

## **UNIVERZITNÁ KNIŽNICA V BRATISLAVE – Centrálny dátový archív**

### **Medzinárodná spolupráca v kontexte projektu OPIS PO2 Univerzitetnej knižnice v Bratislave: Centrálny dátový archív**

**Mgr. Bibiána Žigová**

[bibiana.zigova@ulib.sk](mailto:bibiana.zigova@ulib.sk)

#### **Abstrakt:**

Článok je zameraný na skúsenosti CDA (Centrálny dátový archív) za 5 rokov existencie, z toho 3 roky ostrej prevádzky, so zameraním na medzinárodnú spoluprácu medzi dlhodobými archívami, vyzdvihuje dôležitosť udržiavania toku výmeny informácií a skúseností v LTP medzi knižnicami a archívami z celého sveta, analyzuje stratégiu uchovávania digitálneho obsahu v rôznych formátoch z hľadiska LTP (Long Term Preservation). Informuje o stave LTP v Európe.

#### **Kľúčové slová:**

Medzinárodná spolupráca, digitálne objekty, digitálny obsah, dátové úložisko, dlhodobé uchovávanie digitalizovaného kultúrneho dedičstva, LTP, ochrana digitálneho obsahu, formáty

### **1. Úvod**

Cieľom projektu OPIS PO2 - CDA bolo vybudovať komplexný integrovaný systém dlhodobého uchovávania a ochrany digitálneho obsahu, jeho získavania, spracovania, ochrany a využitia. V rámci projektu sa vybuďoval centrálny dátový archív, ktorý v zmysle predpísaných štandardov zabezpečuje uloženie kópií digitálnych objektov v najmenej dvoch geograficky oddelených lokalitách vzdialených minimálne 50 km. Lokality sú vybavené príslušným hardvérovým i softvérovým vybavením a prostredníctvom dátových telekomunikačných sietí sú vzájomne prepojené. Súčasťou dátového archívu je aj pasívny sklad pamäťových médií (terciárna lokalita). Projekt sa ukončil v roku 2014 a v súčasnosti je CDA v prevádzke tretí rok. CDA je vybudovaný na platforme OAIS v súlade s medzinárodnými normami ISO 16363, ISO 14721. Je držiteľom certifikátu Systému manažérstva informačnej bezpečnosti podľa normy ISO/IEC 27001. K marcu 2017 má uložených viac než 1,5 milióna SIP balíkov v celkovej veľkosti 3,6 PetaByte. (<http://cda.kultury.sk>).

Stratégia uchovávania digitálneho obsahu je kľúčová a túto otázku riešia takmer na celom svete. Uchovávanie zahŕňa širokú škálu úkonov určených na zabezpečenie využiteľnosti obsahu a jeho ochranu pred fyzickým zničením, stratou, nefunkčnosťou média na ktorom je uložený.

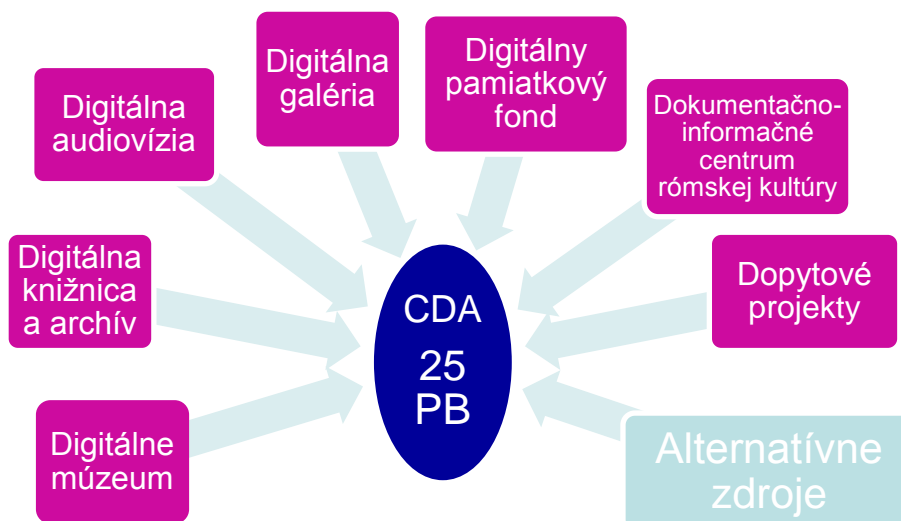
Digitalizácia je finančne a časovo náročný proces a je preto nevyhnutné pracovať efektívne a zabezpečiť trvalé uchovanie digitalizovaného obsahu pre budúce generácie. Nástrojom pre tento proces je štandardizácia jednotlivých krokov digitalizácie od skenovania až po vklad do elektronického archívu. Medzinárodná spolupráca medzi dlhodobými archívami, dôležitosť udržiavania toku výmeny informácií a skúseností v LTP medzi knižnicami a archívami z celého sveta je kľúčová.

