set of custom tags mapped to predefined Acrobat tags producing a unique tag set for the document.



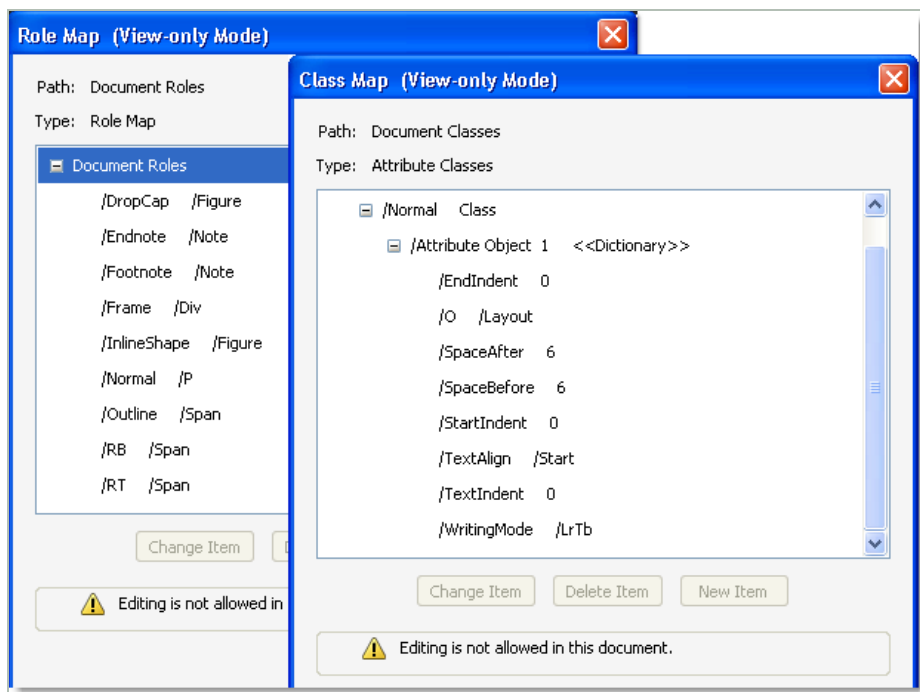


Figure 4. Maps for styles and custom tags are visible, but can't be modified.

Shortcuts to PDF/A File Creation

You can open a file, open Preflight, locate and select the fixups and profiles you need to use for a compliant file, and then run the profile—or you can use one of these shortcuts:

- Rather than opening a file, opening Preflight, locating and running a profile, export the file. Choose File > Export and select PDF/A from the submenu. During the export, Acrobat examines the file, applies necessary Preflight commands automatically and saves the file.
- Choose File > Save As and choose a standards format (such as PDF/A) from the “Save as type” drop-down list. Click Settings to open the Preflight: Convert to PDF/A dialog box, and choose either PDF/A-1a or PDF/A-1b. Then click Save to apply the standard, make modifications, and save the file automatically.
- Choose **Edit > Preferences (Acrobat > Preferences)** to open the Preferences dialog box, and choose Documents from the Categories list. If your workflow involves preparing files for archival purposes, choose Only for PDF/A mode from the “View documents in PDF/A mode” drop-down menu. Your files are automatically evaluated and tested.

What About TIFF Images?

TIFF images have traditionally been used for electronic archival, but not without some drawbacks:

- Construction of large TIFF archives can be hampered by proprietary image tags and a range of format variations
- Unlike PDF files, TIFF files aren't searchable files, and don't contain renderable text in addition to graphics and metadata
- It's simpler to organize PDF files into documents, complete with link and bookmark navigation, and associated metadata
- TIFF files are almost always larger than PDF files

On the upside, TIFF files convert well to PDF/A-compliant files.

Pre-Archival Tips

If your organization may have to comply with PDF/A standards in the future, or you're planning ahead "just in case," here are some measures you can take now:

- Minimize the number of file formats used by your organization. Settle on one document format, one spreadsheet format, defined image formats, and so on. The fewer file formats you need to deal with during an archival process, the better.
- Image transparency may be unsupported, and images flattened. In the example shown in Figure 5, running the Preflight Profile, "Verify compliance with PDF/A-1a" identifies and lists all items in the document using a transparency blend mode other than "Normal."

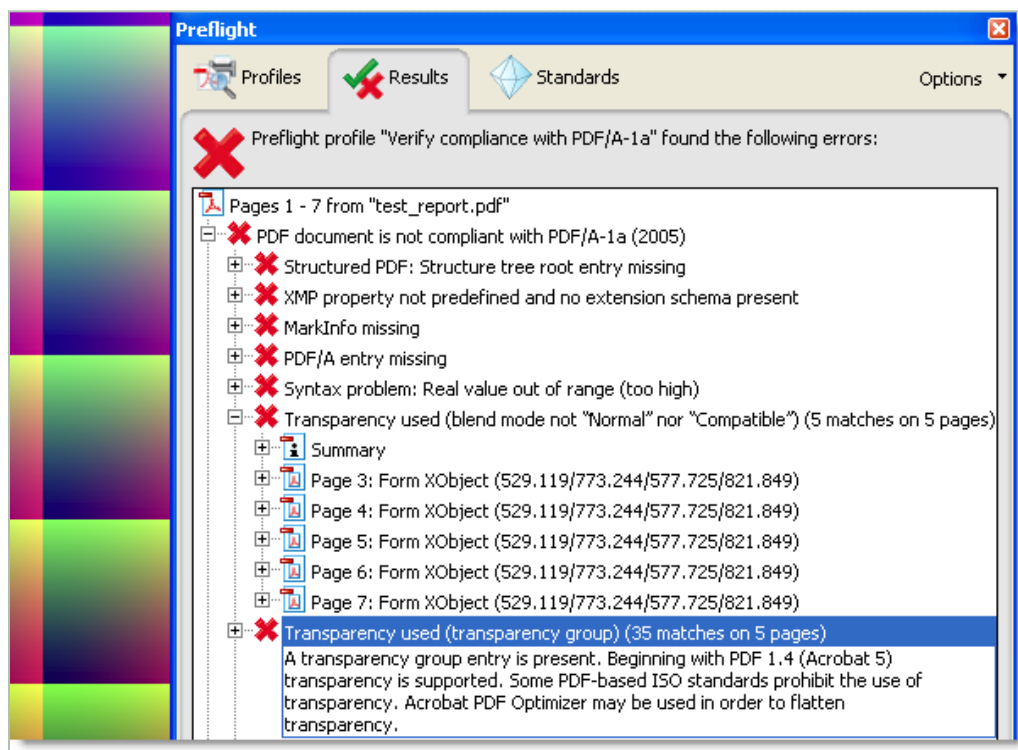


FIGURE 5. Be aware of issues with transparency blend modes.

- Use an in-house list of fonts, embed the fonts and stick to the list. Regularly embedding fonts makes archival efforts much simpler, with predictable visual results.
- Tag documents, either during conversion to PDF or using the Accessibility features in Acrobat.

Note: Although not necessarily an implementation issue, be aware that of the increased file size for PDF/A files due to the embedded fonts.