

## Zpráva ze služební cesty do Švédska

**Místo:** Stockholm, Švédsko

**Termín cesty:** 4. 9. - 7. 9. 2012

**Účel cesty:** pracovní návštěva švédského Národního archivu a Královské (Národní) knihovny

**Účastníci cesty:** PhDr. Karel Koucký, Národní archiv

PhDr. et Ing. Milan Vojáček, PhD., Národní archiv

Mgr. Tomáš Dvořák

**Zprávu podává:** PhDr. Karel Koucký, Národní archiv

**Datum vyhotovení:** 12. 11. 2012

**Podpis ředitelky archivu:** PhDr. Eva Drašarová, CSc., podepsáno elektronicky

## Část všeobecná

Služební cesta do Stockholmu byla zahájena v úterý 4. 9. odletem z letiště Praha-Ruzyně a zakončena v pátek 7. 9. příletem do výchozího místa. Veškeré náklady na cestu (doprava, ubytování, strava, účastnické poplatky) byly pokryty z projektu *Zajištění ochrany archivních dokumentů důležitých pro potřeby státu* realizovaného v rámci programu bezpečnostního výzkumu České republiky 2010–2015 (kód projektu VG20112014054). Poskytovatelem je Ministerstvo vnitra. Řešitelem projektu je Národní archiv, spoluřešitelem Státní oblastní archiv v Praze.

## Část odborná

### **Švédská Národní knihovna (Kungliga biblioteket)**

Představení knihovny se ujala koordinátorka naší návštěvy paní Elisabeth Mannerfeldt (vedoucí odd. digitálních sbírek KB). Ve Švédsku je od r. 1661 zákonem umožněn veřejný přístup do knihovny. Královská knihovna původně sídlila na zámku Tre Kronor, a to až do požáru v r. 1697. Tehdy shořely dvě třetiny sbírek. K reinstalaci knihovny došlo teprve ve 2. polovině 18. století po dostavbě Královského paláce (*Stockholms slott* nebo *Kungliga slottet*) na místě bývalého zámku. Nové účelové budovy uprostřed veřejného parku Humlegården se knihovna dočkala v roce 1879, v roce 1887 byla znovuotevřena veřejnosti. V roce 1997 byla vybudována dvě podzemní křídla, která významně zvětšila ukládací prostory. Národní knihovna nyní opatruje kolem 15 miliónů knihovních jednotek.

Susanne Sellei (koordinátor digitalizace v KB) v krátkosti představila digitalizační aktivity v Národní knihovně. Oblast digitalizace kulturního dědictví je na národní úrovni v kompetenci ministerstva kultury, které ji usměrňuje normativními dokumenty. Knihovně je vládou zadán dlouhodobý úkol provádět digitalizaci všech švédských tisků za účelem jejich uchování a zvýšení dostupnosti prostřednictvím sítě Internet. Proto je v knihovně sestavována digitalizační strategie, která se stane součástí národní digitalizační strategie (představena bude v příštím roce). Problematika digitalizace tisků naráží z pochopitelných důvodů na autorskoprávní ochranu, proto se neprolamuje doba mladší než 140 let od vzniku díla. Jako

řešení daného problému knihovny vidí užívání tzv. kolektivních licencí, tj. užití díla v rámci např. sítě státních a univerzitních knihoven. Financování projektu se snaží doplňovat o soukromé sponzorské zdroje, nicméně vláda systematicky odrazuje státní instituce od takovéto úzké spolupráce se soukromým sektorem. Na digitalizaci se snaží spolupracovat v rámci obvyklé výměny znalostí a zkušeností s dalšími paměťovými institucemi (archivy a muzei), které řeší podobné projekty. Celý projekt má poměrně propracovanou organizaci a technickou infrastrukturu, jak bývá u podobných velkých projektů dobrým zvykem. Důležité pro nás bylo zjištění kritérií kvality digitalizace. V knihovně se razí stanovisko, že nastavení kvality musí být natolik dostatečné, aby se daná publikace nemusela znovu v budoucnu digitalizovat a zároveň, aby vynaložená práce byla adekvátní výsledkům (tj. aby byla efektivní v dalších důležitých ohledech digitalizačního procesu). Proto se opírají o stávající instrukce mezinárodních projektů orientovaných na metodiku digitalizace, zejména Metamorfoze a FADGI. Stanoveny jsou tři úrovně kvality, které se uplatňují podle typu materiálu. Dále jsou určena obecná pravidla pro snímání a úpravu digitalizačních výstupů. Jako datové formáty se uplatňují TIFF a JPEG2000 s rozlišením 300 ppi. Sérií experimentů knihovna dospěla také ke kompromisní úrovni kvality digitálních reprodukcí (úroveň 2), které je méně náročná. Bohužel však nejsou k dispozici propočty finanční nákladnosti digitalizace v daných úrovních digitalizace.

