

PROJEKT HISPRA – ZÁCHRANA HISTORICKÝCH PRAGENSIÍ Z FONDU MĚSTSKÉ KNIHOVNY V PRAZE (MKP) (PROJEKT DIGITALIZACE PODPOŘENÝ TZV. „NORSKÝMI FONDY“)

Eva Měřínská – Kateřina Vojířová, Městská knihovna v Praze

Digitalizace v MKP před projektem HISPRA

V MKP existuje značné množství zajímavých dokumentů, které jsou svým charakterem unikátní a žádané. Jedná se především o obsáhlý a dnes cíleně budovaný fond pragensií, dále o dokumenty divadelní nebo hudební povahy, nemluvě o vzácných tiscích a bibliofilích. Čtenáři jsou tyto specializované fondy poměrně žádané. Bohužel jejich fyzický stav není většinou dobrý. Není to pouze důsledkem častých výpůjček, ale především kvůli nekvalitnímu kyselému papíru, který je charakteristický pro dokumenty 19. století a jeho degradace je prakticky nezadržitelná.

MKP v minulosti úspěšně žádala o dotace MK ČR (VISK 7) na záchranu některých dokumentů prostřednictvím digitalizace. Tím se povedlo do elektronické podoby převést např. Pražskou bibli z roku 1488 (poničenou povodněmi v roce 2002) a další tři dokumenty. Ovšem bilance čtyř digitalizovaných dokumentů za 5 let (2001–2005) nás příliš neuspokojovala. Bylo jasné, že budeme hledat i jiné formy získávání finančních prostředků na digitalizaci (a tím záchranu a zpřístupnění) našich ohrožených dokumentů.

Norské fondy a projekt HISPRA

V roce 2005 se naskytla možnost zpracovat projekt do programu Finančních mechanismů Evropského hospodářského prostoru a Norska (FM EHP/N; „Norské fondy“: http://www.mfcr.cz/cps/rde/xchg/mfcr/hs.xsl/fm_norska.html) v prioritní oblasti „Záchrana evropského kulturního dědictví“.

Projekt HISPRA (záchrana historických pragensijních a dalších vzácných dokumentů z fondu MKP) vznikl takřka dva roky. Byl zpracován rozpočet, přesný časový harmonogram prací, seznam vytipovaných titulů k digitalizaci, plán publicity a další dokumenty. Musela být charakterizována rizika a přislíbena minimální desetiletá udržitelnost projektu po skončení dotování z „Norských fondů“. Proces schvalování projektu byl zdoluhavý, doprovázený mnoha žádostmi o doplnění informací. Projekt postupně putoval od kontaktního místa FM EHP/N (pro MKP Magistrát hlavního města Prahy),

přes Monitorovací výbor MF ČR, Kancelář finančních mechanismů až k Výboru FM EHP/N, který rozhodl o udělení dotace.

Oznámení o schválení grantu obdržela MKP na podzim roku 2006. Rozpočet projektu je vyčíslen na 504 440 Eur, z toho 403 552 Eur, tj. 80 % tvoří dotace z „Norských fondů“.

Projekt HISPRA byl vystavěn na 4 hlavních cílech:

1. vybudovat vlastní digitalizační pracoviště;
2. pořídit digitální kopie vytipovaných dokumentů („záchrana“);
3. vytvořit online digitální knihovnu („zpřístupnění“);
4. rozšířit povědomí o těchto dokumentech a jejich využití („medializace“).

Financování z Norských fondů

Schválený rozpočet projektu je rozdělen do jednotlivých aktivit („budget headings“), u projektu HISPRA to jsou: „Project Management“, „Support of Sub-projects“ a „Services“. Náklady se vykazují čtvrtletně v „Project Interim Report“ a propláceny jsou zpětně na základě těchto zpráv. Přesuny financí mezi jednotlivými kapitolami jsou možné (max. do 15 % hodnoty každé položky), ale nejsou žádoucí a jsou těžko obhajitelné. Projekt HISPRA byl schválen v Eurech v době, kdy 1 Euro bylo 30,50 Kč, nyní je 1 Euro 25 Kč. Peníze tedy ubývají, aniž by byly čerpány. Po skončení projektu HISPRA je přislíbena financování digitalizačního pracoviště MKP od zřizovatele – Magistrátu hlavního města Prahy (MHMP).

