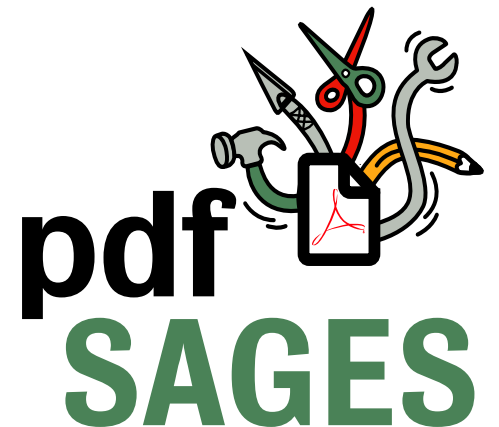


# PDF Standards: PDF/A, PDF/E, PDF/X

---

Leonard Rosenthol  
Chief Technology Officer  
PDF Sages, Inc.



# How I do things

- There is also an electronic copy of this presentation (PDF format, of course!) on my website at <<http://www.pdfsages.com>>
- I've left time at the end for Q&A, but please feel free to ask questions at any time!

# Overview

- Why the need for specific PDF standards?
- PDF/X – PDF for Graphic Exchange
- PDF/A – PDF for Archiving
- PDF/E – PDF for Engineering
- PDF/Access – PDF with Accessibility
- Wrap Up
- Q & A

# PDF - What's in there?

- **PostScript/Adobe imaging model**
  - Text & graphics in a device & resolution independent manner
  - Bitmap Images
- **Prepress Features**
  - Trapping, bleed, overprint, etc.
- **Navigation Tools**
  - Bookmarks, hyperlinks
- **Annotations**
  - Text notes, “MarkUp”, Movies, Sounds
- **Forms**
- **Security & Authentication**
- **And more...**

# PDF - What's NOT in there?

- **PDF is NOT Postscript!**
  - Non-printable elements (hyperlinks, etc.)
  - No programming language constructs
  - Strict file structure allowing for random access
  - Presence of font metrics for viewing fidelity
- **A PDF file can not be directly interpreted by a PS interpreter, though conversion of PDF page descriptions to PS is simplified.**

# A bit of history

- **PDF 1.0 - Acrobat 1.0**
  - Postscript imaging model
- **PDF 1.1 - Acrobat 2.0**
  - Security, annotations, binary files
- **PDF 1.2 - Acrobat 3.0**
  - Interaction, movies/sounds, forms, CJK, web

# History (cont.)

## ■ PDF 1.3 - Acrobat 4.0

- Structure, Digital Signatures, embedding, JavaScript, RTL, color separations, PS3

## ■ PDF 1.4 – Acrobat 5.0

- Transparency
- XML-based Metadata (XMP)
- Tagged PDF
- JBIG2

# History (cont.)

- **PDF 1.5 – Acrobat 6.0**
  - Compressed Objects
  - Optional Content (Layers)
  - JPEG2000
- **PDF 1.6 – Acrobat 7.0**
  - 3D
  - Object metadata
  - Page and view scaling
  - NChannel color

# A few other “features” of PDF

- **11 acceptable colorspaces**
  - DeviceGray, DeviceRGB & DeviceCMYK
  - CalGray, CalRGB, Lab & ICCBased
  - Indexed, Pattern, Separation & DeviceN
- **Font can either be embedded or referenced**
  - Type 1, TrueType, OpenType
  - (even old-style bitmap fonts as Type 3!)
- **OPI (Open Prepress Interface)**
  - Keep low-res images in the original and have them substituted for high res at the printer/output device.
- **Interactive Elements**
  - Hyperlinks, Bookmarks, Forms, Markup/Comments
- **Multimedia**

# What can we learn from this?

- PDF has grown beyond “electronic paper” to become the “digital envelope”
  - Single file that incorporates all standard forms of media (text, images, vectors, sounds, animation/motion, etc.)
  - Security & Authentication
- PDF can represent data types, compression standards and rendering features that Postscript does not.
  - Remember PS hasn’t been updated in over 7 years!
- That’s great for the web, eBooks, etc. – but what about people who just want reliable printing, archiving, etc.?



PDF/X

# What is PDF/X?

- PDF/X is not an alternative to PDF!!