### **2. Spolupráca v rámci Slovenska**

Na Slovensku prebiehala a stále prebieha digitalizácia kultúrneho dedičstva v jednotlivých PFI samostatne v období rokov 2011 – 2015. Cieľom bolo vybudovať Slovenskú digitálnu knižnicu, sieť špecializovaných digitalizačných pracovísk PFI, centrálny archív pre dlhodobé uchovávanie, webharvesting, webarchiving, centrálny portál pre sprístupňovanie zdigitalizovaného obsahu. Samostatnou kapitolou boli legislatívne opatrenia v oblasti autorskoprávnej ochrany diel. Zohľadnené boli medzinárodné normy a odporúčania OAIS, LTP, METS, XFDU, DCMES, PREMIS ako aj skúsenosti a know-how z existujúcich systémov pre dlhodobé uchovávanie digitálneho obsahu. Sprostredkovateľský orgán rezortu kultúry vydal pre projekty pamäťových a fondových inštitúcií v rámci OPIS PO2 (Operačný program informatizácia spoločnosti, Prioritná os 2 - Rozvoj pamäťových

a fondových inštitúcií a obnova ich národnej infraštruktúry. Operačný program je spolufinancovaný zo zdrojov Európskeho fondu regionálneho rozvoja. (<http://www.opis.gov.sk>) niekoľko metodických manuálov. Jedným z nich je metodický manuál č.17 pre zabezpečenie dlhodobej archivácie konvertovaných objektov. Manuál vytvoril rámec pre pamäťové a fondové inštitúcie (PFI) v oblasti digitalizácie kultúrneho dedičstva Slovenska. Stanovil východiská a princípy pre metadáta, formáty, registre, číselníky, SIP/AIP/DIP balíky. Referenčným zdrojom sú normy ISO 16363, ISO 14721, Centrálny dátový archív Univerzitetnej knižnice prevzal a pripravil slovenské verzie STN. Referenčným zdrojom pre formáty je register PRONOM.

Pamäťové fondové inštitúcie sú združené do tzv. „určeného spoločenstva“. V rámci OPIS PO2 je celkovo 9 národných projektov pamäťových fondových inštitúcií z rezortu kultúry a 7 dopytových projektov. To naznačuje aj veľkú rôznorodosť dát, veľa formátov. CDA s určeným spoločenstvom pravidelne komunikuje prostredníctvom cielených emailov o novinkách, odstavkách systému a pod. Určené spoločenstvo má na web stránke CDA (<http://cda.kultury.sk>) vyčlenenú sekciu s dokumentmi, prístupnú len pre členov komunity. Zároveň CDA poskytuje prehľadné informácie o spracovaní dát, štatistiky, prezentácie na vlastnej web stránke (<http://cda.kultury.sk>), ktorú zamestnanci CDA sami denne aktualizujú (redakčný systém Drupal).



Obr.1 Určené spoločenstvo

Slovensko a jeho kultúrne dedičstvo reprezentuje portál „SLOVAKIANA“ (<https://www.slovakiana.sk/>). Bol spustený v novembri 2015, ktorý pre odbornú aj laickú verejnosť sprístupňuje výsledky digitalizácie slovenského kultúrneho dedičstva. Portál je súčasťou siete európskych kultúrnych portálov. Je výsledkom projektu CAIR „Centrálna aplikačná infraštruktúra a registratúra“ Národného osvetového centra NOC ([www.nocka.sk](http://www.nocka.sk)) v rámci OPIS PO2.

### 3. Spolupráca v rámci Európy

- **Česká Republika**

V CDA si nesmierne vážime prácu českých kolegov, ktorí už neraz ukázali, že sú odborníkmi na dlhodobé uchovávanie dát, LTP, či formátovú konverziu. Na pravidelnej báze konzultujeme aktuálne problémy, vymieňame si skúsenosti, know-how. Osvedčila sa nám online video-komunikácia. Národní knihovna ČR v Prahe sprístupnila portál Národní digitální knihovna ČR, ktorý využíva systém „Kramerius“ (<http://www.ndk.cz>) digitalizuje, sprístupňuje a dlhodobo uchováva zdigitalizované kultúrne dedičstvo ČR. Systém Kramerius je postavený na open source riešení Fedora. Pre všetky tieto činnosti je nevyhnutné stanoviť a dodržiavať štandardy vytvárania dát a metadát. Národní

digitální knihovna využíva systém persistentných identifikátorov založených na štandarde URN:NBN (<https://resolver.nkp.cz>). „ČIDLO“ (CZIDLO - CZech IDentification and LOcalization tool) je služba pre knižnice a pamäťové a fondové inštitúcie z oblasti českého kultúrneho dedičstva pre potreby trvalej identifikácie digitálnych dokumentov, alebo ako prostriedok pre zaistenie dôveryhodnosti citačnej praxe (overovanie autenticity citovaných dokumentov).

