Intro to PDF/A

What is PDF/A?

- PDF version for archiving and long-term preservation

- ISO standard

- Full replacement for paper

Why PDF/A?

Goal: long-term reproducibility

- View original document regardless of SW/HW/platform changes

- No need to convert a PDF/A document at a later date

- A PDF/A document is future-proof

$\rightarrow$ A PDF/A document must be self-contained

- PDF/A is a subset of PDF

PDF/A & Memory Institutions

The National Archives & PDF/A

It's a preferred or accepted file format for:

- Digitized Paper Records

- Born-Digital Textual Data

- Born-Digital Posters

- Presentation Formats (slide decks)

The National Archives & Paper

By July 1st, 2024, Federal agencies must transfer *all* permanent records in electronic format with metadata.

Library of Congress & PDF/A

It's a preferred or accepted file format for:

- Textual Works - Digital

- Textual Works - Electronic Serials

- Musical Scores - Digital

PDF/A characteristics

PDF/A forbids certain PDF features

- JavaScript and executable file launches

- Audio and video content

- Encryption

- External content references

- Certain forms of compression

- Certain PDF forms

PDF/A requires certain optional PDF features

- Standards-based metadata

- Embedded fonts without IP restrictions

- Device-independent color specifications

PDF/A versions

version	published	ISO number	PDF version
PDF/A-1	2005	ISO 19005-1	1.4
PDF/A-2	2008	ISO 19005-2	1.7
PDF/A-3	2012	ISO 19005-3	1.7
PDF/A-4	2020	ISO 19005-4	2.0

- older versions are not obsolete!
  - PDF/A viewer must be backward compatible

Generating PDF/A documents

- Supported by many PDF creation tools
    - Commercial:
        - Adobe Acrobat Pro
        - ABBYY FineReader PDF
    - Ghostscript: open source
    - Word (version dependent - not Mac!)

Validating PDF/A documents

- Checking to see if a document is PDF/A compliant
  - version and conformance level checks
- veraPDF
  - open source
  - funded by the European Commission, industry-supported
- Adobe Acrobat Pro
- Query metadata (not guaranteed)

Converting PDF to PDF/A documents

- Adobe Acrobat Pro

- Ghostscript

- Other open-source and commercial options

Related PDF Standards Work

Email Archiving in PDF

- Not part of PDF/A, but related

- Known originally as EA-PDF, now called PDF/mail

- Longer term goal: PDF/M ISO standard

- Working group exists
  - led by Chris Prom, History Lab participates
  - currently developing the technical specification for PDF Mail

- Learn More here

PDF/UA (Universal Accessibility)

- Standard so that people with disabilities can read PDF documents

- Requires specific tags and metadata to provide structural information

- Implies conformance with PDF/A-2a

References

- Library of Congress: PDF/A

- PDF/A in a Nutshell

Discussion

- PDF/A in your institution

- PDF/A personal experiences