

NA CESTĚ K MANUSCRIPTORIUM – VAŠE DOKUMENTY V MODERNÍ DIGITÁLNÍ KNIHOVNĚ

Tomáš Psohlavec, AiP Beroun s.r.o.

Manuscriptorium je databáze zveřejněná na Internetu, směřující k vybudování virtuálního badatelského prostředí pro oblast historických fondů. V evropském měřítku jde o nejrozsáhlejší digitální knihovnu zpřístupňující staré rukopisy. Tento projekt je od počátku koncipován jako otevřený k datům z různých zdrojů. V posledních měsících byly připravovány nové nástroje a postupy, které proces zapojení výsledků individuálních digitalizačních projektů výrazně zjednodušují. Následující text se snaží tyto postupy představit potenciálním přispěvatelům Manuscriptoria.

Současný stav

Manuscriptorium

Manuscriptorium aktuálně zpřístupňuje data z mnoha různých zdrojů – jde o významné české instituce a zhruba desítku zahraničních partnerů. V databázi jsou soustředěny katalogové záznamy, část z nich je doplněna o plně digitalizované dokumenty. V databázi je k 1.9.2007 cca 95 tisíc katalogových záznamů, z toho cca 2 200 dokumentů je digitalizováno. 130 dokumentů je opatřeno plnými texty.

Manuscriptorium pro školy

Manuscriptorium pro školy zvolna přerůstá v samostatný projekt, který je postaven nad společnou datovou základnou a využívá některá zobrazovací rozhraní původního Manuscriptoria. Základní myšlenkou a cílem projektu je vybrat a zpřístupnit ty nejzajímavější dokumenty, významné z hlediska výuky především na základních a středních školách a obohatit tak školní osnovy.

Samotné předpřipravené seznamy jsou doplňovány přípravami pro učitele, které mohou být vodítkem či jen inspirací při práci s konkrétním dokumentem. Tyto podklady připravují zástupci z řad pedagogů.

ENRICH

V letošním roce byl iniciován projekt podporovaný EU, ve kterém spolupracuje 17 evropských partnerů a který si klade za cíl agregovat dostupná digitální data o historických dokumentech, zpřístupnit je na platformě Ma-

nuscriptoria a vytvořit tak evropskou digitální knihovnu pro oblast nejstarších historických fondů.

Přístup do Manuscriptoria

Licenční politika: Placený a bezplatný plný přístup

Manuscriptorium je v současnosti plně přístupné pouze na základě licence a licenční politiky, kterou definovala NKČR. Jednotlivci mohou získat roční přístup za částku 2380 Kč. Během 12ti měsíční platnosti licence lze zobrazit 50.000 obrazů a dalších souborů, neposkytovaných volně.

Knihovny a jiné instituce mohou zakoupit licenci umožňující roční plný přístup do databáze za 4760 Kč. Licenční smlouva umožňuje zprostředkovat přístup pro návštěvníky/klienty instituce a není omezena počtem zobrazených obrazů.

Bezplatný přístup

Partnerům, kteří přispějí do Manuscriptoria svými daty, je bez ohledu na objem spolupráce umožněn trvalý bezplatný přístup k datům poskytnutým ostatními partnery. Stejně tak je do Manuscriptoria umožněn bezplatný přístup všem stupňům škol v České republice a to díky dohodě NKČR a MŠMT.

Zdroje dat pro Manuscriptorium

Katalogové záznamy

Samostatné katalogové záznamy vznikají buď jednotlivě přímo pro Manuscriptorium ve formátu XML dle MASTER DTD, nebo jsou do Manuscriptoria konvertovány v dávkách z jiných formátů – nejčastěji MARC 21, UNIMARC, ale i z proprietárních formátů partnerských institucí. Kromě toho jsou do Manuscriptoria převáděny i tištěné katalogy a další zdroje.

Digitalizované dokumenty

Většina digitalizovaných dokumentů, které jsou v současnosti zpřístupněny v MANUSCRIPTORIU vznikla v rámci programu VISK6 na pracovištích v NKČR. Manuscriptorium je ale samozřejmě otevřeno i dalším zdrojům digitalizovaných dokumentů.

Postup vedoucí ke zpřístupnění dokumentů partnerských institucí se liší podle toho, zda se jedná o zpřístupňování nově vznikajících dokumentů, či o zpřístupňování výsledků existujících digitalizačních projektů.

