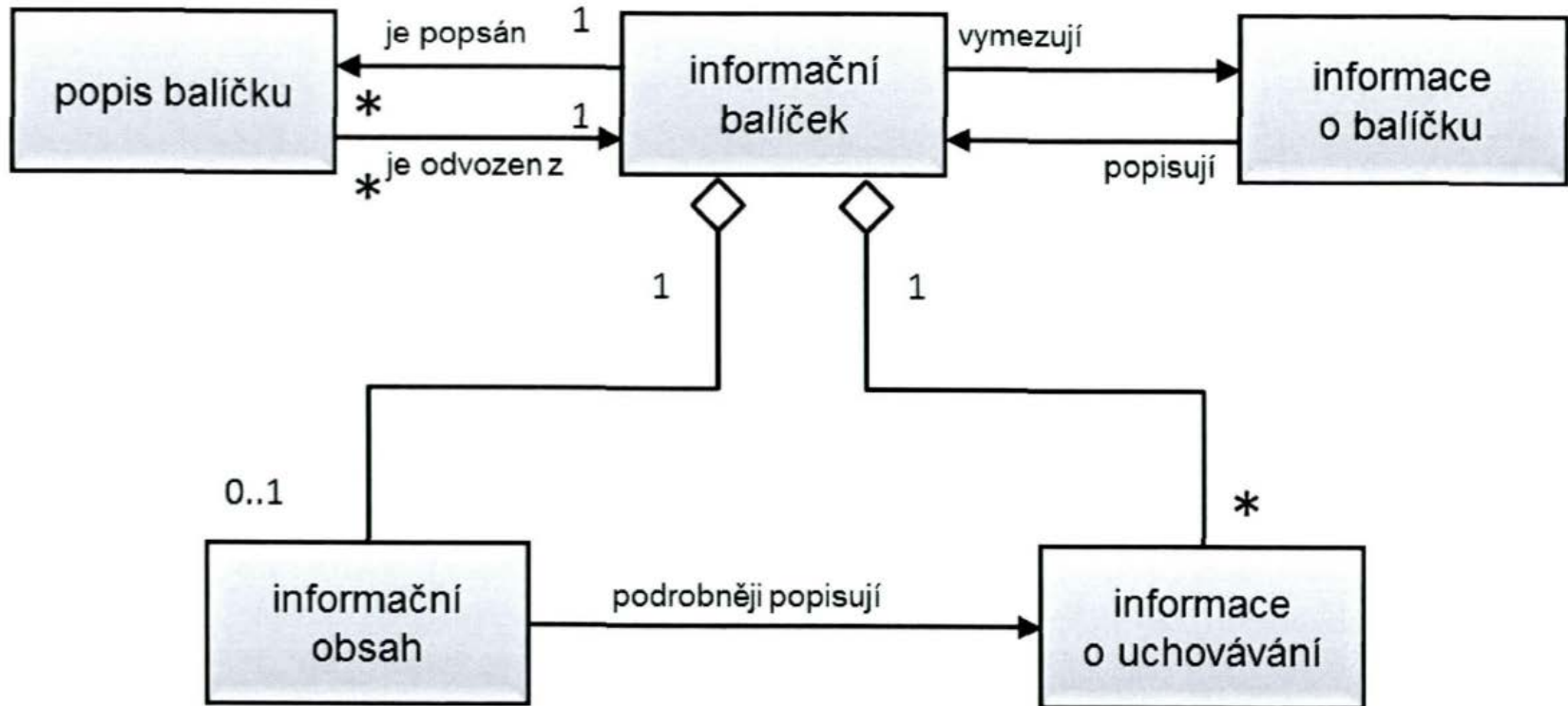


# Formáty a dlouhodobé uložení: Identifikace, extrakce a validace

# Informační balíčky

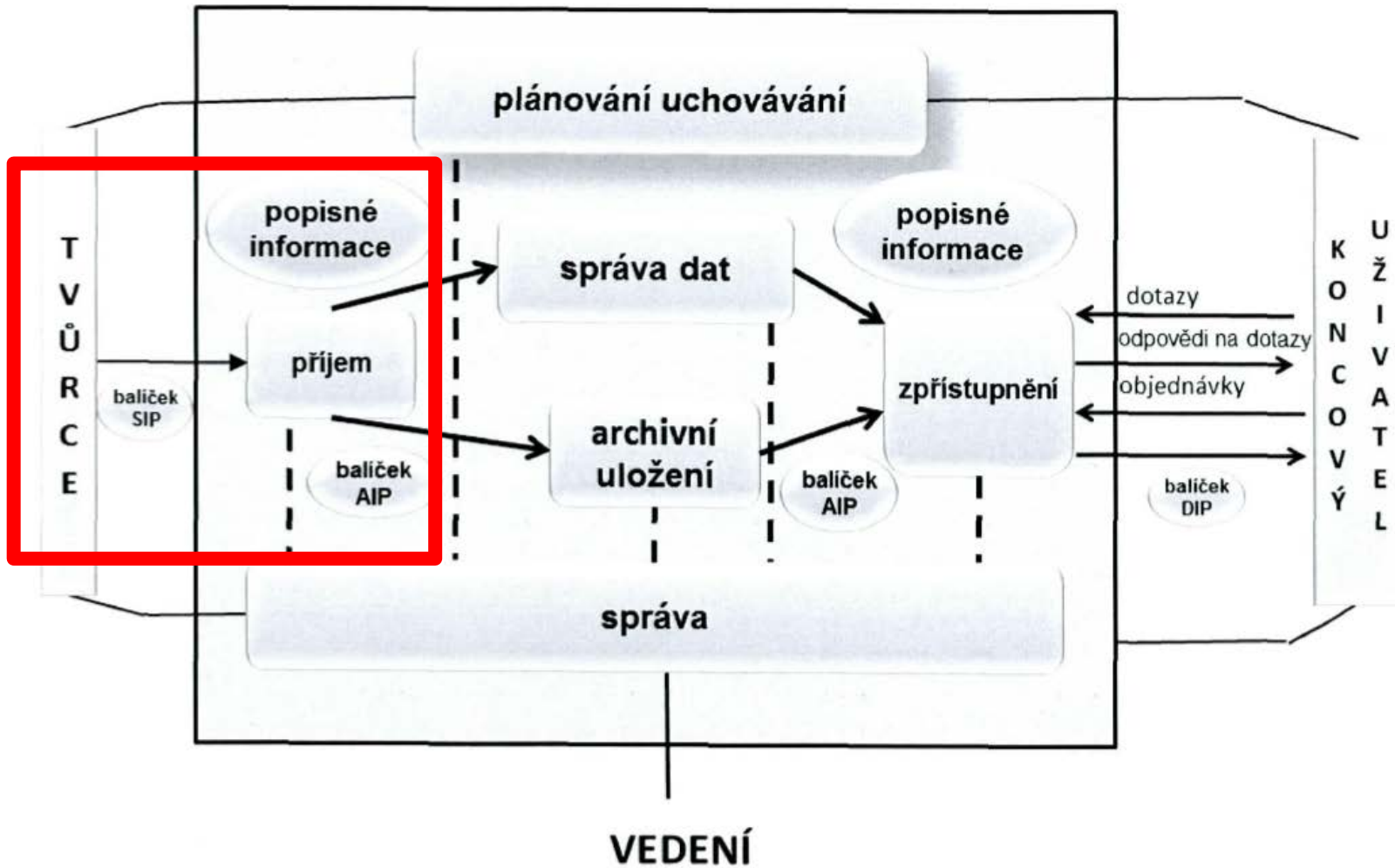
ČSN ISO 14721



Obrázek 4.13 – Obsah informačního balíčku

# Funkční celky digitálního archivu

ČSN ISO 14721



Obrázek 4.1 – Funkční celky archivu OAIS

# Příjem

---

- služby a funkce k balíčům SIP (od tvůrců, tj. původců nebo od vnitřních prvků)
- zajišťuje:
  - a) příjem balíčku SIP
  - b) příprava obsahu na uložení a správu v archivu
  - c) provedení kontroly kvality balíčku SIP
  - d) vytváření archivního informačního balíčku AIP
  - e) výběr popisných informací z balíčku AIP
  - f) sladění aktualizací funkčních celků archivní uložení a správa dat

# Výstupní datové formáty dokumentů v digitální podobě

---

## § 23 vyhlášky č. 259/2012 Sb.

**(1)** Výstupním datovým formátem dokumentů v digitální podobě se rozumí

- a)** datový formát výstupu z elektronického systému spisové služby,
- b)** datový formát dokumentu ukládaného ve spisovně, která je součástí elektronického systému spisové služby,
- c)** datový formát pro předávání do digitálního archivu.

