



# PDF/A ISO-Standard

The new standard for long-term preservation?

Marc Straat

Standards Development Europe

DLM Forum meeting, Den Haag, 18 October 2004



# Agenda

- **Adobe and PDF**
- **PDF – An Introduction**
- **PDF – DeFacto standard and beyond...**
- **Requirements for Electronic Archives**
- **PDF/A - Archive file format ISO standard**



# Adobe's philosophy

"We were never in this for financial gain. We wanted to have an impact. We wanted to make a difference in the world."

Chuck Geschke, Adobe Co-founder and  
Chairman of the Board



Adobe PostScript 3

TIFF





# PDF – Portable Document Format

- **View and print on any platform**
  - UNIX/LINUX, Mac OS, Microsoft Windows, Symbian
- **Preserves fonts, images, graphics, and layout of any source document**
- **Uses the model of free reader with ‘paid’ writer**
- **PDF files are compact and complete, and can be shared, viewed, and printed by anyone with free Reader software**



# PDF – Main Features

- **Free reader to view a PDF file**
  - More features through Adobe server products
- **Tagged PDF files contain information on content and structure**
  - To repurpose content, e.g. Mobile Devices
  - Accessibility - Screen readers
- **Digital rights to Document / Digital Signatures**
  - Document control and fidelity
  - Secure distribution and exchange of electronic documents and forms

# The PDF Standard Revolutionized Document Sharing

- De facto standard
- Universal file format
  - Across platform/application
- Specs are public
- High volume of free readers
  - 700 million
  - Bundled with new PC's
- Appr. 1800 OSS/commercial developers





# What are Standards?

- **Definition of file format, form, or process**
- **Guidelines for submitting information electronically**
- **Three categories:**
  - **De jure:** developed by industry standards groups
  - **De facto:** driven by market adoption
  - **Mandated:** controlled by governments, corporations, or regulatory authorities





# Standards Benefit Business and Government

- **Facilitate interoperability**
  - **Streamline regulatory compliance**
  - **Align software development with industry need**
  - **Raise quality, reliability, efficiency levels**
- 



# PDF standards: Beyond DeFacto...

- **PDF/X - ISO 15929, 15930-1, 15930-3**
  - 15929: Exchange of digital advertising information
  - 15930-1: Using CMYK data
  - 15930-3: Complete exchange for color-management workflows
  - PDF/X-2: under development – replacement of images and XMP
- **PDF/A - ISO/TC171/SC2 N 226 E**
  - Long term preservation and archiving of digital records
- **PDF/E – Under development**
- **PDF/Access – WG launched 3/04**
- **PDF/is—IEEE—Image-Streamable Internet Fax Transmission**
  - Transmission of high quality image-only documents to and from internet fax devices



# Archiving Issues

- **How do you preserve an electronic document today or 30 years from today?**
- **How do you preserve both paper-based and electronic records in a consistent format?**
- **How do you provide consistency in the integrity of your archives?**
- **How do you ease the search for an archived document?**

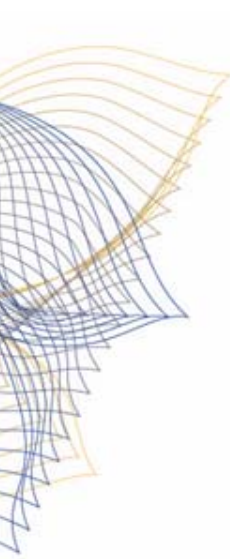




# Which archiving file format?

## Other formats and technologies in use as standards

- TIFF
- SGML
- HTML
- ASCII
- PDF
- But also: word processing formats, spreadsheet formats, etc



# Scope of PDF/A

- **International Standard specifies the use of the Portable Document Format (PDF) suitable for the long-term preservation of electronic documents.**
  - Based on the business and technical needs of governments, regulated industries, corporations, educational institutions and libraries

