ISO 19005-1:2005

PDF/A-1

Date: July 10, 2006

What is PDF/A?

ISO 19005-1, Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1) is the first in a new family of ISO
Standards to address the growing need to maintain information in electronic documents over archival time spans.
Because this initial version of PDF/A is based on PDF 1.4, the standard is being published in parts so that new parts can be added without obsolescing previous parts. For example: PDF/A-1 refers to the format defined by Part 1 (ISO 19005-1) of the standard (or PDF 1.4) while Part 2 (ISO 19005-2) and later parts may be based on a later version of PDF and/or may define archiving requirements for more complex content types.

Why PDF/A?

The feature-rich nature of PDF can create difficulties in preserving information over the long-term, and some useful features of the PDF file format are incompatible with the demands of long-term preservation. For example, PDF documents are not necessarily self-contained, drawing on system fonts and other content stored external to the original file. As time passes, and especially as technology changes, these external connections can be broken, and the dependencies cause information to be lost. Additionally, because of the lack of standardization among the many PDF development tools on the market, there is inconsistency in the implementation of the file format. This lack of standardization could be chaotic for the information managers of the future, especially as it would be difficult (if not impossible) for them to "get under the hood" of the PDF files unless a format specification were put in place that specifically addressed long-term preservation needs. Tremendous quantities of valuable information are currently be created and saved all over the world as PDF, and a specification solution is needed to ensure that digital PDF documents remain readable, renderable and accessible for the long-term. PDF/A is designed to be that specification.

What is the difference between PDF and PDF/A?

The PDF/A-1 (ISO 19005-1:2005) standard is based on Adobe's PDF Reference 1.4, and specifies how to use a subset of PDF components to develop software that creates, renders and otherwise process a flavor of PDF that is more suitable for archival preservation than traditional PDF. PDF/A-1 aims to preserve the static visual appearance of electronic documents over time and also aims to support future access and future migration needs by providing frameworks for: 1) embedding metadata about electronic documents, and 2) defining the logical structure and semantic properties of electronic documents. The result is a file format, based on PDF 1.4 that is more suitable for long term preservation. PDF/A-1 files will be more self-contained, self-describing and more device-independent than traditional PDF 1.4 files.

What does PDF/A-1 allow/disallow?

One of the key differences between PDF and PDF/A is the restrictions that PDF/A places on PDF. PDF/A-1 files must include:

- Embedded fonts
- Device-independent color
- XMP metadata

PDF/A-1 files may not include:

- Encryption
- LZW Compression
- Embedded files
- External content references
- PDF Transparency
- Multi-media
- JavaScript

What does PDF/A mean by "long term"?

PDF/A defines long-term as: "the period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository, which may extend into the indefinite future."

Does PDF/A-1 replace other archival file formats?

No. PDF/A-1 does not replace other archival file formats. Rather, PDF/A-1 was developed to allow PDF to be used as an archival format in a well-defined and robust manner.

What long-term preservation needs does PDF/A-1 address?

Characteristics identified as objectives for PDF/A were:

1. Device Independent - Can be reliably and consistently rendered without regard to the hardware or software platform

2. Self-contained - Contains all resources necessary for rendering

3. Self-documenting -Contains its own description

4. Unfettered - Absence of technical file protection mechanisms

5. Available - Authoritative specification publicly available

6. Adoption - Widespread use may be the best deterrent against preservation risk

How does PDF/A-1 address these long-term preservation needs?

Device Independent. PDF/A-1 requires device independent components so that the static visual appearance can be reliably and consistently rendered and printed without regard to the hardware or software platform used. The graphics clause, for example, incorporates requirements to ensure predictable color rendering. PDF/A-1 also prohibits the use of components not defined in PDF Reference 1.4.

Self-Contained. Everything that is necessary to render or print a PDF/A-1 file must be contained within the file. The fonts clause requires that all fonts used for rendering content are embedded in the file. A PDF/A-1 conforming writer must always embed fonts, and a conforming reader must always use the embedded fonts. This means that the file will be rendered using the fonts intended and not using fonts residing on the local workstation.

Self-Documenting. PDF/A-1 requires Adobe Extensible Metadata Platform (XMP) be used for embedding metadata in PDF files. To allow flexibility of implementation,
PDF/A-1 provides recommendations for documenting file attributes such as: File identifier, File provenance, Font metadata, and allows non-XMP schemas to be included, as long as they are embedded. Implementers can use XMP in a variety of ways to include information about electronic records within the file itself. Having metadata embedded in the file can increase the informational value of electronic documents and enhance the future researcher's understanding of the document.

Unfettered. PDF/A-1 prohibits encryption. This prohibition means that User IDs and/or Passwords are not needed to do anything with a PDF/A-1 file. PDF/A-1 files are open and available to anyone or any software that processes the file. Implementers that require access controls can provide these access controls outside of the file format.

Available. PDF/A-1 is based on an authoritative specification that is publicly available.
Anyone can use the PDF Reference and XMP Specification in conjunction with PDF/A-1 to create applications that read, write, or process PDF/A-1 files. Adobe has granted a general royalty free license to use certain of its patents to create applications that process
PDF/A-1 files. Additionally, Adobe has granted AIIM and NPES the rights to publish these specifications on their respective Internet sites into the indefinite future.

Adoption. PDF/A-1 was designed for flexibility of implementation to promote its wide adoption. If widely adopted, PDF/A software tools will proliferate, and the market will support the file format as long as the demand exists. Market support of PDF/A will help ensure the viability of PDF/A and extend the length of time that PDF documents can be maintained as PDF/A.