

# Koncepce rozvoje Národní knihovny České republiky jako výzkumné organizace na léta 2010 - 2015

---

*Zpráva o řešení za r. 2010*

Koordinátor:

Mgr. Adolf Knoll, ředitel pro vědu, výzkum a mezinárodní vztahy

Hlavní řešitelé dílčích oblastí:

PhDr. Zdeněk Uhlíř

Mgr. Jan Hutař

Ing. Libor Coufal

Ing. Petra Vávrová, PhD.



## Obsah

Oblast 1: Zdokonalení virtuálního badatelského prostředí Manuscriptoria .....	5
Priorita 1: Agregace dat .....	5
Priorita 2: Technický rozvoj .....	5
Priorita 3: Externí nástroje pro srovnávání plných textů .....	6
Priorita 4: Diverzifikace obsahu – rozšíření o edice hudebních dokumentů .....	6
Priorita 5: Externí nástroje pro srovnávání obrazů .....	7
Oblast 2: Dlouhodobá ochrany (LTP – Long-Term Preservation) digitálních dokumentů .....	8
Priorita 1 : Plánování ochrany (2010-2013) .....	8
Podčást 1 – strategie dlouhodobé ochrany digitálních dat .....	8
Podčást 2 – nástroje na plánování dlouhodobé ochrany digitálních dat .....	11
Podčást 3 – cost analysis - náklady na dlouhodobou ochranu jednotlivých typů digitálních dokumentů ...	13
Oblast 2 - priorita 3 : Vývoj podpůrných/návazných aplikací pro chod LTP systému (2010-2012) .....	15
Podčást 1 – mapování DTD monografie a DTD periodika do interních formátů LTP systémů .....	16
Podčást 2 – tvorba aplikací pro převod starých dat do interních formátů LTP systému .....	19
Oblast č. 5: Vývoj nových metod konzervace novodobých knihovních dokumentů .....	23
Ad 1.) Rešerše odborné literatury .....	23
Ad 2.) Vývoj metodiky průzkumu .....	23
Ad 3.) Typologie skladovacích prostor NK ČR .....	25
Ad 4.) Analýza požadavků na funkcionalitu znalostní báze .....	25
Ad 5.) Zahájení průzkumu .....	26
Závěr: .....	27
Upřesnění úkolů na rok 2011 .....	27
Návrh výsledků výzkumu a vývoje k nahlášení do edatabáze RIV .....	29
Publikace .....	29
Výsledek č. 1 .....	29
Výsledek č. 2 .....	29
Výsledek č. 3 .....	29
Výsledek č. 4 .....	29
Výsledek č. 5 .....	30
Výsledek č. 6 .....	30
Výsledek č. 7 .....	31
Výsledek č. 8 .....	31
Aplikované výsledky .....	32
Výsledek č. 1 .....	32

# *Koncepce rozvoje Národní knihovny České republiky jako výzkumné organizace na léta*

*2010 – 2015*

Rok 2010 byl prvním rokem, kdy byl v Národní knihovně ČR financován výzkum a vývoj novým způsobem. Šlo o rok přechodný, ve kterém byly ještě řešeny všechny tři výzkumné záměry, nicméně dva z nich skončily a v r. 2011 pokračuje pouze výzkumný záměr *Výzkum a vývoj nových postupů v ochraně a konzervaci písemných památek*. Tento záměr končí v r. 2011.

Vzhledem k průběhu řešení dochází také k revizi koncepce po prvním roce řešení, zejména jde o sloučení obou oblastí týkajících se ochrany fondů. Z těchto důvodů je jako jedna z příloh předkládána i aktualizovaná Koncepce rozvoje z pohledu r. 2011.

V r. 2010 byly řešeny tři oblasti z celkových pěti, a to:

Oblast č. 1: Zdokonalení virtuálního badatelského prostředí Manuscriptoria

Oblast č. 2: Dlouhodobá ochrany (LTP – Long-Term Preservation) digitálních dokumentů

Oblast č. 5: Vývoj nových metod konzervace novodobých knihovních dokumentů

K řešení oblastí 3 (Archivace a zpřístupnění sklizených webových stránek) a č. 4 (Metodika ochrany a konzervace klasických knihovních a archivních materiálů) nebylo v souladu s původním plánem přistoupeno.

Jako zvláštní přílohy tvoří součást této zprávy:

- Jednotlivé dílčí zprávy, pokud byly výstupem příslušných činností. Tyto zprávy jsou sdruženy do jednoho dokumentu, nicméně jejich formátování se liší dle volby jejich autorů. V některých případech, především v oblasti porovnávání digitálních obrazů (Oblast 1, Priorita 5) nebylo možné dílčí zprávy předat v tiskové formě, resp. na CD, neboť by ztratily vypovídací schopnost. V tomto případě je třeba čerpat z URL z <http://cmp.felk.cvut.cz/~chaluv1/>
- Cestovní zprávy ze zahraničních cest, financovaných z prostředků přidělených na realizaci této koncepce
- Aktualizovaná *Koncepce rozvoje Národní knihovny České republiky jako výzkumné organizace na léta 2010 – 2015. Výhledová verze r. 2011*

## Oblast 1: Zdokonalení virtuálního badatelského prostředí Manuscriptoria

Vývoj virtuálního badatelského prostředí Manuscriptoria v tomto období plynule navazoval na období předchozí v agregaci dat, v technickém rozvoji, ve vývoji externích nástrojů pro srovnávání plných textů i v diverzifikaci obsahu (rozšíření o edice hudebních dokumentů). Dříve, než bylo plánováno, byla zahájena práce na vývoji externích nástrojů pro srovnávání obrazů.

