

Zpráva ze služební cesty do Rakouska – Vídeň

Účel cesty: Digitalizační projekty v Rakouské národní knihovně (ANNO, ALEX), použití programu Scope v Rakouském státním archivu v praxi

Termín cesty: 26. 11. 2008 – 28. 11. 2008

Účastníci cesty: PhDr. Tomáš Kalina; Ing. Miroslav Kunt; PhDr. Ing. Vojáček Milan, Ph.D.; Národní archiv

**S podporou projektu výzkumu a vývoje MV
č. VE2007200904**

“Možnosti a formy zpřístupnění archivních fondů nebo jejich součástí veřejnosti v elektronické podobě.”

Zprávu podává: _____

PhDr. Tomáš Kalina, Ing. Miroslav Kunt, PhDr. Ing. Vojáček Milan, Ph.D.,
Národní archiv

Datum vyhotovení: 5.1.2008

Podpis ředitelky archivu:

Organizační údaje o služební cestě

Odjezd z Prahy: dne 26. 11. 2008 v 13.33 hod. vlakem

Příjezd do Vídně: dne 26. 11. 2008 v 18.03 hod.

Odjezd z Vídně: dne 28. 11. 2008 v 18.33 hod. vlakem

Příjezd do Prahy: dne 28. 11. 2008 v 23.10 hod.

Ubytování v hotelu ve Vídni (2 noci)

Vyúčtování cesty provedeno v NA v předepsané lhůtě po návratu.

Průběh pobytu: Konzultace na předem dohodnutých pracovištích (viz název) proběhly.

Průběh a výsledky jednání

Österreichische Nationalbibliothek (dále ONB)

Datum návštěvy: čtvrtek 27. 11. 2008

Jednáno s: Mag. Christa Müller, vedoucí oddělení digitálních služeb (*Abteilung Digitale Services*)

Hromadná digitalizace v ONB

Digitalizační projekty Rakouské národní knihovny mají pro českého uživatele obzvláštní význam, protože zpřístupňují základní dokumenty pro historická i právní badání v 19. a 20. století (částečně též v 18.). Navíc se přímo nabízí srovnání s obdobným projektem Národní knihovny ČR Kramerius.

Klíčové projekty ONB jsou dva:

- **ANNO – AustriaN Newspapers Online**, elektronická knihovna periodik od roku 1716.
<http://anno.onb.ac.at/>
- **ALEX - Historische Rechts- und Gesetzestexte Online**, digitalizované sbírky zákonů, parlamentaria, judikáty a věstníky od roku 1780.
<http://alex.onb.ac.at/>

Kromě těchto vytvořila ONB ještě několik dalších projektů drobnějších, z nichž pro nás významná je digitalizace fotosbírky (**Bildarchiv Austria**) a plakátů z roku 1848 (**Revolution 1848**). Zvláště u Bildarchiv Austria je zajímavým způsobem řešeno elektronické tržiště z důvodu reprodukčních poplatků – celkově však tyto projekty využívají poněkud odlišné technické a ideové řešení, takže nebyly v rámci diskuse sledovány.

Digitalizace je prioritou ONB. Dříve se od 70. let mikrofilmovalo, od roku 2003 se již pouze digitalizuje s COM výstupem (na mikrofilm) a od roku 2007 se již žádné mikrofilmy nepořizují – ONB se plně soustředila na koncept dlouhodobého uchovávání v elektronické podobě (soubory TIFF s metadaty ve formátu XML). S nástupem nové ředitelky ONB (2001) byl změněn rozpočet, každý rok se na digitalizaci vydává 100 tisíc EUR, což odpovídá 1 milionu skenů/rok.

Základní postoj ONB je zpřístupnit, pokud tomu nebrání autorská práva, bezplatně co nejvíce dokumentů. Na základě finančního rozboru se totiž ukázalo, že jde o levnější řešení, než budování a údržba čtenářského místa ve studovně.