Rozpočet projektu HISPRA

Celkové náklady projektu byly schváleny na 504 440 Euro s dotací 403 552 Eur. Finance začaly být čerpány od druhého čtvrtletí roku 2007, poslední účtovací období je druhé čtvrtletí roku 2010.

Budget heading „Project Management“ je propočten na 65 500 Eur. Z této sumy jsou financovány veškeré služby spojené s pronájmem a provozem pracoviště a administrativou projektu a dále náklady na propagaci a publicitu projektu. Tato položka byla z naší strany bohužel podhodnocena, především proto, že pracoviště vzniklo nakonec v jiných (větších a detašovaných) prostorech, než bylo plánováno.

Budget heading „Support of Sub-projects“ zahrnuje aktivity spojené s vybudováním pracoviště, tato kapitola byla uzavřena v prvním čtvrtletí roku 2008. Přestavba a vybavení prostor, nákup HW, serverů, SW a skenerů stál celkem 215 471 Eur. Nevyčerpáno bylo 12 329 Eur. Tyto finance nebylo povoleno přesunout do jiné aktivity projektu.

Poslední aktivita „Support“ zahrnuje veškeré personální náklady spojené s projektem. Je propočítána na 211 140 Euro.

Zrození pracoviště

Projekt HISPRA původně počítal s umístěním pracoviště v hlavní budově MKP na Mariánském náměstí v Praze. Postupně se ukázalo, že tato myšlenka je nereálná z hlediska dispoziční náročnosti pracoviště ve vztahu k prostorovým možnostem samotné budovy. Po nelehkém hledání nabídl MHMP knihovně jedny ze svých bývalých kanceláří v Provaznické ul. v blízkosti Můstku. Prostory bylo nutné kompletně zrekonstruovat od nové elektroinstalace, pokládky podlah, výmalby až po zavedení vzduchotechniky. Rekonstrukce probíhala v období červenec až září 2007. Během října 2007 bylo pracoviště vybaveno serverem (10 TB), sedmi PC a dvěma „face up“ skenery. Jedná se o skener značky Minolta (PS 5000 C, skenovaný formát do velikosti A3, plnobarevný, 600 dpi) a skener značky Zeutschel (OS 5500 TT, skenovaný formát do velikosti A2, škály šedi, 200–600 dpi dle předlohy). Zpracování naskenovaných obrázků probíhá v programu Sirius od firmy Elsyst Engineering, zpřístupnění dat přes, v českých knihovnách obvyklý, open source SW Kramerius. IT vybavení je doplněno o další dva servery (každý 10 TB) umístěné v budově ústřední knihovny, jeden z toho slouží výhradně jako datové úložiště.

Personální zajištění projektu

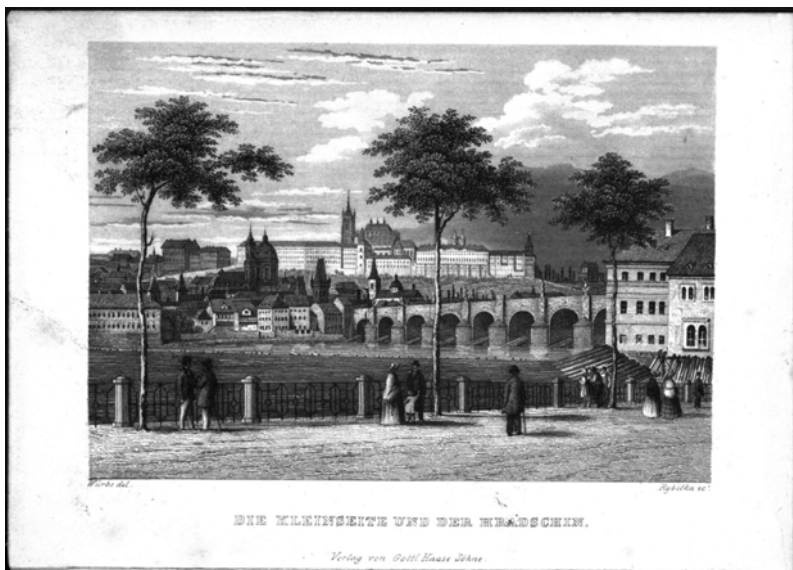
Za plnění projektu a vedení Oddělení digitalizace je zodpovědná jedna osoba. V rámci oddělení dále pracuje 10 zaměstnanců (2 na celý a 8 na poloviční úvazek), kteří se věnují samotnému procesu digitalizace. Projekt je dále podpořen 5 pracovníky, kteří se mu věnují v rámci dohody o činnosti. Jejich starostí je ekonomické řízení projektu, IT problematika a výběr a příprava fondu k digitalizaci. Za rekonstrukci a vybavení pracoviště byli zodpovědní dva pracovníci, celkem v rozsahu půl úvazku v časovém rozmezí realizace příslušných prací.