Ďalšou inštitúciou s ktorou CDA spolupracuje je Moravská zemská knihovna v Brne (<https://www.mzk.cz/>). Podobne ako NK Praha, sprístupňuje Digitální knihovnu v systéme Kramerius. Digitalizuje, sprístupňuje a dlhodobo uchováva zdigitalizované regionálne diela. Bohaté skúsenosti kolegov z MZK Brno sme využili pri internom audite dlhodobého dôveryhodného úložiska dát podľa ISO normy ISO 16363. Analyzovali pre nás dokumentáciu, procesy, technológie, aplikácie v CDA, výstupom bola správa, ktorá bude použitá ako podklad pri príprave externého auditu LTP a Data Seal of Approval, ktorý v CDA pripravujeme.

Nesmieme zabudnúť na Národní technickou knihovnu ČR v Prahe (<https://www.techlib.cz/cs/2919-digitalni-knihovna>). Digitalizácia a dlhodobá archivácia je predmetom projektu Digitální knihovna NTK. Národní technická knihovna sprístupňuje zdigitalizované dokumenty zo svojich fondov online. Digitální knihovna NTK je sprístupnená v systéme Kramerius 4, ktorý je vyvíjaný spoločnosťou Incad v spolupráci s Knihovnou Akademie věd ČR, Národní knihovnou ČR, Moravskou zemskou knihovnou a Národní technickou knihovnou.

S kolegami z NTK Praha sme v kontakte aj z dôvodu riešenia ďalšieho projektu OPIS PO2 v Univerzitnej knižnici v Bratislave, národného projektu Digitálne pramene – Webharvesting a archivácia e-Born obsahu, ktorého cieľom je vybudovanie informačného systému (IS) na záznam, zbieranie, sprístupňovanie a ochranu pôvodného digitálneho obsahu, webharvesting a webarchiving on-line digitálnych prameňov ako integrálnej súčasť kultúrneho dedičstva Slovenskej republiky. (<https://www.webdepozit.sk/projekt-dip/>). Tento cieľ korešponduje s oprávnenou aktivitou:– systematická digitalizácia a sprístupňovanie digitálneho obsahu pamäťovými a fondovými inštitúciami.

Predmetom záujmu IS Digitálne pramene sú pôvodné (e-Born) elektronické pramene, elektronické monografie, periodiká, webové stránky a ďalšie druhy elektronických publikácií a artefaktov, vrátane elektronických tvorivých a publikačných modelov a predlôh.

Zámerom projektu Digitálne pramene – Webharvesting a archivácia e-Born obsahu je ochrana kultúrneho dedičstva, ktoré sa tvorí a publikuje v pôvodnej elektronickej forme, založená na metódach systematického zberu prostredníctvom dátových sietí a elektronických médií. Výsledným stavom je funkčný informačný systém IS Digitálne pramene na zber, primárne spracovanie a sprístupňovanie pôvodných elektronických t.j. digitálnych prameňov, vybavený optimálnymi hardvérovými a softvérovými komponentmi, ktorý bude umiestnený vo vlastných a prenajatých priestoroch prijímateľa (UKB) s odpovedajúcou infraštruktúrou a obsluhou. Systém bude mať zabezpečený servis formou komplexných služieb podpory (HW a SW).

Výstupom projektu Digitálne pramene – Webharvesting a archivácia e-Born obsahu je vybudovanie a prevádzka informačného systému na zber, spracovanie a sprístupňovanie pôvodných elektronických a WWW prameňov vo forme otvorenej digitálnej knižnice a webových služieb s konsolidovaným rozhraním (katalóg, portál, repozitár), ktorá bude formálne harmonizovaná so systémom Centrálnaj aplikačnej infraštruktúry a registratúry a so systémom Centrálného dátového archívu na dlhodobú ochranu kultúrneho dedičstva, ktoré sú realizované v rámci národných projektov NP7 a NP8 programu OPIS PO2.



- **Platforma V4**

Medzi knižnicami V4 prebieha už tradične spolupráca na veľmi dobrej úrovni. Prebieha mnoho aktivít a vynakladá sa značné úsilie v oblasti získavania, ochrany digitálnych zdrojov. Tieto metodické a praktické znalosti majú veľkú hodnotu. Spolupráca v rámci platformy V4 sa týka najmä nasledujúcich oblastí:

- certifikácia dlhodobých úložísk podľa medzinárodných noriem ISO 16363, ISO 14721, ISO 27001,
- harmonizácia legislatívy v oblasti e-born dokumentov, autorského práva online dokumentov,
- metodika uchovávania zdigitalizovaného kultúrneho dedičstva,
- webharvesting.

Digitálne formáty sú jednou zo základných odborných tém v rámci dlhodobého uchovávania údajov. CDA ako dlhodobý archív je s touto problematikou dennodenne konfrontovaný a venuje sa jej systematicky. Odborný seminár a konferencia CDA'2016: Formátové výzvy LTP ([http://cda.kultury.sk/sk/Konferencia\\_Formatove\\_vyzvy\\_LTP](http://cda.kultury.sk/sk/Konferencia_Formatove_vyzvy_LTP)) v novembri 2016 prispel k výmene skúseností, know-how v tejto oblasti a obohatí metodologické základy systému CDA. Je zároveň v súlade so zámermi na organizovanie odborných platforiem diskutovaných na kolokviu Knihovny zemí V4 v nástrahách digitálneho veku v Brne (<https://www.mzk.cz/o-knihovne/odborne-cinnosti/konference-v4>). Platformy sú celkovo 4:

- Know-how Platform,
- Synergy Platform,
- Methodology Platform,
- Usage Platform.