**(2)** Výstupním datovým formátem statických textových dokumentů a statických kombinovaných textových a obrazových dokumentů je datový formát Portable Document Format for the Long-term Archiving (PDF/A, ISO 19005).

**(3)** Výstupním datovým formátem statických obrazových dokumentů je

- a)** datový formát Portable Network Graphics (PNG, ISO/IEC 15948),
- b)** datový formát Tagged Image File Format (TIF/TIFF, revize 6 – nekomprimovaný),
- c)** datový formát Joint Photographic Experts Group File Interchange Format (JPEG/JFIF, ISO/IEC 10918).

# Výstupní datové formáty dokumentů v digitální podobě

---

**(4)** Výstupním datovým formátem dynamických obrazových dokumentů je

**a)** datový formát umožňující uložení komprimovaných dat kódovaných podle standardu Moving Picture Experts Group Phase 2 (MPEG-2, ISO/IEC 13818),

**b)** datový formát umožňující uložení komprimovaných dat kódovaných podle standardu Moving Picture Experts Group Phase 1 (MPEG-1, ISO/IEC 11172),

**c)** datový formát Graphics Interchange Format (GIF).

**(5)** Výstupním datovým formátem zvukových dokumentů je

**a)** datový formát umožňující uložení komprimovaných dat kódovaných podle standardu MPEG-1 Audio Layer II nebo MPEG-2 Audio Layer II (MP2),

**b)** datový formát umožňující uložení komprimovaných dat kódovaných podle standardu MPEG-1 Audio Layer III nebo MPEG-2 Audio Layer III (MP3),

**c)** datový formát Waveform audio format (WAV), modulace Pulse-code modulation (PCM).

**(6)** Výstupním datovým formátem pro databáze je datový formát Extensible Markup Language Document (XML), kde součástí předávaného dokumentu v datovém formátu XML je popis jeho struktury pomocí schématu XML nebo Document Type Definition (DTD), o kterém veřejnoprávní původce vede dokumentaci.

**(7)** Výstupním datovým formátem metadat, jimiž jsou opatřovány dokumenty v elektronickém systému spisové služby, je datový formát Extensible Markup Language Document (XML) podle schématu XML pro výměnu dokumentů a jejich metadat mezi elektronickým systémem spisové služby stanoveného národním standardem nebo datový formát Extensible Markup Language Document (XML) podle schématu XML pro vytvoření datového balíčku SIP stanoveného národním standardem, který obsahuje metadata podle schématu XML pro zaznamenání popisných metadat uvnitř datového balíčku SIP stanoveného národním standardem.

