

Projekt MEMORIA, rukopisy a staré tisky na internetu

MEMORIA je název sdružující iniciativy, které se zrodily v souvislosti s provozem a řešením programu Memoriae Mundi Series Bohemica (MMSB). Projekt MEMORIA směřuje k vybudování virtuálního badatelského prostředí pro oblast historických knižních fondů. Projekt podporuje vznik nových a využívání existujících informací, zajišťuje jejich dlouhodobou životnost a trvalou použitelnost. Zpřístupňuje výsledky dosavadní digitalizace dokumentů, výsledky detailních popisů historických dokumentů a přináší přístup k bibliografickým informacím z různých zdrojů v databázi Manuscriptorium. Náhledy do digitalizovaných rukopisů jsou volně přístupné, kvalitní zobrazení je licencované, přičemž pro aktivní účastníky projektu je licence zdarma.

CÍLE PROJEKTU

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 podstatně liší, mění se názory, poznání, z něhož se vychází, mění se tedy počáteční podmínky při jejich vzniku. Z principu nevěříme na definitivní, správné, konečné podoby dat. Nehodnotíme jejich rozsah nebo kvalitu, přijímáme je. Jsme přesvědčeni, že každá dostupná informace je pro badatele přínosem a vodítkem pro další samostatnou práci.

Praktickým cílem projektu MEMORIA je vybudování virtuálního badatelského prostředí pro oblast historických fondů:

- formou sdíleného **otevřeného katalogu těchto fondů**
- pořizováním digitálních forem těchto fondů (obrazových, textových) a jejich zpřístupňováním, a tedy vy-

tvořením **digitální knihovny obrazových kopií a plných textů**

- zabezpečením propojení k dalším datům majícím vztah k těmto fondům (studie, další popisná metadata, další elektronické dokumenty/objekty).

Současným výsledkem snahy o dosažení těchto cílů je databáze Manuscriptorium, která je dostupná buď přes stránku představující celý projekt MEMORIA www.memoria.cz, nebo přímo přes www.manuscriptorium.com. Tento první krok je statický, a nelze ho proto zatím nazvat badatelským prostředím. Tím se stane až v další etapě, kdy bude umožněno tyto zveřejněné informace řízeně měnit, doplňovat a provazovat.

RUTINNÍ AKTIVITY

Zahrnují rutinní popisy a digitalizaci vzácných originálů dokumentů, výrobu digitálních dokumentů, archivaci a ochranu digitálních dat. Tyto převážně výrobní činnosti probíhají především v rámci projektů **VISK6**, částečně **VISK4**, případně dalších projektů, které koordinátor a provozovatel projektu MEMORIA iniciovali, nebo se na nich podílejí.

Podíl firmy AiP Beroun na zajišťování rutinních aktivit má zajímavý vývoj. Na samém počátku v letech 1992–93 byla firma AiP Beroun řešitelem iniciálního úkolu UNESCO – vydání prvního CD „Paměť světa“, později se stala dodavatelem digitalizační technologie a zajišťovatelem servisu. Postupně byla AiP Beroun vyzvána Národní knihovnou k zajištění dalších činností včetně samotné digitalizace a manipulace se vzácnými dokumenty. Jako výhoda se ukázala lepší možnost soukromé firmy vybírat a motivovat pracovníky. To dalo během deseti let vzniknout spolehlivému specializovanému kolektivu.

Popisy

Popisy digitalizovaných dokumentů musí být dosti podrobné, protože jsou určeny nejen ke katalogizaci, ale také k propojení digitálních obrazů do formy elektronického dokumentu.

I zde lze sledovat zajímavou genezi prostředků, které se pro popisy dokumentů využívají. Prvním pokusem bylo vytvoření prostředí, v němž se popisovaly přímo digitální obrazy. Toto pro techniky jasné a logické řešení jsme prověřovali na popisu Sedleckého antifonáře. Technologii s nepříliš lichotivým závěrem prověřila prof. Zdeňka Hledíková. Ve spolupráci s ní jsme zavedli princip pevně strukturovaného popisu, který je vytvářen před digitalizací a vede až k formulaci DOBM formátu využívajícího SGML. Tento postup byl přijat jako doporučení projektu UNESCO „Memory of the world“. Nástroje jsou volně k dispozici a dosud se používají pro svou jednoduchost. Při popisu dokumentu se současně kontroluje, zda je rukopis bez rizik způsobilý pro digitalizaci.