Spolupráca odborných pracovísk knižníc zemí V4 (Slovensko, Čechy, Maďarsko, Poľsko) je naplánovaná na obdobie 2016 – 2020.



- **Maďarsko**

V Maďarsku je projekt digitalizácie národného kultúrneho dedičstva a dlhodobého uchovávania dát v stave prípravy projektovej dokumentácie (Országos Széchényi Könyvtár - Digitális Könyvtár) (<http://oszkdk.oszk.hu/>). Kolegovia z Maďarskej národnej Szechenyi knižnice v Budapešti pravidelne konzultujú so zamestnancami v CDA funkčné a nefunkčné požiadavky na repozitáre, žiadajú o metodické posúdenie navrhovaných postupov. Zúčastňujú sa na seminároch a konferenciách k LTP, V4 v Univerzitetnej knižnici v Bratislave.

- **PREFORMA**

Iniciatíva PREFORMA (<http://www.preforma-project.eu>) je projekt „PREservation FORMAts“ uchovávania formátov pre elektronické archívy. Projekt začal 1.1.2014 a spolutorcom je Európska komisia a jej program „FP7-ICT Programme“ ([http://cordis.europa.eu/fp7/ict/home\\_en.html](http://cordis.europa.eu/fp7/ict/home_en.html)). Cieľom projektu je vyriešenie problému zavádzania kvalitných štandardizovaných formátov súborov pre uchovanie obsahu dát v dlhodobom horizonte. Hlavným cieľom je poskytnúť pamäťovým inštitúciám plnú kontrolu nad priebehom validácie súborov pri vklade do archívu.

Iniciatíva PREFORMA poskytuje priestor expertom, inštitúciám a iným projektom diskutovať, nájsť riešenia a poskytnúť ich verejnosti. Cieľom iniciatívy je spojiť pamäťové a fondové inštitúcie s vývojármi, softvérovými spoločnosťami a normalizačnými agentúrami, poskytnúť všetkým zainteresovaným stranám spätnú väzbu. Výsledkom by mal byť súbor formátov s adekvátnymi validátormi, vhodnými pre dlhodobú archiváciu. Členmi iniciatívy sú pamäťové a fondové inštitúcie, archívy, knižnice, softvérové firmy z Belgicka, Švédska, Holandska, Nemecka, Rakúska, Grécka, Talianska, Španielska, Estónska, Veľkej Británie, Slovenska. CDA je členom iniciatívy PREFORMA od júla 2016. Konkrétnymi výstupmi z iniciatívy sú voľne dostupné validátory pre formáty TIFF, PDF (validátor VeraPDF) či video formáty (validátor Matroska) (<https://github.com/veraPDF/veraPDF-library/releases/tag/v1.0.2>). Ukončenie projektu sa očakáva na jeseň 2017 na záverečnej konferencii v Taline (EST).



Obr. 3 PREFORMA EXPERIENCE WORKSHOP

- **UNESCO**

Charta o zachovaní digitálneho dedičstva, prijatá na 32. Zasadnutí Generálnej konferencie UNESCO v roku 2003 v časti Digitálne dedičstvo ako spoločné dedičstvo položila základy digitalizácie hmotného a nehmotného kultúrneho dedičstva. Následne Vancouverská deklarácia UNESCO z roku 2012 „Pamäť

sveta v digitálnom veku: digitalizácia a uchovávanie“ stanovila rámec medzinárodnej spolupráce pri uchovávaní zdigitalizovaného kultúrneho dedičstva sveta. Konzorcium pamäťových a fondových inštitúcií stanovené deklaráciou na prácu v rokoch 2013 – 2015 usporiadalo workshopy, konferencie a panelové diskusie, kde odborníci z oblasti knihovníctva, archívniectva, múzeí, štátnej správy a komerčných firiem diskutovali o urgentnej potrebe uchovávaní digitálnych informácií pre budúcnosť. Jednou z tém diskusií bola problematika otvorených formátov, štandardizácia formátov, binárna ochrana dát a ochrana konzistentnosti dát. Diskusie mali za cieľ nájsť riešenia a ponúknuť ich odbornej verejnosti. Na základe posledných zistení sú softvérové riešenia dlhodobej ochrany digitálnych dát veľmi drahé a preto takmer nedostupné pre mnohé inštitúcie. Zároveň open source riešenia nie sú dostatočné, potreby kurátorov a používateľov sú nad možnosti týchto riešení.

Diskusie konzorcia sa týkali aj skúmania dopytu po digitalizovaných dátach. Skúmalo sa kto, aké dáta využíva, v akej forme ich potrebuje. Na základe posledných zistení je využívanie väčšiny digitálnych dát spoplatnené, čo obmedzuje prístup k nim.