# SIP (NSESSS)

```
<mets:dmdSec ID="dmd001">
  <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="NSESSS" MDTYPEVERSION="1.0" MIMETYPE="text/xml">
    <mets:xmlData>
      <nsecs:Spis ID="CRDUX000GEVI" xmlns:nsecs="http://www.mvcr.cz/nsecs/v2">
        <nsecs:EvidencniUdaje>
          <nsecs:Identifikace>
            <nsecs:Identifikator zdroj="Gordic.Ginis.CRDU.X">CRDUX000GEVI</nsecs:Identifikator>
          </nsecs:Identifikace>
          <nsecs:Popis>
            <nsecs:Nazev>D-Hostivice, 9. května č.p. 1607, ÚPS</nsecs:Nazev>
            <nsecs:Komentar>D-Hostivice, 9. května č.p. 1607, ÚPS</nsecs:Komentar>
          </nsecs:Popis>
          <nsecs:Pristupnost>
            <nsecs:BezpecnostniKategorie>
              <nsecs:Identifikator zdroj="Gordic.Ginis.CRDU.BK">10</nsecs:Identifikator>
              <nsecs:Nazev>Běžný dokument</nsecs:Nazev>
              <nsecs:BezpecnostniStupen>Běžný dokument</nsecs:BezpecnostniStupen>
            </nsecs:BezpecnostniKategorie>
          </nsecs:Pristupnost>
          <nsecs:JineUdaje>
            <nsecs:IniciacniDokument>CRDUP00085D7</nsecs:Ini /> </mets:dmdSec>
          </nsecs:JineUdaje>
          <nsecs:Evidence>
            <nsecs:EvidencniCislo>MP-SOP0425/09</nsecs:Evide /> <mets:fileSec>
            <nsecs:PoradoveCislo>425</nsecs:PoradoveCislo> <mets:fileGrp USE="Input">
              <nsecs:UrceneCasoveObdobi> <mets:file ID="CRDUAWL0G610" MIMETYPE="application/msword" CREATED="2009-03-19T09:43:57"
                <nsecs:Rok>2009</nsecs:Rok> <mets:fileGrp CHECKSUM="9D437A8275157A26BEC83F654817DAE4" USE="Original">
                <nsecs:UrceneCasoveObdobi> <mets:FLocat LOCTYPE="URN" p7:type="simple" p7:href="CRDUAWL0G610" xmlns:p7="
                  <nsecs:NazevEvidenceDokumentu>SOP</nsecs:Nazev /> </mets:fileGrp>
                <nsecs:UrceneCasoveObdobi> </mets:FLocat LOCTYPE="URN" p7:type="simple" p7:href="CRDUAWL0G610" xmlns:p7="
                <nsecs:NazevEvidenceDokumentu>SOP</nsecs:Nazev /> </mets:fileGrp>
              </nsecs:Evidence> </mets:fileGrp>
            <nsecs:Puvod> </mets:fileGrp>
            <nsecs:DatumVytvoreni datum="2009-03-19T08:42:54.00 /> </mets:fileGrp>
          </nsecs:Evidence> </mets:fileGrp>
          <nsecs:Puvod> </mets:fileGrp>
          <nsecs:DatumVytvoreni datum="2009-03-19T08:42:54.00 /> </mets:fileGrp>
        </nsecs:Spis>
        <mets:structMap>
          <mets:div DMDID="CRDUX000GEVI_Gordic.Ginis.V.S.CRDU" TYPE="Spisový plán">
            <mets:div DMDID="CRDUX000GEVI_Gordic.Ginis.V.S.CRDU-30.8" TYPE="Věcná skupina">
              <mets:div DMDID="CRDUX000GEVI_Gordic.Ginis.V.S.CRDU-30" TYPE="Věcná skupina">
                <mets:div DMDID="CRDUX000GEVI" TYPE="Spis">
                  <mets:div DMDID="CRDUP00085D7" TYPE="Dokument"/>
                  <mets:div DMDID="CRDUX000GFCY" TYPE="Dokument">
                    <mets:div LABEL="g0023355.DOC" TYPE="Elektronický obraz">
                      <mets:fptr FILEID="CRDUAWL0G610"/>
                    </mets:div>
                  </mets:div>
                </mets:div>
              </mets:div>
            </mets:div>
          </mets:div>
        </mets:structMap>
      </nsecs:Spis>
    </mets:xmlData>
  </mets:mdWrap>
</mets:dmdSec>
```