Zavedení popisů ve formě XML v rámci projektu MASTER vedlo k prvotnímu popisu dokumentů ve volné a badatelským potřebám přizpůsobivější struktuře MASTER a k následnému přepisu dat do pevné formy DOBM. Nyní se dokončují prostředky využívající jen XML. I tyto prostředky budou dostupné.

Původně dosti rozsáhlý kolektiv spolupracujících odborníků se omezil na pracovníky dlouhodobě garantující včasnost a kvalitu popisů. Podstatné množství si nyní popisují majitelé digitalizovaných dokumentů sami. K úspěšné

spolupráci je nutné jen velmi krátké a jednoduché zaškolení do formálních pravidel, jež je nutno dodržovat.

Digitalizace

Technologické vybavení má ve své historii několik „nej“.

První digitální kamera KODAK 460 RGB byla první kamerou tohoto typu v ČR a v zemích východní Evropy.

Také první digitální kamera BetterLigth 6000 byla první v ČR, a to v době, kdy tato kamera byla ještě takřka neznámá. Byla vybrána na základě mimořádné kvality produkovaných obrazů v rozlišení až 48 milionů pixelů.

Současnou špičkou je Special BookScanner 145 CRUSE. Jde o modifikaci scannerů určených především pro snímání obrazů a map. Vznikl v přímé spolupráci firmy CRUSE a AiP Beroun, která se na vývoji této modifikace podílela. Je ve všech ohledech optimalizován pro náročnou digitalizaci vzácných historických dokumentů. Ke vzniku tohoto zařízení přispělo paradoxně nerovnoměrné a nejisté financování projektů, které způsobilo, že bylo nezbytné rychle reagovat na neočekávané zvýšení požadavků na digitalizaci. Zařízení vzniklo během tří měsíců a bylo plně financováno firmou AiP Beroun, která leasing dosud splácí.

Poslední zařízení založené na kameře BetterLigth je již plně vyvinuto firmou AiP Beroun a nahradilo kameru KODAK. Výhodami jsou zúročení vlastních praktických zkušeností a výrazně nižší náklady oproti dřívějším nákupům univerzálních zařízení. Zařízení jsou optimalizována z hlediska bezpečnosti dokumentů, snadné manipulace s nimi a jejich ochrany před UV a IR zářením. Tyto primární požadavky neovlivňují vysokou produktivitu a kvalitu na úrovni špičkové studiové práce.

Výroba digitálních dokumentů

Digitální obrazy jsou spojovány s dříve připravenými popisnými daty do formy dokumentu, který jednak obsahuje všechny popisné a technické informace s využitím standardů MASTER (dříve DOBM) a navíc vygenerované HTML soubory svazující obrazy do formy umožňující prohlížení dokumentu běžně dostupnými internetovými prohlížeči.

Archivace a ochrana digitálních dat

Archivace dat prošla zhruba třemi obdobími:

1. V počátcích, koncem 90. let, nebyla k dispozici jiná levná možnost archivace větších objemů dat než CD-R disky. Pracovníci AiP Beroun zajistili pro bezpečnou archivaci na CD-R technologii měření kvality záznamu, zajistili vyhodnocení stárnutí vytvořených médií (viz dále – programy Věda a výzkum). Strategie archivování byla založena na existenci sice drahých, ale kvalitních a stabilních médií (KODAK Ultima Gold).
2. Boom využívání CD-R způsobil řádové zlevnění CD-R médií, a tím odstranil z trhu drahá média vhodná pro archivaci. Kvalita médií velmi poklesla, protože konkurenční boj si vynutil snižování nákladů na výrobu i za cenu nižší kvality. Začalo se více využívat tolerovatelné chybovosti při zápisu. Také souboj o co nejvyšší záznamové rychlosti není v souladu s požadavky na spolehlivost záznamu. Přitom měřicí technika pro kontrolu kvality zůstávala stále velmi drahá a její cena dále stoupala. To nedávalo dlouhodobou perspektivu pro využívání CD-R jako archivačního média.

