



NACRX003JBGU

Dne: 10.12.2019 | Naše značka: NA- 3295-3/05-2019 | Vyřizuje/tel.: Stodůlka / 974 847 343  
Vaše značka:

## ZPRÁVA ZE ZAHRANIČNÍ PRACOVNÍ CESTY

MÍSTO:	Helsinki, Finsko
ÚČEL CESTY:	Účast na zasedání DLM Fórum
ÚČASTNÍCI CESTY:	Mgr. Zbyšek Stodůlka
ZPRÁVU PODÁVÁ:	Mgr. Zbyšek Stodůlka
NAVŠTÍVENÉ INSTITUCE:	Národní archiv, Finsko
TRVÁNÍ CESTY:	13. 11. 2019 - 16.11.2019
DATUM VYHOTOVENÍ:	10. 12. 2019

SCHVALUJE: PhDr. Eva Drašarová, CSc., ředitelka Národního archivu  
podepsáno elektronicky

## TEXT ZPRÁVY ZE ZAHRANIČNÍ PRACOVNÍ CESTY

### ČÁST VŠEOBECNÁ:

13. 11. 2019 11.35 odlet Praha, přelet Helsinky 14:45 hod.

16. 11. 2019 9.40 hod odlet Helsinky, přelet Praha 10:50 hod.

### ČÁST ODBORNÁ:

Předseda DLM Fórum Jan Dalsten Sørensen z dánského národního archivu připomněl, že v letošním roce slaví tato asociace 25. výročí. Tématy letošního podzimního setkání jsou důvěryhodnost digitálních repozitářů/digitálních archivů a dále prohloubení informovanosti stran iniciativy Evropské komise v rámci infrastrukturního programu Connecting Europe Facility (dále jen CEF) Building Blocks.

Zástupkyně ředitele Národního archivu Finska Päivi Happonen připomněla, že digitální agenda je klíčovou součástí aktuálního finského předsednictví s cílem zajistit bezpečné zpracování dat v rámci životního cyklu a lepší služby státu v oblasti veřejných služeb - zejména s principem „data only once“, Finsko plánuje do roku 2023 dosáhnout plné digitální komunikace státu s občany a firmami – inspirací jsou skandinávské země a Austrálie. Cestou je nová směrnice EU 2019/1024 z června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru, která předpokládá interoperabilitu mezi veřejným a soukromým sektorem, také nový projekt EARK3 a iniciativa EAG Archiving by design.

Fulgencio Sanmartín z Evropské komise představil současnou podobu CEF Building Blocks v rámci budování jednotného digitálního trhu s tím, že v rámci bloku eArchiving se nečeká dokonalé technologické řešení (např. jednotný formát), ale je třeba předvídat budoucí vývoj a přizpůsobit se mu (např. migrace formátu). Pilotně bylo ověřeno využití bloků eArchiving a eDelivery na DG TAXUD. Smlouva mezi EK a konsorciem projektu EARK3, která naváže na v dubnu ukončený EARK4all, bude uzavřena do konce roku 2019.

Janet Anderson (Highbury IVS) a Anders Bo Nielsen and Jan Dalsten Sørensen z dánského národního archivu představili za konsorcium EARK3 hlavní cíle – poskytnout více specifikací včetně některých redefinic, větší zapojení aktérů i mimo archivní instituce v podobě více zemí a více poskytovatelů služeb. Nový koncept připraví model zralosti (CMM), zohlední i analogové dokumenty a rozšíří přístup a vyhledávání. Další z členů Kuldar Aas z estonského národního archivu nastínil práce na software v rámci projektu - různé druhy validátorů, které budou poskytnuty formou open-source. Je možné využít service desk EARK na jeho webové stránce. Noví dodavatelé budou piql a Sydarkivera. Zcela nová oblast školení bude vycházet z aktuálních potřeb uživatelů – např. soukromý sektor k využití E-ARK specifikací a nástrojů. Plán využít různé formáty – osobní, e-learning (moodle, youtube atd.). Istvan Alfoldi, který vede v projektu skupinu k vytvoření referenční architektury, přiblížil, jak rozšířit a prohloubit architektonické řešení nad rámec OAIS. Skupina zkoumá stávající rámce jako EIRA (European Interoperability reference architecture), TOGAF, DERA a (Digital Heritage Reference Architecture), SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg). Ukázal obecné návrhy v rovině strategické, business, aplikační, technologické a fyzické vrstvy.