# Identifikace formátů

---

- k rozpoznání formátů datových souborů při příjmu (ingestu) do digitálního archivu
- předpoklad pro další úkony:
  - a) extrakci metadat
  - b) migraci do formátu 1) pro dlouhodobé uložení (dle uchovávací strategie) nebo 2) při vytváření objektů v rámci modulu pro přístup (datové balíčky DIP v jiném formátu – např. TIFF -> JPG)
- využití i předchozích částech životního cyklu dokumentu – např. v eSSL (příjem na podatelně, převod do výstupního formátu při vyřízení/uzavření atd.)



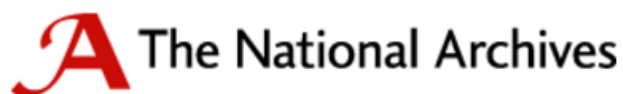
# Identifikace formátů

---

Metody:

- dle přípony (extension)
- dle obsahu (signatures)

# PRONOM



Search our website ▾



About us

Education

Records

Information management

Archives sector

You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > PRONOM



## The **technical registry** PRONOM

[Welcome](#) : [About](#) [Add an entry](#)  
 [Search](#) [Help](#) [Information resources](#)

### **Welcome to PRONOM**

[PRONOM changes and DROID signature file release notes.](#)

[DROID signature files.](#)

Work has been undertaken to create a linked data version of PRONOM. We hope to revisit this model for PRONOM in the future. As part of a broad reorganisation of The National Archives web services the current iteration of Linked Data PRONOM has been removed.

The online registry of technical information. PRONOM is a resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value. Find out about the future of PRONOM on our [Information resources](#) page.

**Search PRONOM** >

# PRONOM

- vyvíjí Národní archiv Velké Británie
- od roku 2002
- v současnosti 1493 formátů
- celosvětová akceptace
- <https://www.nationalarchives.gov.uk/PRONOM/default.htm>
- využívají programy: DROID, FIDO, Sigfried aj.

You are here: [Home](#) > [Information management](#) > [Our projects and work](#) > [Digital preservation](#) > [PRONOM](#) > Search: PRONOM



The **technical registry**  
**PRONOM**

[Welcome](#) : [About](#) [Add an entry](#)  
 [Search](#) [? Help](#) [Information resources](#)

[Search : Simple Search](#)

[? Help : Simple Search](#)

[Simple search](#) [File format](#) [PRONOM Unique Identifier](#) [Software](#) [Vendor](#) [Lifecycles](#) [Migration Pathways](#)

## 1. Search

Enter a simple search string and then click 'search'.

[Search >](#)

# DROID

---

- Digital Record Object Identification
- vyvíjí Národní archiv Velké Británie
- identifikace pomocí :

1) přípony (extension)

2) obsahu (signature) pomocí  
tzv. magického čísla

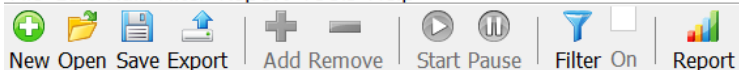


# Magické číslo

- hodnota typická pro daný souborový formát

DROID v6.3

File Edit Run Filter Report Tools Help



Untitled-2 \*

Resource	Extension	Size	Last modified	Ids	Format	Version	Mime type	PUID	Method
D:\DATA\PVH_Start - Filozofická fakulta MU.html	html	26,7 KB	1.5.17 18:57		Extensible Hypertext Markup Language	1.0	application/xhtml+xml	<a href="#">fmt/102</a>	Signature