3. Rok 2003 přinesl radikální novinku. Pokrok ve vývoji vypalovacích jednotek přinesl další generaci IC obvodů, umožňující monitorování činnosti signálového procesoru a přímé získávání informací o vadných blocích (BLER) na první a druhé úrovni rekonstrukce zaznamenaných dat. Tyto informace jsou základními veličinami pro hodnocení kvality vytvořeného disku a jsou srovnatelné s informacemi dosud zjistitelnými jen velmi drahými měřicími přístroji. To dává reálnou možnost zajistit a ověřit kvalitu produkovaných disků s vypálenými daty a monitorovat jejich stárnutí. V případě signifikantního poklesu jejich kvality lze informace přepsat včas na nová média.

Kvalita médií se dalším vývojem opět stabilizovala, i když jsou využívány levnější technologie. Navíc AiP Beroun spolupracuje s výrobcem médií. To dovoluje dále využívat CD-R jako archivační médium.

Vznikající CD-R disky se uchovávají ve dvou sadách na dvou oddělených místech. Samozřejmě se připravuje paralelní uložení všech dat na vznikajících hromadných datových úložištích v Národní knihovně ČR.

VÝZKUMNĚ VÝVOJOVÉ AKTIVITY

V rámci těchto aktivit jsou řešeny programové projekty Výzkumu a vývoje:

- národní, viz <http://digit.nkp.cz/projekty/ProjektyVaV.htm>
- mezinárodní, viz http://digit.nkp.cz/Projects/index_cz.htm.

V průběhu řešení dosavadních úkolů vznikly mnohé prostředky pro speciální oblast historických fondů, vhodné ke zpracování a prezentaci souvisejících informací. Prakticky jsme ověřili postupy umožňující vytvoření katalogu otevřeného pro jakákoli existující data, nezávisle na jejich původní formě.

V krátké době budou na stránkách projektu MEMORIA volně zpřístupněny prostředky pro export/import dat do UNIMARC/MARC 21 a jejich konverzi do formy MASTER.

Z 39.50

Připravuje se poskytování dat prostřednictvím protokolu Z39.50. K systému bude realizován Z39.50 server podporující Bath profil Functional Area A level 1. Data budou předávána ve formátu UNIMARC.

OAI

Pro výměnu dat mezi kooperujícími systémy bude použit protokol Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Data budou poskytována a přebírána ve formátu XML ve struktuře podle DTD OpenM (NK ČR).

DRUHY DAT

V databázi Manuscriptorium se setkávají data různého typu.

Data zastupující originály

Digitální obrazy

Tato data si kladou za cíl co nejvěrněji nahradit kontakt s originálními dokumenty, a tím tyto originály ochrá-

nit. Při pořizování obrazových dat je úzkostlivě dbáno na úplnost a věrnost vznikajících informací včetně zachování informací o barevné kalibraci. V současnosti je zdigitalizováno 1100 dokumentů, z nichž **879** je nyní **zpřístupněno**. To představuje cca **450 000** obrazů.

Plné texty dokumentů

Plné texty jsou další cestou, jak zpřístupnit informace obsažené v historických dokumentech. Předpokládáme, že plné texty budou využívat standardu TEI a budou zpřístupňovat zpravidla:

- pragmatické edice historických dokumentů reprezentovaných v otevřeném katalogu historických fondů, případně i v podobě digitální kopie
- texty sekundárních dokumentů, vztahující se k originálům.

Tento postup se zatím zkouší.

Popisná data

Prostřednictvím dat, která popisují dokumenty, je umožněn přístup k originálům, respektive k datům zastupujícím originály. Jsou to především:

Popisná data vzniklá v souvislosti s digitalizací

Popisná data dosud vznikala ve formátu DOBM. Nyní proběhla konverze z DOBM na MASTER a začíná se s popisem dokumentů výhradně v XML s využitím výsledků projektu MASTER. Dříve vzniklá data splynou s následujícími daty.

Popisy dokumentů v MASTER

Dokumenty vzniklé na základě projektu MASTER představují data, která jsou způsobem vzniku a použitým formátem blízká předchozím. Tato data představují nyní **cca 5000 záznamů**.

Další data

Existuje množství dalších dat, která mají charakter od podrobných popisů až po stručné inventární seznamy. Nyní je dostupných **cca 23 000 záznamů** z těchto zdrojů. Tato data jsou vesměs snadno převoditelná do formy MASTER a mohou tedy využít stejných prostředků pro zpřístupnění jako výše uvedené typy.