V odpolední části věnované aktuálním výzvám digitální archivace nastínil John Sheridan (Národní archiv VB) otázku technologických změn v kontextu některých ekonomických teorií, kdy hlavním motorem vývoje a změn v oblasti digitální archivace je ekonomika provozu (s posunem ke komoditizaci – snazší dostupnosti). Sebastian Gleixner z německého spolkového archivu

představil postup pracovní skupiny konsorcia nestor věnující se standardizaci balíčku SIP (AG SIP-Konkretisierung) v německém prostředí. Nelichotivá situace je způsobena absencí standardizace v uplynulých více než deseti letech (dříve-XBARCH, X-DOMEA atd.). Nová specifikace bude respektovat principy dle doporučení E-ARK, do prosince bude k dispozici německý překlad. Spolkový archiv plánuje využít specifikaci AIP dle standardu TR-ESOR, u filmů bude kontejnerem BagIt, DCDM, DCP a pro televizní vysílání MXF. Erik Loklindt z dánského národního archivu hovořil o probíhající analýze a přípravě nové generace řešení LTP ve své instituci do roku 2023. Předpokládá se využití domácích zdrojů. Aktuálně uloženo 300 TB (+ 500 TB digitalizátů), uchovávací strategie je migrace, úložiště na třech místech. Remke Verdegem z nizozemského národního archivu pak shrnula vývoj a stav implementace v prostředí své instituce v uplynulých dvou desetiletích. Řešení e-depot, které se v produkci používá od roku 2012, slouží jak pro trvalé uchovávání, tak i jako meziarchiv pro ústřední úřady a samosprávu. V současnosti je uloženo 1,2 PB dat.

Závěrečná část prvního dne se věnovala pilotnímu digitalizačnímu projektu finského národního archivu, který měl za cíl otestovat proveditelnost masové digitalizace analogových dokumentů a umožnil následně implementovat toto řešení u ústřední státní správy. Mikko Eräkaski finského národního archivu představil vládní zmocnění (2017), které vychází z předpokladu, že digitální archivace je nákladově efektivnější a digitalizace zároveň zlepšuje dostupnost, přičemž může být využita jako samoobslužná služba pro občany s využitím strojově čitelných dat. Digitalizace proběhne u původců a archiv převezme již jen data. V rámci pětiměsíčního pilotu, kdy bylo vybudována digitalizační linka s několika stanicemi (průběžný skener, knihovnický skener atd.). Při pilotu skenováno 400 bm (6,5 mil. obrázků), příští rok proběhne analýza nákladů. V pilotu byl jako ukládací formát zvolen JPEG, ale plánuje se využít i TIFF.

Po ukončení jednání proběhla krátká schůzka nově ustavené platformy DLM Fóra pro archivaci prostorových dat Geoforum, kterou vede Gregor Završnik ze slovinské společnosti Geoarh. Národní archiv se přihlásil k aktivní účasti na práci této skupiny.

Druhý jednací den začal podpisem dohody o porozumění mezi DLM Fórem a Open Preservation Foundation. Remco van Veenendaal z nizozemského národního archivu pak pokračoval představením výsledků certifikace jejich LTP-systému a získáním certifikace CoreTrustSeal. Proces trval tři roky a nejvyšší náklady si vyžádal předklad dokumentace do anglického jazyka (13.000 €). Dalším cílem je Pečeť nestoru, o ISO 16363 se zatím neuvažuje vzhledem k enormním nákladům a malé komunitě. Zajímavostí je, že v původním návrhu počítali se šesti určenými skupinami a po konzultacích vymezili nakonec pouze dvě – badatelé a obecná veřejnost. Norské řešení pro ukládání a popis relačních databází představil Torbjørn Aasen z KDRS. Uložili do formátu SIARD 13.000 databází samosprávných úřadů a s využitím vlastního SW DECOM je popisují. Problémem je nutná znalost výchozích systémů, činnost je časově náročná a problémy jsou třeba zavádějící názvy tabulek a sloupců. DECOM poskytuje přehledně název tabulky, kolik je tam řádek, zda je v nich LOB a umožňuje migraci objektů do PDF/A pomocí LibreOffice/Microsoft Office. Remco van Veenendaal poté v roli člena rady ředitelů OPF představil další plány na rozvoj aktivit této organizace. Portugalský projekt z oblasti bioinformatiky BioData.pt, na kterém aktivně participuje i české konsorcium akademických institucí z oblasti přírodních věd ELIXIR, představil Diogo Proença.

Dopolední blok zcela opanoval Espen Sjøvoll z norského národního archivu, který představil vize pracovní skupiny EAG Archiving by design. Tato skupina si klade za cíl nalézt nové způsoby získání dokumentů, neboť se ukazuje, že spisová služba je v digitální oblasti překonaný koncept, což doložil na příkladu masivních ztrát dokumentů na nejvyšší úrovni rozhodování norských ministerstev v oblasti zahraniční politiky a obrany – komunikace realizována jinými kanály a není tak možné zkoumat tvorbu politického rozhodnutí. V rámci projektu financovaného norskou výzkumnou radou se předpokládá využití institutu pre-commercial procurement k vývoji řešení, které zatím nejsou na trhu (inovační soutěž). V organizované diskusi však byl významně zpochybněn cíl této iniciativy s tím, že důraz by se měl klást na výraznější právní rámec správy dokumentů u původců, silnější mandát a zdroje archivů a zároveň využít možných technologických změn (např. umělá inteligence). Kati Saltiola a Noora Talsi z několika finských akademických institucí pak představily projekt Memory campus, který se zabývá digitalizací,

vytěžováním a zpřístupňováním v návaznosti na výzkum na několika finských univerzitách. Národní archiv, Národní knihovna Finska a Centrální archiv finských obchodních dokumentů ELKA zastupují nejdůležitější paměťové instituce. Součástí je i výzkumné centrum Digitalia (založily Univerzita jihovýchodního Finska pro aplikované vědy, Národní knihovna Finska a Univerzita Helsinky), které se zabývá výzkumem managementu digitálních informací vč. vývoje řešení digitální archivace. Vedle této akademické platformy tak vyrůstají různé obchodní příležitosti zaměřující se na technologii, vybavení, software, zabezpečení informací atd.

