

# ZPŘÍSTUPNĚNÍ OHROŽENÝCH FONDŮ POMOCÍ HYBRIDNÍ TECHNOLOGIE REFORMÁTOVÁNÍ

*Jiří Polišenský, Národní knihovna ČR*

Digitalizace knihovních dokumentů má v NK ČR krátkou, ale bohatou historii. První aktivity v této oblasti v rámci projektu Paměť světa, realizované ve spolupráci s firmou Albertina icome Praha, vyvrcholily vytvořením společného pracoviště přímé digitalizace vzácných rukopisů a schválením navrženého formátu DOBM pro tvorbu metadat, založeném na standardu SGML.

V další fázi byly realizovány dva rozsáhlé projekty, které umožnily v široké míře uplatnit hybridní technologii reformátování knihovních dokumentů. V rámci prvního projektu ochranného mikrofilmování, nazvaného „Kramerius I“, byla vedle pracoviště mikrografie v NK ČR vzkříšena k novému životu a doplněna technickým vybavením další tři v Moravské zemské knihovně v Brně a v SVK v Plzni a v Olomouci. Během tří let bylo mikrofilmováno 18 nejvíce ohrožených novinových titulů. V rámci realizace projektu bylo třeba zavést do praxe požadavky mezinárodních norem ISO v oblasti formálního vybavení mikrofilmů a kontroly kvality. V r. 1999 se NK ČR stala členem evropského sdružení knihoven EROMM (European Register of Microform Masters), které bude databázi záznamů mikrofilmovaných dokumentů.

Druhý projekt „Digitalizace mikromédií“, realizovaný také ve spolupráci s firmou Albertina icome Praha, řešil problematiku skenování mikrofilmů, vývoje programových nástrojů pro tvorbu metadat ve standardu SGML a zpřístupňování digitálních dokumentů prostřednictvím CD-R médií a sítí. V oblasti tvorby metadat projekt navázal na projekty „Paměť světa“ řešené již v minulosti, a využil jejich výsledků. Oba projekty byly ukončeny v r. 1999.

V současné době existují ve výše uvedených knihovnách čtyři mikrografická pracoviště zhotovující mikrofilmy podle požadavků mezinárodních norem a v kvalitě, která je zárukou jejich dalšího elektronického zpracování. Kromě nich existuje v NK ČR výkonné pracoviště digitalizace mikrofilmů, které bude schopno skenovat poměrně značnou část produkce mikrofilmů těchto čtyř pracovišť a zpřístupňovat digitální dokumenty prostřednictvím internetu.

## **Popis stávající technologie**

### *Hybridní technologie reformátování*

Hybridní technologie je založena na využití výhod mikrofilmů a digitálního záznamu. Ohrožený dokument se nejprve snímkuje na mikrografický film, který

garantuje dlouhodobé dochování v termínech staletí. Skenováním mikrofilmu ve druhé fázi se získává digitální dokument, který slouží pro zpřístupňování. V současné době se hybridní technologie uznává vedle ochranného mikrofilmu jako metoda reformátování vhodná pro dlouhodobé nebo trvalé dochování ohrožených dokumentů.

#### *Kriteria výběru dokumentů pro reformátování*

Za výběr dokumentů pro mikrofilmování a příp. následnou digitalizaci odpovídá Pracovní skupina pro ochranné reformátování CASLIN, kterou tvoří zástupci zatím třinácti knihoven. Hlavními kritérii pro výběr jsou fyzický stav (stupeň poškození), frekvence výpůjček a unikátnost dokumentu. Především se jedná o novinové tituly tištěné na dřevitém kyselém papíru, podléhajícím degradaci, které jsou ve všech knihovnách nejvíce ohroženy. Při výběru se také přihlíží ke skutečnosti, pokud některá knihovna financuje reformátování z vlastních zdrojů.

#### *Příprava dokumentů a tvorba metadat*

V první fázi přípravy dokumentů pro mikrofilmování a digitalizaci se vybírá a kompletuje určený titul za pomoci všech dostupných exemplářů. Nejde jen o dosažení maximální úplnosti, což je v případě novin velmi obtížné, vzhledem k častým a nepravidelným přílohám a specifickým vydáním (např. večerním, mimořádným a pod.), ale současně je třeba mikrofilmovaný dokument složit z nejlépe dochovaných exemplářů. Nejvíce postižené bývají první strany a první čísla svazků, uvolněné listy a pod. Na prvních obrazových polích mikrofilmu se snímkuje technické předlohy (identifikační, bibliografická a obsahová) usnadňující orientaci uživatele v dokumentu. Informace, které uvádějí, se získávají velmi podrobným průzkumem, při kterém se zjišťují všechny změny názvů, podnázvů, vydavatelů a pod., všechny chyby a nepravidelnosti jako chybná paginace, číslování výtisků, chybné údaje o ročnících a pod., a chybějící části dokumentu (strany, výtisky), které se nepodařilo dohledat.