Zejména v případě ANNO jsou pro digitalizaci voleny dokumenty, ke kterým dosud neexistuje mikrofilm, jsou ve špatném stavu, popř. jsou řešeny s kooperujícím partnerem. Všechny dokumenty se skenují dodavatelským způsobem, u firmy "EMD - Elektronische u. Mikrofilm-Dokumentationssysteme Gesellschaft. m.b.H.", první výběrové řízení na dodávku proběhlo v roce 2003, druhé v roce 2007. V prvním kontraktu byly z digitalizovaných image současně vytvářeny bezpečnostní mikrofilmy (COM), v případě druhé zakázky bylo od mikrofilmů ustoupeno s ohledem na dostatečné zajištění bezpečnosti digitalizátů Originály jsou restaurovány – při digitalizaci dochází totiž k jejich rozvázání a skenování průchozím skenerem s výjimkou opravdu cenných tisků (v 1 exempláři).

Jako výstup z digitalizace byl zvolen formát TIFF, 300 dpi černobíle (2 bitově), uspořádání dat je ve stromové struktuře. Soubory jsou tak organizovány na disku – u periodik je základní struktura periodikum-rok-den. V rámci dne jsou v jedné složce všechna vydání (Ausgabe) daného periodika (např. ranní + večerní). Celá stromová struktura až po úroveň jednotlivého souboru (TIFF) je popsána v databázi Access, odkud jsou čerpána data pro webový prohlížeč.

Technické řešení zpřístupňovacího webového portálu řešila ONB vlastními silami, programátorem tu byl vystudovaný filosof. Základem technologie jsou cgi skripty. I když řešení působí technicky archaicky (mj. použití klasických rámců), je uživatelsky nesmírně přívětivé a pracuje velmi rychle.

Přímo naskenovaný soubor TIFF je on-line zpřístupněn také, ostatní čtyři prezentované soubory (ve formátu PNG) ve velikostech 1280-1024-800-500 dpi jsou generovány dynamicky serverem při prohlížení. Obdobně (dynamicky) je generován z vybraných stránek (nejvíce celé "Ausgabe") soubor PDF, kterým lze listovat, lze ho uložit, tisknout apod. Pracuje se s ním daleko snadněji než s jednotlivými stránkami.

V aplikaci pro digitalizované divadelní plakáty (všechny rakouské od r. 1855), které jsou skenovány barevně, jsou image prezentovány ve formátu JPEG s kompresí 85% (TIFF by měl cca 50 MB nejméně).

Knihovní systém Aleph má k dispozici proprietární systém pro správu image (DigiTool). Barevné originály (TIFF) se duplikují po digitalizaci do formátu JPEG2000, který ONB považuje za progresivní.

ONB by ráda umožnila fulltextové vyhledávání v periodikách, avšak naráží to na technické problémy (nečistoty na stránkách, fraktura) zvyšující chybovost OCR a ekonomické (rozpočet). V rámci projektu EU "Impact" (firma EBA) je testován OCR FineReader. Problém místních názvů a osobních jmen je snaha řešit digitalizací Staatshandbuch, telefonních seznamů, lexikonů apod., vlastně vytvořením slovníku autoritních záznamů (autoritní databázi využívá ONB německou, v Deutsche Bibliothek). Digitalizuje se na průběžném skeneru za cenu 10 centů/stránku včetně následných úprav. Katalogy knihovny i obsahy zákoníků v projektu ALEX jsou zadávány k přepisu do Vietnamu, kde vycházejí cenově příznivě: 1000 písmen za 40 centů (němčina) nebo 60 centů (ostatní).

V rámci ALEX se budou digitalizovat zákoníky (i zemské) i v jiných jazycích než je němčina. Sponzorem je zde Erste Bank, která má pobočky ve všech zemích býv. monarchie.

