EDTA, iText and INBATEK Conference

Bangkok, July 27, 2017
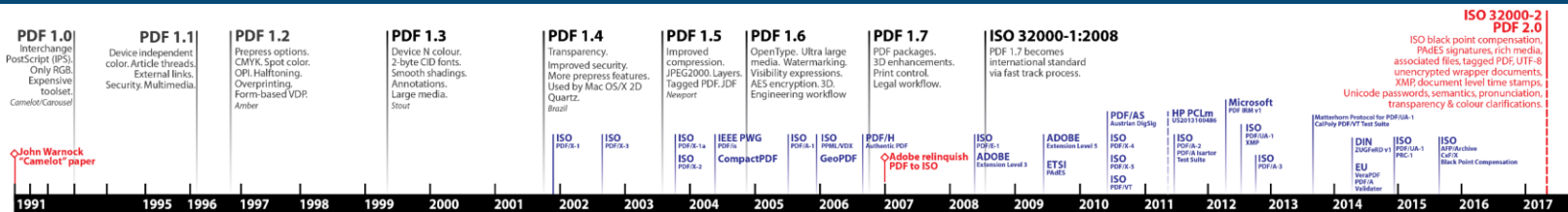
How standards drive business

- History of PDF
- Umbrella of standards
- Focus on: PDF/A, PDF/UA, PAdES, PDF 2.0, next-generation PDF

Speaking the same language

≡ Not being able to understand each other is a punishment, NOT a business model!

≡ Standards are about speaking the same language!

iTEXT

History of PDF

Version	Date	# pages	Content
Adobe PDF 1.0	June 1993	230	43 tables, 42 figures
Adobe PDF 1.1	23 January 1996	302	20 references
Adobe PDF 1.2	12 November 1996	394	137 tables, 86 examples
Adobe PDF 1.3	July 2000	696	223 tables, 73 figures
Adobe PDF 1.4	December 2001	978	277 tables, 20 color plates
Adobe PDF 1.5	August 2003	1172	333 tables, 70 figures
Adobe PDF 1.6	November 2004	1236	370 tables, 80 figures
Adobe PDF 1.7	October 2006	1310	389 tables, 98 figures
ISO 32000-1:2008 (PDF 1.7)	1 July 2008	756 (A4)	78 Normative References
ISO 32000-2:2017 (PDF 2.0)	2017	970 (A4)	5836 "shall", 411 "should"