- U nově vznikajících dokumentů lze doporučit postupy aplikované v Manuscriptoriu a pro produkci metadat o jednotlivých dokumentech použít volně šířené programy z produkce Manuscriptoria (např. M-Tool).

- U projektů, které již dosáhly v oblasti digitalizace většího objemu výsledků, pravidelně spolupracují a splňují pravidla Manuscriptorium Compatible, jsou realizovány konektory, které rutinně konvertují předávaná metadata do MASTERu za účelem prezentace v Manuscriptoriu.

U projektů, které navíc provozují vlastní digitální knihovnu s vytěžovacím rozhraním, je možné vyjednat napojení do Manuscriptoria opatřené automatickými konverzemi (harvester pro Manuscriptorium se právě dokončuje).

PROČ zpřístupňovat data prostřednictvím Manuscriptoria?

- **Manuscriptorium je snadnou cestou k modernímu prezentačnímu systému**

Vývoj či nákup vlastních kvalitních rešeršních nástrojů je nákladnou záležitostí, stejně jako jejich údržba či rozšiřování funkčnosti. Partneri, kteří se rozhodli přispívat svými daty do Manuscriptoria získávají bez finančních investic moderní prezentační a rešeršní platformu, která se plynule přizpůsobuje aktuálním požadavkům badatelů a v poslední době i potřebám laických uživatelů.

Tyto nástroje získá partner trvale při importu svých dat - bez ohledu na jejich objem.

- **Manuscriptorium akceptuje existující data v libovolných formátech**

Formát vhodný pro ukládání dlouhodobě životných dat je podle našeho názoru XML, například struktura dle MASTER DTD. Nechceme však partnerské instituce a jejich pracovníky omezovat v oblasti pořizování dat – proto Manuscriptorium akceptuje data v různých formátech (MARC21, UNIMARC a další). Badatelé se tak mohou soustředit na svou badatelskou činnost a používat ty nástroje, které jim vyhovují a na které jsou zvyklí. Výsledky jejich práce jsou pak převzaty a teprve uvnitř Manuscriptoria jsou pro účely prezentace převedeny do formátu MASTER.

Pokud si však partner přeje vytvářet data přímo v MASTERu, je možné používat volně dostupné nástroje z produkce AiP Beroun s.r.o., které jsou volně ke stažení na adrese www.manuscriptorium.cz.

- **Zvýšení užítka z pořizovaných dat**

U historických dokumentů je jejich současné uložení pro badatele spíše podružnou informací a zpřístupnění v lokálních systémech neumožňuje plně využít potenciál dat pořizovaných při digitalizaci.

Důležitý je obsah, časové souvislosti, okolnosti vzniku atd. Zveřejněním v široké množině dokumentů se informace o konkrétním exempláři

(ať už vlastní obsah či informace vytvořené při digitalizaci) dostávají do nových souvislostí – tím je užitek z práce vložené do jejich pořizování významně navýšen.

Zároveň si ale Manuscriptorium neklade za cíl nahradit existující lokální digitální knihovny. Ty mohou případně lépe řešit některé specifické potřeby konkrétních institucí. Proto mohou být záznamy v Manuscriptoriu opatřeny i odkazem na lokální prezentace.

Konektivita na ostatní systémy

Jednou z významných funkcí Manuscriptoria je napojení na významné projekty typu JIB, TEL a další. Manuscriptorium disponuje vytěžovacím rozhraním, ve kterém předává nejdůležitější data obsažená v katalogových záznamech pomocí protokolů Z39.50 (MARC21) a OAI-PMH (Dublin Core, MODS, OpenM a další formáty).

Informace zpřístupněné v Manuscriptoriu jsou tedy viditelné i prostřednictvím portálů, využívání pořízených dat narůstá (a roste případně i návštěvnost lokálních digitálních knihoven).

Pokud jsou data partnera přijata do Manuscriptoria, je to pro partnerské projekty automaticky zárukou kvality a není potřeba zapojení do těchto projektů vyjednávat individuálně.

Jak zpřístupňovat data prostřednictvím Manuscriptoria

abízená data by měla splňovat vlastnosti, které popisuje dokumentace Manuscriptorium Compatible (dostupná ke stažení jako soubory PDF na www.manuscriptorium.cz v části „Partnerům / Dokumentace“).

