



národní
úložiště
šedé
literatury

PDF a ISO standardy pro dlouhodobé archivování

Střálka, Vladimír
2008

Dostupný z <http://www.nusl.cz/ntk/nusl-112152>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 02.07.2022

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

PDF a ISO standardy, dlouhodobé archivování

Vladimír Střálka

Sr. Territory Account Manager

Adobe Systems, 8.10.2008



Agenda

- Co je PDF
- Adobe a standardy
- PDF/A.....ISO Standard
 - Proč vznikly standardy PDF/...
 - PDF/A – o co vlastně jde?
 - PDF/A, A-1a, A-1b, A-2....jak se v tom vyznat
 - Použití PDF/A
 - Rozšíření PDF/A
- Jak archivovat v PDF/A
- Závěr

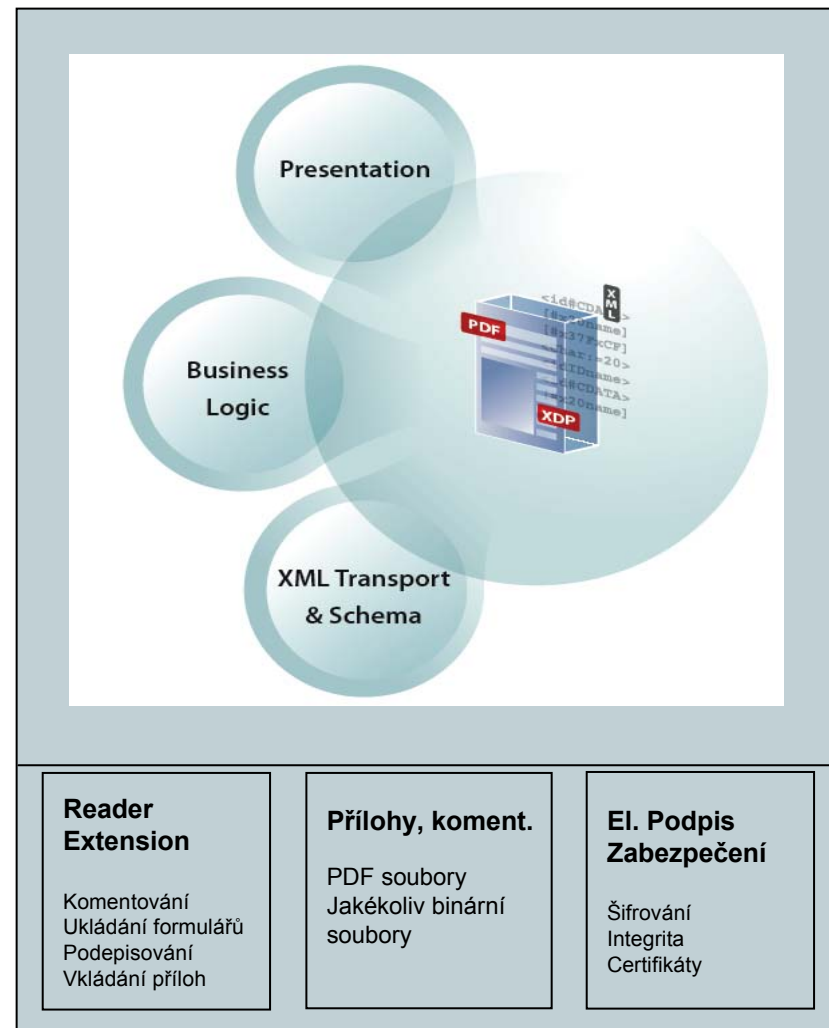
Elektronický dokument a požadavky na něj

- Technologická nezávislost - nezávislý na aplikaci
- Jasná, na technických prostředcích nezávislá interpretovatelnost obsahu
- Schopnost vyhovět legislativně
 - Není možné ohýbat právo dle potřeb technologií, technologie musí umět vyhovět právu
 - Důvěrnost
 - Oprávnění
 - Zodpovědnost
 - Integrita
 - Pravost
 - Nezpochybnitelnost



PDF – plnohodnotný elektronický dokument (spis)