Všechna výše uvedená data mají společného jmenovatele – formát MASTER, v němž vznikají nebo do kterého jsou do značné míry převoditelná.

ZPŮSOBY ZPŘÍSTUPŇOVÁNÍ DAT

Základem jsou doposud CD-R disky, které jsou použitelné bez instalace speciálních programů jako dostatečná elektronická náhrada přístupu k obsahu originálních dokumentů pro velkou většinu badatelů. Na CD-R discích jsou nyní uchovávány také archivní kopie dat. Stále větší význam si získává **zpřístupnění na internetu**, kde jsou digitální dokumenty zpřístupňovány v rámci elektronického on-line katalogu – databáze Manuscriptorium. Katalog je opatřen výkonnými vyhledávacími nástroji, které jsou uzpůsobené specifikám oboru.

KOLEKCE DIGITALIZOVANÝCH DOKUMENTŮ

Všechna data v projektu jsou v současné době nabízena a zpřístupněna v rámci jediné souhrnné kolekce. Tato kolekce je poněkud nehomogenní, protože zahrnuje všechny dokumenty, které jsou v projektu MEMORIA

zpřístupněny. Výběr digitalizovaných dokumentů je poznamenán množstvím nezávislých přispěvatelů, růzností jejich odborných specializací i motivací k digitalizaci. Díky projektu VISK6 a podpoře Ministerstva kultury ČR je možnost digitalizace a zpřístupnění dokumentů mnoha institucemi skutečně využívána.

Častou motivací je ochranná digitalizace – nahrazení přístupu k příliš využívanému originálu přístupem k jeho digitálnímu obrazu. U institucí, které již zahájily spolupráci s projektem MEMORIA, je dále zřejmá snaha o postupné úplné zpřístupnění lokálních kolekcí významných dokumentů.

Příznivě se zde projevuje, že návrhy na zařazení do digitalizace schvaluje jmenovaná komise odborníků tvořící poradní orgán Ministerstva kultury ČR. Celostátně řízený výběr dokumentů přispívá v delším čase ke vzniku tematicky souvisejících kolekcí. Tato odborná koordinace spolu s počtem kompletně digitalizovaných dokumentů dává záruku postupného vzniku kolekcí, které přesahují hranice fyzického uložení dokumentů.

Přestože jsou hranice tematických kolekcí již viditelné, rozhodli jsme se pro začátek ponechat jediný kompletní digitalizovaný fond. Rozsáhlé rešeršní nástroje vyhledávacího systému dovolují snadno dospět ke specifické kolekci a konkrétním dokumentům.

Již nyní jsou zřetelné snahy některých badatelů ovlivnit další postup digitalizace konkrétními požadavky na **doplňování vznikajících kolekcí**. Vyhovět těmto potřebám je jedním z úkolů projektu MEMORIA. Je pravděpodobné, že využíváním databáze Manuscriptorium se tento trend prohloubí.

Nejvíce digitalizovaných rukopisů v databázi Manuscriptorium pochází z Národní knihovny ČR. Výběr zde má již od počátku svou strategii a soustřeďuje se především na specifické fondové celky a na středověké monistické a scholastické texty. Postup digitalizace a vznik tematických kolekcí určuje za Národní knihovnu především PhDr. Zdeněk Uhlíř.

Podrobnější informace lze najít na www.memoria.cz, v sekci Kolekce. Je zde zpřístupněn také fundovaný pohled na dokumenty pocházející z institucí mimo Národní knihovnu od paní profesorky PhDr. Zdeňky Hledíkové, CSc. a pohled na zastoupení hudebních rukopisů z hlediska hudebního vědce PhDr. Stanislava Tesaře.

Významná kolekce orientovaná na metternichovské dokumenty vzniká z iniciativy PhDr. Miloše Říhy, správce kynžvartského zámku a tamní zámecké knihovny. Charakteristika této práce je dostupná na stránkách Kynžvartské digitální knihovny <http://www.kynzvalt.cz/libra.html>

ŘÍZENÍ PŘÍSTUPU K DIGITÁLNÍM OBRAZŮM

Projekt umožňuje snadné a bezplatné zpřístupnění veškerých dat přinášejících informace o existenci historických dokumentů. Na druhé straně vznikají data, která není obvyklé poskytovat v plné míře zcela volně a bezplatně. Radou projektu proto bylo rozhodnuto a schváleno, že část informací bude zpřístupněna na základě přidělování nebo prodeje licencí.