Pro partnery, kteří s projekty pořizování digitálních dat teprve začínají, existují a jsou volně dostupné nástroje usnadňující vytváření potřebných informací.

Obecné zásady importu dokumentů do Manuscriptoria

Cílem Manuscriptoria je vytvořit virtuální badatelské prostředí pro oblast historických fondů – neomezuje se na zpřístupnění dat konkrétních institucí, ale chce zpřístupnit veškerá dostupná existující i nově vznikající data. To od Manuscriptoria předpokládá značnou flexibilitu pokud jde o zpracování importovaných dat a zároveň implikuje i určité technické vlastnosti řešení.

Hlavní zásadou Manuscriptoria je, že do systému jsou importována pouze metadata, tj. bibliografické popisy a v případě digitalizovaných dokumentů také informace o struktuře dokumentu informace o umístění obrazových dat případně další informace (např. technické o digitalizaci atp.)

V Manuscriptoriu nemusí být fyzicky uložena obrazová data pořizovaná v rámci individuálních digitalizačních projektů jednotlivých institu-

cí. Manuscriptorium disponuje vlastním obrazovým serverem, odkud jsou prezentovány především obrazy odvozené z dat pořizovaných v rámci programu VISK6 na pracovišti v NKČR. Výstupy jiných projektů jsou v Manuscriptoriu zpřístupňovány z webových serverů partnerských institucí a jedinou technickou podmínkou je, aby obrazové soubory byly dostupné pomocí protokolu HTTP.

Technické řešení Manuscriptoria zaručuje, že uživatel vizuálně nepozná, z kterého místa jsou obrazová data do prohlížeče přenášena.

Toto řešení má především tyto důvody:

- objem dat v evropském měřítku (snaha budovat centrálně spravované místo pro uvažovaný objem dat nemá ekonomické, ale ani technické opodstatnění)
- různé přístupy k vytváření a uložení obrazů a jejich kvalitativních hladin (Manuscriptorium nechce předepisovat formu, což by v případě centrálního uložení bylo nezbytné)
- řada partnerů trvá z různých důvodů na správě obrazových dat (což lze při uvažovaném řešení zajistit)

Příprava digitalizovaných dokumentů

Příprava obrazových dat

V Manuscriptoriu jsou obrazy zpřístupňovány v několika uživatelských kvalitách, které jsou všechny odvozeny od archivních dat – samotná archivní data nejsou prezentována on-line.

Obvykle se v Manuscriptoriu prezentují tyto pojmenované kvality:

Normal	nejvyšší on-line rozlišení, v současnosti ještě opatřené vodoznakem,
Low	nižší rozlišení, bez vodoznaku,
B&W	černobílé obrazy optimalizované za účelem čtení a s ohledem na snížení objemu přenášených dat,
Preview	malý náhledový obrázek pro rychlou orientaci a navigaci ve větších obrazech,
Gallery	miniatura stránky určená pro náhledové galerie.

Manuscriptorium nepředepisuje počet kvalit, jejich konkrétní vlastnosti ani konkrétní formáty jednotlivých hladin.

Formáty:

Zobrazovací rozhraní Manuscriptoria umožňuje zobrazovat běžně v prohlížeči podporované formáty (GIF, JPG, PNG ...). Pro Manuscriptorium tedy můžete jednotlivé kvality připravit v libovolném z těchto formátů.

Kvalitativní hladiny: záleží na zpracovateli dat, jaké kvality zvolí pro zobrazení v Manuscriptoriu. Optimální je s ohledem na uživatele odvodit z archivních dat alespoň tři kvality obrazů:

- Nejnížší, která se použije pro prezentaci v galerii stránek
- Náhledovou, která se zobrazí v náhledech stránek
- Vyšší, která je určena pro čtení a detailní prohlížení obsahu stránek

Manuscriptorium se k vzdáleně uloženým obrazovým datům chová tak, že bez ohledu na počet kvalitativních úrovní použije nejnížší kvalitu pro zobrazení galerií, druhou nejnížší jako náhledový obrázek. Pro detailní prohlížení lze zvolit jakoukoliv kvalitu. Systém může teoreticky pracovat i s jedinou obrazovou kvalitou – je však lépe brát ohled na uživatele a kvality pro jednotlivé účely připravit.

