Florida Digital Archive

Guidelines for Creating Archival Quality PDF Files

Carol Chou

version 1.1, June 2006
supercedes version 1.0, May 2006

This document provides guidelines for creating preservation-quality PDF files. It is based in part on the PDF/A-1 specification, ISO 19005-1. It is strongly recommended that PDF documents submitted to the Florida Digital Archive conform to at least the PDF/A-1b standard.

This document provides general guidelines for creating PDF documents that are optimized for long-term digital preservation. The guideline is based in part on the PDF/A-1 specification, ISO 19005-1 [ISO 19005-1]. It is strongly recommended that PDF documents submitted to the Florida Digital Archive conform to at least the PDF/A-1b standard.

The most recent version of Adobe Acrobat Professional, version 7.0.7 (released on Feb 14, 2006), supports the official PDF/A-1b standard by providing two configuration settings: PDF/A-1b:2005(RGB) and PDF/A-1b:2005(CMYK). When using the built-in support in Acrobat Professional 7.0.7 to convert documents into PDF format, it is recommended that you use one of these two PDF/A-1b standard settings without modification.

Tool	Conversion Support	Recommended PDF Settings
Acrobat Professional	MS Office, MS Project, MS Visio, MS Publisher and Postscript to PDF/A-1b	PDF/A-1b:2005(RGB) or PDF/A-1b:2005(CMYK)
Acrobat Distiller	Postscript to PDF/A-1b	PDF/A-1b:2005(RGB) or PDF/A-1b:2005(CMYK)
Adobe PDF Maker plug-in Microsoft Word	MS Word to PDF/A-1b	PDF/A-1b:2005(RGB) or PDF/A-1b:2005(CMYK)

Users of MS Word should use the PDF Maker plug-in for Word to create PDF/A files from Word documents, rather than Acrobat Professional. PDF/A-1b files converted from Word by Acrobat Professional lack the required "ID" field in the trailer dictionary and thus will not pass the PDF/A validation.

However the PDF/A file is created, users should use the preflight feature available in Adobe Acrobat Professional 7.0.7 to validate the file against the PDF/A-1b standard and remove all non-conformances prior to submission.

The following sections describe the PDF/A requirements to aid in understanding the built-in PDF/A-1b settings during the PDF/A creation

process. Some of the sections also contain applicable suggestions for configuring Adobe Acrobat PDF settings when it is required. Nevertheless, this document is not intended to be served as the user manual for Adobe Acrobat nor as a comprehensive how-to document for creating PDF/A documents.

1. Color Space
The submitted documents must use only device-independent color spaces, such as CalGray, CalRGB, Lab or ICCBased. Any ICCBased color space must be embedded and conform to ICC specification. Alternatively, the device independence may be achieved by using an uncalibrated color space (such as, DeviceRGB or DeviceCMYK) which also contains an OuputIntent to specify the color characteristics of the device that is intended for rendering.

Suggestion: To ensure that all device-dependent color will be converted by using either device-independent color space or in a device-independent manner, it is recommended that the PDF/A-1b standard settings related to colorspace to be used unaltered. However, if it is necessary to configure the Adobe PDF settings, select one of the "Tag Everything for Color Management," "Convert All Colors to sRGB," or "Convert Everything to CMYK" options [Adobe]. Do not set the applicable working space for the selected color management policy to "None".

2. Compression
The submitted documents must use only lossless compression algorithms which are not subject to intellectual property constraints. The use of LZW compression is prohibited.

Suggestion: For compression on monochrome images, choose either "Off," "Run Length," "CCITT Group 3" or "CCITT Group4" in Adobe Acrobat. For color or grayscale images, choose either "Off," "Zip," "JPEG" or "automatic (JPEG)" in the compression settings.

3. External References
For the purpose of long-term preservation, it is mandatory that the submitted document is self-contained with no external dependencies. Therefore, those PDF features allowing for external references must not be used. These features include:
- Actions to launch external applications
- Actions to execute Java scripts. The use of Java script may create external dependency and interfere with normal rendering of the document.
- Actions to import form data from external files (ImportData action)
- Annotations to contain references to external files (FileAttachment

annotation)

4. Font
All fonts used in a document, including standard fonts, must be legally embedded for unlimited, universal rendering. Font embedding ensures that the submitted document will be rendered in the way it was originally created. To reduce the size of the PDF file, the embedded fonts must also be subsetted such that it contains only the required glyphs to render the document.

