

Manuscriptoium – základy a kompatibilita

Verze 1.2

23.2.2006

Obsah

Historie Manuscriptoria.....	3
Základy Manuscriptoria	3
Úcta k práci našich předků	3
Oddělení trvalého a proměnného	3
Nalézání společných vlastností informací, zachování rozdílných	3
Otevřenost jiným projektům	3
Přístupnost a užitečnost pro partnery	4
Bezpečné a věrohodné zachování informací pro budoucnost	4
Otevřené možnosti uložení dat	4
Otevřené možnosti rozsahu volného zpřístupnění	4
Licenční politika	4
Manuscriptorium kompatibilita	4
Aktuální stav.....	5
Cíle Manuscriptoria	5
Dokumenty v Manuscriptoriu	5
Technická kompatibilita	6
Uložení dat a související metadata	6
Neveřejné uložení dat	6
Veřejné vystavení dat.....	6
M-tool.....	7
Kvalitativní kompatibilita a věrnost dat	7
Konektivita Manuscriptoria	7
Reference a literatura.....	7

Manuscriptorium je systém shromažďující a zpřístupňující na Internetu informace o historických knižních fondech, provázaný s digitální knihovnou digitalizovaných dokumentů. Provoz Manuscriptoria je financován Národní knihovnou ČR, provozovatelem této služby je firma AiP Beroun s.r.o.

Historie Manuscriptoria

Manuscriptorium je výsledkem aktivit, které se započaly v souvislosti s iniciativou UNESCO „Memory of the World“ v roce 1992. Aktivity pokračovaly vytyčením úkolů navazujícího českého programu Memoriae Mundi Series Bohemica (MMSB), řešením souvisejících aktuálních požadavků a praktickým provozem. Záhy byly rozpoznány požadavky na dlouhodobou udržitelnost projektu a stanoveny cíle směřující k vybudování elektronického badatelského prostředí pro oblast historických knižních fondů. V Manuscriptoriu jsou soustředěny informace poskytnuté spolupracujícími partnery, přispěvateli z knihoven, archivů a muzeí.

Praktické zkušenosti a spolupráce odborníků z technické a knihovnické oblasti vedli k formulaci základních thesís, které jsou pilíři Manuscriptoria.

Základy Manuscriptoria

Úcta k práci našich předků

Vycházíme z toho, že nové, stejně jako staré informace o historických dokumentech vznikly v nejlepší víře a úmyslu. Doba i podmínky, v nichž vznikly, se liší. Mění se názory i rozsah poznání, na které se při jejich vzniku navazovalo. Každá dostupná informace však může být přínosem, je-li nezkrácená a zdokumentovaná. Smyslem naší práce je zajištění nezkrácující dostupnosti všech existujících významných informací.

Oddělení trvalého a proměnného

Rozpoznali jsme, že je nutno tvrdě oddělit vznikající informace od jakýchkoli nástrojů souvisejících s jejich vznikem, správou a také s formou a způsobem zpřístupnění. Existence a použitelnost informací musí být zaručena při jakékoli úpravě nebo výměně nástrojů.

Nalézání společných vlastností informací, zachování rozdílných

Informace o historických fondech se nalézají v nejrůznějších systémech, obsahově je vždy část velmi podobná obsahově i strukturací. Detailní formu a obsah přirozeně významně ovlivňuje zaměření badatele, archiváře, či knihovníka a jimi používané technické nástroje. Zatímco volnost používání a tvorby informací nechceme jakkoli omezovat a ovlivňovat, chceme v principu odstranit nutnost podřizovat se v odborné práci technickým prostředkům.

Otevřenost jiným projektům

Naše aktivity pokládáme za užitečné, nikoli však za jedinečné. Jsou službou zajišťující zachování svěřených dat pro budoucnost a zároveň pro jejich aktuální zpřístupnění. Pokládáme proto za nezbytné budovat aktuálně požadované nástroje zajišťující přístup ke shromážděným datům, a to jak přímo koncovými uživateli, tak prostřednictvím služeb poskytovaných jiným systémům.