# ISO/TC171/SC2 – Document Imaging Applications



**Proposal** (to be completed by the proposer)

**Title of proposal** (in the case of an amendment, revision or a new part of an existing document, show the reference number and current title)

English title      **Document management - Long-term electronic preservation - Use of PDF (PDF/A)**

French title  
(if available)

## **Scope of proposed project**

This International Standard specifies the use of the Portable Document Format (PDF) for the long term preservation of black and white and color compound documents as electronic data. Compound documents may contain combinations of character, raster, vector, and other data. This International Standard also specifies methods for creation from these data of an exact visual reproduction of the document as it appeared at the time it was submitted for preservation. It also enables the preservation and retrieval of appropriate metadata.



# ISO PDF/A Participants and Process

- **Joint initiative of accredited standards bodies**
  - AllIM International (co-ordination)
  - NPES
- **Working group with many vendors, universities, and government agencies, e.g. UK, US, Sweden, Japan, Australia, France**
- **International Organization for Standardization (ISO)**
  - ISO work initiated Fall 2002
  - ISO Draft International Standard (DIS) March 2005



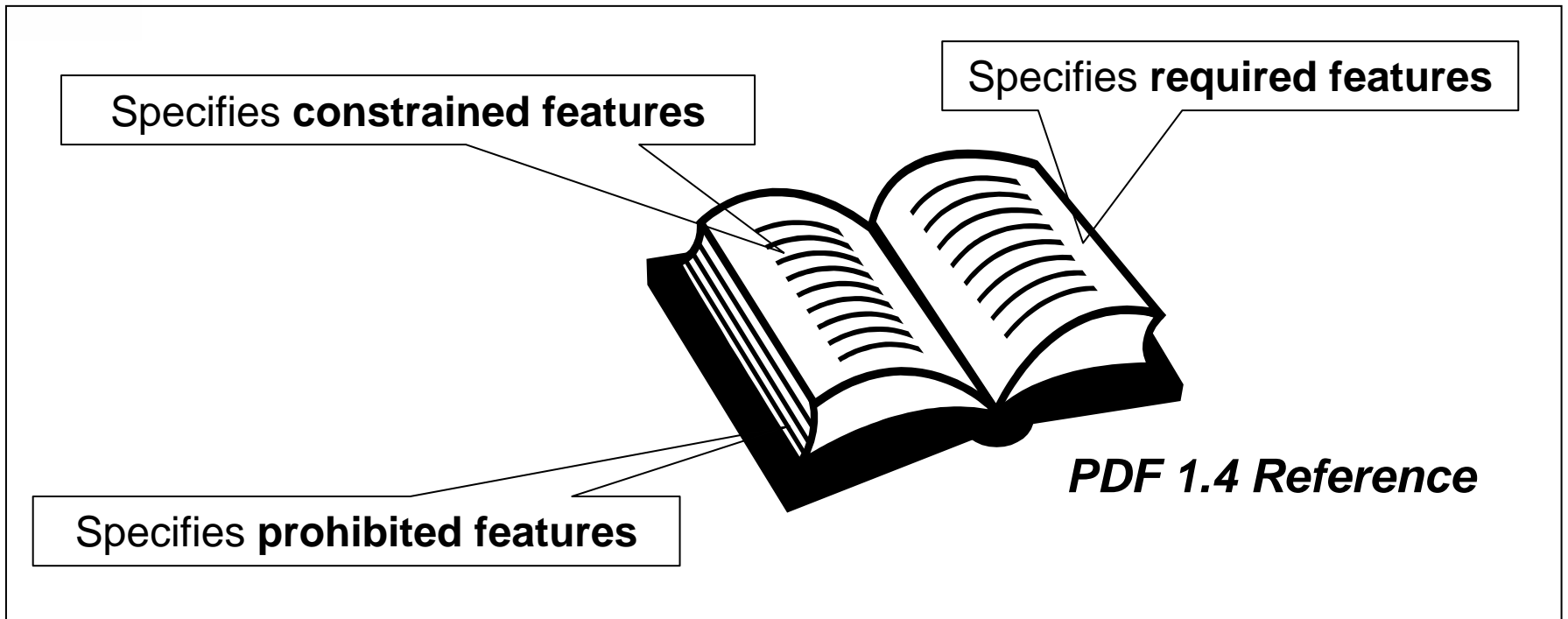


# Why a “Standard” Version of PDF

- PDF is too powerful and flexible
- Higher degree of reliability than required by the published specification
- Compatibility into the future
- Reliable migration
- Developed and maintained by an external organization