Digitalizace vlastními silami

Pro zpracování čtenářských požadavků na reprodukce (např. rukopisy, velké formáty) a menší digitalizační práce má ONB k dispozici skenovací pracoviště vybavené dvěma skenery book2net, kde je hned prováděn i základní postprocessing skenů. Skenovat je zde

možné dokumenty až do formátu A0 na zařízení Zeutschel US40000 (pořizovací náklady 30 až 40 tisíc EUR). Na skenery je placen základní servis formou technické podpory firmě Zeutschel (1000 EUR/rok), opravy jsou účtovány zvlášť.

Autorskoprávní otázky digitalizovaných dokumentů

Rakouský autorský zákon¹ je až na nevýznamné detaily prakticky totožný s naším (oba vycházejí z tzv. Bernské úmluvy o autorském právu). Autorská práva (resp. jejich majetková složka) tedy trvají 70 let po smrti (posledního) autora, fakticky by tak bylo možné zpřístupnit mimo studovny (po internetu) jen periodika starší roku 1880. Po konzultaci s právníkem knihovny však ONB zákon fakticky “porušuje”, protože zveřejňuje periodika 70 let po vydání. U monografií je uplatňována striktně přístupnost 70 let po úmrtí autora. Dosud nebyla v tomto směru zaznamenána jediná stížnost – ty lze teoreticky očekávat až se zveřejněním periodik z období anšlusu/okupace Rakouska. S vydavatelstvími, které mají právní nástupce byly uzavírány dohody, takže v některých případech jsou periodika přístupná i novější.

Österreichische Staatsarchiv

Datum návštěvy: pátek 28. 11. 2008

Jednáno s: Herbert Hutterer, Allgemeine Verwaltungsarchiv (AVA)

Program Scope byl již představen pracovníkům archivu na služebních cestách v minulosti (naposledy v Bernu a Basileji). Při současné prezentaci v počítačové učebně Rakouského státního archivu se jednalo spíše o využívané služby, praktické zkušenosti a postup prací při implementaci systému. Archiv aktuálně přešel na novou verzi aplikace Scope 4.3.0.5.

Každá součást Rakouského státního archivu má svého administrátora, pro AVA je jím H. Hutterer. Jednotliví pracovníci mají za úkol zadávat do systému postupně popisy skupin archivních fondů (Bestandsgrupe), archivních fondů až jednotlivin. Řada z nich se úkolu zhostila velmi laxně, v některých případech jsou data nepřístupná pro veřejnost (zpracovávají se). Systém má výhodu – prakticky on-line dokáže publikovat/zakázat publikování jednotlivin i celků, velmi podrobně jsou řešena uživatelská práva. Principiálně jsou fondy rozdělovány na knihy a spisy.

Administrátor definuje uživatelům formuláře pro zápis dat – poskládá je z množiny prvků, které má program k dispozici. V současnosti je takovýchto formulářů 40; ty lze upravovat nebo doplnit. Některá pole jsou při vyplňování kontrolována na správnost zápisu – v případě datace je dokonce k dispozici retrospektivní kalendář; kontrola správnosti zápisu data je prováděna i při importu.

Vytvářet rejstříková hesla a připojovat je k záznamům může každý pracovník, ale rejstříky prakticky nejsou v Rakouském státním archivu využívány. Jejich problémem je mj. nemožnost používat “neněmeckou” diakritiku – např. háčky a čárky. Je vidět počet záznamů, ke kterým je heslo připojeno. Rejstříky jsou hierarchicky strukturovány, při vkládání hesla zaznamenám, do které úrovně patří. Např. v zeměpisném rejstříku “Česká Třebová” bude ve

1) Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz). StF: BGBl. Nr. 111/1936. Základní popis viz Wikipedia, Die freie Enzyklopädie. Dostupné z [http://de.wikipedia.org/wiki/Urheberrechtsgesetz_\(%C3%96sterreich\)](http://de.wikipedia.org/wiki/Urheberrechtsgesetz_(%C3%96sterreich)), aktuální znění zákona dostupné z <http://www.ris2.bka.gv.at/GeltendeFassung.wxe?QueryID=Bundesnormen&Gesetzesnummer=10001848>>

větví “Česko”. U záznamu se následně objeví za názvem obce v závorce “(Ort/Tschechien)”. Rejstřík je generální, není možné ho vymezit pro jeden archivní fond nebo jinou část.