Pär Nilsson (senior knihovník) představil tři významné digitalizační projekty zabývající se skenováním novin, na kterých byl demonstrován postupný odklon knihovny od mikrofilmování a příklon k obecné digitalizaci. Knihovna schraňuje švédské noviny od poloviny 17. století, což aktuálně představuje množství 2 500 titulů o 122 miliónech stran. Se snímáním novin se začalo v roce 1949 prostřednictvím komerčních mikrofilmovacích společností. Postupem doby se podařilo převést 70 miliónů stránek na mikrofilm a proces mikrofilmování je naplánován až do roku 2014. V průběhu posledních deseti let začala být uplatňována kontrola kvality snímků podle mezinárodních standardů, která odhalila některé nedostatky. Další pokračování mikrofilmování se aktuálně nezdá jako udržitelné, třebaže jde o ověřenou a funkční technologii s očekávanou vysokou trvanlivostí výstupů. Problém se objevuje s dostupností mikrofilmovacích zařízení, která ustupují ze švédského trhu ve prospěch digitalizační techniky. První dva projekty (Tiden, TELplus ) se neukázaly pro Národní knihovnu jako příliš úspěšné, zatímco třetí (Digidaily) spustil perspektivní spolupráci se

Švédským Národním archivem. V rámci této spolupráce mají obě instituce za cíl v letech 2010–2013 digitalizovat velkoformátové noviny. Zatímco knihovna poskytuje noviny a bibliografické informace, archiv materiály připravuje, digitalizuje a provádí post-processing snímků (OCR, segmentaci obrázků). Náklady se v přepočtu pohybují kolem 5 – 30 Kč za stránku v závislosti na velikosti předlohy. Zajímavá je celková struktura nákladů – 47% zaujímá příprava materiálů, 39% skenování, 11% OCR a segmentace a jen 3% logistika a kontrola kvality. Letos by měly být první výsledky digitalizace zpřístupněny na Internetu v takovém rozsahu, který umožní autorskoprávní ochrana. Pro účely prezentace projektu funguje stránka <http://digidaily.kb.se/>. Další financování digitalizace novin je pochopitelně otázkou (stávající projekt je financován Evropskou unií, krajem Västernorrland, univerzitou Mid Sweden a společností Schibsted Media Group). Ale zdá se, že skenovací středisko Národního archivu ve městě Fränsta nabízí efektivní (nákladově přiměřenou) digitalizaci, proto by knihovna ráda pokračovala v digitalizaci všech 122 miliónů stránek novin.

Prohlídka knihovny jsme absolvovali v doprovodu pana Bengta Neisse. Navštívili jsme hlavního sál knihovny, depozitáře s uloženými novinami, místo uskladnění mikrofilmů, studovny vzácných rukopisů a studovny audiovizuálních dokumentů a nakonec i části digitalizačního střediska v budově.