PDF: an umbrella of standards
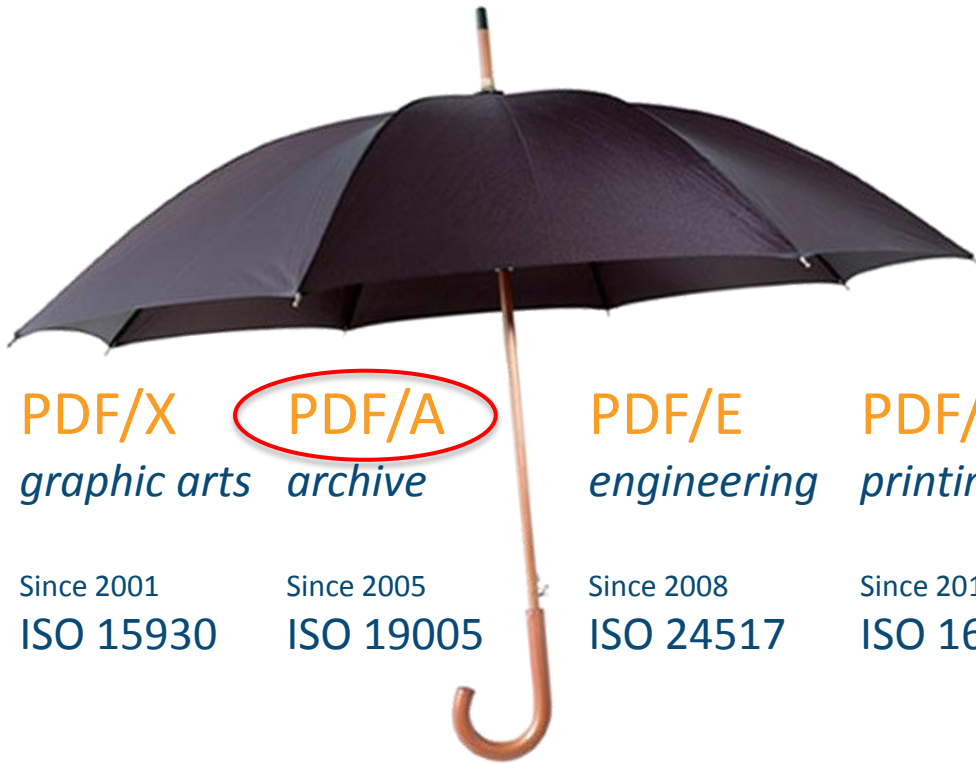
PDF

Portable Document Format
First released by Adobe in 1993
ISO Standard since 2008
ISO 32000

PDF/X	PDF/A	PDF/E	PDF/VT	PDF/UA	Related:
graphic arts	*archive*	*engineering*	*printing*	*accessibility*	• EcmaScript (ISO)
					• PRC (ISO)
Since 2001	Since 2005	Since 2008	Since 2010	Since 2012	• PAdES (ETSI)
ISO 15930	ISO 19005	ISO 24517	ISO 16612	ISO 14289	• ZUGFeRD (DIN)

ITEXT

PDF/A

ISO 19005: long-term preservation

Goals and concept

- ISO-19005,
    - Long-term preservation of documents,
    - Approved parts will never become invalid,
    - Individual parts define new, useful features.
- Obligations and restrictions,
    - Metadata: ISO 16684, eXtensible Metadata Platform (XMP),
    - The document must be self-contained:
        - All fonts need to be embedded,
        - No external movie, sound or other binary files.
    - No JavaScript allowed,
    - No encryption allowed.

iTEXT

Versions

≡ PDF/A-1 (2005):

    ≡ based on PDF 1.4,

    ≡ Level B ("basic"): visual appearance,

    ≡ Level A ("accessible"): visual appearance + structural and semantic properties (Tagged PDF).

≡ PDF/A-2 (2011):

    ≡ Based on ISO-32000-1,

    ≡ Features introduced in PDF 1.5, 1.6, and 1.7:

        ▪ Added support for JPEG2000, Collections, object-level XMP, optional content,

        ▪ Improved support for transparency, comment types and annotations, digital signatures.

    ≡ Level U ("unicode"): visual appearance + all text is in Unicode.

≡ PDF/A-3 (2012):

    ≡ Based on PDF/A-2 with only 1 difference: attachments do not need to be PDF/A.

ITEXT

PDF/UA

ISO 14289: Universal Accessibility

Accessible PDF?