- It's a focused subset of PDF designed specifically for reliable prepress/print-oriented data exchange.
- PDF/X defines both the file format and the way the files should be processed.

- Design Goals

- reliable blind transmission
- reduce processing errors
- improve proof and press matches
- reduce the cost and complexity of software tools
- all software should produce identical results

# Who Created PDF/X?

- Committee for Graphics Art Technical Standards (CGATS)
- Digital Distribution of Advertising for Publications (DDAP)
- Newspaper Association of America (NAA)
- International Standards Organization (ISO)
- Software vendors
- Users like you

# Timeline

- **DDAP User Requirements - 1990**
- **ANSI CGATS - 1991**
- **ANSI IT8.8:1993 (TIFF/IT)**
- **ISO 12639:1998 (TIFF/IT-P1)**
- **ANSI CGATS12.1:1999 (PDF/X-1)**
- **ISO 15930-1:2001 (PDF/X-1 & PDF/X-1a:2001)**
- **ISO 15930-3:2002 (PDF/X-3:2002)**
- **ISO 15930-4:2003 (PDF/X-1a:2003)**
- **ISO 15930-5:2003 (PDF/X-2:2003)**
- **ISO 15930-6:2003 (PDF/X-3:2003)**

# PDF/X in a nutshell

- **Subset of PDF 1.3/1.4 (2001 vs. 2003)**
  - Removes things that can be problematic for printing
    - ◆ Annotations, forms, links, security, etc.
    - ◆ No transparency or JBIG2
  - All print elements prepared for a single intended printing condition
  - Requires presence of page boxes (trim or bleed)
  - Requires trapping status (but not necessarily be trapped)
  - Requires all fonts embedded
- **X-1a**
  - Gray, CMYK and Spot colors only
- **X-3**
  - Color managed via ICC profiles

# What about this new PDF/X-2?

- Designed for “non-blind” transmission
- Subset of PDF/X-1 and PDF/X-3
- Allows fonts and images to be outside of PDF
  - Fonts have to be uniquely identified but they do not have to be embedded
    - ◆ Why? font license restrictions
- Allows OPI comments
  - supports late binding of images and layout
  - automated image replacement on server
  - centralized digital asset management
  - transmission to printer prior to completion of all work

# How is this different then preflighting?

- Not much...
- EXCEPT that a file that has been verified and “marked” as PDF/X is known to meet a certain set of internationally agreed upon criteria that you can rely on.
- And you can still preflight for additional requirements beyond the standard PDF/X specifications that you might have.
  - Image resolution
  - Font types
  - etc.

# Beyond the Standard

## ■ PDF/X Plus workflow checks

- color image resolution must be => 300 DPI
- font restrictions - no TrueType or CID
- process color only – no spot colors
- hairline rules
- white set to overprint
- etc..

## ■ This aims to combine preflighting and/or local policy rules with PDF/X creation

- However, the details of the “workflow checks” aren’t stored in the PDF for downstream comparison or validation.

# De-facto vs. Accredited standards

## ■ De-facto standards

- one vendor
- quick to market
- often private - not documented
  - example: Pitstop Certified PDF

## ■ Accredited standards

- developed through consensus process
- involvement of vendors and users
- independent testing
- user demand drives process