Eva Lundgrenová (projektový manažer v KB) a Boel Larssonová (vedoucí odd. akvizic v KB) nás ve své prezentaci seznámily s projektem E-legal deposit. Jedná se o iniciativu zaměřenou na uchování digitálního obsahu vznikajícího v současnosti ve Švédsku nebo vytvářeného švédskými autory, resp. ve švédském národním jazyce. Tato iniciativa vyústila v přijetí speciálního zákona (E-legal Deposit Act, platný od 1. 7. 2012), který ukládá veřejným institucím včetně univerzit, nakladatelům novin a časopisů, provozovatelům rádiových a televizních stanic, zpravodajským agenturám a obecně profesionálním producentům a nakladatelům povinnost předávat vybraný elektronický obsah ze své činnosti k trvalému uložení do Národní knihovny. V Národní knihovně se na implementaci zákona chystali od roku 2009, v letošním roce spustili činnost, která navazuje na dlouhodobou snahu uchování webu (od r. 1997). V průběhu let sestavovali metodiku a technické řešení, přičemž se inspirovali řešením z Britské knihovny. Zkušenosti se sklízením webu celé švédské domény nebývaly dobré, proto se přistoupilo k výběrovému sklízení. V rámci implementace E-legal zákona se počítá s přechodným obdobím (od 31. 3. 2013 do 31. 12. 2014), ve kterém budou

povinni odevzdávat elektronický obsah pouze vybraní původci (konkrétně 15 státních institucí, 10 deníků a časopisů s celostátním nákladem, všechny rádiové a televizní společnosti s celoplošným vysíláním). Odevzdávat se bude elektronický obsah, který je na webu unikátní, konkrétně novinové a časopisecké články, části webových stránek, vybraná inzerce, letáky, průvodce, zprávy, elektronické knihy, webové vysílání, streamované audio a video, obrázky, fotografie apod. Naopak do výběru nebudou zařazeny kompletní webové stránky a databáze, programový kód, živé internetové vysílání, stránky s průběžně aktualizovaným obsahem (Wiki), podnikové intranety, soukromě publikované obrázky, hudba, film, blogy, komentáře, kalendáře, články zpravodajských agentur, komerční reklama apod.

Švédská Národní knihovna má dále zákonem danou možnost požadovat společně s digitálním obsahem dodání také příslušného popisu (metadat), konkrétně kde a kdy byl digitální obsah poprvé publikován, datový formát souborů, přístupové kódy k chráněným dokumentům, strukturální popis vztahů mezi jednotlivými soubory apod.

Původním záměrem bylo předávat data knihovně na off-line médiích (USB, CD). Ze strany producentů je ovšem vyvíjen tlak na možnost on-line přenosu, který přináší komplikace s přetížením síťové linky a též bezpečnostní rizika.

Pan Bengt Neiss (IT architekt v KB) nastínil pohled informatika na aktuální stav technické infrastruktury knihovny a obecně také svůj soukromý pohled na realizované digitalizační projekty. Zatímco před deseti lety se uchováváním digitálních dokumentů vnímalo pouze uložení dat (bitová ochrana), dnes se má na mysli celý komplex činností popsany v modelu OAIS; v roce 1997 se při sklizení webu ukládalo 200-300 GB dat, dnes se ukládá kolem 20 TB. Roční datový přírůstek při digitalizaci publikací je 40 TB, digitalizace audiovizuálních dokumentů představuje objem 350-500 TB ročně (1,5 TB denně). V oblasti dlouhodobého uchovávání digitálního obsahu představuje výzvu péče o výsledky webových sklizní (2 miliardy malých souborů) nebo zajištění cca 2 PB audiovizuálního obsahu (velké soubory).

Je zjevné, že největší objem představují právě audiovizuální dokumenty. Na otázku, kdy a proč začali řešit dlouhodobé uložení dat, bylo odpovězeno, že vše začalo se sklizením webu a prvním hromaděním dat. Tehdy bylo překročeno k vývoji vlastního software pro správu dokumentů na bázi OAIS s názvem MIMER. Byly to dva roky práce, na které pracovali místní programátoři a další dočasně najatí vývojáři. Cílem byl vývoj systému s okamžitou

dostupností dat při udržení autenticity obsahu. Cenová náročnost softwaru však není známá. Systém MIMER využívá otevřené technologie Java EE, GlassFish, MySQL a digitální úložiště Fedora Commons. Pan Neiss také popsal model entit a jejich vztahů v systému. Data jsou uchovávána v balíčkách AIP s využitím standardů METS, MODS, PREMIS, MIX apod. O hardwarové uložení se stará datové centrum EMC Atmos, zálohování je prováděno na magnetické pásky v rámci HSM od společnosti IBM.