V této fázi přípravy se vytváří podklad pro metadata vyplněním papírového formuláře, do kterého se zaznamenávají všechny potřebné údaje pro každou stranu dokumentu. Ve druhé fázi přípravy se údaje z formuláře přepisují do pracovního listu počítače a konvertují do formátu DOBM.

#### *Mikrofilmování*

Je-li dokument zkompletován a připraven pro mikrofilmování, provádí se snímkování na mikrografických kamerách v tomto pořadí: nejprve předepsané technické předlohy, poté zkušební obrazce sloužící pro vyhodnocování kvality, a potom vlastní dokument. Pro snímkování se používají speciální mikrografické jemnozrné 35mm filmy v délce 30,5m, které mohou obsahovat až 600 polí se dvěma stranami dokumentu (celkem až 1200 stran). Kvalita výsledného produktu závisí zejména na úrovni expozice a vyvolání. Vyvolávací automaty musí svou konstrukcí umožnit nejen správné nastavení provozních parametrů (teploty provozních lázní, rychlosti posuvu filmu), aby bylo možné dosáhnout optimálních optických vlastností, ale musí také garantovat dokonalé chemické zpraco-

vání citlivé vrstvy (zejména ustálení a vyprání), které je důležité pro dlouhodobou chemickou stabilitu filmu. Snímkováním předlohy na mikrografické kameře se zhotovuje první generace mikrofilmu - archivní negativ, který slouží pro dlouhodobé dochování dokumentu. Druhá generace - matriční negativ, slouží ke zhotovování všech uživatelských kopií (pozitivních) a měla by sloužit i pro skenování.

#### *Digitalizace*

Skenování se provádí na přístroji SunRise, který umožňuje digitalizaci až do rozlišovací schopnosti 600 dpi v černobílém zobrazení i v šedé škále, rychlostí jedno pole mikrofilmu za vteřinu při rozlišení 200 dpi. Přístroj je vybaven bohatým programem umožňujícím vylepšování kvality, zejména kontrastu, odstranění „šumu“, otáčení obrazu, dělení polí na jednotlivé strany a řadu dalších funkcí. Dva adaptéry umožňují skenovat svitkové filmy a mikrofiše. Získané obrazové soubory je možné ukládat ve formátech TIFF, GIF a JPEG a v různých kompresních faktorech. V jedné pracovní směně je možné skenovat dva až čtyři svitky, v závislosti na použití šedé škály a rozlišovací schopnosti. Získané obrazové soubory se propojují s metadaty a vytvářejí se komplexní digitální dokumenty, které garantují nezávislost obrazových dat na přístrojovém a programovém vybavení.

#### *Zpřístupňování digitálních dokumentů*

Dokumenty v digitálním obrazovém formu se zpřístupňují formou CD-R médií, pomocí lokální sítě NK ČR a prostřednictvím internetu. Pro zpřístupňování pomocí sítí slouží systém, jehož součástí je robotická pásková knihovna Adic Scalar 1000, diskové pole, páskový server Sun 450, programový systém (HSM) SAM FS a speciální programová aplikace AIP Safe pro práci s digitálními dokumenty. Pásková knihovna má v současné době kapacitu 2 TB, kterou je možné rozšířit až na 8 TB a po dokoupení dalších tří modulů se kapacita může zvýšit až na 100 TB. Diskové pole je využíváno jako paměťová cache pro zpřístupňování nejvíce žádaných dokumentů, pro uložení metadat a jako pracovní prostor pro dočasné ukládání obrazových souborů. Všechna obrazová data jsou primárně ukládána na magnetických páskách.

V případě výše popsaného systému se hovoří o zpřístupňování near-line, to znamená, že tento systém pracuje s delšími časy. Uživatel má poměrně rychle k dispozici metadata, ve kterých může vyhledávat příslušnou část dokumentu. Čas pro zpřístupnění obrazových souborů o velikosti do 500 KB (větší soubory budou zpřístupňovány pouze na lokální síti NK ČR) je závislý na několika faktorech, z nichž nejvýznamnější je kvalita připojení uživatele, momentální propustnost sítí a počet uživatelů přistupujících v daném okamžiku k robotické knihovně. Ten je limitován počtem mechanik, které jsou k dispozici, což je v současnou chvíli 6, a může být rozšířen na maximální počet 12.

