

PDF / Archive

Preserving Electronic Assets

NACRC – Washington, DC

March 6, 2010

Betsy Fanning, AIIM

Agenda

- ▶ Background
- ▶ What is PDF/A?
- ▶ Proposed legislation and adoption



Background for PDF/A Judiciary Use Case

- ▶ Court documents protect citizen's rights
- ▶ Access is assured in trial courts for 20 to 40 years for the Judiciary
- ▶ Access is often time sensitive
- ▶ On-site courthouse storage not cost effective
- ▶ Court decisions are permanent records held “until the end of the republic” by the National Archives
- ▶ Document format conveys critical information, which must be rendered accurately
- ▶ Cases – New York Southern, Enron, etc.
- ▶ 20 years of filings are in PDF



Role of AIIM and NPES

- ▶ Joint sponsors of the US PDF/A committee
 - AIIM, Association for Information and Image Management
 - Secretariat to ISO/TC 171 and ISO/TC 171/SC2
 - Secretariat to US Technical Advisory Group (TAG) for ISO/TC 171
 - NPES, The Association for Suppliers of Printing, Publishing, and Converting Technologies
 - Secretariat to ANSI Committee for Graphic Arts Technologies Standards (CGATS)
 - Secretariat to US TAG for ISO/TC 130



Role of ISO

- ▶ ISO Joint Working Group (JWG) for PDF/A
 - ISO/TC 171 /SC 2, *Document management applications – Application issues*
 - ISO/TC 130, *Graphic technology*
 - ISO/TC 46/SC 11, *Information and documentation – Archives/records management*
 - ISO/TC 42, *Photography*



Business Need

- ▶ Cradle to grave
- ▶ Compatibility into future
- ▶ Migration
- ▶ Independence



Sustainable Formats

- ▶ NARA defines:
 - “...the ability to access an electronic record throughout its lifecycle, regardless of the technology used when it was originally created”
- ▶ Characteristics of Sustainable Formats
 - Published documentation and open disclosure
 - Widespread adoption and use
 - Self-describing formats
 - External Dependency
 - Impact of Patents
 - Technical Protection Mechanism



Electronic Document Types

- ▶ TIFF
- ▶ XML
- ▶ Native File Formats
- ▶ PDF



Electronic Document Types

▶ TIFF

- Well known
- Difficult to create digitally born documents
- Indexing documents may be difficult
- Access is through an index



Electronic Document Types

▶ XML

- Many schema exist
- Preserves content not the structure
- Depends on DTD for visual appearance



Electronic Document Types

- ▶ **Native File Formats**

- Several file formats
- May render differently depending on the device or platform used
- May be proprietary



Electronic Document Types

- ▶ PDF
 - Widely adopted
 - Feature rich
 - Reliable and secure
 - Maintains visual appearance



Why Standardize a Version of PDF

- ▶ PDF is powerful and flexible
- ▶ May be too flexible for some applications
- ▶ Restrict subset of PDF
- ▶ Need higher degree of reliability
- ▶ May want standard in hands of neutral non-commercial body – Internationally recognized standards body such as ISO
- ▶ Focus on archive needs of government, corporations, libraries
- ▶ Resolve issues with font embedding replacement



PDF/A

- ▶ PDF/A is intended to address three primary issues:
 - Define a file format that preserves the static visual appearance of electronic documents over time
 - Provide a framework for recording metadata about electronic documents
 - Provide a framework for defining the logical structure and semantic properties of electronic documents



ISO 19005-1:2005

- ▶ “This International Standard specifies how to use the Portable Document Format (PDF) 1.4 for long-term preservation of electronic documents”
 - Applicable to documents containing character, raster, and vector data
 - The standard does not address:
 - Processes for generating PDF/A files
 - Specific implementation details of rendering PDF/A files
 - Methods for storing PDF/A files
 - Hardware and software dependencies



The PDF/A standard

- ▶ Multi-part ISO International Standard
 - ISO 19005-1:2005, *Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1)*
 - Part 2 (19005-2) intended to bring PDF/A into conformance with ISO 32000
 - Part 3 (19005-3) Dynamic documents
 - And additional future parts, as necessary



