



NACRX004Q3JZ

Dne: 3.11.2022 | Č. j.: NA- 3788-5/05-2022 | Vyřizuje/tel.: Rehtorik / 974 847 111
Vaše č. j.: | |

ZPRÁVA ZE ZAHRANIČNÍ PRACOVNÍ CESTY

MÍSTO:	Dánský národní archiv Rigsarkivet Kalvebod Brygge 34, 1560 København, Dánsko
ÚČEL CESTY:	Navázání kontaktů, představení postupů a metod ukládání geodat a databází v Národním archivu ČR k trvalému uložení v digitálním archivu; a reciproční detailní představení metod a postupů používaných v Dánském národním archivu.
ÚČASTNÍCI CESTY:	Mgr. Pavlína Nimrichtrová Mgr. Martin Rehtorik
ZPRÁVU PODÁVÁ:	Mgr. Pavlína Nimrichtrová Mgr. Martin Rehtorik
NAVŠTÍVENÉ INSTITUCE:	Dánský národní archiv
TRVÁNÍ CESTY:	17. 10. 2022, 6.15 až 19. 10. 2022, 21.59
DATUM VYHOTOVENÍ:	3. 11. 2022

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

TEXT ZPRÁVY ZE ZAHRANIČNÍ PRACOVNÍ CESTY

ČÁST VŠEOBECNÁ:

Oba účastníci cesty nastoupili cestu dne 17. 10. 2022 na letišti Václava Havla, s odletem v 6.15 a příletem do Kodaně v 7.40, letem EW4210. Pro pohyb po Kodani (mezi letišťem, archivem a ubytováním) byla využívána místní MHD a za tím účelem byla zakoupena jízdenka s platností na tři dny.

Ubytování bylo zajištěno v hotelu Zoku Copenhagen, Amagerfælledvej 108, 2300 København, Dánsko.

Jednání probíhalo každý den mezi 9 – cca 16.00 v administrativní budově Rigsarkivet.

Zpáteční cestu nastoupili účastníci dne 19. 10. 2022 odpoledne, návrat do ČR proběhl letem EW4211, mezi 20.40 a s příletem do Prahy ve 21.59.

ČÁST ODBORNÁ:

Na každý den byla připravena jednání dle přiloženého rozvrhu. Během plánovaných prezentací probíhala diskuse a výměna zkušeností na dané téma.

Program dne 17. 10. 2022 byl zahájen setkáním nad tématem *Zpřístupnění a přístup občanů k archiváliím* od vedoucí oddělení Anne Bækgaard, DNA nepoužívá automatický anonymizační nástroj, ale umožňuje občanům přístup k datům na základě požadavku "co o mě máte" reflektující obecně známé požadavky GDPR. Prvotní představa náročnosti ve výši 1 celého úvazku se ukázala jako nereálná, nyní má oddělení 17 osob a je to nedostatečné. Oddělení řeší stovky dotazů během roku, které obnáší nelehká rozhodnutí týkající se informací o rodičovství, domácím násilí ve střetu s právem na zapomení, a co z toho má kdy přednost.

Za hojně účasti kolegů a kolegyně napříč odděleními věnujících se digitální archivaci jsem představil postupy, jakým způsobem vybíráme a archivujeme databáze, a to včetně těch, která obsahují prostorové informace. Sdíleli jsme zkušenosti o původcích, požadavcích, které jsou nezbytné pro úspěšné vytvoření souboru/souborů pro archivaci databáze a problémy, se kterými se v praxi potkáváme. Dále byla pozornost věnována námi využívané aplikaci dbDIPview, kterou již nyní využíváme ke zpřístupnění archivovaných databází, ale také pro potřeby archivářů, např. k dohledávání různých informací, a zda je možnost jejího využití i ze strany Dánského národního archivu.

Po společném obědě se nám věnoval vedoucí digitálního oddělení Anders Rahauge a debatovali jsme nad strategiemi digitální archivace, zda je vhodnější komerční software, vlastními silami vyvíjený nebo open source, kterému se věnuje širší komunita. Vynechány nebyly ani možnosti automatizace procesů v digitálním archivu, plány na budoucí vývoj a problém s vendor lock-in u původců i archivů.

Poslední část prvního dne byla věnována formátům využívaných ke sdílení a archivaci prostorových dat. Kol. Ann-Kristin Egeland nám ukázala, jaké způsoby a postupy používá, když provádí formátové migrace do pro archivaci vhodného formátu GML pomocí software FME. Zvláštní část byla věnována validaci souborů GML v software ADA, který si Dánský Národní archiv vyvíjí vlastními silami. Kromě validace ADA nám byla představena také validace pomocí dalšího software jako např. FME, Altova.