Suggestion: In the Adobe PDF settings, select both the "Embed all fonts" and "Subset embedded font when percent of characters used is less than 100%" options to embed and subset all the fonts used in the file. Do not select the "Optimize for fast web view" option as it will cause Acrobat not to embed the standard fonts in the created PDF file.

5. Form
To ensure consistent rendering on form fields, any action that could potentially change the visual appearance of the form is prohibited.

Suggestions: Remove all form actions from a document prior to converting it into the PDF format.

6. Images
Alternative images, allowing for variant representations of an image, must not be used. The use of transparency feature in an image is also prohibited. Instead, flattening multiple-layer images in the PDF file is recommended.

Image down sampling during the PDF creation process should be avoided, as it could result in loss of quality. (Note this is not a formal PDF/A requirement.)

7. Metadata
To uniformly describe the descriptive, administrative and technical metadata, the PDF file must contain a Metadata stream that conforms to the XMP specification. The following XMP metadata properties must be provided in the PDF file:

Metadata	XMP Property
Document Title	dc:title
Document Author	dc:creator
Subject	dc:subject
Keywords	pdf:keywords
Creator	xmp:CreatorTool

Program	
Producer	pdf:Producer
Creation Date	xmp:CreateDate
Modification Date	xmp:ModifyDate
File Identifier	either through the use of xmp:Identifier, xmpMM:DocumentID, mpMM:VersionID, or xmpMM:RenditionClass
File Provenance Information	xmpMM:History

Every XMP metadata property which has a corresponding entry in the document information dictionary must be defined with the value equivalent to the one in the document information dictionary. The XMP metadata stream must also contain properties to indicate the PDF/A version and conformance level of the document. Those properties must be specified using the PDF/A Identification extension schema.

The description of all extension schemas used in the document must be embedded. The descriptions must be specified following the PDF/A extension schema description schema. In addition, the XMP metadata stream must not be filtered.

Suggestion: Using a PDF/A conformed writer will embed a XMP metadata stream according to the PDF/A standard. Do not use any PDF writer that is not PDF/A conformed.

8. Multimedia
Embedding of multimedia content, including the use of Sound Annotation, Sound Action, Movie Annotation and Movie Action, is prohibited.

Suggestion: Remove all multimedia content from a document prior to converting it into the PDF format.

9. Security
The document must not be encrypted or password protected. Encryption or password protection hinders the accessibility of the documents. Any encrypted or password protected document would not be migrated in the future.

Suggestion: When configuring Adobe PDF settings to create PDF documents in Acrobat, Distiller or PDF Maker, please select "None" for the "Adobe PDF Security" setting to avoid possible password protection in the

created PDF file.

10. PDF/A-1a Conformance

PDF/A-1a conformance includes the full set of PDF/A features. In addition to the PDF/A-1b features, a document which is PDF/A-1a compliance must also adhere to the following constraints:

- The use of tags to specify the logical structure of the document. (Note, however, that tagging could significantly increase the size of the PDF file.)
- The use of Unicode Character Map to map character codes into Unicode values.

As Adobe currently does not support PDF/A-1a conformance level, the FDA does not require submitted documents to conform with PDF/A-1a. However, if the submitting institutions use a tool that supports PDF/A-1a, creating PDF/A-1a conformed PDF files is strongly recommended.

References

[Adobe] "Create Adobe PDF Online," http://createpdf.adobe.com/cgi-feeder.pl/help_color?BP=&LOC=en_US

[Adobe 2004] "XMP Specification," Adobe Systems Incorporated, January 2004, http://www.aiim.org/documents/standards/xmpspecification.pdf.

[ISO 19005-1] "Document management – Electronic document file format for long-term preservation – Use of PDF 1.4 (PDF/A-1)," International Organization for Standardization, 2004.

[ICC] "ICC.1: 1998-09, File Format for Color Profiles," International Color Consortium, http://www.color.org/ICC-1_1998-09.PDF.

[PDF Tools] "White Paper PDF/A – The Basics," PDF Tools AG, February 7, 2006, http://www.pdf-tools.com/public/downloads/whitepapers/whitepaper-pdfa.pdf.