Pokud jde o problematiku evidence subjektů – jako podpora předarchivní péče, spisové služby a tvorby autoritních záznamů o původcích, zabezpečuje ji modul “Partner”, který však není v Rakouském státním archivu implementován. Modul pro předarchivní péči “Ablieferung” není zatím spuštěn.

Data z programu Scope je možné exportovat do Excellu nebo PDF. Stejně tak tisk je definován pevnými sestavami (sestavěnými zkušeným administrátorem), které se dají připojit jako plug-in. Zásadní informací je, že export do standardu EAD v současnosti Scope nemá a je nově programován na žádost a ve spolupráci s městským archivem v Amsterdamu.²

Jako samostatné utility (vedle vlastního Scope) jsou provozovány “Bildassistent” a “Findmittelassistent.”

“Findmittelassistent” umožňuje import databází (Access, Excell, text-csv). V případě, že je v importovaných datech chyba, je přesně identifikována. Naproti tomu velkým nedostatkem je nemožnost importu více úrovní najednou tímto způsobem.

“Bildassistent” slouží k pořizování a zpracování image. Skenuje se do formátu TIFF 300 dpi. Z takto získaných souborů jsou pomocí Bildassistenty vytvořeny 3 soubory JPEG s rozdílnou kompresí tak, že jeden z nich je tiskové kvality. Obrázky jsou uchovávány ve file-systemu, kde např. celý karton tvoří jednu složku. ID image je bezvýznamové, v rámci Bildassistenty jsou však přiřazeny názvy (např. signatura), které image napojí na data ve Scope s názvy stejnými.

Jelikož je využíván hosting na vládním serveru, je obtížná koordinace s ohledem na technickou podporu. Vedle produkčního serveru je provozován ještě server testovací. Zálohuje se jedenkrát denně. Zakoupeno bylo pro archiv celkem 50 licencí (cena SW není nejnižší).

Zhodnocení a přínos cesty

Nedostatky a problémy: nevyskytly se.

Přínos cesty/závěr:

V případě Rakouské národní knihovny bylo její pracoviště vybráno jednak proto, že s jejími výsledky čeští badatelé denně pracují, jednak proto, že technologická jednoduchost, razantní postup při digitalizaci a vcelku přijatelné náklady na něj, mohou být vzorem pro některé projekty v ČR. Současně se nabízí přímé srovnání (zejména uživatelské) s projektem Národní knihovny Kramerius.

Rakouský státní archiv je příkladem velmi obtížného zavádění informačního systému v paměťové instituci, kde doposud nic takového neexistovalo. Konzultace osvětlila některá úskalí přijatého řešení a předvedení funkčnosti některých částí systému Scope bylo inspirativní s ohledem na řešení naše. I když struktura vídeňského archivu je jiná než v archivech, kde jsou archivní fondy nebo jejich části jasně provenienčně ukotveny “na jednom místě,” ukazuje nezbytnost přímého hierarchického propojení struktury fondů a jejich částí s vlastními záznamy o archiváliích. Tedy nikoli oddělené “archivní pomůcky” a “evidence fondů”, jak je tomu dosud u nás.

V průběhu výkladu byly pořizovány snímky práce s programem Scope, které jsou k dispozici u řešitele (doplňují tuto zprávu).

²) Viz též Kalina, Tomáš – Kunt, Miroslav: Služební cesta Holandsko, Národní archiv 2005, dostupné z <<http://www.nacr.cz/Z-Files/sc2005-14.rtf>>