# Advantage of PDF/X vs. TIFF/IT

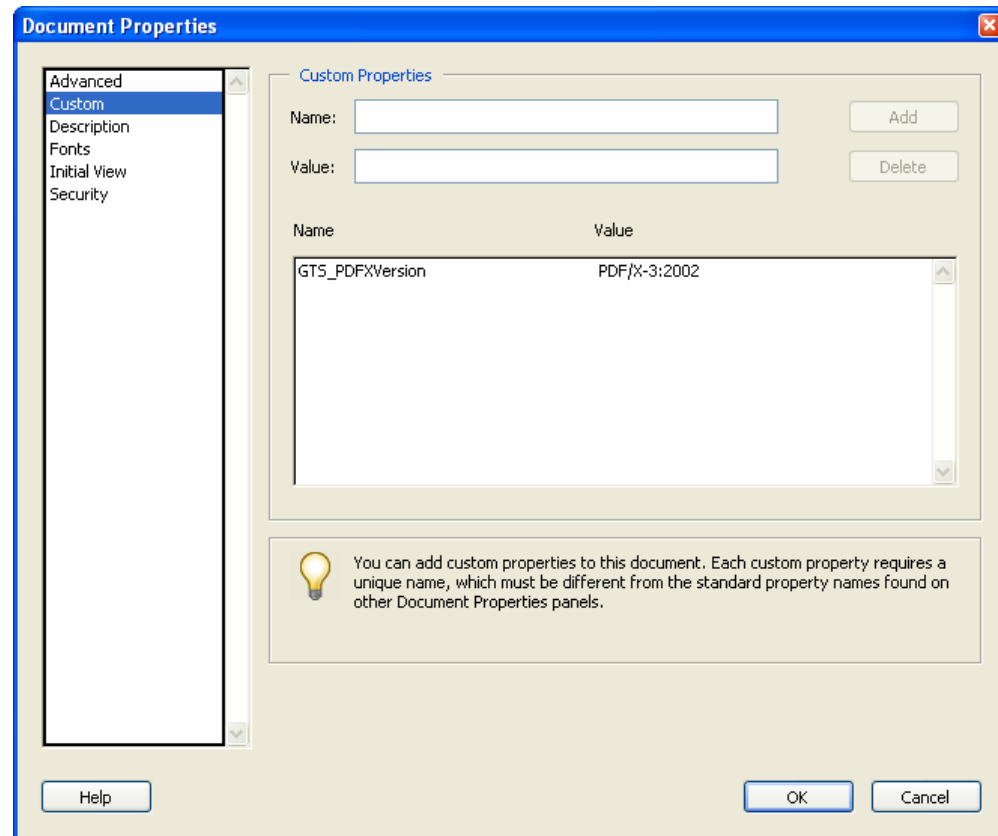
- **Designed to be predictable and reliable**
  - Ability to specify trim & bleed
    - ◆ Enables more intelligent and automated compositing & imposition
  - Ability to specify the intended printing condition (eg. SWOP)
  - Ability to specify if the file has already been trapped
- **Cost effective solutions are available**
  - Including a freely available and widely deployed viewer
- **Smaller files (better compression: ZIP & JPEG)**
- **Support for spot colors (incl. Duotones & DeviceN)**
- **Searchable**
- **Support for late-stage editing**
- **Can be archived and repurposed**

# Possible Disadvantages of PDF/X

- You must have PDF-native workflow tools
- Printing may require PostScript Level 3
- Supports DeviceN colors (duotone, spot+process)
- Supports smooth shading
- Potential overprint issues
- Complex transition from CT/LW and separated workflows
- Quark trapping is lost
- Legacy DCS files have to be recomposed
- Can't address local policies such as image quality, no spot colors, no CID fonts, etc.
- "Valid" bleed and trim boxes can be useless

# How to check if file is PDF/X?

- Acrobat Document Properties
- Acrobat 6/7 Professional Preflight
- Apago Box Editor



# Tools for PDF/X

- **AdobePDF Printer & Acrobat Distiller 6/7**
  - Create PDF/X from any program/application
  - Can perform limited corrections on provided data to meet requirements
- **Acrobat 6/7 Professional**
  - Great new preflight engine based on Callas pdfInspector
  - NO support for fixing problems – just reporting!
- **Adobe InDesign 3/CS**
  - PDF/X compliance is one of the options when doing direct export
  - Illustrator and Photoshop in CS2 also support direct export
- **Mac OS X**
  - Quartz Filters allow for native creation of PDF/X-3 from any application in Mac OS X.

# 3<sup>rd</sup> Party options

- **Apago's PDF/X Checkup**
  - Acrobat plugin for PDF/X creation & validation
- **Callas pdfInspector**
  - Acrobat plugin for preflight & PDF/X creation
- **Enfocus PitStop**
  - Available in Desktop & Server versions
  - Provides “Action Lists” to fix most problems
- **Workflow solutions**
  - Creo, Dalim, etc.

# Tip for correcting viewing PDF/X

- I recommend Acrobat 5.05 or later
  - enable “Overprint preview”
  - disable “Use local fonts”
- One nice feature in Reader 7 is support for OverPrint Preview
  - So now even Reader users will see incorrect OP issues!

