

PDF Guideline for DAITSS

By Carol Chou
April 18, 2006

This document provides a general guideline for creating PDF documents, to be processed by DAITSS for long-term preservation. It is based in part on the draft of PDF/A-1 (ISO 19005-1) specification.

The most recent version of Adobe Acrobat Professional, version 7.0.7 (released on Feb 14, 2006), supports the official PDF/A-1b standard. It is highly recommended to use the preflight feature in the latest version of Adobe Acrobat Professional to validate a PDF against PDF/A-1b conformance prior to saving it. To use Acrobat Distiller to convert a document from other formats into PDF, please be sure to use the PDF/A standard settings to create PDF/A-1b conformance document. It is highly recommended that the submitted PDF document conforms to at least PDF/A-1b.

Color Spaces

It is recommended to use only device-independent color spaces, such as CalGray, CalRGB, Lab or ICCbased color space in the submitted document. Any ICCBased color spaces shall conform to ICC specification. Alternatively, the device independence may be achieved by using an uncalibrated color space which contains an OutputIntent to specify the color characteristics of the device that is intended for rendering.

Compressions

It is recommended to use only lossless compression algorithms and those compression algorithms that are not subject to intellectual property constraints. The use of LZW compression is prohibited.

External References

For the purpose of long-term preservation, it is recommended that the submitted document is self-contained with no external dependency. Therefore, those PDF features allowing for external references shall not be used. They include:

- Launch action to launch an external application.
- Java Script action to execute a Java script. Use of Java script may create external dependency. It may also interfere with normal rendering of the document.
- ImportData action to import form data from an external file.
- FileAttachment annotation to contain a reference to an external file.

Fonts

All fonts used in a submitted document shall 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 document file, the embedded fonts shall (or may?) also be subsetted such that it contains only the required glyphs to render the document.

Q: How about standard fonts? Shall they be embedded also?

Forms

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

Images

- It is not recommended to use the alternative images.
- It is recommended to avoid image down sampling during PDF creation process as it could result in quality loss.
- The use of transparency feature in a image is prohibited. Instead, it is recommended to flatten those multiple layers images prior to embedding them in a PDF document.

Metadata

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

Document Title	dc:title
Document Author	dc:creator
Subject	dc:subject
Keywords	pdf:keywords
Creator Program	xmp:CreatorTool
Producer	pdf:Producer
Creation Date	xmp:CreateDate
Modification Date	xmp:ModifyDate
File Identifier	xmp:Identifier or xmpMM:DocumentID, xmpMM:VersionID and xmpMM:RenditionClass
File Provenance Information	xmpMM:History

If a document information dictionary is present in the PDF, all of its entries which has corresponding XMP metadata properties shall also be defined with the equivalent value. The XMP metadata stream shall not be compressed.

Multimedia

Embedding of multimedia content, including the use of Sound Annotation, Sound Action, Movie Annotation and Movie Action, is discouraged rather than prohibited. The use of multimedia content is prohibited in PDF/A, however, it is perhaps impossible for submitters to always comply to this restriction, especially in ETD.

Security

The submitted document shall not be encrypted or password protected. Encryption or password protection hinders the accessibility of the document. An encrypted or password protected document would not be migrated in the future.

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.
- The use of Unicode character map to map character codes to Unicode values.

Q: Shall we require the submitted PDFs to conform with PDF/A-1a even though Adobe doesn't support it yet?

References

[Adobe] “XMP Specification”, Adobe Systems Incorporated, January 2004

[ISO 19005-1] Draft, “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

[PDF Tools] “White Paper PDF/A – The Basics”, PDF Tools AG, February 7, 2006