Digitální dokumenty s většími obrazovými soubory budou moci uživatelé získat na CD-R médiích. Stávající způsob zpřístupňování pravděpodobně neu-

možní v plné míře pracovat stejně s digitálním dokumentem jako s původními novinami. Na druhé straně však přináší nové možnosti, které uživatelé ocení. Při formulování projektu bylo prvotním záměrem zpřístupnit na internetu dva větší novinové tituly, jeden český a jeden vydávaný v němčině, který by byl určen také pro zahraniční uživatele, pro které čeština představuje jazykovou bariéru. Ty měly být postupně doplněny několika menšími zajímavými tituly. Volba padla na Bohemii a Národní listy, ty však byly později nahrazeny Národní politikou s ohledem na formát a fyzický stav dochování. Dnes je však již jisté, že bude možné zpřístupnit podstatně větší počet dokumentů.

#### *Archivace a zálohování*

Hybridní technologie reformátování garantuje dlouhodobé dochování dokumentů díky mikrofilmu. Archivní negativ musí být uložen v prostředí s kontrolovanými parametry mikroklimatu. Relativní vlhkost by se měla podle většiny doporučení pohybovat do 40%, teplota do 18° C. Existují však doporučení uvádějící podstatně přísnější požadavky. Optimální chemické zpracování filmu a uvedené podmínky pro skladování garantují dochování mikrofilmu po dobu 500 až 800 let. Archivní a matriční generace by měly být skladovány na dvou různých místech, nejlépe ve dvou různých budovách.

Digitalizace je nákladná a časově náročná činnost, proto je třeba chránit před zničením nebo poškozením i digitální dokumenty zálohováním jejich kopií. Kopírování dokumentů v elektronickém formátu má velkou výhodu v tom, že nedochází k žádné degradaci informace. Zálohování je důležitější v těch případech, kdy se skenování provádí přímo z archivního negativu.

Zálohování digitalizovaných periodik se řeší pomocí robotické knihovny a systému SAM-FS. Každý obrazový soubor je v knihovně uložen dvakrát na dvou identických médiích. Třetí kopie je uložena off-line mimo budovu. V případě poškození nebo zničení první magnetické pásky se dokumenty automaticky zpřístupňují z identické kopie, v případě zničení celého zařízení se dokumenty zachovávají na kopii uložené off-line. Tento systém by měl maximálně garantovat dochování dokumentů jak na mikrofilmech, tak i v digitální podobě, a současně chránit mikrofilmy před opakovaným skenováním.

#### **Další vývoj**

##### *Kvalita mikrografických a digitálních záznamů*

Další vývoj v této oblasti řeší projekt „Optimalizace hybridní technologie reformátování ohrožených knihovních fondů“ realizovaný v letech 2000 a 2001. Zlepšení kvality mikrografického záznamu bylo dosaženo úpravami kamer Dokumator DA 5, které jsou provozovány na všech čtyřech pracovištích. Byly provedeny změny v systému zaostřování, které umožní nastavit přesný faktor zmenšení pro všechny formáty a byla seřizena celková geometrie kamery. Úpravami se zvýšila rozlišovací schopnost na 120 čar na mm.

Velkým problémem limitujícím možnosti mikrografického a digitálního zpracování periodik je jejich velmi špatná optická kvalita daná jednak nekvalitním tiskem již v době vzniku dokumentu a stupněm jeho degradace a opotřebení. Další aktivity jsou zaměřeny na zlepšení výsledného kontrastu. Proběhlo testování barevných filtrů, které však neprokázalo podstatné zlepšení, jsou zkoušeny jiné druhy osvětlení, filmy s různou citlivostí a optimalizuje se pracovní režim vyvolávacího automatu. Kromě toho se připravuje návrh nového zkušební obrazce, který by se mohl používat i pro diagnostikování chyb nastavení skeneru a pro vyhodnocování kvality digitálního obrazového souboru. Zvýšení kvality mikrofilmu by se mělo projevit ve zlepšených výsledcích skenování.

Poslední zkušenosti ukazují, že pro zachování maximálního množství informace, je optimální použít při skenování šedou škálu (276 úrovní), která však velmi zvětšuje výslednou velikost obrazového souboru. Proto je velmi důležité zvolit takový kompresní formát, který při zachování univerzální dostupnosti (pomocí běžných prohlížečů) a dostatečné čitelnosti bude vykazovat co největší kompresní poměry. To je již obsahem řešení dalšího projektu řešeného v NK ČR „Optimalizace archivace a zpřístupnění digitálních dat“.