- **Prezentační vrstva**
 - Plnohodnotná presentace dokumentu tak, jak vypadá v papírové formě
 - **Nestrukturovaná data**
- **Business logika**
 - Programový kód – Java Script
 - Logika nad XML v dokumentu
 - Konektory na externí datové zdroje
- **XML Transport**
 - Možnost nést XML data pro integraci s dalšími systémy
 - **Strukturovaná data**
- **Ostatní**
 - Vrstva pokynů pro Reader
 - Poznámky, připomínky, připojené soubory
 - Vrstva bezpečnosti



Elektronický spis a jeho význam

- **Spis**
 - Obálka, nesoucí různé typy zaručených informací k danému případu
- **Jak dnes vypadá**
 - Soubor složek s množstvím dokumentů a dalších přiložených materiálů
 - Převážně papírové dokumenty
- **Proč je spis klíčový prvek pro sdílení a ukládání dokumentů**
 - Umožňuje sdílet a předávat informace důležité k rozhodování mezi různými organizacemi a držet je „pohromadě“
 - V souladu se zákony zajišťuje integritu a pravost informací, definuje zodpovědnost za informace a dává informaci o stavu věcí v čase
 - Archivační role
 - Má integrační roli

Co zajišťuje nový ISO standard

- Uchovává věrně vzhled dokumentu nezávisle na zařízení a platformě
- Spojuje obsah z více různých typů zdrojů
- Umožňuje sdílenou práci s dokumentem na různých platformách
- Podporuje digitální podpis z důvodu hodnověrnosti
- Zajišťuje zabezpečení a oprávnění k ochraně a řízení kontroly obsahu
- Zajišťuje přístup k obsahu lidem s různým postižením
- Umožňuje předávání obsahu mezi aplikacemi
- Vytváří zabezpečenou a důvěryhodnou obálku pro libovolná binární data

Využití PDF

- Publikace dokumentů
- Zaručené elektronické dokumenty
- Elektronický spis - složka dokumentů (ISO 32000)
- Plně interaktivní a inteligentní formuláře
- Digitalizace papírových dokumentů a následné OCR
- Archivace (ISO PDF/A)
- Ochrana důvěrných dokumentů proti zneužití



Pár příkladů

- PDF – typický dokument šedé literatury
 - Dokument v MS-Word
 - Standardní PDF
 - PDF/A verze
- PDF – Scannovaný dokument včetně OCR
 - Scannovaný dokument – PDF/A
 - Scannovaný dokument s OCR – PDF/A – Fulltext
- Vyšší verze PDF
 - Převod z CAD – 3D model
 - Převod z CAD – Mapy
 - Elektronický spis
 - Portfolio

Adobe a jeho cesta ke standardům

- Rozšíření PDF a jeho všeobecná akceptovatelnost
 - Univerzální elektronický dokument
 - Vizualizace + DATA
 - Zveřejněný dokumentový formát, mohou s ním pracovat všichni
 - Vysoké rozšíření
 - Akceptován v GOV na celém světě
 - 100ky mil. dokumentů na internetu
 - Skoro 1/3 dokumentového obsahu na internetu
- Široká podpora aplikací
 - Kancelářské balíky
 - Grafické a CAD aplikace
 - Výstupy aplikací různých výrobců



Co jsou standardy a kdo je vytváří

- Pravidla nebo předpisy, jak bude elektronická informace vizualizována, zaznamenána a archivována
- Definuje
 - Formát souboru
 - Formu interpretace nebo proces
- Standardy jsou klasifikovány do třech typů
 - **De jure** – vytvářen komerčními standardizačními instituty
 - **De facto** – vytvářen přirozenou pozicí na trhu
 - **Povinné, předepsané** – řízeny státními organizacemi a regulátory
- Národní úřady státní správy nebo jejich organizace
 - FDA, BUND, EC Archive, US Army Corps
 - Český normalizační institut
- Komerční organizace nebo jejich asociace
 - FSI: MISMO, PISCES, ACORD, XBRL
 - Výroba: RosettaNet
- Nezávislé standardizační organizace
 - ISO, ANSI, W3C
- Adobe je účastníkem v několika standardizačních skupinách
 - ISO, W3C, OASIS, XML.gov, etc.