# PDF/X Futures

- **Currently a proposal for PDF/X:2006 has been put forth for discussion at ISO**
  - ISO Meeting in May in London
  - AIIM Meeting in May in Philadelphia
- **Some specific proposals**
  - Based on PDF 1.6
  - Layers
  - Transparency (PDF/X-2 & 3 only)
  - Improved metadata
  - PDF/X-1a -> PDF/X-1
  - Grayscale PDF/X1

# PDF/X Wrap Up

- **PDF/X is the standard for PDF distribution in the prepress community**
  - Many publications have already standardized on it for advertisement delivery
  - Many printers are now requesting that delivered material be PDF/X, so as to have some form of “guarantee of quality”
- **PDF/X was never designed to solve all the problems in creating good pdf files.**
  - there are simply too many different production requirements
  - provides a well designed baseline upon which other requirements can be added
    - ◆ PDF/X plus and Ghent Workgroup



# PDF/A

# What is PDF/A?

- A file format for the “Long Term Electronic Preservation” of multi-page documents that may contain a mixture of text, raster images and vector graphics.

# Long Term?

- **Court documents protect citizens rights**
  - Access is assured in trial courts for 20 to 40 years for the Judiciary
- **Permanent records held “until the end of the republic” by NARA (National Archives & Records Administration)**
  - NARA enables people to inspect for themselves the record of what government has done. NARA enables officials and agencies to review their actions and helps citizens hold them accountable for those actions. And NARA ensures continuing access to essential evidence that documents the rights of American citizens, the actions of Federal officials, and the national experience.

# Some previous attempts...

- **Federal courts began imaging for Central Violations Bureau in 1993 in TIFF**
  - Short retention schedule on tickets and citations (5 years)
  - National system allowed for electronic shipment of calendars and tickets
  - Operational savings from reduced number of files

# More History

- **Maritime Asbestos cases - 1995**
  - Format important and retention long-term
  - PDF chosen
  - Born electronic & paper
- **New York Southern Bankruptcy Pilot - 1996**
  - Multi-state filings
  - Sophisticated Court and Bar
  - Seven plus years of filings now in PDF
- **SEC/EDGAR allows submissions of PDF - 1999**

# So what is needed?

- A document format that
  - Can convey critical information
  - Can be rendered accurately
    - ♦ “consistent and predictable”
  - Can incorporate associated “marginalia”
    - ♦ Notes scrawled in the margins on documents
    - ♦ Today perhaps as version history, general metadata, etc.
- A definition of how retrieval devices (Readers) will behave

# What PDF/A does NOT solve...

- Hardware migration issues
- OS and application migration issues
- Document management system issues
  - Document access tracking
  - Digital Rights Management

# What is PDF/A?

- **International Standard**
  - Initiated by AIIM & the US Gov't (Courts, NARA, IRS)
  - Now in final approval by ISO
- **Based on Portable Document Format (PDF)**
- **Long term preservation of black and white and color compound documents as electronic data**
  - Combinations of character, raster, vector and other data
  - Preservation and retrieval of appropriate metadata
- **Initial version “emulates static paper”**
  - Plus annotations, signatures, marginalia and approvals

# Details

- **Based on PDF 1.4 & PDF/X (ISO 15930)**
  - All PDF/X documents can potentially be minimally conforming PDF/A documents without any changes
  - Reduces ambiguity between different vendor's implementations
- **More restrictive “coding” of PDF details**
  - Ensures less ambiguity when implementing
  - Improved searching/indexing
- **Removal of any complex or potentially confusing graphic concepts**
  - No transparency
  - Limited colorspace

# More Details

- **No Security/Encryption**
- **All data must be self-contained**
  - No external references
  - Fonts **MUST** be embedded(!!)
- **Limited annotation support**
  - No movies and sounds
  - No JavaScript
  - Links are stored but not to be executed
- **Digital Signatures are for verification and not authentication**
- **Metadata based on Adobe XMP**

# Two “levels”

## ■ Minimal Conformance

- Meet the standard/basic requirements

## ■ Full Conformance

- Improved searchability via Unicode mappings
- Comprehensive metadata requirements
  - ◆ Font data
  - ◆ Document “pedigree”
  - ◆ Audit trail