# PDF/A Specifies

- **Subset of the PDF Reference Version 1.4**



**PDF/A**



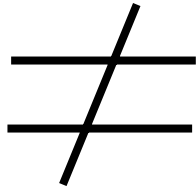
# Requirements of PDF/A

- **Constrains, requires, prohibits features**
  - No external resources needed
  - All fonts embedded
  - No encryption
  - No embedded executables
  - Meta data
- **Guidance documents**
  - Business overview
  - Technical overview
  - Guidance for Industry
  - Implementation documents

# Scope of PDF/A

- PDF/A does not address specific physical methods of storing these documents such as the media and storage conditions, required computer hardware and/or operating systems

**PDF/A**





# Other ISO-PDF Standards Initiatives

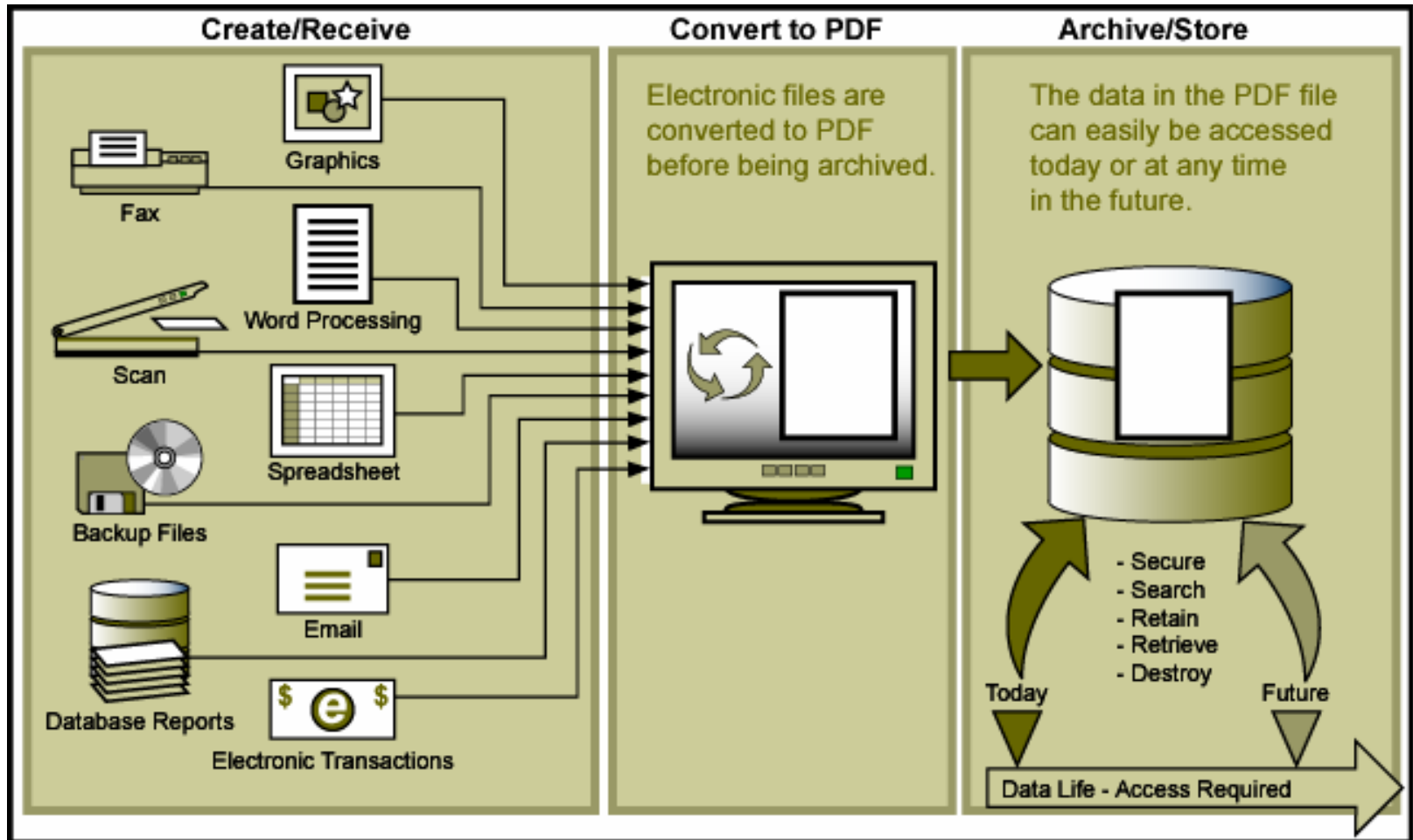
## ■ PDF Access

- Standard to define the required properties of a PDF file for supporting accessibility and third party assistive technologies (e.g. screen readers)
- Launch March 2004 - AIIM Standards Week NY
- NFB, RNIB, BSI Germany, Section 508 Coordinators, FreedomScientific