Přístupnost a užitečnost pro partnery

Poskytnutí dat do Manuscriptoria chceme natolik zjednodušit, aby pro velké instituce představovala spolupráce s Manuscriptorem přinejmenším zajímavou a užitečnou alternativu a doplněk vlastních aktivit. Pro menší instituce i jednotlivce je pak Manuscriptorium cestou, jak pro svá data zajistit bezpečné zachování a moderní zpřístupnění. Současným výsledkem snahy o dosažení těchto cílů je prezentace databáze Manuscriptorium, která je dostupná buď přes stránku představující celý projekt www.memoria.cz nebo přímo přes www.manuscriptorium.com. Připouštíme i možnost *individuálních forem* prezentace dat partnerů.

Bezpečné a věrohodné zachování informací pro budoucnost

Jsme si vědomi vysoké odpovědnosti za bezpečné uchování shromážděných dat a to nejen ve smyslu jejich fyzické existence, ale též ve smyslu zachování jejich obecné použitelnosti, věrohodnosti a také zachování informací o podmínkách jejich vzniku.

Jsme si vědomi vysoké zranitelnosti elektronických dat a jejich závislosti na morální a technické úrovni a kontinuitě naší společnosti.

V Manuscriptoriu jsou uložena a zpřístupněna data a informace s garantovaným původem a kvalitou.

Otevřené možnosti uložení dat

Manuscriptorium je vytvořeno primárně pro data vzniklá v ČR v projektu **VISK6**, tedy pro data, za jejichž uchování a zpřístupnění nese odpovědnost především NKČR. Manuscriptorium ale respektuje také jiná uložení dat a umožňuje zpřístupnit shodnými nástroji data ležící v nezávislých, **vzdálených úložištích**.

Otevřené možnosti rozsahu volného zpřístupnění

Manuscriptorium zpřístupňuje široké veřejnosti zcela volně metadata (katalogové záznamy).

Manuscriptorium zpřístupňuje zásadně volně veškerá data (obrazová, zvuková, plné texty,...) svým **partnerům**, ostatním pak prostřednictvím **licencí**.

Rozsah volného poskytování dat široké veřejnosti v plném rozsahu respektuje rozhodnutí partnera.

Licenční politika

Licenční politiku určuje Národní knihovna České republiky. Jejím cílem je kontrolovat zacházení se svými a svěřenými daty tak, aby byl zajištěn snadný přístup odborné veřejnosti k veškerým shromážděným informacím a aby byly zároveň respektovány individuální zájmy a pravidla nakládání s daty rozdílných partnerů.

Manuscriptorium kompatibilita

Partner, jehož informace budou shledány jako věcně a obsahově vhodné k zařazení do Manuscriptoria a jehož data splní technické a formální požadavky importu obdrží certifikát **Manuscriptorium Compatible**. Tím zároveň získá právo na bezplatnou licenci na neomezený přístup do Manuscriptoria, tedy přístup k úplným datům poskytnutým ostatními partnery.

Partneři, jejichž data navíc splní stanovené kvalitativní požadavky, obdrží certifikát **Manuscriptorium Image Quality**. Touto značkou jsou označeny jeho dokumenty v Manuscriptoriu, a jsou pro další partnery informací o garantované kvalitě existujících primárních dat.

Aktuální stav

Existuje digitální knihovna rukopisů, starých tisků a dalších vzácných dokumentů. Jako každá knihovna se skládá z katalogu, kterým je v tomto případě **OKHF** (Otevřený Katalog Historických Fondů) a vlastních digitálních dokumentů, které jsou uloženy v datovém úložišti. **OKHF** soustřeďuje informace (metadata) o fyzických dokumentech (rukopisech atd.) ve formě tzv. evidenčních záznamů ve formátu XML. Úložiště pak obsahuje digitální kopie podмноžiny z těchto katalogizovaných dokumentů ve formě tzv. komplexních digitálních dokumentů (**KDD**).

Principiálně systém Manuscriptorium centrálně soustřeďuje metadata v **OKHF** a zajišťuje přístup k digitalizovaným dokumentům na datových úložištích provozovatele i vzdálených úložištích partnerů.