Medializace projektu

U grantů udělených evropskými institucemi je důležitá propagace projektů i donátora. Zavázali jsme se projekt HISPRA a Norské fondy propagovat formou informačního plakátu, brožury, CD nebo DVD a uspořádáním mezinárodní konference, která proběhne v roce 2010. V lednu 2008 byl slavnostně zahájen provoz pracoviště za přítomnosti velvyslance Norského království, zástupců MHMP, ministerstva kultury i financí a dalších institucí. Událost byla poměrně dobře mediálně pokryta jak v tisku, tak v rozhlase a televizi. (viz přehled: http://www.mlp.cz/dokumenty/pdf/tzp/TZ_Hispra_ohlasy.pdf)

Na cestě k digitalizaci

Dokumenty nutné k digitalizaci jsou po pečlivém výběru odbornými knihovníky ze Studovny pragensii Ústřední knihovny MKP předány do Oddělení digitalizace. Vybírány jsou přednostně, jak již bylo řečeno, historické pragensijní tisky nejvíce poškozené degradací nekvalitního papíru. Digitalizace se prozatím týká monografií, posléze přijdou na řadu i periodika. Výběr titulů je koordinován se seznamy poskytovanými NK ČR, které obsahují tituly určené k digitalizaci NK ČR nebo jinými knihovnami. Podmínkou poskytnutí dotací Norských fondů je totiž jedinečnost zpracovávaných předloh – nemohou se tedy zpracovávat dokumenty, které již byly digitalizovány jinou institucí. Jako způsob digitalizace byla v případě MKP vybrána přímá metoda, bez mikrofilmování. Všechna data jsou ukládána na třech místech, na různě lokalizovaných serverech – tím je docíleno jejich vysoké zabezpečení. Ve chvíli, kdy bude v NK ČR vybudováno národní datové úložiště pro potřeby paměťových institucí ČR, bychom chtěli svá data ukládat i tam.



Ukázka skenu z knihy KLUTSCHAK, Franz. *Der Führer durch Prag*. Praha, 1843.

Proces digitalizace

Na rozdíl od jiných pracovišť, jakými jsou NK ČR či Knihovna AV ČR se na zpracování jednoho určitého exempláře podílí v Oddělení digitalizace MKP vždy jeden operátor, který je za proces korektní digitalizace zodpovědný od počátku do konce. Tento postup se jeví jako optimální pro dané

pracoviště, kde se střídá 10 různých operátorů, a proto je nutné mít přehled o jednotlivých exemplářích. Tímto opatřením se také předchází určitému rutinnímu způsobu zpracování, a tím chybovosti při stabilním vykonávání jedné činnosti.

Skenování

Samotný proces digitalizace začíná skenováním, které je realizováno na dvou přístrojích – Minolta PS 5000C a Zeuschel OS5000 TT, které byly vybrány podle zvolených kritérií ve výběrovém řízení.



Práce se skenerem Minolta, foto: Sylva Hanzalová

Minolta skenuje plnobarevně, do velikosti formátu A3, a proto bývá využívána především pro práci s barevnými předlohami obsahujícími ilustrace. Zpočátku byla více využívána díky své všestrannosti a snazšímu způsobu ovládní. V průběhu skenování se ovšem objevily problémy s kvalitou skenovaných obrázků – především s obrysovou ostrotí. Čas od času se objevují také skeny, u kterých dochází k rozmazání částí stránek a vzniká tzv. rozsypaný obraz (jedná se o chybu při skládání jednotlivých částí obrázku při jeho ukládání).

Druhý skener Zeuschel umožňuje skenovat ve škálách šedi předlohy do velikosti A2, takže je používán především na velkoformátové knihy a mapy. Má také velmi dobře vyřešenou kolébku na knihy, díky níž lze skenovat také dokumenty s problémovou vazbou, případně objemnější svazky. Tento skener je možno ovládat automaticky pomocí nožních pedálů.