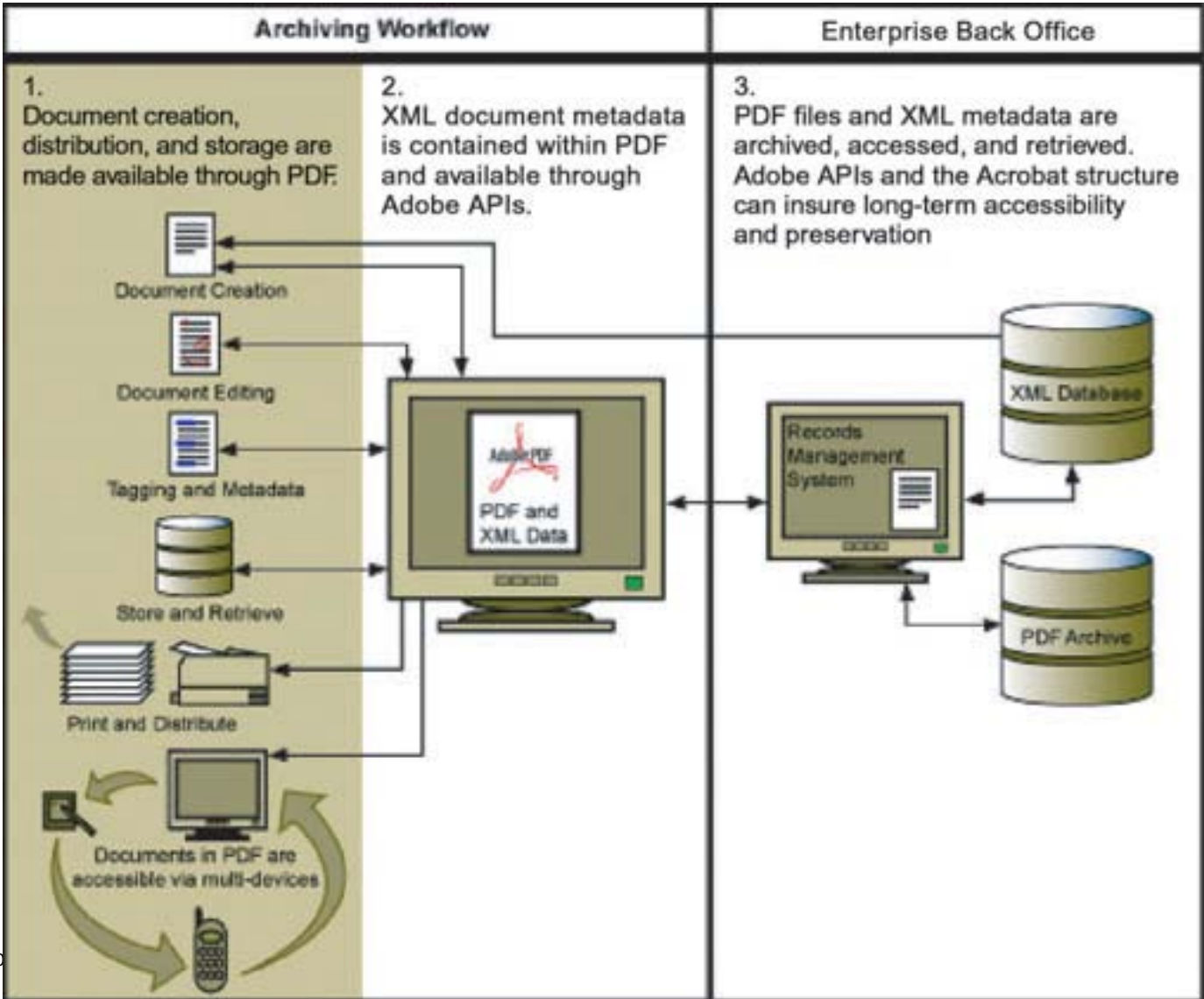
## ■ PDF/E

- Reliable exchange of engineering documentation
- Launch March 2004 – AIIM Standards Week NY
- Army Corps of Engineers, Bentley, Boeing, Halliburton, HP, Intel, NIMA

# PDF/A – Conversion, Archive, Store



# PDF/A in an archiving system environment





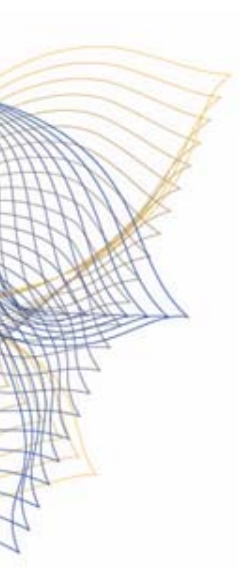
# Summary

- **PDF/A represents not only the data contained in the document, but also the exact form of the document**
- **A PDF/A file can be viewed without the originating application**
- **Document Fidelity and Accessibility through XML Metadata (RDF-XML, XMP)**
- **PDF/A has a natural connect to an ERMS**



# For More Information

- **PDF/A standard Web site:**
  - [www.aiim.org/pdf\\_a](http://www.aiim.org/pdf_a)
- **Contacts:**
  - Marc Straat  
Standards Development Europe  
[mstraat@adobe.com](mailto:mstraat@adobe.com)
  - Colin van Oosterhout  
Senior consultant  
[cvanoost@adobe.com](mailto:cvanoost@adobe.com)





# Archiving

- **Create documents, archive documents**
- **No consistent platform for paper and electronic archiving**
- **Archiving Electronic Documents**
  - Emulation
    - Use the originating application and its specific release
  - Migration
    - With every new release of the originating application, the electronic document will be upgraded



# Archiving Requirements Addressed by PDF/A

- Engineered for reliable migration and future compatibility
- Developed in cooperation with potential users, government agencies, and technology companies





# ISO PDF/A Standards Process & Timeline

- **Work organized by accredited standards bodies**
  - AllM International (the Association for Information and Image Management)
  - NPES (The Association for Suppliers of Printing, Publishing and Converting Technologies)
- **International Standards Organization (ISO) status**
  - March 2003 Issued as a New Work Item (NWI)
  - October 2003 ISO Working Draft Issued December 2003 ISO Committee Draft (CD)
  - 31 March 2004 ISO International Meetings (New York)
  - April 2004 second ISO CD
  - 4-5 October 2004 Int'l Meeting, BSI/National Archives, London
  - Target ISO ratification March 2005