Více informací o obrazových datech lze opět nalézt v dokumentaci „Manuscriptorium Compatible“.

Aby bylo možno připravené obrazy zobrazovat v prostředí Manuscriptoria, je potřeba do systému importovat potřebná metadata.

Příprava metadat digitalizovaných dokumentů

Nejdůležitější a povinné informace, které musí metadata digitalizovaných dokumentů obsahovat, aby byla importovatelná do Manuscriptoria jsou:

- informace o struktuře dokumentu (někdy nazýváno jako elektronická vazba – určuje správné pořadí stránek atp.)
- informace o místě výskytu jednotlivých obrazů všech dostupných kvalit (které umožní vlastní odkazování na prezentované obrazy konkrétních stránek)
- skutečná (aplikovatelná) paginace / foliace (umožní uživateli identifikovat/odkazovat konkrétní folio či stránku dokumentu)

Při převodu existujících metadat pomocí konektorů jsou tato data převáděna automaticky dle definovaných pravidel do interního formátu Manuscriptoria – MASTER+.

Pro jednotlivě vznikající dokumenty lze použít program M-Tool, který tato data generuje na základě zadaných informací. Tento program je volně k dispozici na www.manuscriptorium.cz v sekci „Zdroje“.

Příprava popisných metadat

Popisná metadata tvoří především bibliografický popis dokumentu, případně lze do formátu MASTER+ zapsat i popis jednotlivých stránek či folií.

Při převodu existujících metadat pomocí konektorů jsou tato data převedena automaticky.

Pro jednotlivě vznikající dokumenty lze opět použít program M-Tool, který umožňuje tato data zadat do připravených formulářů. Objem a struktura in-

formací, které tento program dovoluje zapsat, je vybrán jako „optimální minimum“ pro vytváření kvalitních záznamů. V případě, že autor popisu potřebuje zapsat širší okruh informací, je vhodné použít některý obecný XML editor.

Kontrola připravených dokumentů před importem do Manuscriptoria

Pro kontrolu nabízených metadat bylo vytvořeno on-line rozhraní M-Can (Manuscriptorium kandidátů, <http://candidates.manuscriptorium.com>), které umožňuje připravené dokumenty zkontrolovat – po nahrání metadat je zobrazen digitální dokument tak, jak bude prezentován v Manuscriptoriu.

Příklad praktického postupu – import jednotlivých dokumentů

Tento příklad předpokládá, že jednotlivé digitalizované dokumenty vznikají postupně a k jejich tvorbě jsou využity přímo nástroje poskytované Manuscriptoriem.

- **Inicializace spolupráce:** nejprve je nutno kontaktovat pracovníky NKČR, případně správce Manuscriptoria (info@manuscriptorium.com) a oznámit úmysl zpřístupňovat svá data v Manuscriptoriu. Ve spolupráci s pracovníky AiP Beroun je možné předběžně posoudit technickou kvalitu obrazových dat, aby byly případně vyloučeny hrubší chyby. S NKČR lze předběžně diskutovat vhodnost zařazení uvažovaných dokumentů z hlediska věcného.

AiP Beroun - správce Manuscriptoria pro Vás vytvoří uživatelský účet v prostředí aplikace M-Can. Po inicializaci spolupráce bude partnerovi také k dispozici technická podpora ze strany správců Manuscriptoria.

- **Vytvoření popisných metadat:** první záložka „Popis“ v programu M-Tool se týká vlastností fyzického exempláře a předpokládá se, že by je měl vyplnit odborník – badatel. Podle informací zadaných v této záložce bude vygenerován bibliografický popis ve formátu MASTER.
- **Vytvoření metadat o struktuře:** data v záložce „Číslování“ slouží pro vygenerování struktury digitalizovaných stránek. Předpokládá se, že podklady pro vyplnění formuláře vytvoří odborník – badatel, případně pracovník digitalizačního pracoviště.
- **Vytvoření metadat o digitálních obrazech:** v záložce „Soubory a cesty“ lze definovat seznam dostupných kvalit, jejich jmen a adres, kde budou obrazy pro systém Manuscriptoria dosažitelné pomocí protokolu HTTP. Podklady pro vyplnění těchto informací dodává obvykle technický pracovník digitalizace, případně správce serveru partnera.
- **Export formulářů do XML:** výstupem formulářů programu M-Tool je XML soubor ve formátu MASTER+, který obsahuje všechna metadata potřebná pro import digitalizovaných dokumentů do Manuscriptoria

(podrobnější informace jsou dostupné v uživatelském manuálu programu).