Součástí prohlídky knihovny byla i návštěva datového centra. Jde o moderně instalovanou serverovnu s dvojitou podlahou a podstropními lávkami pro kabelové vedení. Na místě je instalováno 6 věží EMC Atmos, server s diskovým polem sloužícím jako cache a starší zálohovací zařízení na LTO-3. Stávající pásková knihovna LTO-5 je uložena v jiné místnosti a skládá se z 19 samostatných jednotek. Na otázku, jak dlouho probíhala migrace 2 PB dat z LTO-3 na LTO-5 byl uveden údaj 1,5 roku (výrobce technologie prezentuje rychlost migrace 14 měsíců). V diskusi zaznělo, že bitová ochrana dat je plně v rukou IT oddělení, ovšem technické řešení se vytváří na základě zadání kolegů z oddělení digitálních sbírek. Zajímavý je výběr technologií pro dlouhodobé uložení dat. Po zkušenostech knihovny je prováděna s ohledem na relativně krátkodobou udržitelnost technologií (zejm. problém morálního zastarávání) pravidelná obnova technologií, a to přibližně v sedmiletých cyklech. Je to dáno prudkým vývojem ICT, během kterého starší technologie přestávají podporovat výrobci a je nutné migrovat na novější (udržitelné) technologie. Tuto skutečnost knihovna zohledňuje ve svém finančním plánu. Z diskuze vyplynulo, že aktuálně nemá cenu hledat dlouhodobé (a nákladné) řešení ukládání dat, ale spíše se soustředit na neustálou udržitelnost technologií.

### **Švédský Národní archiv (Riksarkivet)**

Úvodní schůzka v Národním archivu byla koncipována jako diskuze v širším plénu o tématech, se kterými jsme švédskou stranu seznámili v předchozí e-mailové komunikaci. Jednání se zúčastnili pánové Magnus Geber, Börje Justell, Martin Bjersby (všichni odd. IT) a paní Désirée Veschetti Holmgren (konzultant spisové služby). S většinou zúčastněných jsme se dále průběžně setkávali po celou dobu návštěvy archivu.

V roce 1991 byl Národní archiv pověřen vládou systematickým shromažďováním matričních záznamů a jejich mikrofilmováním pro zajišťovací účely. Zároveň byly tyto mikrofilmy

konvertovány do mikrofišů a předkládány badatelům. V průběhu deseti let bylo zhotoveno cca 50 mil. snímků. Postupně se ovšem objevila potřeba obnovit technické vybavení mikrografického pracoviště, a to ve chvíli, kdy alternativní způsob zhotovování kopií archiválií - digitalizace - procházela překotným technologickým rozvojem. Externí dodavatelé techniky a služeb začali preferovat digitalizační technologii a na základě podrobného zhodnocení se Národní archiv rozhodl pokračovat nadále již pouze v digitalizaci (r. 2003). Třebaže mikrofilmování ještě nějaký čas pokračovalo, nyní jsou všechny kopie archiválií zhotovovány pomocí skenerů a digitálních kamer. Paralelně probíhá skenování mikrofilmové sbírky. Při této činnosti archiv výhodně spolupracuje s mormonskou církví ze Salt Lake City.

Mikrofilmování v Národním archivu tedy již neprobíhá. Ačkoli role výstupů z digitalizace archiválií je primárně v usnadnění distribuce archiválií uživatelům, tedy ve zpřístupnění dokumentů, digitální výstupy postupně nahrazují zajišťovací funkci mikrofilmů. Digitalizací jsou zajišťovány všechny požadavky na snímkování archiválií.