How standards drive business

ITEXT
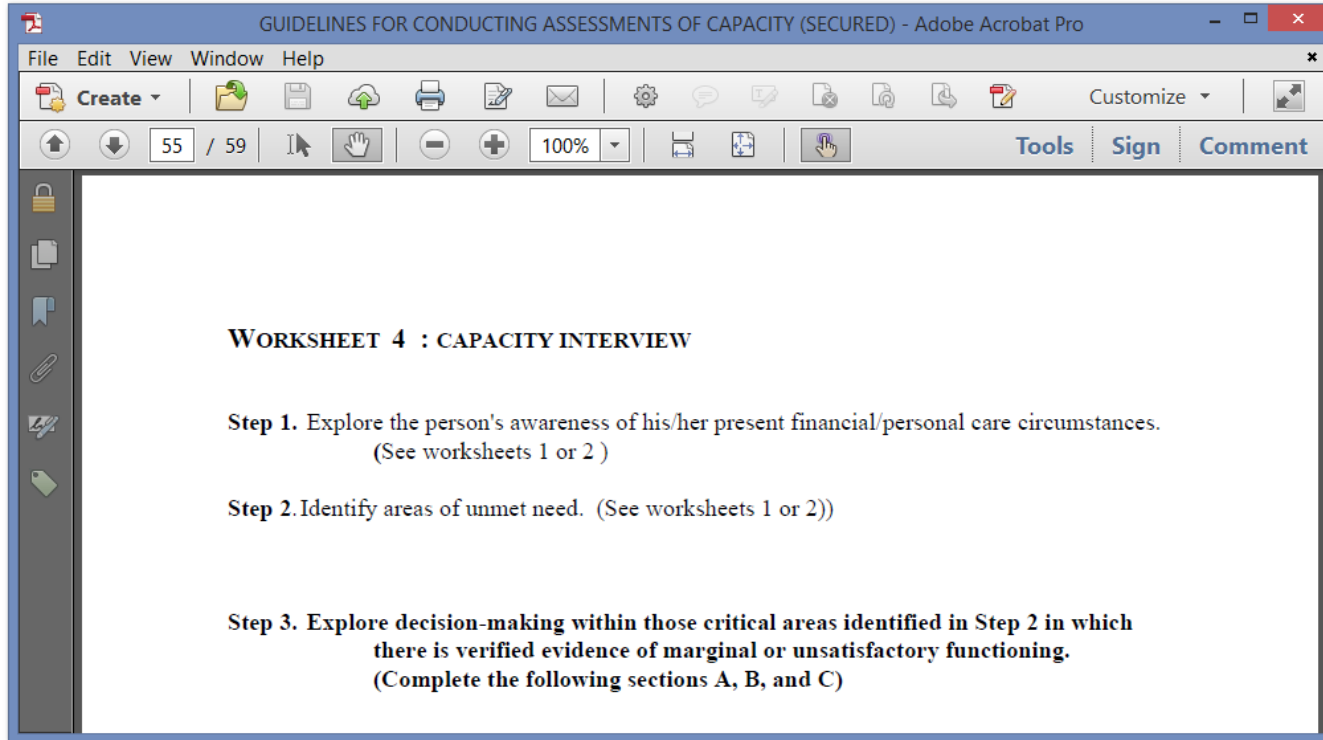
Structure and Semantics

How standards drive business

© 2015, iText Group NV, iText Software Corp., iText Software BVBA

How to read this chart?

How standards drive business

Structure of the data

What goes into rows / columns?

Resolve abbreviations

Make info color independent

	Functional Leadership	Risk Management	Structured Finance	Mergers & acquisitions	Governance & Internal Control	Accounting Operations	Treasury operations	Management Information & Business Decision Support	Business Planning & Strategy	Finance Contribution to IT Management	Commercial Activities	Taxation
Person	2	1.7	1.4	1.3	1.9	2.1	0.8	2.2	1.8	1.6	1	1.4
Position	3	2.3	2	1.8	3.2	3.9	2.1	3.1	3.2	2.3	2	2.1

ITEXT

Is this a better way to read the data?

	Person	Position
Functional Leadership	2	3
Risk Management	1.7	2.3
Structured Finance	1.4	2
Mergers & Acquisitions	1.3	1.8
Governance & Internal Control	1.9	3.2
Accounting Operations	2.1	3.9
Treasury Operations	0.8	2.1
Management Information & Business Decision Support	2.2	3.1
Business Planning & Strategy	1.8	3.2
Finance Contribution to IT Management	1.6	2.3
Commercial Activities	1	2
Taxation	1.4	2.1

ITEXT

Concept

- **ISO 14289: Universal Accessibility**
  - PDF/UA is a technical specification intended for developers implementing PDF writing and processing software,
  - PDF/UA provides definitive terms and requirements to allow people with/without disabilities the same rights,
  - For those equipped with appropriate software, conformance with PDF/UA ensures accessibility for people with disabilities who use assistive technology such as screen readers, screen magnifiers, joysticks and other technologies to navigate and read electronic content.