Ďalším problémom, ktorý sa riešil, je kvalita digitalizovaných dát a ich dôveryhodnosť. Na základe zistení mnoho digitálnych informácií nemá základnú štruktúru, nemá, alebo má len minimum metadát, čo znemožňuje vyhľadávanie, prácu s nimi. Sú de facto nedostupné.

Odborníci z oblasti digitalizácie upozornili na fakt, že sa stále mylne interpretuje digitalizácia a uchovávanie digitálneho obsahu. Nevyhnutná je osвета aj v tomto smere. Digitalizácia zahŕňa nielen transformáciu obsahu do digitálnej formy. Zahŕňa mnoho aktivít, od akvizície, identifikácie, čistenia, reštaurovania, katalogizácie, vytvorenia metadát až po uloženie digitalizovaného obsahu na špeciálnom médiu (magnetickom, optickom) do trezora či archívu.

Uchovávanie to nie je len fyzické zachovanie kópie, kópií v trezoroch. Ide hlavne o možnosť využiť digitalizované dáta v budúcnosti za podmienky zachovania úplnej integrity dát. (<http://www.unesco.sk/O-UNESCO>)

Pre pamäťové a fondové inštitúcie je kľúčové stanovenie stratégie digitalizácie a uchovávaní digitalizovaných dát. Stratégia by mala obsahovať nielen obsah dát, ale hlavne analýzu efektivity nákladov, prípadné výnosy, mala by identifikovať potenciálnych donorov, možnosti finančnej podpory vlády daného štátu.

- **Dánsko**

Dánsky národný archív (<https://www.sa.dk/en/about-us/danish-national-archives>) uchováva a sprístupňuje zo zákona v papierovej aj digitálnej forme historické dokumenty Dánska, dáta štátnej správy a justície. Zákon jasne stanovuje formu, formáty, metodiku tvorby SIP balíka, dlhodobé uchovanie až po sprístupňovanie. Zákon stanovuje nasledujúce možné formáty v SIP balíku: TIFF, MP3, MPEG-2, MPEG-4, JPEG2000, GML, WAVE. Pre zaujímavosť dodávam, že v SIP balíky do Dánskeho národného archívu môžu byť dodané na CD-R, DVD-R prípadne na USB médiu. V jednom SIP balíku môžu byť zastúpené všetky spomenuté prípustné formáty. Spolupráca CDA s Dánskym národným archívom je na úrovni konzultácií pri ukladaní problematických formátov do CDA.

- **Švédsko**

Švédsky národný archív (<https://riksarkivet.se>) uchováva a sprístupňuje zo zákona všetky dokumenty štátnych inštitúcií a historické dokumenty kultúrneho dedičstva vo Švédsku. Metodicky usmerňuje, koordinuje vedu a výskum v oblasti formátov pre dlhodobé uchovávanie v rámci iniciatívy PREFORMA. Vďaka nesmierne silnej legislatíve v oblasti uchovávaní dokumentov (RA-FS 2009:2) sú stanovené formáty pre text: HTML, XHTML, XML, CSS, XSD, DTD, XSL, TXT, JavaScript, dokumenty: PDF/A, ODF, obrázky: GIF, PNG, TIFF, JPG/JPEG, audio: WAVE, BWF Linear BCM, video: Matroska. Spolupráca CDA so Švédsky národným archívom je na úrovni konzultácií pri ukladaní problematických formátov do CDA.

- **Belgicko**

V Belgicku vznikla pozoruhodná organizácia „PACKED“- centrum analýz a expertíz v oblasti uchovania digitálneho dedičstva (<http://www.packed.be/en/>). Je to platforma pre spoluprácu organizácií archivujúcich a dlhodobo uchovávajúcich audiovizuálne diela. Cieľom práce organizácie je zlepšovať kvalitu a efektivitu pri digitalizácii a archivovaní všetkých typov dokumentov, najviac však audiovizuálnych, mediálnych diel. PACKED úzko spolupracuje s inštitúciami spojenými do projektu „EUROPEANA“ ([www.europeana.eu](http://www.europeana.eu)). CDA sleduje aktivity organizácie tak, by sa nové poznatky dali aplikovať do interných procesov, do aplikácií CDA.

- **EUROPEANA**

Europeana je digitálna knižnica, ktorá umožňuje širokej verejnosti prezerať si digitálne zdroje z európskych múzeí, knižníc, archívov a audiovizuálnych zbierok. Portál Europeana ([www.europeana.eu](http://www.europeana.eu)) bol spustený 20. novembra 2008 Viviane Redingovou, eurokomisárkou pre informačnú spoločnosť a médiá. Projekt je realizovaný centrálnym tímom, sídliacim v národnej knižnici Holandska - Koninklijke Bibliotheek. Europeana je internetový portál, ktorý funguje ako rozhranie k miliónom kníh, obrazov, filmov, múzejných predmetov a archívnych záznamov v celej Európe. Do Europeanei prispieva viac ako 2000 inštitúcií. (Louvre či British Library až po regionálne archívy a miestne múzeá Európskej únie). Umožňuje prístup k rôznym typom obsahu (digitálne objekty nie sú uložené na centrálnom počítači, ale zostávajú v jednotlivých inštitúciách). Europeana zhromažďuje kontextové informácie (metadáta) o položkách, vrátane malého obrázku. Pre prístup k celému obsahu, musíte prejsť na pôvodný web, ktorý obsah drží.