Dopoledne **dne 18. 10. 2022** jsme věnovali prezentaci Národního archivního portálu, jeho vlastnostem a schopnostem, dále jsme představili Metodiku „Transformace prostorových dat...“, a to včetně nové webové podoby, která je takto snadno dostupná všem původcům v České republice. V praktické ukázce jsme ukázali, jak funguje nahrání a předání běžných balíčků SIP ze systémů elektronické spisové služby, tak i zjednodušenou tvorbu SIP přímo na portále. Ta je sice velmi univerzální, ale neumožňuje plné vytěžení dostupných popisných metadat pro balíčky s prostorovými daty, a proto jsme se zmínili o našich očekáváních od nástroje ArchiGIS, který v rámci projektu z podpory TAČR vyvíjí naši kolegové z Výzkumného ústavu meliorací a ochrany půdy. Dotazy od dánských kolegů pak byly směřovány především na bezpečnost přenosu data přes internet, které vůbec nevyužívají, dále na naše postupy a způsoby ukládání digitálních dat, využívání archivního popisu v mezinárodním standardu EAD a zájem byl i o námi uplatňovanou formátovou politiku, která je daleko méně přísná než ta dánská.

Téma formátů rezonovalo i v navazující části věnované formátům v oblasti prostorových dat, kdy se Dánský národní archiv hodně zaměřuje na "screening" nových formátů, kdy bylo nám naznačeno, že zajímavý výsledek by měl být výhledově zveřejněn. Dále se vedla debata nad způsobem akvizice celých GIS systémů, přijímání a zpracování metadat, možností jejich sklizení z veřejně dostupných zdrojů a mapování standardu INSPIRE na archivní standard INSPIRE.

V odpolední části jsme probírali možnosti zpřístupnění prostorových dat. Dánský archiv nedisponuje žádným speciálním nástrojem, kterým by byli schopni prezentovat archivovaná prostorová data široké veřejnosti, což sami vnímají jako nedostatek, ale vhodné řešení se teprve hledá. Nyní je tak badatel odkázán na populární open-source software QGIS, který však není úplně vhodný pro začátečníky a vyžaduje alespoň elementární obeznámenost s prací nad digitálními mapami. Navazující na toto téma, jsem stručně kolegy seznámil s informacemi o fungování mapového portálu Archivu slovinské republiky a také o úplně novém řešení vyvíjeném soukromým sektorem v ČR s názvem Hub4Everybody. Tento mapový portál má zajímavý potenciál pro archivy a byl vyvinut z podpory programů EU a NAČR tento zcela nový produkt bedlivě sleduje.

Poslední část dne byla zaměřena čistě prakticky, kdy jsme se věnovali technické stránce ukládání geodat v DNA na základě praktických ukázek, validací datových struktur a možnostem automatizace. Značné obtíže a problémy činí kolegům ukládání georeferencovaných rastrových dat; pro ukládání se nevyužívají podkladové vrstvy a další velkou výzvou bude plánované využití EARK GEOSIPU, neboť prozatím ukládají digitální data ve standardu SIARDDK, což vyžaduje vysoké nároky na celý proces digitální archivace a stále hledají dostupné nástroje pro tvorbu a validaci geodat.

Poslední den naší pracovní stáže začal představením pracovním postupů k akvizici informačních systémů v dánské státní správě. Byli jsme seznámeni s povinnostmi původců, kteří musí notifikovat archiv o nově zaváděném systému a postupech archivu, kdy se průběžně kontroluje, zda plán původce odpovídá reálnému stavu a systémem spravovaná data jsou v souladu s požadavky, které vznesl archiv, aby byla vlastní archivace v budoucnu realizovatelná. Jsou to právě průběžné kontroly u původců, které umožňují archivu zvolit vhodnou strategii, a to jsou: úplná archivace systému, archivace v podobě časových snímků, aniž by původce musel převzatá data odmazat, převzetí pouze uzavřených entit např. v 5 letém cyklu a nebo povolené k úplnému vymazání dat. Nemalá část prezentace byla věnována přípravě na předání archiválie do digitálního archivu, neboť vhodné procesy musí být dobře zavedeny už na straně původce a samotné převzetí dat a potvrzení ze strany archivu může zabrat i několik měsíců.