iTEXT

PAdES

ETSI TS 102 778: PDF Advanced Electronic Signatures

Integrity

How standards drive business

Authenticity

How standards drive business

© 2015, iText Group NV, iText Software Corp., iText Software BVBA

Non-repudiation

How standards drive business

Business requirements for signatures

One author signature, multiple approval signatures
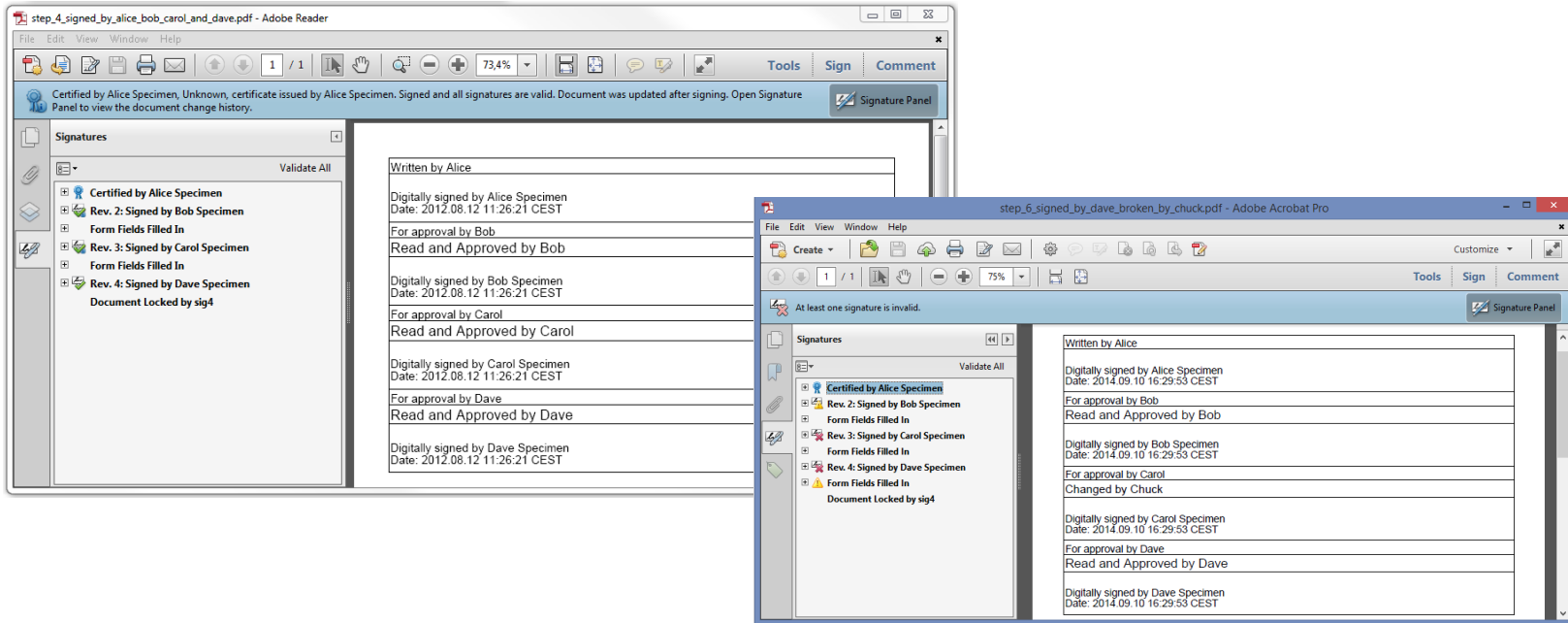
Chronology of the standards