Cíle Manuscriptoria

Jsou položeny základy, dovolující shromažďovat a zpřístupňovat existující data. Dalším cílem je budování virtuálního badatelského prostředí pro oblast historických fondů. Probíhá rozšiřování digitálních prezentací těchto fondů (obrazových, textových) a jejich zpřístupňování, vzniká virtuální digitální knihovna obrazových kopií a plných textů. Zabezpečuje se propojení k dalším datům majícím vztah k těmto fondům (studie, další popisná metadata, další elektronické dokumenty/objekty) a také poskytování informací jiným systémům standardizovanými postupy (komunikačními rozhraními) (**Z39.50, OAI-MPH**)

Dokumenty v Manuscriptoriu

V následujícím stanovujeme vlastnosti, které musí splňovat přijaté digitální dokumenty. Rozhodující je:

1. význam dokumentu a účelnost jeho zařazení do Manuscriptoria,
2. technická použitelnost informací o dokumentu (metadata) a informací prezentujících dokument, tj. obrazy, zvuky, plné texty (data).
3. kvalita a věrohodnost dat

Zařazování dokumentů do Manuscriptoria řídí Národní knihovna ČR tak, že orientuje Manuscriptorium ke zpřístupňování písemného kulturního dědictví do roku 1800 s primárním zaměřením na dokumenty knižní povahy, sekundárně je však otevřeno pro dokumenty povahy archivní (zejména knihy, listiny a mapy).

Denotátem všech digitálních dokumentů v Manuscriptoriu je zásadně originální fyzický dokument, ať dosud existující či již zaniklý.

Metodika výběru je podrobněji popsána v dokumentu [\[11\]](#).

Technická kompatibilita

Vycházíme z toho, že elektronické informace o historických fondech jsou stále častěji přímo použitelné v Manuscriptoriu nebo jsou přetvořitelé do očekávané formy XML rutinně, případně jen s malým úsilím. Kompatibilita lze snadno dosáhnout zvláště u vznikajících projektů a u projektů jejichž dosavadní zaměření směřovalo především k digitalizaci a tvorbě dat.

Základní formou uchování informací v Manuscriptoriu jsou XML soubory se strukturou odpovídající standardu **METS** (Library of Congress). Pro uchování popisných metadat v rámci **METS** je použit standard **MASTER**, speciálně vyvinutý mezinárodní pracovní skupinou ve spolupráci s Oxford University Computing Services pro oblast popisu rukopisů, starých tisků a dalších historických dokumentů.

Za přímo kompatibilní se považují metadata vytvářená libovolnými nástroji, jejichž výstupem jsou záznamy v XML, využívajícím **DTD MASTER** [5]. Mírné varianty jsou přípustné, provozovatel si však vyhrazuje právo na posouzení.

V případě, že konverze existujících informací není snadná nebo přispěvatel nemá důvod jejich formu měnit, a přesto má zájem je v Manuscriptoriu zveřejnit, je možno vytvořit nástroje pro rutinní konverzi za účelem dosažení kompatibility. Přispěvatel si zajistí ke svému systému tzv. **konektor**, a dále spravuje data nezávislým způsobem. Informace poskytuje do Manuscriptoria ve formě kompatibilní přes tento **konektor**.

Konektory pro obvyklé typy záznamů jsou budovány na straně Manuscriptoria nebo na straně přispěvatelů. Využívání existujících konektorů nebo budování nových je předmětem individuálních jednání mezi budoucím partnerem, Národní knihovnou ČR a správcem Manuscriptoria.

NKČR zásadně podporuje a financuje přímý import dat do Manuscriptoria. Informace, pro které existuje **konektor** nebo pro která bude vytvořen, jsou považována za kompatibilní.

Konvence použité v Manuscriptoriu jsou popsány v dokumentu [12].

Akceptování těchto konvencí je zárukou úplné technické kompatibility.

Uložení dat a související metadata

Připojení obrazů k metadatům vyžaduje zavedení a dodržování konvencí popisujících cestu k uložení obrazů.

