

PDF as an Archive Standard

Macduff Hughes
Sr. Eng. Manager
Adobe Systems
October 4, 2002

Topics to be covered

- What is PDF
- Why a “standard” version of PDF
- The example of PDF/X
- The idea of PDF/A
- ~~The standards process~~
- Issues

What is PDF?

- Definition
- Design goals
- History
- Ownership
- Implementation

What is PDF: Definition

- “Portable Document Format”
- A specification for electronic files representing documents

What is PDF: Design Goals (primary)

- Reliable consistent viewing and printing
- Mix text, raster images, lineart, colors
- Basic unit is the Page
- Easy navigation, fast access to any page
- Small file size

What is PDF: Design Goals (secondary)

- Interactivity
- Structured content
- Reliable text extraction
- Accessibility for vision-impaired
- Encryption, digital signatures, modification detection

What is PDF: History

- Version 1.0 published 1993
- Version 1.1 (1995)
 - Encryption, Articles, Transitions, Calibrated Color, improved Hyperlinks, Actions
- Version 1.2 (1997)
 - Prepress features, External Streams, Flate compression, Forms, Asian Fonts, more Annotations

What is PDF: History

- Version 1.3 (1999)
 - PostScript 3 imaging model, ICC color, Logical Structure, JavaScript, more Annotations, Digital Signatures
- Version 1.4 (2001)
 - Transparency, 128-bit encryption, XML metadata, Tagged PDF (Accessibility)

What is PDF: Ownership

- Adobe Systems owns copyright on the specification
- Documented methods for private extensions
- No trademark on “Portable Document Format” or “PDF”
- Anyone can implement software using the specification

What is PDF: implementation

- Adobe Acrobat family of products
- Other Adobe products
- OpenSource: GhostScript and GhostView
- Third party commercial products
 - Apple Computer
 - Global Graphics
 - 583 PDF products listed on www.planetpdf.com

Why a “standard” version of PDF

- PDF is powerful and flexible, ...
- But too flexible for some applications
- Need higher degree of reliability than required by published spec
- Also, may want standard in hands of neutral non-commercial body (e.g. ISO)

The example of PDF/X

- PDF for eXchange (not eXtended!)
- Targeted at print publishing, especially the exchange of digital advertisements
- Restricts PDF in certain ways
 - Fonts must be embedded
 - Some color spaces, annotation types forbidden
 - Some prepress information required
- Published by ISO (International Standards Organization)
- Adopted by Time Inc. as digital ad standard

The idea of PDF/A

- A restricted subset of PDF
- Published by recognized standards body such as ISO
- Focused on Archive needs of governments, corporations, libraries

PDF/A Issues

- Define predictable appearance – main goal, well understood
- Role of structure, content extraction?
- Should standard address security, signatures?

PDF/A: Reliable Appearance

- Prohibit external content
 - External files, OPI, non-embedded fonts, Reference XObjects
- Restrict programming
 - No JavaScript
- Restrict annotations
- Restrict fonts? (Hazards of TrueType)
- Restrict device-dependent operators (transfer functions, overprint, halftones).