Převzetí digitálních dat a jejich validaci má na starosti speciální tým technicky vzdělaných archivářů, kteří nám na právě probíhajících příkladech živě předvedli, co všechno jejich každodenní činnost obnáší. Data, jak bylo zmíněno výše, nedostávají dálkový způsobem, ale pouze na zašifrovaných nosičích, které musí být důkladně prověřeny na přítomnost škodlivého kódu. Vlastní validace struktury a obsahu jsou prováděny pomocí domácího nástroje s názvem ADA, který umožňuje nejen např. vizuální kontrolu statických obrazových dat, ale také funkčnost relací mezi tabulkami a zřejmě i další drobné kontroly. Měli jsme možnost nahlédnout i do stroje, v níž jsou klonovány na archivní nosiče (magnetické pásky, optické disky), aby mohly být rozmístěny do různých úložišť v Dánském království.

V poslední části jsem nejprve otevřel otázku komunikace mezi informačními systémy, sdílená data a otázku vhodné strategie. Následně se rozvinula debata nad možností archivace pomocí formátu SIARD u NoSQL databází a sdílených pracovních prostředí. Na úplném konci nám Anders Bo Nielsen, který je hlavní stratég archivace databází představil plány na další vývoj formátu SIARD, který by měl v budoucnu daleko lépe a efektivněji zvládat práci s externě uloženými datovými objekty, ať už v extrémních datových objemech, s extrémním množstvím těchto objektů, příp. v kombinaci obojího. Jaké je plánováno sekvenování SIP na části a rizika, které dělení obnáší.

Celou tuto stáž s kol. Nimrichtrovou můžeme označit jako mimořádně přínosnou, neboť Dánský národní archiv je neskutečně pokrokový, ať už v mimořádných datových objemech, které je schopen akvizovat, ale také v profesionalitě a obeznamenosti všech archivářů s doporučenými postupy. Dále je nutné zmínit, že většina osob, které se nám věnovali, disponují různými kombinacemi vzdělání humanitního a technického. Většinou se jednalo o historii a nějaký obor z oblasti informačních technologií. Získali jsme velmi zajímavou zkušenost, a ačkoliv jsme si vědomi, že zásadní změnu a zvýšení našich schopností a znalostí nelze provést ze dne na den, tak nemálo drobných kroků a zlepšení můžeme reálně a rychle uplatnit nebo přímo zavést do praxe, a už to nás nepochybně výrazně posune dále. Ačkoliv se jednalo o velmi vyčerpávající program, kdy jsme začínali již před půl desátou dopoledne a končili až okolo páté odpoledne, tak celkové hodnocení možnosti toto vše se dovědět a vidět v praxi, dotázat se, má nesmírnou hodnotu, kterou on-line forma přenosu nikdy nedokáže nahradit.

Všechny kolegyně a všichni kolegové z Dánského národního archivu si zaslouží naprosto bezprecedentní velké poděkování za úžasnou vstřícnost, sdílnost a za obrovské množství informací, které jsme měli možnost získat a nakonec i velký respekt za to, jak vše mají precizně zvládnuto a zavedenou do naprosté rutiny.

Konkrétní stručná náplň všech setkání je obsažena v tabulce níže.

čas	Pondělí říjen 17, 2022	úterý říjen 18, 2022	středa říjen 19, 2022
10-11	<p>GDPR a ochrana osobních údajů, workflow pro přístup v DNA Anne Bækgaard, Zpřístupnění a přístup občanů</p> <ul style="list-style-type: none"> • Workflow pro přístup k archiváliím • Ochraňování citlivých dat - DNA nepoužívá automatický anonymizační nástroj - DNA umožňuje občanům přístup k datům na základě požadavku "co o mě máte" - Prvotní představa náročnosti ve výši 1 úvazku se 	<p>Archivace geodat v NA a portál NDA Martin Rehtorik, P. Nimrichtrová</p> <ul style="list-style-type: none"> • Představení metodiky Transformace geodat, popis posledního stavu • Vzhledem k tomu, že NDA dosud nemá uložené reálné datové balíčky s geodaty, byl předveden portál, postup nahrávání SIP z ESSL, tvorba datových balíčků z volných souborů • Dotazy DNA byly směřovány zejména na bezpečnost dat přenášených a předávaných internetem; postup ukládání v IS NDA, využití více úložišť NDA • Struktura datových balíčků, použití EAD, ukládací formáty – výstupní datové 	<p>Pre ingest, evidence IT systémů a schválení IT systémů v DNA Jette Holmstrøm Kjellberg and Lærke Friis Neergaard, Akvizice archiválií</p> <ul style="list-style-type: none"> • Postup notifikací o existujících a plánovaných systémech původců, formulář pro nahlášení systému a postup rozhodnutí o archivní hodnotě IS https://www.retsinformation.dk/eli/retsinfo/2013/9290. • Postup práce s původcem na přípravě předání archiválií, ověření