Co jsou standardy PDF a proč vznikly

- PDF jako takové je vynikající pro digitální prezentaci na webu, jako náhrada papíru, jako složka na elektronické dokumenty
- Je schopno plnit požadavky na spolehlivý tisk a nezávislou interpretaci obsahu -> archivování
- Proto standardizace PDF
 - PDF/A (ISO 19005) – Dlouhodobé archivování
 - Spolehlivý formát pro archivaci a uchovávání
 - PDF/X (ISO 15930) – Tiskové publikování
 - Predikovatelný tisk digitálních souborů kdekoliv na světě
 - PDF/E (ISO 24517) – Inženýrská dokumentace (stavebnictví, průmysl, geografické aplikace)
 - Převádí komplexní technickou a výkresovou dokumentaci do jednotného tvaru
 - Veřejné informace jsou dostupné na AIIM
 - www.aiim.org/standards

PDF 1.7 - ISO 32000

- Elektronický spis



PDF/A

Archivace

ISO 19005

(PDF 1.4)

PDF/X

Grafická dok.

ISO 15930

(PDF 1.4 & 1.6)

PDF/E

Technická dok.

ISO 24517

(PDF 1.6)

PDF/UE

Dostupnost

AIIM proces

→ ISO

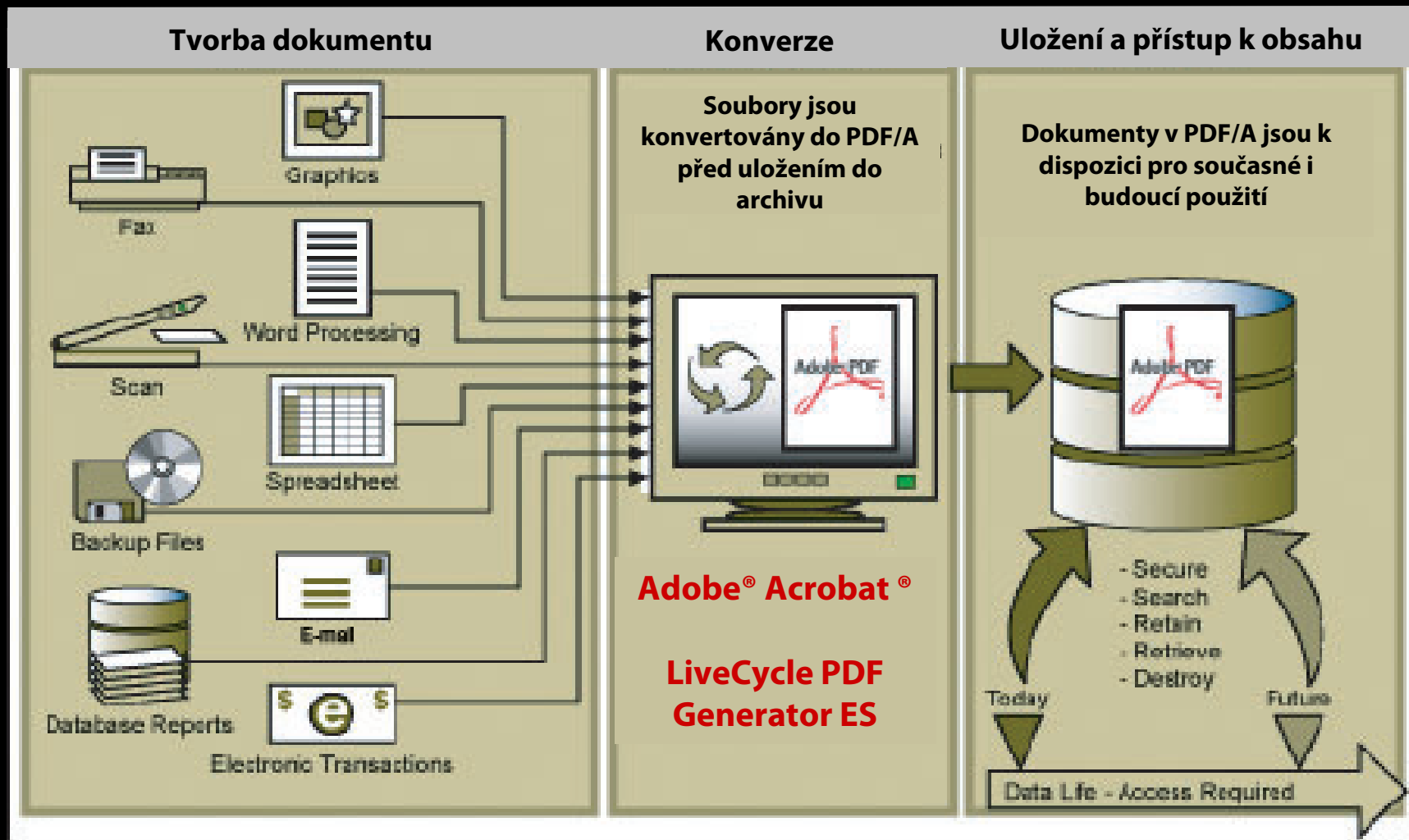
PDF/A, A-1a, A-1b, A-2....jak se v tom vyznat