# Some PDF/A history

- **NARA announcement March 31, 2003**
  - PDF/A defining future technical standards
- **Lots of meetings here in the US to define the requirements and then start on the technical points to achieve them**
- **First presented to ISO in October, 2003**
- **Revised by US committee**
- **Represented to ISO, March, 2004 during AIIM**
- **Currently in final ISO approval (May 2005)**

# Where are we going?

## ■ 3<sup>rd</sup> party tools

- Verifiers
- Creators
- NONE at this time ☹

## ■ Acrobat Integration

- Available in Acrobat 7 as BETA (since standard not final yet)
  - ♦ PDF/A job options for Distiller
- Available in Adobe Creative Suite 2
  - ♦ “Save as PDF/A”
- Changes to viewer to meet PDF/A requirements
  - ♦ NOT present in Acrobat 7

# What can you do now??

- **Make good PDFs using high quality tools**
  - ALWAYS embed ALL fonts
- **Use preflight tools to check for PDF/X compatibility**
  - If you meet PDF/X-3 compliance, you are also PDF/A complaint
- **Try out the beta PDF/A support in Acrobat 7**
- **If creating PDF from scans, be sure to use OCR to enable searching/indexing**



PDF/E

# PDF/E Background

- Idea originated with customers and Adobe partners
- Problem
  - Cost of managing distribution and change throughout project/product development team
  - Lack of reliable engineering drawing representation for downstream lifecycle management
- Why PDF as a solution
  - Published specification
  - Selected as a standard within many organizations
  - Supported by many developers (1800+ in Adobe developer network alone)

# So what is it?

- **Addresses need for reliable exchange of engineering documentation**
  - Secure distribution of intellectual property
  - Reliable exchange and change management (multiple application types and platforms)
  - Reduces costs associated with paper (distribution as well as storage/archive)
  - References existing PDF 1.6 specification
- **Covers 3 primary areas:**
  - Compact, accurate printing of engineering drawings
  - Support for exchanging/managing annotation and comment data
  - Incorporation of complex data into PDF (3D, object-level data, etc.)

# Key Elements & Issues

- **Key Elements for Engineering Workflows**
  - CAD and GIS presentation
    - Layers
    - 3-Dimensional data and information
  - Redlining (or markup) capture and consolidation (i.e., electronic comments)
  - Security
  - Embedded files and multimedia
  - Complex data and document metadata
- **Optimized in order to ensure transmission over low performance platforms, low bandwidth and high latency communication lines commonly found in the engineering world.**

# More specifics for PDF/E

- **Based on work from PDF/A**
  - Goal to enable a PDF/E document to be easily converted to PDF/A for archival storage
  - Leverage work already done by that committee
- **However enables support for limited interactivity and multimedia content**
- **Support for new 1.6 features related to Eng.**
  - 3D
  - Measurements
  - Print Scaling

# Industry Involvement

- Actify
- Agile
- Army Corps of Engineers
- Autodesk
- Bentley Systems
- Boeing
- Caterpillar
- Documentum
- e3open
- GE
- Global Graphics
- Halliburton/KBR
- Hewlett Packard
- Honeywell
- Intel
- Layton Graphics
- NGA (National Geospatial – Intelligence Agency)
- Océ
- Parkview International
- PFS
- PTC
- Solidworks
- UGS PLM Solutions
- Xerox
- And more...

# Timeline

- AIIM/NPES issued press release announcing formation of PDF/E initiative –March 1<sup>st</sup>, 2004
- First meeting -March 10th, AIIMExpo2004
- Draft submitted to AIIM emTAG for balloting
- Next meeting: May 17 2005, AIIM Standards Week-Philadelphia

# PDF/Access

- A very recently started working group to address the concerns of the “differently-abled” community and their needs with PDF
- Beyond what PDF offers today in terms of accessibility – what can be done to improve support for assistance technologies
  - Screen readers
  - Braille printers
  - Audio styling
  - Other...

# Wrap Up

- PDF is a rich and powerful language in which many media types can be described
- But that richness, means the potential for inconsistency and incompatibility
- SO by creating subsets (dialects) of PDF for specific market segments – PDF becomes a more focused tool.
- Using the specific versions of PDF, when you can, is only a benefit for all involved.

# Questions and Answers