- ISO:
  - ISO-32000-1 (2008) based on PDF 1.7 (2006),
  - ISO-32000-2 defines PDF 2.0 (2017).
- ETSI: TS 102 778 (2009 - 2010):
  - PAdES 1: Overview,
  - PAdES 2: Basic – CMS based (ISO-32000-1),
  - PAdES 3: Enhanced – CAdES based (ISO-32000-2),
  - PAdES 4: LTV – Long Term Validation,
  - PAdES 5: XAdES based (XML content),
  - PAdES 6: Visual representation guidelines.
- ETSI: TS 103 172 (2011 - 2013):
  - PAdES Baseline Profile.

iTEXT

PDF 2.0

ISO 32000-2: New version of the core PDF standard

Cleaning up ISO 32000-1

≡ Detailed technical review,

≡ Provide clear references to other specifications

  ≡ Refer to newer versions of other specifications (e.g. Unicode 4.0 -> Unicode 8.0),

  ≡ Avoid duplication of information available in new or existing RFCs, ISO standards, etc. (e.g. XMP is an ISO standard since 2012),

  ≡ Ensure that all reference materials are available (e.g. no more links to a web sites that no longer exist).

≡ Open vendor-neutral, non-proprietary standardized technologies,

  ≡ E.g. remove all references to "Adobe" (no more "Reader-enabling").

≡ Correct typos, errors (e.g. ColorDodge and ColorBurn blend mode formulas) and inconsistencies (e.g. fix wrong examples).

iTEXT

Substantially rewritten for PDF 2.0

- 10.1 – 10.3: Rendering
- 11: Transparency
- 12.8: Digital Signatures
- 14.3: Metadata
- 14.8: Tagged PDF
- 14.9: Accessibility support

iTEXT

PDF 2.0 is more demanding

- File IDs are no longer optional, but required,
- Length (/L) required for inline images,
- Restricting the visual appearances of a signature field to one and only one,
- Every byte shall be signed by the signature (except for the signature itself),
- Appearance streams (/AP) are now required for annotations,
- Making optional entries in a Type 1 font dictionary required in PDF processors,
- Only 256-bit AES encryption is allowed,
- …

iTEXT

New features in ISO 32000-2

- Unencrypted wrapper,
- Rich Media annotations, 3D annotations (PRC), projection annotations,
- Geospatial features,
- PAdES functionality (CAdES signatures, DSS, LTV),
- Namespaces for Tagged PDF,
- Pronunciation hints,
- Document parts, fragment identifiers,
- Associated files,
- …

ITEXT

Deprecated in PDF 2.0

- The XML Forms Architecture (XFA)
- "NeedAppearances" (Appearances are mandatory)
- Movie and sound annotations (replaced by Rich Media annotations)
- Info dictionary (replaced by XMP)
- Assistive technology restrictions via DRM
- PostScript aspects: ProcSet, PS Xobject
- CharSet for Type 1 fonts; CIDSet for CID fonts
- Page-piece dictionaries
- OS-specific file specifications
- SHA-1 for security; all encryption algorithms except AES-256
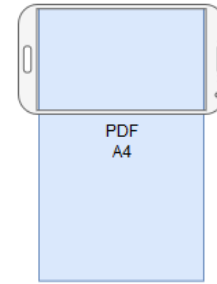- Security handlers adbe.pkcs7.sha1 and adbe.x509.rsa_sha1
- …

ITEXT

Next-generation PDF

What comes next?