Master soubory ve formátu TIFF jsou uloženy ve dvou kopiích na LTO4 magnetických páskách ve dvou geograficky vzdálených lokalitách. Zároveň byly definovány i souborové formáty pro příjem dokumentů od původců. Tento přehled formátů se však neustále mění. Bylo deklarováno, že Národní archiv je i nadále schopen zpracovávat a ukládat velké množství dat, které se v současnosti pohybuje kolem 2,3 PB (tj. 140 milionů dokumentů). K tomu používají systém ESSArch, vyvinutý švédskou společností ES Solutions AB na bázi OAIS (<http://www.essolutions.se/>). Tento software je volně ke stažení na serveru SourceForge a používá ho např. také Národní archiv v Norsku v rámci projektu DIAS. Na otázku, zda mají k dispozici strategii udržitelnosti digitálních výstupů, bylo odpovězeno kladně, byť bez dalších podrobností.

Désirée Veschetti Holmgren nás seznámila s legislativou v oblasti archivnictví a e-governmentu. Švédský archiv má výrazné možnosti ovlivňovat legislativní rámec. Paní Holmgren mj. upozornila, že legislativní kodifikace formátů vhodných pro předávání dokumentů do archivu není dostatečně možná, a to především s ohledem na svižný vývoj v této oblasti oproti strnulým legislativním procesům.

Pan Karl-Magnus Drake (manažer digitalizace) je v současné době jediným pracovníkem archivu, který se systematicky věnuje práci a výzkumu mikrografické technologie. Představil projekt MECOM (*Methods of long-term storage of Electronic archives on COM*), který si klade

za cíl zhodnotit navrženou správu mikrofilmů vytvořenou pomocí technologie COM v návaznosti na masovou produkci digitálních snímků v archivu, porovnat kvalitativní rozdíly mezi digitálními vstupy a mikrofilmovými výstupy a finanční nároky na jejich dlouhodobé uložení. V rámci tohoto projektu bylo vyzkoušeno snímkování několika typů dokumentů na různých digitalizačních přístrojích a paralelní využití různých variant přístroje COM. K porovnávání kvality používali běžné, subjektivní vizuální srovnání, spektrofotometrické měření a konečně i proměřování technické terče podle normy ISO12233. Výsledek nebyl jednoznačný. Zajímavé jsou poznatky v otázce finančních nákladů. Zatímco náklady na ukládání mikrofilmů se rok od roku zvyšují o 2,5%, náklady na uložení stejného množství dat klesá každým rokem o 37%. Nutno však poznamenat, že u digitálních snímků je nutné počítat v pětileté periodě s obměnou technické infrastruktury, která si vyžaduje nové investiční náklady. Při srovnání ukládání snímků oběma způsoby ukazují zkušenosti Národního archivu na skutečnost, že náklady na uchování mikrofilmových snímků jsou ve střednědobé a dlouhodobé periodě (5-30 let) 3x menší, než náklady na uchování digitálních snímků.

Dalším představeným projektem byl projekt MiLoS (*Migration Free Long-Term Storage and Archival Medium and Process Chain*). Zaměřený je na hledání cesty, jak snížit náklady na migraci dat při jejich dlouhodobém uložení. Projekt si klade za cíl vyvinout nové světlocitlivé médium pro uložení snímků, které bude levné, obrazově kvalitní a trvanlivé. Vedle Národního archivu na projektu spolupracuje několik firem z oblasti mikrografického průmyslu. Rozpočet činí 4 milióny EUR a je plánován na 30 měsíců.

Pan Drake se dodatečně vyjádřil také k aktuální podobě nastavení digitalizačního programu v archivu. Národní archiv v současné době provozuje (sám nebo prostřednictvím externích dodavatelů) ve svém „skenovacím závodě“ ve Fränstě více než 50 digitalizačních stanic (digitální kamery, knižní skenery formátů A3 - A0, filmové skenery, mapové skenery). Na všech stanicích jsou při práci uplatňovány stejné principy směřující ke zhotovení kvalitativně porovnatelných výstupů (aplikace opto-elektronických měření s pomocí technických terčů a následné kalibrace skenovacího zařízení a vizuální kontroly kvality snímků, a to v pravidelných cyklech). Digitální snímky jsou zhotovovány ve dvojí kvalitě - tzv. master kopie ve formátu TIFF (Revision 6.0) s připojenými ICC profily (barevný prostor CIE Lab) a tzv. webové kopie ve formátu DjVu (deriváty z master kopií).