Volný přístup

Projekt MEMORIA předpokládá volné zpřístupnění veškerých dostupných dat nesoucích informaci o existenci

dokumentu a jeho základní popis. U dokumentů, u nichž existuje digitální kopie, je volně dostupný kompletní náhled do celého dokumentu v kvalitě nezbytné pro orientaci v dokumentu.

Licencované zpřístupnění

Podrobnější informace, zejména obrazy vyšší kvality a tzv. plné texty dokumentů, budou zpřístupněny na základě udělení licence opravňující k jejich využívání. Licence reguluje nakládání s těmito informacemi v zájmu projektu MEMORIA.

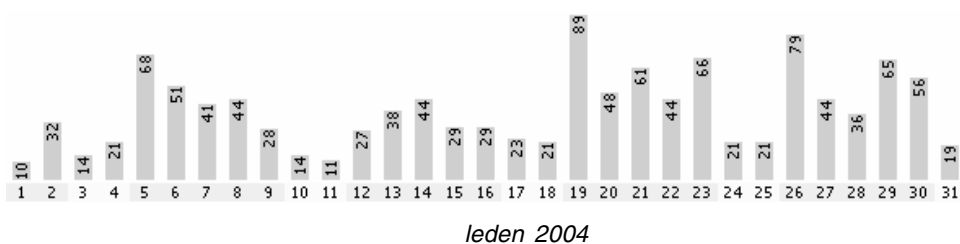
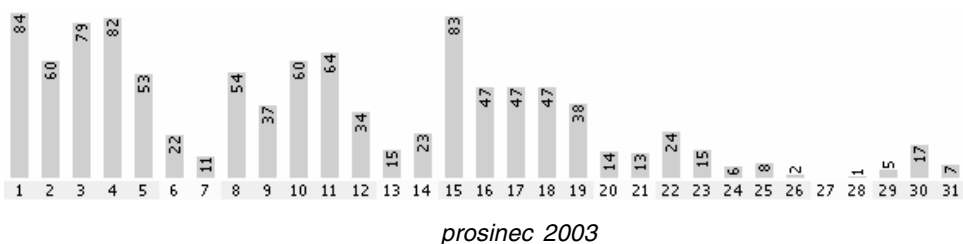
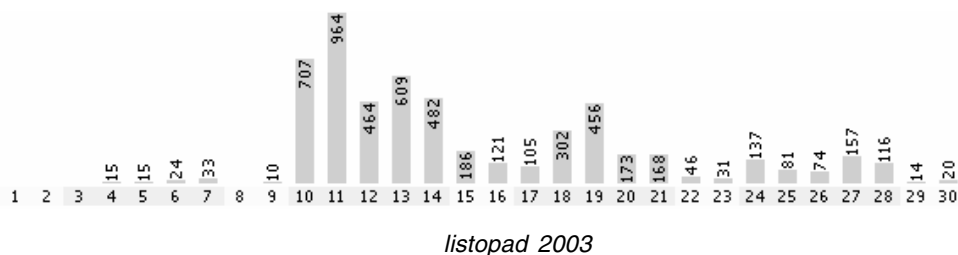
Volně jsou plné digitální obrazy zpřístupněny všem aktivně spolupracujícím subjektům.

Příjmy z prodeje licencí budou používány na podporu dalšího rozvoje projektu a na podporu digitalizace, zejména na aktivní doplňování kolekcí.

Zájem o přístup na internetu

Projekt MEMORIA a prezentace jeho výsledků na internetu v databázi Manuscriptorium prošla v roce 2003 rychlým vývojem. První reakce jsou velmi pozitivní, množství návštěv WWW stránek překročilo naše očekávání, přestože databáze není nijak intenzivně propagována. Podstatná část těchto návštěv je ze zahraničí, trvale tvoří cca 50 % přístupů.

V následujících grafech jsou přístupy k databázi v měsících listopadu, prosinci 2003 a lednu 2004.



Věříme, že projekt bude pro badatele zajímavý a užitečný i v době, kdy ztratí punc novosti a stane se nástrojem běžné práce.

Stanislav Psohlavec, AiP Beroun s.r.o.