	ukázala jako nereálná, nyní má oddělení 17 osob a je to nedostatečné	formáty v ČR. <ul style="list-style-type: none"> Migrace formátů v IS NDA/geodat 	skutečností a požadavky DNA na systémy a jejich implementace u veřejnoprávního původce.
11-12	Ukládání databází v NA <i>Martin Rehtorik,</i> <ul style="list-style-type: none"> Postup ukládání DB, využití DBPTK a tvorba SIARD Příprava DIP za pomoci dbDIPview, předvedení databází připravených k nahlížení v digitální badatelně pro databáze. Práce s původci a požadavky na předávané DB, výběr archiválií z IS. 	Akvizice geodat v DNA <i>Lærke Friis Neergaard, Acquisition of Archival Data</i> <ul style="list-style-type: none"> Postup akvizice archiválií z GIS systémů, přijímání geodat v DNA, struktura SIP „Screening“ nových formátů pro geodata Debata o harvestování geometadat, využití INSPIRE, mapování INSPIRE na EAD 	Workflow pro ukládání databází (SIARD), část 1 <i>Thomas Bolbroe and Henrik Clausen, Digital Preservation</i> Detailní workflow ukládání databází - popis <ol style="list-style-type: none"> Ingest Validace Ukládání
12-13	oběd	oběd	oběd
13-14	Diskuse digital preservation v NDA <i>Anders Raahauge and Ann-Kristin Egeland, Digital Preservation</i> Témata: <ul style="list-style-type: none"> Jaká je uchovávací strategie NDA Jaké formáty NDA přijímá Migrace dat v archivu, používají se, jaké, jsou automatické? LTP systém Archivematica v NDA, LTP systém DNA – 	Přístup ke geodatům v DNA <i>Jakob Humlegaard, Communication of Archival Data and Ann-Kristin Egeland, Digital Preservation</i> <ol style="list-style-type: none"> Postup zpřístupnění geodat Využití QGIS Využití geometadatového katalogu v DNA/metadatový katalog DNA Slovinský geoportal/mapportal <i>Martin Rehtorik, NA</i> Prezentace slovinského mapového geoportálu	Workflow pro ukládání databází (SIARD), část 2 <i>Thomas Bolbroe and Henrik Clausen, Digital Preservation</i> Detailní workflow ukládání databází – praktická ukázka jednotlivých postupů. <ol style="list-style-type: none"> Ingest Validace Ukládání <ul style="list-style-type: none"> Předvedení přebírání SIP, antivirová kontrola, kontrola validity SIP pomocí sw ADA a následně operátorem; nahrání připravených AIP a

	<p>využívá vlastní sw</p> <ul style="list-style-type: none"> • Plány NDA do budoucna • Co jsou největší výzvy pro digital preservation v NDA, jaké jsou společné myšlenky, plány a cíle – obava z vendor lock-in; dlouhodobá udržitelnost řešení 		<p>úložiště. Vše probíhá v prostorách DNA, SIP jsou archivem předávány na hardwarových nosičích.</p>
14-16	<p>Tvorba GML pomocí FME <i>Ann-Kristin Egeland, Digital Preservation</i></p> <ul style="list-style-type: none"> • Ukázka přípravy GML pomocí podle požadavků DNA • Validace GML oproti schématu – v FME, Altova • GML validace v systému DNA - ADA 	<p>Uchování geodat v DNA <i>Ann-Kristin Egeland, Digital Preservation</i></p> <ul style="list-style-type: none"> • Ukládání geodat v DNA – podrobná ukázka, plány na automatizaci validace • Problematické ukládání georeferencovaných rastrových dat; pro ukládání se nevyužívají podkladové vrstvy • Plánované využití EARK GEOSIPU • Open source nástroje, využití dostupných nástrojů pro tvorbu a validaci geodat 	<p>Database preservation – SIARD and external documents <i>Anders Bo Nielsen and Ann-Kristin Egeland, Digital Preservation</i></p> <ul style="list-style-type: none"> • Ukládání externích objektů v SIARD – vysvětlení DNA, jak se pracuje v SIARD DK, rozdíly vůči EARK SIARD. • Vysvětlení, jak jsou externí objekty vkládány do SIP pro SIARD 2.2, možnost rozdělení objemných a tvorba sekvencí SIP (relativní URL v SIP)