Odpolední program zahájilo krátké setkání v plénu za účasti pana Jonase Palma, který je v současné době považován mezinárodní odbornou veřejností za přední kapacitu v oboru digitalizace v prostředí paměťových institucí. Pan Palm se vyjádřil k vnímání hlavních cílů digitalizace švédskými archiváři, kterými jsou podle něho především: 1. zpřístupňování informací veřejnosti s dopadem na omezování přístupu k papírovým originálům a 2. nahrazování papírových originálů, které jsou ve špatném fyzickém stavu a hrozí jejich zánik (anebo k němu již dochází).

Pan Magnus Geber s částečným přispěním pana Justell ozřejmil historii a aktuální stav přejímání a uchovávání digitálních archiválií v Národním archivu. První přejímky elektronických archiválií se dají vystopovat již v 70. letech 20. století. Stávající složení dat (elektronických archiválií) ukazuje, že největší produkci zahrnují výstupy ze „skenovací továrny“ ve Fränsta, která denně vyprodukuje 120 tisíc obrázků. V současnosti Národní archiv ukládá 140 miliónů obrázků ve formátu TIFF, což představuje řádově asi 1,4 PB dat. Další objem dat představují multimedialní záznamy v řádu 100 TB, zbývající část pak prezentační (badatelské) snímky ve formátu DjVu (18 TB) a digitální dokumenty od původců (3 TB). Tento objem (2,3 PB dat) je uložen převážně na magnetických páskách. Za základě referenčního modelu OIAS a s využitím běžných metadatových standardů byla v archivu vybudována systémová informační platforma RADAR, která zahrnuje archivní systém ARKIS, zmiňovaný software ESSArch, testovací nástroj pro vytváření a kontrolu balíčků SIP tzv. RALF a konverzní nástroj a program pro příjem a výdej balíčků KRAM. Pro uchovávání databází upřednostňují norský metadatový model ADDML před švýcarským modelem SIARD (z důvodu celkové jednoduchosti řešení). Jednotlivé komponenty RADARu byly představeny a popsány, bohužel ukázka vzájemného fungování komponent předvedena nebyla.

Poslední prezentaci přednesl pan Matts Berggren (IT - architekt). Za zajímavé lze označit sdělení, že švédský Národní archiv zatím vůbec neoperuje s formátem JPEG2000 a pro účely zpřístupnění drží poněkud problematický formát DjVu. Pro účel výměn dat se však používá běžný JPEG. Další dotazy směřovaly k použití metadatových standardů v jednotlivých informačních balíčcích SIP a AIP. Pokus o získání náhledů informačních balíčků s faktickou strukturou a obsahem se však neseťkal s dopředu předpokládaným úspěchem.

## **Zhodnocení pracovní cesty**

Švédská Národní knihovna a Národní archiv patří v celosvětovém měřítku k institucím, které disponují dlouhodobými praktickými zkušenostmi z oblasti digitalizace a dlouhodobého uchovávání elektronických dokumentů. Zároveň byly do nedávné doby vnímány jako významná a stabilní centra, v nichž byly provozovány mikrografické technologie. Situace se ovšem v posledních dnech dramaticky proměnila. Technologický progres v oblasti digitalizace, důsledky zvýšeného zájmu veřejnosti o zpřístupňování v elektronické podobě a v tuto chvíli neodhadnutelná budoucnost výrobců mikrografické techniky, to vše má pravděpodobně vliv na postupné opouštění mikrofilmování ve švédských archivech a knihovnách. Tyto instituce se ale na druhé straně masovou digitalizací a překotným přejímáním digitálního obsahu dostávají do svízelné situace, ve které musí následně v rámci komplikovaných strategií nákladně řešit přístup k uloženým informacím a jejich trvalé a udržitelné uchování. Aktuální stav švédských institucí spravujících kulturní obsah se v této oblasti proto příliš neliší od situace archivů a knihoven v jiných zemích, které jsme zatím měli možnost v rámci projektu navštívit a studovat.