A challenge: PDF on devices

How standards drive business

The problem with PDF on devices

How standards drive business
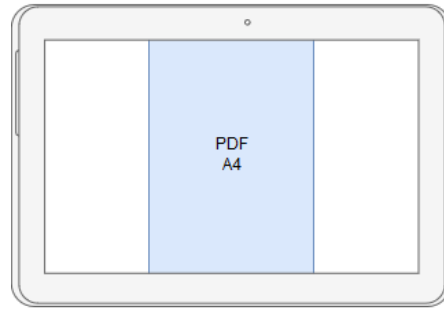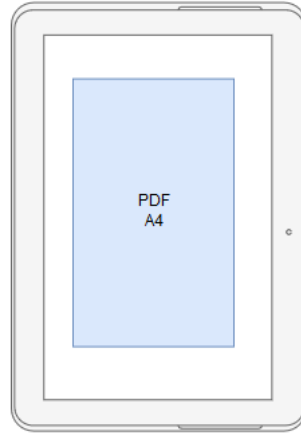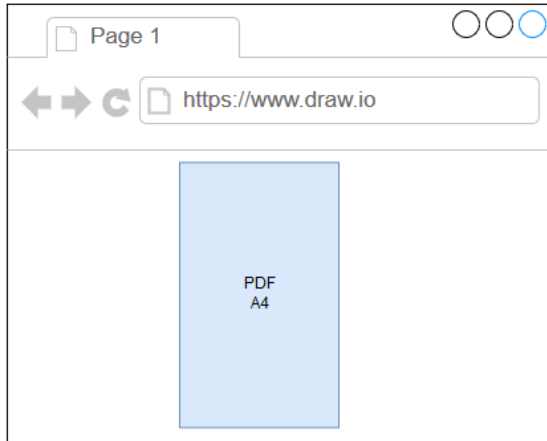
Technical Working Group

- Organized by the PDF Association,

- Different participants work together on the spec:

  - Adobe

  - iText

  - Callas

  - Accessibil-IT

  - Dual Lab

  - …

- Broad and open industry participation.

Premises and concepts

- Premises:
  - Fully compliant with PDF 2.0,
  - Relies on properly tagged PDF (the PDF needs to be accessible),
  - Preserve the qualities of PDF: reliable, portable, self-contained, secure,…
  - Enable presentation adaption to the circumstances and device used to render the PDF,
  - Extend PDF into a variation of new and different workflows and use cases.
- Concepts:
  - Derivation from Tagged PDF: standardized model for representing PDF content as HTML,
    - Logical reading-order, Semantic appropriateness, Distinction from artifacts,
    - Using namespaces, standard structure elements (tags), unambiguous rules,
    - Associations with embedded or external files, e.g. for alternative representations, images,…
  - Support the Open Web Platform (OWP) technology in derived HTML content.

ITEXT

Advantages

- No more pinching and panning: easy derivation of PDF content to HTML5
- Device-optimized navigation: "responsive" PDF
- Rich user preferences in a device-optimized context:
  - Support for Media Queries: for screen (which size?), for print,…
  - Different CSS files can be applied to the same content.
- Search operates flawlessly,
- Introduction of improved interactivity in PDF:
  - E.g. traditional PDF form fields can be replaced by sliders, resizable text areas,…
- Using the best of both worlds: PDF and HTML5

iTEXT

Following Government rules

- Accessibility is required in the US, EU, Canada, Australia,…
- Currently, there's support for HTML in Assistive Technology, but much for PDF,
- PDF as a file format is still crucial in Government agencies,
- But governments also prefer "technology-neutral" specifications such as WCAG 2.0

- Users will have the choice: consume as PDF or consume as HTML,
- No need for the government to give up on PDF,
- For broad consumption, the derived (and accessible) HTML may be used.

ITEXT

Summarized

- Mobile-friendly representation of PDF documents,

- Improved navigation:

  - Richer table of contents,

  - Collapsible headings,…

- Improved Search capabilities,

- Interactive "smart" objects,

- Leverage embedded Associated Files.

iTEXT

Implementation

- Specification still under development,
- Desire to have at least three vendors with a commercial implementation,
  - iText wants to be one of these vendors,
- Ongoing iText projects with respect to Next-Generation PDF:
  - Working with Adobe to create well-tagged PDFs for testing purposes,
  - Working with Dual Lab to derive HTML from PDF,
  - Working on better HTML to PDF conversion (pdfHTML + the iText DITO project).

ITEXT

Thank you!