## Details for: Extensible Hypertext Markup Language 1.0

Save as... XML | CSV

Print

Go to: [Summary](#) > | [Documentation](#) > | [Signatures](#) | [Compression](#) > | [Character encoding](#) > | [Rights](#) > | [Reference files](#) > | [Properties](#) >

### Signatures

**External signatures** File extension: html  
File extension: htm

#### Internal signatures

**Name** XHTML 1.0

**Description** With DOCTYPE declaration for 1.0

#### Byte sequences

**Position type** Absolute from BOF

**Offset** 0

**Maximum Offset** 1024

**Byte order**

**Value** 3C21444F43545950452068746D6C205055424C494320222D2F2F5733432F2F445444205848544D4C20312E30

**Position type** Variable

# Magické číslo

D:\DATA\PVH\_Start - Filozofická fakulta MU.html - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

PVH\_Start - Filozofická fakulta MU.html

```
1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
2 <!-- saved from url=(0029)http://www.phil.muni.cz/wpvh/ -->
3 <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="cs" lang="cs"><head><meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
4
5
6 <title>
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21 </title>
```

PVH: D:\DATA\PVH\_Start - Filozofická fakulta MU.html - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

Sta

PVH\_Start - Filozofická fakulta MU.html

Filo	Address	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f	Dump
	00000000	3c	21	44	4f	43	54	59	50	45	20	68	74	6d	6c	20	50	<!DOCTYPE html P
	00000010	55	42	4c	49	43	20	22	2d	2f	2f	57	33	43	2f	2f	44	UBLIC "-//W3C//D
	00000020	54	44	20	58	48	54	4d	4c	20	31	2e	30	20	54	72	61	TD XHTML 1.0 Tra
	00000030	6e	73	69	74	69	6f	6e	61	6c	2f	2f	45	4e	22	20	22	nsitional//EN" "
	00000040	68	74	74	70	3a	2f	2f	77	77	77	2e	77	33	2e	6f	72	http://www.w3.or
	00000050	67	2f	54	52	2f	78	68	74	6d	6c	31	2f	44	54	44	2f	g/TR/xhtml1/DTD/
	00000060	78	68	74	6d	6c	31	2d	74	72	61	6e	73	69	74	69	6f	xhtml1-transitio
	00000070	6e	61	6c	2e	64	74	64	22	3e	0a	3c	21	2d	2d	20	73	nal.dtd">.<!-- s
	00000080	61	76	65	64	20	66	72	6f	6d	20	75	72	6c	3d	28	30	aved from url=(0
	00000090	30	32	39	29	68	74	74	70	3a	2f	2f	77	77	77	2e	70	029)http://www.p
	000000a0	68	69	6c	2e	6d	75	6e	69	2e	63	7a	2f	77	70	76	68	hil.muni.cz/wpvh
	000000b0	2f	20	2d	2d	3e	0a	3c	68	74	6d	6c	20	78	6d	6c	6e	/ -->.<html xmln
	000000c0	73	3d	22	68	74	74	70	3a	2f	2f	77	77	77	2e	77	33	s="http://www.w3
	000000d0	2e	6f	72	67	2f	31	39	39	39	2f	78	68	74	6d	6c	22	.org/1999/xhtml"

# Extrakce metadat


---

- pro strojové zpracování
- nejen technická, ale i obsahová (např. u fotografií)
- dle ukládací strategie se při zpracování vytěžuje do správy dat a/nebo ukládá do datové balíčku AIP
- další informace pro zpřístupnění
- podporují nástroje: JHOVE, Apache Tika, ExifTool, MediaInfo atd.