Předlohy se skenují ve formátu JPG v rozlišení 600 dpi a výstupy se ukládají ve stejném formátu.



Pracoviště digitalizace MKP, foto: Václav Jiroušek

Ořez, metadata, OCR, XML

Po naskenování jsou obrázky načteny do programu Sirius od firmy Elsyst Engineering, kde probíhá jejich úprava a zpracování. Bibliografické údaje se importují přímo z AKS Městské knihovny – z Koniáše. Sirius automaticky analyzuje okraje stránky, ty jsou ručně zkontrolovány a případně upraveny. Princip analýzy spočívá v ponechávání 2–3 mm okrajů kolem naskenované stránky, aby měl uživatel jasnou představu o tom, že při zpracování nedošlo ke ztrátě cenných informací. Pokud je třeba, dochází ještě k vyrovnání textu. Následuje ořez dokumentů, jehož výsledkem jsou jednotlivé stránky, které se ručně indexují – vkládají se čísla stránek, typy stránek. Po zpracování celého dokumentu je provedeno optické rozeznání textu. Na závěr procesu je vygenerován XML soubor, který je doplněn v programu ADM Creator o technická metadata. Poté již probíhá import dat do SW Kramerius, jehož prostřednictvím jsou dokumenty zpřístupňovány na www. Dále bude ke kontrole a následné opravě XML kódu používán SW XMetaL.

Zpřístupňování

Výsledky práce budou zpřístupněny ve formátu JPG v open source softwaru Kramerius, který pro zpřístupňování digitalizovaných dokumentů užívají i další knihovny včetně NK ČR. Nyní se digitální knihovna MKP postupně zaplňuje již zpracovanými dokumenty. Dokumenty, které již nejsou chráně-

ny autorským zákonem, budou pravděpodobně zpřístupněny bez omezení na internetu, takže si je bude moci kdokoli a odkudkoli prohlédnout (v SW Kramerius je také možnost si z maximálně 20 stránek vytvořit soubor ve formátu pdf). Dokumenty chráněné autorským zákonem budou přístupné pouze v rámci počítačové sítě MKP. V digitální knihovně je dostupná možnost textového vyhledávání využitím textových souborů vytvořených metodou OCR.

Následná ochrana dokumentů

Poté, co jsou dokumenty naskenovány a zpracovány, s nimi práce zdaleka nekončí. Součástí celého procesu digitalizace je také následná péče o dokumenty. Poškozené exempláře putují do restaurátorské nebo knihařské dílny, kde jsou opraveny a je na ně zhotoven ochranný obal z nekyselého lepenky archivní kvality. Tyto dokumenty se vrací zpět do knihovny a jsou opět k dispozici uživatelům. Dokumenty, které jsou velmi poškozené, se do studovny pragensií již nevrátí a budou uchovávány v úseku vzácných tisků, kde jim bude poskytnuta následná ochrana v klimatizovaném depozitáři.

Digitální knihovna MKP

Během příprav projektu bylo vytipováno okolo 20 tis. dokumentů vhodných k převodu do elektronických kopií. V rámci tříletého projektu bude z časové náročnosti zdigitalizována pouze část těchto svazků. Odhadujeme, že během hodiny bude při dokonalém zvládnutí procesu digitalizace převedeno do elektronické podoby (včetně metadat a OCR) okolo 20 stran jedním operátorem. Pracoviště bude ve své práci samozřejmě pokračovat i v následujících letech tak, aby byla postupně zdigitalizována potřebná část fondu MKP. Uvažujeme rovněž o nabízení digitalizace dokumentů v rámci služeb jiným institucím. Během projektu toto není povoleno.

Veřejnosti bude naše digitální knihovna pragensií přístupná přes webové stránky MKP. První výsledky budou dostupné na podzim roku 2008.

Závěr

MKP se vydala na složitou ale neodkladnou cestu digitalizace vzácných a historických dokumentů. Digitalizace umožní nejen záchranu jedinečného fondu, ale také jeho mnohonásobně lepší, širší a kvalitnější využití. Čtenářům a uživatelům naší knihovny tak nabídneme novou službu – vzdálený přístup do Digitální knihovny MKP a možnost studia jedinečných dokumentů z pohodlí svého domova.