- **Rakúsko**

Dlhodobej ochrane digitalizovaných dokumentov sa venuje v Rakúsku viac inštitúcií. Z pohľadu CDA je zaujímavá práca digitálnej mediatéky „Die Österreichische Mediathek“ (<http://www.mediathek.at/>). Nielen preto, že digitalizujú a uchovávajú rakúske audiovizuálne dokumenty, ale preto, že ponúkajú tieto služby aj širokej verejnosti. Úzko spolupracujeme s vývojármi a pravidelne ich pozývame na odborné podujatia organizované v UKB. Od roku 2000 majú v mediatéke k dispozícii digitalizačný systém (DVA Profession), ktorý vyvinuli na mieru v spolupráci so softvérovými spoločnosťami. Systém ich digitalizácie pozostáva z katalogizácie, kde sa produkujú popisné metadáta a katalogizačný záznam s jedinečnou signatúrou do databázy DABIS. Následne prichádza fyzické, mechanické čistenie pôvodného média, reštaurátorské práce napríklad zlepenie roztrhnutej pásky audio kazety. Proces pokračuje produkciou technických metadát o type média, type pásky, platne, rýchlosti prehrávania, o výrobcovi, prehrávači. Samotná digitalizácia je automatický workflow pozostávajúci z validácie formátu, kontroly celistvosti, kontroly checksum MD5, výroby náhľadu. Digitalizácia sa končí uložením do dlhodobého úložiska typu RAID v dvoch kópiách. Prvá kópia je archívny master, druhá je náhľadová, určená pre prezerať a bádanie. Pre audio využíva rakúska mediatéka: Master – Broadcast Wave v kvalite 96 kHz/24 Bit, Náhľad – MP3 v bitrate 128 kbit/s. Pre video využíva rakúska mediatéka: Master – video codec FFV1, audio codec PCM bez komprimácie, container AVI, Náhľad – MPEG-2.

- **Nemecko**

Priekopníkom v Nemecku je nemecká národná knižnica „Die Deutsche Nationalbibliothek“ ([http://www.dnb.de/DE/Home/home\\_node.html](http://www.dnb.de/DE/Home/home_node.html)). Už od roku 2004 zastrešuje niekoľko projektov s témou digitalizácie a LTP. Najznámejší je projekt „KOPAL“, ktorého výsledkom bol formát pre elektronické zdroje „LMER“, softvér pre ingest a dissemináciu dát „koLibRi“, či projekt „AREDO“ pre LTP. Významným projektom je „Nestor“- centrálny systém pre metodiku, certifikácie, hodnotenie dlhodobých dôveryhodných úložísk. Archívny systém nemeckej národnej knižnice bol certifikovaný ako dôveryhodné úložisko pečatou kvality „Data Seal of Approval“. Vo svojich politikách uvádza ingest niekoľkých fixných formátov: PDF, TIFF, JPEG, PS, EPUB. V CDA sledujeme prácu nemeckých

kolegov, pretože zámerom CDA je v blízkej budúcnosti certifikovať archív pečaťou kvality, resp. certifikovať podľa normy LTP ISO 16363.

- **Albánsko a Kosovo**

Technické riešenie projektu Centrálny dátový archív je produktom spoločnosti Tempest a.s. Centrálna archivačná platforma je ponúkaná ako komplexné know-how v podobe definovania legislatívneho, procesného a metodického rámca pre dlhodobé uchovanie obsahu, jeho sprístupnenie a spracovanie. Predstavitelia ministerstiev kultúry a národných knižníc Albánska a Kosova navštívili v roku 2016 Univerzitnú knižnicu v Bratislave za účelom prehliadky Centrálného dátového archívu, oboznámenia sa s projektom, finančnými nárokmi archívu, technickým zabezpečením, personálnym pokrytím jedinečného riešenia dlhodobej archivácie dát. Celkovo si tak spravili obraz o robustnom systéme, ktorým CDA nepochybne je.