Jamie Kaminski z Highbury IVS představil složitou oblast zpřístupňování dat z 3D snímání uměleckých předmětů pro různé strategie (konzervování, restaurování, modelování atd.). Jedná se o relativně komplexní objekty, ale metadatové standardy jsou nekonzistentní a je třeba napřít tímto směrem úsilí mezinárodní komunity. Připomněl, že v tomto případě nemusí být migrace jedinou uchovávací strategií, ale reálně uvažovat také o emulaci. Anssi Jääskeläinen z projektu Digitalia v rámci Univerzity jihovýchodního Finska pro aplikované vědy představil projekt identifikace a rozdělení 75000 skenů v JPEG (60 GB), původně se jednalo o 7 bm materiálu. Cílem bylo identifikovat a rozdělit rozhodnutí a přílohy. K vylepšení kvality použít ImageMagick, k OCR Tessaract, součástí bylo i odstranění prázdných stránek, uložení PDF, jejich kombinace a uložení s využitím ID z původního systému. Postupnou optimalizací dosaženo rychlosti 1,5s/stránka, do budoucna snaha integrovat dělení slov pomocí API/UI a zpracování v clusteru.

Na problematiku životního cyklu dokumentu a správy dokumentů, ke kterému se DLM Fórum po letech opět vrací, se rozhodl soustředit Peter Anderton z britské společnosti Preservica. Připomněl neuralgické body, které vytvářejí myšlenkové hranice mezi vnímáním správy dokumentů a digitální archivací v oblasti, kdy záznamy jsou dynamické a rychle se mění v čase. Jörgen Rejström z Fujitsu nejprve obecně nastínil problematiku ochrany osobních údajů v životním cyklu dokumentu. Prakticky pak předvedl v rámci redakčního systému nastavení anonymizace osobního údaje po uplynutí doby výkonu veřejné funkce v prostředí webu – reprezentace v XML, PDF/A a pak v PDF/A po redakci.

Závěrečná část odpoledního programu se tradičně soustředila na představení nových členů. Platform of Trust je finská společnost, která poskytuje zabezpečenou platformu pro komercializaci dat mezi poskytovateli dat, uživateli těchto dat a zákazníky. Aktivita Mezinárodního výboru červeného kříže představila Brigitte Troyon. Archiv, který funguje od roku 1863, byl sloučen v roce 2010 se sekci správy informací, aby pokryl celý životní cyklus dokumentu. Spravují 18 km archiválií, elektronicky od 2010 celkem 7,5 mil. dokumentů a 8 mil. karet z 2. sv. války. V audiovizuálních sbírkách 104.000 fotek, 3.200 filmů a 3.900 audio záznamů zveřejněných on-line od 2016. Cílem jsou uživatelsky přívětivé služby. V současnosti řešen mj. projekt k hledání identit (identity matching), kdy při digitalizaci dokumentů např. z války v bývalé Jugoslávii se algoritmus snaží ztotožnit jména. Představení finské komercializace aplikovaného výzkumu v oblasti strojového zpracování dokumentů a digitální archivace zakončil Aki Lassila z Disec Oy, což je spin-off z výzkumných projektů k zobrazování zdravotnické dokumentace (PACS atd.), který dnes poskytuje vlastní cloudovou infrastrukturou včetně archivu.

Závěrem představil Jozo Ivanović z chorvatského státního archivu stav řešení digitální archivace v Chorvatsku, neboť příští jednání se bude konat v Záhřebu na konci května 2020 a to s tématem umělé inteligence. V současnosti se v Chorvatsku připravují dva hlavní projekty – centrální digitální repozitář pro digitální reprodukce archivů, knihoven a muzeí. Chorvatský státní archiv pak bude s ministerstvem zahraničních věcí, ministerstvem kultury a úřadem e-croatia budovat digitální archiv v rámci připravovaného státního cloudu.

Setkání umožnilo intenzivní výměnu zkušeností v oblasti digitální archivace, kdy trendem je věnovat pozornost opět způsobu a kvalitě předávání dat od původců, standardizaci procesů a jak ukázali finští kolegové na příkladu své země, využívat synergie regulace, výzkumu a komercializace výsledků pro rozšíření služeb v rámci efektivnější správy a využívání dat v elektronické podobě. Nově byla dohodnuta výměna zkušeností s nizozemským národním archivem v oblasti legislativní úpravy a metodiky k akreditaci digitálních archivů, stejně tak byl vyjádřen zájem ze strany zástupců Národního archivu Gruzie navázat na proběhlou studijní návštěvu v loňském roce hlubší kooperací při designu aplikace pro výběr dokumentů v digitální podobě.