- ISO 19005-1 definuje „souborový formát založený PDF, známý jako PDF/A“, který je:
 - Otevřený formát pro elektronické dokumenty
 - Nezávislý na zařízení a platformě
- PDF/A představuje mechanismus pro reprezentaci elektronického dokumentu způsobem jeho vizualizace
- Standard umožňuje interpretovat vizualizaci dokumentu tak, aby mohla být vytvořena nezávisle na čase
- PDF/A byl předložen jako řada definic skládající se z několika částí
- V současnosti byla certifikována pouze první část, tedy PDF/A-1
- Tato definice ještě byla definována ve dvou úrovních
 - PDF/A-1a – ukazuje plnou shodu se současně schváleným PDF/A ISO
 - PDF/A-1b – zajišťuje pouze dlouhodobou reprodukovatelnost
 - Hlavní rozdíly jsou ve schopnosti extrakce textu
 - PDF/A-1a – zajišťuje, že extrahovaný text bude čten v přirozeném pořadí (Section 508 of the US Rehabilitation Act), čitelnost na PDA a jiných zařízeních, s malým display (tagged PDFs) nebo pro zdravotně postižené
 - PDF/A-1b – zajišťuje, že dokument bude korektně zobrazen
- PDF/A-2 je nový návrh normy, obsahující nové vlastnosti s plnou kompatibilitou normy PDF/A-1

Použití PDF/A

- PDF/A ISO 19005-1 standard lze stáhnout z www.iso.org/iso/
- Samotná norma není dlouhá, ale odkazuje na další definice fontů, XML, atd. takže nakonec jde o asi 1000 stran definice
- Nástroje pro tvorbu a validaci PDF dle standardu jsou na trhu od roku 2006
 - Adobe nabízí ucelenou sadu produktů zajišťující plný převod dokumentů nebo jejich transformaci do souladu s normou v předarchivačním procesu
 - Acrobat 7.x a vyšší
 - PDF Generator
 - Existují i nástroje třetích stran zajišťující tvorbu PDF
- PDF/A není celé řešení
 - Samo o sobě nevytváří archivní řešení
 - Je potřeba jej zakomponovat do celkového řešení, které se sestává i z technologií ostatních výrobců

Řešení archivu s využitím PDF/A



Ukázka tvorby PDF/A a verifikace PDF/A

ARCHIVY

- Předepsáno
 - Victoria Archives – Australia
 - National Archives – Holandsko
 - Library of Congress – USA
 - State of Florida Archives – USA
- De Facto
 - National Central Library – Taiwan
 - National Archives – Norsko
- Akceptován
 - Govt Archiving – Itálie
 - National Archives – Norsko
 - National Archives & Records Administration (NARA) – USA

Ostatní instituce v EU

- Předepsáno
 - Ministerstva financí a spravedlnosti – Itálie
 - Ministerstvo obrany – Francie
 - Ministerstvo vnitra – SRN
- De Facto
 - Ministerstvo financí – Francie
- Akceptován
 - European Commission
 - Ministerstvo zahraničí – Francie
 - eDay – Dánsko

Závěr

- Standardizace PDF pokračují – specifikace PDF 1.7 byla dodána do AIIM ke standardizaci
 - <http://www.adobe.com/aboutadobe/pressroom/pressreleases/200701/0129070penPDFAIIM.html>
- Informace
 - PDF standardy: www.aiim.org/standards
 - Adobe: www.adobe.com/standards
 - PDF/A: www.pdfa.org
 - ISO: www.iso.org



Revolutionizing
how the world engages
with ideas and information