- **Kontrola metadat v prostředí Manuscriptoria kandidátů:** prostředí M-Can umožňuje nezávazně nahrát metadata do Manuscriptoria a kontrolovat jejich správnost. Výsledný digitální dokument – jeho katalogový záznam i samotnou digitální kopii – lze zobrazit v podobě, v jaké bude prezentována v Manuscriptoriu a odhalit tak případné chyby v textech i v dostupnosti obrazů.

Při práci v prostředí M-Can nejsou nahrávané dokumenty dostupné ostatním uživatelům. Stávají se viditelnými až po nabídnutí k importu.

- **Nabídnutí metadat k importu do MANUSCRIPTORIUManuscriptoria:** přímo z prostředí M-Can lze metadata digitálních dokumentů nabídnout k zařazení do MANUSCRIPTORIUManuscriptoria. Po odeslání metadat následuje jejich kontrola odpovědným pracovníkem NKČR (kontrolní fázi lze u dlouhodobě spolupracujících partnerů ze zpracovatelského procesu vypustit). V případě schválení jsou dokumenty zařazeny do další aktualizace (MANUSCRIPTORIUManuscriptorium je aktualizováno v měsíčních dávkách).

- **Aktualizace existujících dat v MANUSCRIPTORIUManuscriptoriu:** M-Can je rozhraní pro předávání nových záznamů a neumožňuje provádět aktualizaci záznamů. Aktualizace záznamů se provádí ve spolupráci s administrátory MANUSCRIPTORIUManuscriptoria.

V rámci projektu ENRICH vzniknou nástroje pro snadnou manipulaci s prezentovanými dokumenty, které umožní podobné zásahy přímo konkrétním autorům.

Příklad praktického postupu – dávkový import dokumentů

V případě importu výsledků pokročilých digitalizačních projektů je Manuscriptorium připraveno převzít existující metadata a do MASTER+ je převádět pomocí individuálních konektorů. Je důležité, aby v dostupných metadatach byly dostupné výše popsané informace (o struktuře dokumentů, skutečná paginace/foliace a umístění existujících obrazů), nebo aby tyto informace byly z dostupných metadat vytvořitelné pomocí jednoznačně definovatelných pravidel.

- **Inicializace spolupráce:** nejprve kontaktujte pracovníky NKČR případně správce Manuscriptoria (info@manuscriptorium.com) a oznamte úmysl zpřístupňovat svá data v Manuscriptoriu. NKČR posoudí vhodnost zařazení souboru do Manuscriptoria z věcného hlediska. AiP Beroun pak posoudí metadata z technického pohledu.
- **Vytvoření konektoru:** pokud metadata vyhovují dle Manuscriptorium Compatible (tj. jsou dostatečně strukturovaná a obsahují potřebné in-

formace), jsou ve spolupráci s pracovníky partnera vytvořeny příslušné konektory, které slouží i k dalším opakovaným importům a aktualizacím.

Během přípravy konektoru je vytvořena zvláštní testovací instance Manuscriptoria mimo oficiální serverový systém, s jejíž pomocí lze bezpečně kontrolovat a doladit výsledky převodů.

Podle míry individuality daných metadat je možné pro partnera vytvořit konektor bezplatně (v rámci provozu Manuscriptoria). V případě náročnějších převodů (data nejsou Manuscriptorium kompatibilní) je nutno pro výrobu konektoru zajistit vhodným způsobem financování.

- **Předání metadat pro import:** předání balíčků takto připravených podkladů probíhá přímo mezi pracovníkem partnera a administrátorem Manuscriptoria, který zajistí provedení převodů a zařazení výsledků do další aktualizace.

Literatura

Dostupné podklady naleznete na www.manuscriptorium.cz, kde je ke stažení dokumentace, aplikace, licenční smlouvy a odkazy na volně dostupný SW třetích stran (např. doporučitelné XML editory atp.).