

Vývoj procesov formátovej validácie v Centrálnom Dátovom Archíve (CDA)

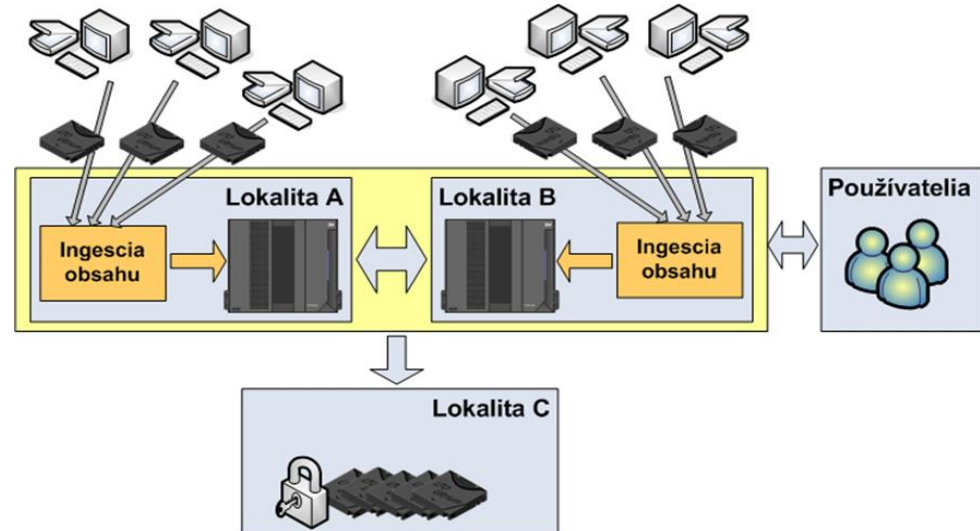
Ing. Juraj Strnisko

O projekte CDA

projekt UKB v rámci OPIS PO2 (Operačný program informatizácia spoločnosti, prioritná os 2 - „Rozvoj pamäťových a fondových inštitúcií a obnova ich národnej infraštruktúry)

podľa noriem ISO/IEC 27001:2013 (Informačná bezpečnosť) a ISO 14721:2012 (OAIS)

- príprava **2007 – 2011**
- realizácia **2012 - 2014**
- prevádzka od **2015**



Design CDA

- cieľom vytvoriť dlhodobé dôveryhodné úložisko dát
- **Redundancia**
 - Lokálna (A + B + C)
 - Technologická
- **Komponenty HW**
 - Servery IBM Power + HP DL3xx
 - Diskové polia IBM Storwize (200+ TB)
 - Páskové knižnice IBM JAG 25 PB (Archív) + IBM LTO5/6 (I/O)
- **Komponenty SW**
 - AIX, Linux, Win/VMware
 - HSM, TSM, LTFS, GPFS + APP moduly (dodávateľ)

Zdroje údajov

PFI - Pamaťové fondové inštitúcie (napr. MSNP, SNG, NOC, SNK, ...)

Dohoda o vklade

Volumetria (GB / TB, počet SIP, DIP)

vklad SIP → AIP

výber AIP → DIP

Profil vkladateľa

Certifikáty

autentifikácia (prístupy)

autorizácia (balíky)

Vklad do archívu (Ingest)

CDA implementované v súlade s OAIS STN ISO 14721:2014 a
STN ISO 27001:2013

Proces pozostávajúci z cca 40 krokov

- Kontrola: profil / certifikáty / formáty / antivírus
- Transformácia SIP na AIP
- Uloženie v 3 kópiach

Formátová validácia CDA súčasťou

- **Droid** identifikátor formátov,
- využívajúci PRONOM databázu

- **Jhove** (ver 1.7.2) java validátor formátov
- (JSTOR / Harvard Object Validation Environment)

Validácia v procese testu/implementácie

- PREFORMA = **PRE**servation **FORMA**ts
- projekt spolufinancovaný z fondov EU
- nástroje na kontrolu nad procesom testu zhody súborov pred vkladom

<http://www.preforma-project.eu/>



Preforma open-source nástroje

VeraPDF na validáciu súborov PDF/A

<http://verapdf.org/>

DPFmanager na validáciu súborov TIFF

<http://dpfmanager.org/>

MediaConch na validáciu súborov MKV / FFV1 / LPCM

<https://mediaarea.net/MediaConch/>



DPF Manager

CONFORMANCE CHECKER

File

vzorka01.tif

Browse ...

Configuration

- Baseline HTML.dpf
- Baseline JSON.dpf
- Baseline PDF.dpf
- Baseline XML.dpf
- Custom config...

Check files

Online validator / Report



DPF Manager

MULTIFILE REPORT

1 file processed

1 Conforms to Baseline TIFF 6.0



1 passed
0 failed
Global score 100%



vzorka01.tif

Conformance checker

Baseline TIFF 6.0

0 errors

0 warnings



Passed
Score 100%

Záver

- validácia súborov a výber vhodného formátu pre dlhodobé uchovanie dát sú témy, ktoré sú z pohľadu CDA a digitálnych archívov stále aktuálne
- nasadenie nástroja MediaConch na validáciu video formátov MKV
- implementácia ostatných 2 validátorov z dielne PREFORMA, (VeraPDF a DPFmanager) príp. o nasadení vyššej verzie Jhove (1.16 vydané v marci 2017 podporujúce aj PDF vo verzii 1.7)

juraj.strnisko@ulib.sk