PDF/A Conformance

▶ Two Conformance Levels

- PDF/A-1 a
 - Compliance with all requirements of 19005-1
 - Including those regarding structural and semantic tagging
- PDF/A-1 b
 - Compliance with all requirements of 19005-1 minimally necessary to preserve the visual appearance of a PDF/A file




Key Elements for Rendering

A PDF/A document must have:

- ▶ Embedded fonts and images
- ▶ Embedded graphics
- ▶ No audio or video
- ▶ No direct or indirect references to external resources
- ▶ Metadata
- ▶ No security or encryption



PDF/A Caveats

- ▶ However...
 - PDF/A is a file format standard
 - PDF/A *alone* does not guarantee preservation
 - PDF/A *alone* does not guarantee exact replication of source material
 - The intent of PDF/A is *not* to claim that PDF-based solutions are the best way to preserve electronic documents
 - But once you have decided to use a PDF-based approach, PDF/A defines an archival profile of PDF that is *more* amenable to long-term preservation
 - Governance is needed – Policies, procedures and quality auditing
- 

Proposed Legislation

- ▶ Alabama
- ▶ Alaska
- ▶ California
- ▶ Connecticut
- ▶ Florida
- ▶ Idaho
- ▶ Kentucky
- ▶ Massachusetts
- ▶ Missouri
- ▶ Nevada
- ▶ New York
- ▶ Ohio
- ▶ Wisconsin



PDF/A Adoption

- ▶ Europe
 - Standard eBilling (Organisation for Promotion of Automated Accounting)
 - Germany, France, Austria, Switzerland, Poland, Norway
- ▶ Brazil
- ▶ China
- ▶ MoREQ2
- ▶ NARA



California Proposed Legislation

- ▶ Update to existing legislation – ANSI or AIIM Standards
- ▶ Emphasis on trusted system/repository
- ▶ Electronic documents created or stored as official record
- ▶ Requires use of standards recommended in AIIM ARP-1
- ▶ Proposes use of PDF/A for electronic documents
 - Documents created after a date certain by entities subject to the regulations
 - Particularly when destroying the original document and digital born documents



Massachusetts Proposed Legislation

- ▶ Proposed changing the archiving standard to PDF/A
- ▶ Document scanning and conversion be done by Massachusetts small businesses
- ▶ Coordinated with an effort to reduce paper and accessibility
- ▶ Convert all microfilm records to PDF/A
- ▶ Proposed storing on DVD



NARA Guidelines

- ▶ PDF/A-1 compliance is not enough
 - Comply with NARA's transfer instructions for records in PDF
 - Provide transfer documentation
 - Must comply with image quality specifications for transfer of permanent records
 - Must use OCR processes that do not alter the original bit-mapped image



Sampling of PDF/A Products

- ▶ Acrobat 8
- ▶ 3-Heights PDF Producer
- ▶ LuraDocument PDF Compressor
- ▶ PDFlib 7, 3-Heights Image to PDF Converter
- ▶ DocBridge Mill, pdfaPilot, PDF Appraiser

More products:

<http://www.aiim.org> or <http://www.pdfa.org>



PDF Standards

- ▶ PDF/X, ISO 15930
 - Pre-press data exchange
- ▶ PDF/E (Engineering), ISO 24517-1
 - For engineering, architectural, and GIS documents
- ▶ PDF/E (Engineering), ISO/NWP 24517-2
 - Archive engineering, architectural, and GIS documents
- ▶ PDF/UA (Universal Access), ISO/CD 14289
 - Intended to address Section 508 concerns
- ▶ PDF Healthcare
 - Exchange of electronic health records
- ▶ PDF, ISO 32000-1 (ISO/CD 32000-2)
- ▶ PDF/VT, ISO 16612 (2 parts)
 - Variable data exchange
- ▶ PRC, Product Representation Compact (ISO/WD 14739-1)



Trusted Systems

- ▶ Technology – Hardware and media storage methodologies
- ▶ Records policies and procedures
- ▶ Auditable
- ▶ Unalterable state



Thank you

- ▶ Questions ???????
- ▶ Contact: Betsy Fanning at bfanning@aiim.org or 301-755-2682
- ▶ PDF Wiki - <http://pdf.editme.com>
- ▶ AIIM Standards - Twitter, Facebook, LinkedIn
- ▶ PDF Standards - <http://www.informationzen.org>