Ke zmenšení obrazového souboru přispívá ořezávání obrázku, při kterém se odstraňují části protilehlé strany a okolí dokumentu. Současně s tím se provádí pootočení obrázku tam, kde je to třeba. V letošním roce probíhá testování programů, které jsou schopny provádět tyto operace dávkově a pracují s formáty TIFF a JPEG.

*Konverze obrazového formátu do plného textu.*

Zpřístupnění reformátované kopie umožňuje ohrožený dokument vyřadit z cirkulačních služeb a uložit v ochranném obalu (krabici nebo deskách). Zpřístupněním v obrazovém formátu však naše snahy o lepší přístup uživatelů k dokumentům nekončí. Obrazový formát považujeme z několika důvodů za základní. Především zachovává původní grafickou podobu díla včetně obrazových příloh, grafické výzdoby, typů písma a pod. Dále je důležité, že zachovává původní kontext dokumentu, který se jinak ztrácí při převedení jednotlivých článků do plného textu a jejich slitím do databáze obsahující tisíce statí z různých periodik. Konečně třetím důvodem je garantování větší autenticity dokumentu. Pozměnit obrazový soubor je přece jen poněkud obtížnější, než pozměnit textový dokument.

Na druhé straně jsou výhody textového formátu natolik významné, že je třeba se zabývat možností poskytnout dokumenty uživatelům i v této podobě. Textový formát umožní rychlé vyhledávání pomocí fulltextových nástrojů, analýzy textů, vytváření anotací atd. Objem informací uchovávaných v novinách a časopisech je tak mimořádný, že při klasické práci s dokumenty nelze v reálném čase zpracovat více titulů. Textový formát a vyhledávací nástroje jsou v tomto směru velkou nadějí.

Konverze obrazového souboru do plného textu je však časově velmi náročná. I zahraniční zkušenosti ukazují, že se jedná o nejnáročnější část celého procesu: skenování - tvorba metadat - konverze do plného textu. Druhou negativní okolností je vysoká chybovost těchto technologií, která je úměrná kvalitě předlohy. U novin je samozřejmě vyšší. Záměrem NK ČR není poskytnout dokonalý plný text, ale pouze výstup z konverze bez oprav a zvolit takový vyhledávací nástroj, který dokáže překonat chybovost použité technologie. Dokument v plném textu nebude prezentován uživatelům, ale bude sloužit pouze pro vyhledávání. V současné době probíhají testy zaměřené na výběr technologie OCR a vyhledávacího systému, které budou nejlépe vyhovovat našim záměrům.

Je samozřejmé, že se nepodaří konvertovat všechny digitalizované dokumenty. Zde se otevírá pole pro širokou spolupráci institucí. Lze si představit, že se tento úkol podaří snáze realizovat pokud další instituce naváží na aktivity NK ČR a např. v rámci společných projektů se budou podílet na konverzi dokumentů odpovídajícího obsahového zaměření.

#### *Autorské právo*

Zpřístupňování digitalizovaných periodických dokumentů je limitováno nejednotností právních stanovisek, týkajících se výkladu autorského zákona. Zpřístupňování na internetu má velkou flexibilitu a může být přizpůsobeno do té míry, že může být paralelou k využívání knižních dokumentů. Bude-li např. umožněn přístup pouze registrovaným uživatelům v budovách knihoven, jedná se o obdobu MVS. Pokud si uživatel objedná dodání jinak běžně nedostupného dokumentu elektronickou poštou, jde o obdobu reprografických služeb. Tento postup však není naším cílem. Potenciální výhodou elektronických dokumentů je právě ta skutečnost, že mohou být dostupné všem uživatelům připojeným na internet bez ohledu na to, zda jsou momentálně v knihovně, doma nebo na pracovišti. Tomu však v současné době brání absence příslušného ustanovení v autorském zákonu, a na jeho základě vybudovaného mechanismu plateb a jejich odvodů. Řešení tohoto problému si vyžádá složitá konsorciální jednání mezi knihovnami, vydavateli, autory a autorskými svazy. To už však přesahuje možnosti uvedeného projektu.

### **Koordinace a kooperace knihoven**

#### *Pracovní skupina CASLIN pro ochranné reformátování.*

Aktivity spojené s mikrofilmováním a digitalizací knihovních dokumentů již nelze vykonávat v jednotlivých izolovaných knihovnách bez vzájemné koordinace a kooperace. Tomu slouží „Pracovní skupina CASLIN pro ochranné reformátování“, která je česko-slovenskou pracovní skupinou zabývající se strategickými a dlouhodobými otázkami mikrofilmování a digitalizace, výměnou zkušeností, pořádáním seminářů a stáží. Národní sekce této skupiny schvalují programy mikrofilmování a digitalizace na příslušné roky, navrhují a realizují jednotlivé projekty, zabývají se otázkami kontroly kvality a optimalizace technologií.