- **Veľká Británia**

Národný archív Veľkej Británie (<http://www.nationalarchives.gov.uk>) plní funkcie štátneho archívu. Metodicky zastrešujú digitálne kurátorstvo. Uchováva len stabilné formáty vhodné pre dlhodobé uchovanie. „The National Archives“ preferujú ingest týchto formátov: TIFF, JPEG, JPEG2000, PDF. Akceptujú aj ingest iných formátov, ale neuchovávajú proprietárne formáty, otvorené štandardy pre dokumenty, dáta v ASCII, Unicode. (<http://www.nationalarchives.gov.uk/about/commercial-opportunities/digitisation-services/our-digitisation-services/>) Spolupráca s NA VB je založená na aktualizácii registra PRONOM (<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>), register formátov, voľne dostupných validátorov a nástrojov na identifikáciu DROID, založil a spravuje Národný archív práve pre PFI a dlhodobé archívy na celom svete. CDA sleduje aktivity organizácie tak, aby sa nové poznatky dali aplikovať do interných procesov, do aplikácií CDA. Kolegovia z Národného archívu aktualizovali register PRONOM o nové formáty aj na náš podnet. CDA prijíma dáta vo formátoch, ktoré sú zverejnené v registri PRONOM, sú stabilné a majú dlhodobú podporu zo strany softvérových firiem.

- **Austrália a Nový Zéland**

Novátormi v oblasti dlhodobého uchovávanía digitalizovaných dát sú nepochybne kolegovia z opačnej strany zemegule. Národná knižnica Austrálie (<https://www.nla.gov.au/>) digitalizuje, dlhodobo archivuje a sprístupňuje historické, kultúrne dedičstvo nielen pre Austráliu, ale aj pre Indonéziu. Ich digitálna knižnica a depozit web stránok PANDORA obsahuje 4,8 Petabyte dát (jún 2016). Program PANDORA je funkčný od roku 1996 a skúsenosti kolegov z Austrálie sú pre nás jedinečné.

Národný archív Nového Zélandu (<http://archives.govt.nz>) plní funkciu štátneho archívu. Výskum a vývoj dlhodobej ochrany a uchovávanía dát je jednou z priorit tejto inštitúcie. Základným poslaním je nielen dáta uchovať a sprístupniť, ale zaručiť aj kontinuitu procesov v inštitúcii ako takej. Program „Digital Continuity“ je natoľko kľúčový, že bol vytvorený samostatný odbor, ktorý sa venuje zachovaniu existencie digitálneho archívu „Digital Continuity Action Plan“. Národný archív pracuje s odlišnými systémami ako CDA, majú implementované komerčne dostupné riešenie Fedora a Rosetta. Avšak riešia podobné problémy s validáciou dát, s certifikáciou archívu. Podrobne sledujeme prácu tejto inštitúcie.

Pridanou hodnotou k práci dvoch českých kolegov žijúcich a pracujúcich v Austrálii a na Novom Zélande sú pravidelné prednášky, semináre v Českej republike. Touto formou nám sprostredkujú vždy najnovšie poznatky „sharing good practice, lessons learned, quick winns, and complete know-how“.

#### 4. Formy spolupráce



V Európe je spolupráca medzi pamäťovými a fondovými inštitúciami na príkladnej úrovni. CDA UKB úzko spolupracuje, konzultuje, metodicky sa podieľa na tvorbe pracovných postupov pre LTP (Long Term Preservation). Foriem spolupráce je veľmi veľa, spomeniem najčastejšie využívané v CDA:

- Emailová a telefonická komunikácia,
- Interná spolupráca v rámci UKB s Odborom digitalizácie, Odborom informačných technológií a informačných systémov, s Odborom depozitu digitálnych prameňov,
- Osobné stretnutia, návštevy, konzultácie, stáže,
- Online video-konferencia - (NK Praha, MZK Brno, Archív Nového Zélandu),
- Prehliadky a exkurzie v CDA - Príkladná spolupráca s Univerzitou Komenského v Bratislave - Filozofickou fakultou, s Katedrou knižničnej a informačnej vedy, s Katedrou archívniectva a pomocných vied historických,
- Blog - <http://www.digitalpreservation.cz/>,
- Web - <http://cda.kultury.sk/>,
- Konferencie - Konferencia knižníc V4, PASIG, PREFORMA, INFOS, INFORUM, Digitálna Knižnica, Konferencia CDA'2016: Formátové výzvy LTP,
- Semináre - medzinárodný odborný seminár CDA'2016 Formátové výzvy LTP.

## 5. Aktuálne témy

Na základe skúseností z troch rokov prevádzky Centrálného dátového archívu, poznajúc kvalitu, množstvo, rôznorodosť digitalizovaných dát zo slovenských pamäťových a fondových inštitúcií, v súčasnosti riešime s kolegami, PFI, dodávateľmi, aj v spolupráci s kolegami zo zahraničia nasledujúce témy:

- Ukladanie proprietárnych formátov bez podpory vývojárov,
- Validácia video formátov,
- Validácia PDF 1.7, PDF/A,
- Bitová ochrana,
- Formátová ochrana,
- Starnutie médií,
- Certifikácia archívu ako bezpečného a dlhodobého úložiska dát.