# Apache Tika

- pro formát PDF
- zobrazuje:
  - metadata
  - oprávnění
  - text

<https://tika.apache.org>

 Apache Tika: text\_only\_fontsEmbeddedAll.pdf

File View

Author: van der Knijff  
Comments:  
Company: Koninklijke Bibliotheek  
Content-Length: 70062  
Content-Type: application/pdf  
Creation-Date: 2012-11-23T13:20:31Z  
Keywords:  
Last-Modified: 2012-11-23T13:20:33Z  
Last-Save-Date: 2012-11-23T13:20:33Z  
SourceModified: D:20121123115302  
X-Parsed-By: org.apache.tika.parser.DefaultParser  
X-Parsed-By: org.apache.tika.parser.pdf.PDFParser  
X-TIKA:digest:MD5: f4fc8d3b59938e1a83a8443066b6fda7  
X-TIKA:digest:SHA256: 8ee33b2b2dee93bf44c5426e12e12064bfff001a808ff4d2feea87c7beea4e4ba  
access\_permission:assemble\_document: true  
access\_permission:can\_modify: true  
access\_permission:can\_print: true  
access\_permission:can\_print\_degraded: true  
access\_permission:extract\_content: true  
access\_permission:extract\_for\_accessibility: true  
access\_permission:fill\_in\_form: true  
access\_permission:modify\_annotations: true  
cp:subject:  
created: Fri Nov 23 14:20:31 CET 2012  
creator: van der Knijff  
date: 2012-11-23T13:20:33Z  
dc:creator: van der Knijff  
dc:format: application/pdf; version=1.7  
dc:subject:  
dc:title: This is a test document  
dcterms:created: 2012-11-23T13:20:31Z  
dcterms:modified: 2012-11-23T13:20:33Z  
meta:author: van der Knijff  
meta:creation-date: 2012-11-23T13:20:31Z  
meta:keyword:  
meta:save-date: 2012-11-23T13:20:33Z  
modified: 2012-11-23T13:20:33Z  
pdf:PDFVersion: 1.7  
pdf:docinfo:created: 2012-11-23T13:20:31Z  
pdf:docinfo:creator: van der Knijff  
pdf:docinfo:creator\_tool: Acrobat PDFMaker 9.1 voor Word  
pdf:docinfo:custom:Comments:  
pdf:docinfo:custom:Company: Koninklijke Bibliotheek  
pdf:docinfo:custom:SourceModified: D:20121123115302  
pdf:docinfo:keywords:



# ExifTool

- výborný pro obrazové soubory (metadata standardu EXIF)
- u textových souborů mj:
  - autor
  - datum vytvoření
  - poslední změna
  - statistika atd.
- export možný i do TXT

```
D:\EXIFTOOL\exiftool(-k).exe
Author                : van der Knijff
Keywords              :
Comments              :
Template              : Normal.dot
Last Modified By     : van der Knijff
Revision Number       : 3
Software              : Microsoft Office Word
Total Edit Time       : 6.0 minutes
Create Date          : 2012:11:22 12:28:00
Modify Date          : 2012:11:23 11:53:00
Pages                 : 1
Words                 : 34
Characters            : 198
Security              : None
Code Page             : Windows Latin 1 (Western European)
Company              : Koninklijke Bibliotheek
Lines                 : 1
Paragraphs           : 1
Char Count With Spaces : 231
App Version           : 11.9999
Scale Crop            : No
Links Up To Date     : No
Shared Doc            : No
Hyperlinks Changed   : No
Title Of Parts       : This is a test document
Heading Pairs         : Title, 1
Comp Obj User Type Len : 31
Comp Obj User Type   : Microsoft Office Word Document
-- press RETURN --
```

<https://sourceforge.net/projects/exiftool/>

# MediaInfo

---

- pro multimedialní formáty:
- video: MPEG-1/2 Video, H.263, MPEG-4, H.264/AVC, H.265/HEVC aj.
- audio: MPEG Audio (i MP3), AC3, DTS, AAC, Dolby E, AES3, FLAC aj.
- kontejnery: MPEG-4, QuickTime, Matroska, AVI, MPEG-PS (DVD), MPEG-TS (Blu-ray), MXF, GXF, LXF, WMV, FLV aj.
- <https://mediaarea.net/cs/MediaInfo>