Manuscriptorium respektuje široké spektrum způsobů uložení dat.

Neveřejné uložení dat

Jsou-li data uložena uvnitř systému znemožňujícího přímou dostupnost dat, je nutno připojení dat řešit individuálně konektorem.

Veřejné vystavení dat

Mnoho dat je nebo může být zpřístupněno na Internetu volně – každý obraz je dosažitelný unikátní adresou URL prostřednictvím protokolu HTTP.

Takto dostupná data jsou do Manuscriptoria připojitelná obvykle jednoduše.

M-tool

Pro usnadnění tvorby metadat byl poskytnut ke stažení a volnému použití **M-tool** – nástroj dovolující vytvořit základní metada a spojit je s obrazovými daty uloženými na volně dostupných adresách HTTP. Informace o nástroji **M-tool** jsou k dispozici na adrese <http://www.memoria.cz/mtool/>.

Kvalitativní kompatibilita a věrnost dat

Data zpřístupněná na Internetu v Manuscriptoriu jsou obvykle přizpůsobena technickým požadavkům Internetu a bývají odvozena od primárních dat. Manuscriptorium vychází z požadavku, aby bylo garantováno bezpečné a věrohodné zachování informací pro budoucnost. Touto garancí je spolehlivě zajištěná existence kvalitních primárních dat, nikoli samotné zpřístupnění na Internetu. Pro data vzniklá v projektu české republiky VISK6 je tomuto požadavku věnována mimořádná pozornost. Data vznikající s podobnou péčí a zajištěním jsou v Manuscriptoriu označena značkou kvality. Požadavky nezbytné pro toto označení dat v Manuscriptoriu jsou popsány v dokumentu [13].

V praxi však není vždy možné zaručit optimální podmínky při pořizování dat. Proto do Manuscriptoria lze zařazovat i data vzniklá za horších podmínek.

O zařazení dat nesplňujících kvalitativní požadavky si vyhrazuje právo rozhodnout Národní knihovna ČR.

Konektivita Manuscriptoria

Pro využívání dat mimo Manuscriptorium jsou dostupné nástroje pro zpřístupnění a sdílení metad ve standardních formátech (UNIMARC, MARC21, MASTER, Dublin Core...) pomocí standardních komunikačních rozhraní, **Z39.50**, **OAI-PMH**.

Rozsah těchto služeb a zařazování dalších určuje Národní knihovna ČR.

Reference a literatura

V dokumentu se odkazujeme na následující literaturu:

- [1] Manuscriptorium v.2.0 - analýza systému, prosinec 2004
- [2] Manuscriptorium v. 2.0 – komplexní digitální dokument, říjen 2005
- [3] Memoriae mundi series Bohemica, dostupné z URL: <http://digit.nkp.cz>
- [4] Manuscript Access through Standards for Electronic Records (MASTER), dostupné z URL: <http://xml.coverpages.org/master.html>
- [5] Reference Manual for the MASTER Document Type Definition, dostupné z URL: <http://www.tei-c.org.uk/Master/Reference/oldindex.html>
- [6] MEdit, dostupné z URL: http://www.memoria.cz/download/medit_cz.asp
- [7] TorXmlValid, dostupné z URL: http://www.memoria.cz/site_cz/download.asp
- [8] jEdit, dostupné z URL: <http://www.jedit.org/>
- [9] Emacs, dostupné z URL: <http://www.gnu.org/software/emacs/emacs.html>
- [10] NoteTabLight, dostupné z URL: <http://www.webmasterfree.com/notetabligh.html>
- [11] Manuscriptorium – Výběr a popis dokumentů:
http://www.memoria.cz/docs/manuscriptorium_document_description_CZE.pdf

[12] Manuscriptorium – Technická kompatibilita:

http://www.memoria.cz/docs/manuscriptorium_compatibility_technical_CZE.pdf

[13] Manuscriptorium – Kvalita obrazu:

http://www.memoria.cz/docs/manuscriptorium_image_quality_CZE.pdf