

Indiana Commission on Public Records

Guidelines on the Use of PDF/A

What is PDF/A?

Portable Document Format (PDF) is a file format that can be used to capture information from almost any application and create documents that can be easily accessed and shared. However, the feature-rich nature of the format can create difficulties in preserving PDF files over the long-term, and some useful features of the format are incompatible with the demands of long-term preservation. For example, PDF documents are not necessarily self-contained, sometimes 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, which also impedes long-term preservation.

A PDF/A file is more suitable for long term preservation as it is more self-contained, self-describing and more device-independent than traditional PDF files. PDF/A aims to preserve the static visual appearance of electronic documents over time whilst also supporting future access and migration needs.

What are the characteristics of PDF/A?

There are two versions of the PDF/A standard: **PDF/A-1** (issued in 2005, based on PDF 1.4) and **PDF/A-2** (issued in 2011, based on PDF 1.7). For both PDF/A-1 and PDF/A-2, there are also two levels of conformance: **PDF/A-1b** and **PDF/A-2b** ensure basic, visual integrity of a document; **PDF/A-1a** and **PDF/A-2a** offer enhanced accessibility, including flexible rendering for virtually any device, and compatibility with screen readers.

Although all varieties of PDF/A are “preservation-friendly”, PDF/A-1 offers some constraints including:

- Maximum document size is 5 meter by 5 meter (approximately 16 feet by 16 feet)
- Audio and video content are forbidden
- JavaScript and executable file launches are prohibited
- All fonts must be embedded and also must be legally embeddable for unlimited, universal rendering
- Encryption is disallowed, i.e. user IDs and/or passwords are not possible
- LZW compression prohibited
- PDF Transparency forbidden
- Hyperlinks are allowed but may not function depending on how the document is created/viewed; this seems to be a particular issue where hyperlinks are “masked,” meaning that the full address of the external link/file is not written out; e.g. clicking a word/phrase - “the IDEM website” - opens an external link rather than having it spelled out as www.in.gov/idem

PDF/A-2 removes some of these limitations and offers a number of new features:

- JPEG2000 image compression
- Support for transparency effects and layers - helpful for technical documents such as construction drawings and plans, where content can be shown or hidden
- Embedding of cross-platform OpenType fonts
- Page sizes on a 1:1 scale of up to 236 miles by 236 miles, allowing larger physical page sizes to be rendered to match original scanned documents
- Provisions for digital signatures meeting the PDF Advanced Electronic Signatures standard
- Ability to compile a number of PDF/A files into a single PDF/A document - could be used to package sections of large or irregularly shaped documents that cannot be scanned as a single image

Text in both PDF/A-1 and PDF/A-2 files can be searched. If the PDF file was created from a document created electronically (e.g. a Word file), the text should automatically be searchable. If the PDF was created from a scanned version of a paper document, Optical Character Recognition (OCR) will need to be applied to make the text searchable.

When should PDF/A be used?

ICPR recommends the use of the PDF/A file format for the long-term preservation of text and graphic based records, whether electronic originals or digital reproductions of analog originals. It is also a suitable format for preserving digital versions of other types of record, such as blueprints and other technical documents. Where records are being digitized/scanned, the ICPR strongly recommends the application of OCR to ensure that the resulting PDF/A files support full-text search functionality wherever the content allows it.

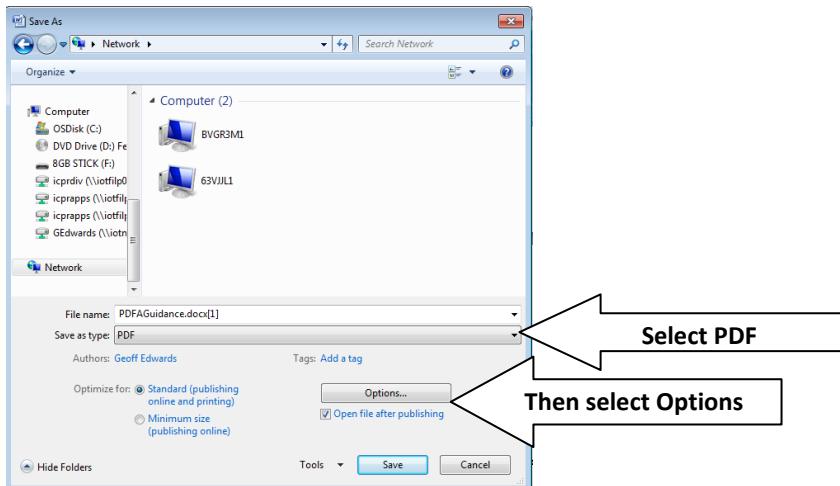
As noted, any variety of PDF/A is an acceptable preservation format; PDF/A-2 does not replace or supersede PDF/A-1 but simply adds additional features. The decision as to which PDF/A to use depends on whether the extra features of PDF/A-2 would preserve or enhance the functionality or representation of a document in a way that PDF/A-1 would not.

As PDF/A-2 is a relatively recent development, it may be more difficult to find applications that create files in that format. PDF/A-1 documents, on the other hand, can be created by a variety of applications (including directly from MS Office programs). There are also conversion and validation programs available that will turn standard PDF files into PDF/A and validate their conformance with the standard.

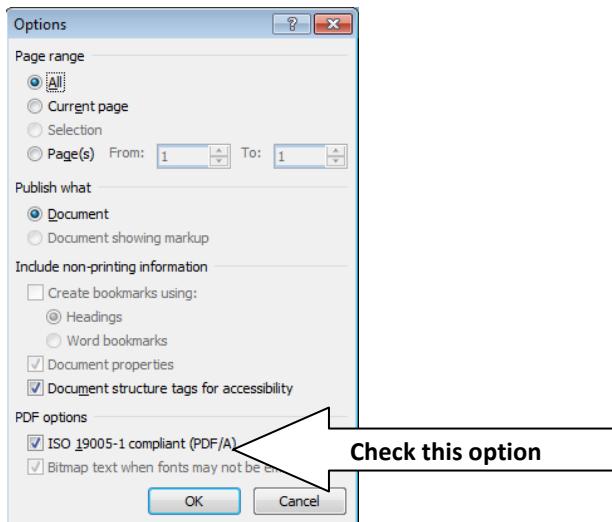
How to save a document as PDF/A in Microsoft Office programs:

This guideline was originally created as a Microsoft Word document; it was then saved as a PDF/A-1 compliant file as follows:

Go to Save As > Save As Type - choose PDF; then select Options



From the Options menu, check ISO 19005-1 compliant (PDF/A); then click OK and Save:



Other Resources:

Further details about PDF/A-1 and PDF/A-2 are available from:

- http://www.aiim.org/documents/standards/19005-1_FAQ.pdf
- <http://www.digitalpreservation.gov/formats/fdd/fdd000318.shtml>
- <http://www.pdfa.org/2011/08/pdfa-%E2%80%93-a-look-at-the-technical-side/>
- <http://www.pdfa.org/wp-content/uploads/2011/10/Flyer-PDFA2-Overview-EN.pdf>
- <http://www.pdfa.org/publication/4th-international-pdfa-conference-pdfa-forever/>

For additional guidance, please contact ICPR.