# Validace formátů

---

- ověření souladu se stanovenou normou
- standardizace zpracování
- odhalení chyb – např. nemožnost zobrazit v budoucnu
- pro různé formáty specializované programy

# PDF/A

---

- formát vhodný pro dlouhodobé uložení (zejm. textu)
- oproti PDF omezena možnost změnit zobrazení (zakázány skripty, šifrování atd.)
- 2 úrovně požadavků
  - úroveň A: součástí je vždy textová vrstva, struktura aj.
  - úroveň B: bez textové vrstvy, pouze ztvárnění

Formát	Odvozeno z	Změny
PDF/A-1 ISO 19005:2005	PDF 1.4	- výchozí verze
PDF/A-2 ISO 19005:2011	PDF 1.7	- podpora JPEG 2000 - fonty OpenType - podpora elektronického podpisu PAdES
PDF/A-3 ISO 19005:2012	PDF 1.7	- vložení jiného souboru (XML, DOC, XLS aj.)

# PDF/A

- různé validátory – odlišné výsledky (<http://digi.nacr.cz>)

## Validace PDF

Soubor:  Soubor nevybrán

Norma: PDF/A-1a ,  
PDF/A-1b ,  
PDF/A-2a ,  
PDF/A-2b ,  
PDF/A-2u ,  
PDF/A-3a ,  
PDF/A-3b ,  
PDF/A-3u

Zaslat

### Výsledek\*

Soubor file002.pdf je nevalidní dle normy pdfa-1b

### [ - ] výstup 1. validátoru

Soubor file002.pdf je nevalidní dle normy pdfa-1b

#### kompletní výstup

```
0:Validating file temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a
1:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 0, 0, 0x0341851C, "The ID in the 1st page and last trailer of a linearized file are different.", 1
2:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 0, 0, 0x03418533, "The file contains cross reference streams.", 4
3:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 0, 0, 0x00410106, "The file header version does not conform to the standard.", 1
4:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 0, 8, 0x0A09C882, "The property 'xmpMM:subject' is not defined in schema 'XMP Media Management Schema'.", 1
5:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 0, 8, 0x0A09C881, "The schema description for namespace 'pdfx:' (http://ns.adobe.com/pdfx/1.3/) is missing.", 3
6:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key SA was ignored.", 2
7:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key BM was ignored.", 2
8:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key SMask was ignored.", 2
9:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key CA was ignored.", 2
10:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key ca was ignored.", 2
11:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 23, 0x00410602, "The key AIS was ignored.", 2
12:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 34, 0x00410602, "The key FontFamily was ignored.", 5
13:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 34, 0x00410602, "The key FontStretch was ignored.", 5
14:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 34, 0x00410602, "The key FontWeight was ignored.", 5
15:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 14, 0x0041060B, "The key Metadata is recommended.", 2
16:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 35, 0x00418704, "The font ArialMT must be embedded.", 1
17:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 37, 0x00418704, "The font Arial-BoldMT must be embedded.", 1
18:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 39, 0x00418704, "The font TimesNewRomanPSMT must be embedded.", 1
19:"temp/b04da9622059df68e94ec98617ff59a61036d5b5cfb36463065348be304e1e1a", 1, 0, 0x83410612, "The document does not conform to the requested standard.", 1
20:Done.
21:
```

# JHOVE

---

- JSTOR/Harvard Object Validation Environment
- umožňuje jak extrakci určitých metadat, tak jejich validaci
- podpora více formátů: AIFF, ASCII, Bytestream, GIF, HTML, JPEG, JPEG 2000, PDF, TIFF, UTF-8, WAV a XML
- pro některé formáty méně vhodný (např. PDF/A-1)
- <http://jhove.sourceforge.net>

---

[zbysek.stodulka@nacr.cz](mailto:zbysek.stodulka@nacr.cz)

<http://digi.nacr.cz>