V rámci aktivit PSpOR CASLIN byly zpracovány a uskutečněny projekty „Kramerius I“ a „Digitalizace mikromédií“ a další dva dodatkové projekty realizované z prostředků nadace Open Society Institute.

#### *Program VISK 7*

Financování těchto nových aktivit představuje velký problém. Zpočátku byly náklady hrazeny z prostředků projektů vědy a vývoje, po jejich skončení jsou však zcela závislé na běžném rozpočtu zúčastněných knihoven. Není, myslím, třeba dodávat, že ani jeden z rozpočtů nebyl navýšen o prostředky potřebné pro udržení provozu mikrografických pracovišť a pracoviště digitalizace. Náklady na běžný provoz pracovišť by měly být v příštích letech hrazeny z programu VISK 7 Kramerius - národní program mikrofilmování a digitálního zpřístupňování dokumentů ohrožených degradací kyselého papíru. Jedná se o podprogram širšího programu známého pod názvem Veřejné informační služby knihoven (VISK), který je součástí Akčního plánu Státní informační politiky.

Bohužel nebyly v r. 2000 na jeho realizaci přiděleny žádné finanční prostředky. Očekáváme však, že již od příštího roku bude tento program financován. Větší část peněz bude věnována na zajištění provozu zmíněných pracovišť, zbývající bude použita na realizaci předložených projektů. Jejich výběr bude provádět PSpOR CASLIN a žadatelé se musí zavázat ke splnění některých podmínek: respektovat normy a doporučení, koordinovat svoji práci s činností ostatních pracovišť, poskytovat bibliografické záznamy do společné centrální databáze záznamů reformátovaných dokumentů, zpřístupňovat uživatelské kopie reformátovaných dokumentů a zajišťovat služby, napomáhat při ochranném reformátování ohrožených titulů např. zápůjčkami dokumentů nebo poskytnutím bibliografických informací, dlouhodobě uchovávat archivní kopie reformátovaných dokumentů v souladu s doporučeními normy ISO.

Mikrofilmování může zajišťovat některé ze čtyř mikrografických pracovišť. V případě dalších pracovišť, je třeba doložit naplnění požadavků norem ISO (formální vybavení mikrofilmu, použití předepsaných zkušebních obrazců, atd.) a dodržení kvalitativních parametrů. Pro digitalizaci mikrofilmů vyčlení NK ČR určitou část vlastní kapacity. Při digitalizaci mimo NK ČR je třeba akceptovat požadavek vytvoření metadat ve formátu DOBM. Knihovny zajišťující tyto činnosti dodavatelsky budou účtovat pouze vynaložené náklady, přičemž se mohou na realizaci podílet např. úhradou režijních nákladů a pod. Předběžná kalkulace nákladů v NK ČR se pohybuje okolo 5 Kč za mikrofilmování, indexaci (tvorba metadat) a digitalizaci jedné strany dokumentu.

#### **Závěr**

Realizací uvedených projektů se českým knihovnám dostává výkonný nástroj pro konverzi papírových dokumentů do elektronického formátu, jejich dlouhodobou archivaci a zpřístupňování uživatelům. Současná kapacita čtyř mikrografických pracovišť je cca 1 milion stran dokumentů ročně, což představuje

1 až 2 tis. svazků. Odhadujeme, že jen v samotné NK ČR je cca 200 tis. svazků periodik tištěných na kyselém papíru a tím ohrožených degradací, které potřebují nebo budou v blízké budoucnosti potřebovat převést na jiný, trvanlivý nosič. Za stávajících podmínek by reformátování vyžadovalo 100 až 200 let. Je nereálné spoléhat, že za tu dobu budou tyto dokumenty ještě existovat, nebo budou ve stavu, kdy bude možné jejich reformátování. Jediné řešení je postupné zvyšování kapacity pracovišť mikrofilmování a digitalizace, aby se celková doba snížila na 50 až 100 let. Kromě periodik je však velké množství monografií z období od poloviny 19. stol. do druhé poloviny 20. stol. které sice nejsou tak intenzivně využívány, ale stejně jako u periodik, mění v důsledku nevratných chemických procesů své fyzikální vlastnosti, a již brzy budou na seznamu ohrožených dokumentů. Jejich počet může být vyšší než u periodik. Příklad jiných, nám blízkých zemí, v tomto případě Maďarska, nám však ukazuje, že je to úkol zvládnutelný, pokud je obecná vůle k záchraně národního intelektuálního dědictví uloženého v knihovních sbírkách.