## 6. Záver

Jednou z priorít Centrálného dátového archívu je budovanie a udržiavanie kontaktov s knižnicami archívami, s odbornými pracovníkmi zaoberajúcimi sa dlhodobou archiváciou dát. UKB sa aktívne zapája do medzinárodnej odbornej spolupráce a výmeny odborných poznatkov, kultivuje a rozvíja pracovné kontakty a pracovné siete, aktívne sa zúčastňuje na odborných podujatiach v zahraničí a prezentuje dosiahnuté výsledky zahraničným návštevníkom UKB. Môžeme konštatovať, že prístup inštitúcií k dlhodobému uchovávaniu dát sa líši s ohľadom na know-how, best practice, legislatívu v danej krajine, poslanie inštitúcie, rôznorodosť projektov a pestrosť zdigitalizovaných dát. Až budúcnosť ukáže, či prístupy boli správne, či vieme spoľahlivo uchovávať dáta pre nasledujúce generácie.

## 7. Zoznam skratiek

AIP - Archival Information Package - Archívny informačný balík

CAIR - Centrálna aplikačná infraštruktúra a registratúra

CDA – Centrálny dátový archív

IANA – The Internet Assigned Numbers Authority - register

DIP - Dissemination Information Package - Výberový informačný balík

DPX – typ formátu súboru pre video

DROID - Digital Record Object Identification

DVD - typ kompaktného disku

EPUB - typ formátu súboru pre elektronické publikácie

EST – Estónsko

GIF - typ formátu súboru pre obrázky

GDFR - Global Digital Format Register

HTML – Hypertext Markup Language - hypertextový jazyk

ISO – International Organization for Standardization

IT - Taliansko

JAG – typ magnetickej pásky

JPEG – typ formátu súboru pre obrázky

LTO – typ magnetickej pásky

LTP – Long Term Preservation – dlhodobé uchovávanie

MD5 – Message-Digest Algorithm je skupina kryptografických hašovacích funkcií

MP3 – typ formátu súboru pre audio

MPEG - typ formátu súboru pre video

NOC – Národné svetové centrum

OAIS – Open Archival Information System

OPIS PO2 - Operačný program informatizácia spoločnosti, Prioritná os 2 - Rozvoj pamäťových a fondových inštitúcií a obnova ich národnej infraštruktúry

PFI – pamäťová a fondová inštitúcia

PREFORMA – platforma, názov projektu

PRONOM – technický register formátov

PS - typ formátu súboru

PUID – jedinečný identifikátor formátu

RAID - Redundant Array of Independent Disks – súbor diskov - úložisko

SIP – Submission Information Package - Vkladový informačný balík

STN – slovenská technická norma

TAR - súborový formát slúžiaci na uloženie mnohých jednotlivých súborov

TIFF - typ formátu súboru pre obrázky

UC3 - California Curation Center

UDFR - Unified Digital Format Registry

UKB – Univerzitná knižnica v Bratislave

UNESCO - United Nations Educational, Scientific and Cultural Organization -Organizácia Spojených národov pre vzdelávanie, vedu a kultúru

URN:NBN – Uniform Resource Names: National Bibliography Numbers - Identifikačný systém je určený pre pamäťové a ďalšie inštitúcie na účely trvalej identifikácie digitálnych objektov v systémoch na správu, uchovávanie a sprístupňovanie dokumentov a informácií

V4 – Vyšehradská skupina alebo Vyšehradská štvorka, skrátene V4, je spoločenstvo štyroch stredoeurópskych štátov: Česka, Maďarska, Poľska a Slovenska.

ZIP - súborový formát slúžiaci na uloženie mnohých jednotlivých súborov

## **8. Zoznam odkazov**

<https://sk.wikipedia.org/wiki>

[www.iana.org](http://www.iana.org)

<http://www.udfr.org/>

<http://www.nationalarchives.gov.uk/pronom/>

<https://polona.pl>

<https://academica.edu.pl>

<http://www.ndk.cz>

<https://resolver.nkp.cz>

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

[http://cordis.europa.eu/fp7/ict/home\\_en.html](http://cordis.europa.eu/fp7/ict/home_en.html)

<http://www.unesco.sk/O-UNESCO>

<https://www.sa.dk/en/about-us/danish-national-archives>

<https://riksarkivet.se>

<http://www.packed.be/en/>

[www.europeana.eu](http://www.europeana.eu)

<http://www.mediathek.at/>

[http://www.dnb.de/DE/Home/home\\_node.html](http://www.dnb.de/DE/Home/home_node.html)

<http://www.nationalarchives.gov.uk>

<http://www.nationalarchives.gov.uk/about/commercial-opportunities/digitisation-services/our-digitisation-services/>

<http://www.opis.gov.sk>

<https://www.slovakiana.sk/>

<https://www.nocka.sk>

<http://cda.kultury.sk>

<https://github.com/veraPDF/veraPDF-library/releases/tag/v1.0.2>

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

<https://www.webdepozit.sk/projekt-dip/>

<https://www.mzk.cz/o-knihovne/odborne-cinnosti/konference-v4>

<http://oszkdk.oszk.hu/>

<http://www.digitalpreservation.cz/>

<https://www.nla.gov.au/>

[http://cda.kultury.sk/sk/Konferencia\\_Formatove\\_vyzvy\\_LTP](http://cda.kultury.sk/sk/Konferencia_Formatove_vyzvy_LTP)