### Priorita 1: Agregace dat

Agregace dat, která se výrazně zvýšila v průběhu řešení evropského projektu ENRICH v letech 2007-2009, pokračovala plynule i v tomto období, třebaže v menší kvantitě. Někteří partneři, kteří se k síti Manuscriptoria připojili už dříve, pokračovali v dodávání dat. To se týká zejména Rumunské národní knihovny v Bukurešti, která dosáhla počtu 306 poskytnutých dokumentů, 109 z Bukurešti a 197 z pobočky v Alba Iulia. Získání dokumentů z Alba Iulia pro Manuscriptorium je obzvláště důležité, protože jde o nejvýznamnější rukopisnou sbírku v Rumunsku, jež je typická tím, že obsahuje materiál týkající se nikoli jen Rumunska, ale celé střední Evropy. Týká se to také Univerzitní knihovny v Budapešti, která má největší fond středověkých rukopisů v Maďarsku a která k dosavadním 4 dodala dalších 20 plně digitalizovaných dokumentů. V rámci řešení projektu REDISCOVER (2009-2010) se do sítě Manuscriptoria zapojili noví partneři, kterými jsou Litevská národní knihovna ve Vilně (dodala 168 plně digitalizovaných dokumentů) a Polská národní knihovna ve Varšavě (dodala 46 plně digitalizovaných dokumentů). Pro obohacování obsahu Manuscriptoria nebylo nutno v tomto období vyvíjet žádnou novou konverzi, protože bylo možno využít postupů vytvořených a ověřených v rámci projektu ENRICH.

V dalším období se předpokládá završení spolupráce s partnery, kteří projeví zájem o spolupráci s Manuscriptoriem, jakož i získání nových partnerů.

### Priorita 2: Technický rozvoj

Technický rozvoj Manuscriptoria se v tomto období soustředil na dokončení, resp. dotažení dosavadního vývoje, zejména pokud jde o nástroje vytvořené v průběhu řešení projektu ENRICH. Práce na této prioritě byla v tomto období koordinována s výzkumným záměrem MK00002322101 *Vytvoření virtuálního badatelského prostředí pro zpřístupnění a ochranu digitálních dokumentů*, resp. řešena v jeho rámci.

Byla definitivně zpřístupněna nová platforma Manuscriptoria založená na TEI P5 ENRICH Specification včetně modulů MTool (nástroj pro vytváření komplexních digitálních dokumentů, resp. distribuovaných komplexních digitálních dokumentů, jakož i virtuálních dokumentů) MCan (nástroj pro testování distribuovaných komplexních digitálních dokumentů a pro jejich nahrávání do Manuscriptoria), a My Library (individuální uživatelský účet umožňující heuristickou práci se statickými a dynamickými virtuálními kolekcemi a virtuálními dokumenty) a byl ukončen provoz předchozí platformy založené na TEI P4 MASTER+.

Byly zdokonaleny možnosti práce s obrazovými daty (zpráva Zpřístupnění objemných obrazových datových souborů s cílem zrychlit odezvu a snížit datové přenosy) a byl položen základ pro vytváření reprezentativních náhledů, jež by dobře ilustrovaly obsah konkrétních digitálních kopií a které by byly

využitelné pro Europeana, pro něž je Manuscriptorium subagregátorem (Zpráva o automatizovaném vytváření nových hodnotných metadat pro digitalizované dokumenty zpřístupněné v Manuscriptoriu).

Pro uživatelský komfort Manuscriptoria byly zpracovány analýzy využití grafických variant (zpráva Řešení problematiky grafémů při vyhledávání v Manuscriptoriu), ontologií (zpráva Ontologie: Analýza ontologických dat Národní knihovny ČR, jejich použitelnost pro Manuscriptorium a ověření pilotním řešením) a tezaurů a autoritních databází pro geografické názvy, resp. osobní jména (zpráva Autoritní databáze a řízené slovníky: Pilotní řešení s využitím tezauru historických měst).

V dalším období se výzkum a vývoj soustředí na praktickou implementaci grafických variant, ontologií, resp. ontologických instancí a tezaurů, resp. autoritních databází do vyhledávacího stroje Manuscriptoria.

### **Priorita 3: Externí nástroje pro srovnávání plných textů**

V tomto období byl vytvořen poloprovoz on-line nástroje pro porovnávání plných textů. Webová aplikace NKCatalog umožňuje uživateli vyhledat podobné segmenty textu mezi dokumentem uživatele a dokumenty v předem stanoveném repozitáři. Podobnost se hledá na úrovni různě dlouhých segmentů textu na základě statistického vyhodnocení vektorové analýzy. Aktuální verze aplikace, jejíž poloprovoz byl zpřístupněn pro soustavné testování, pracuje s repozitářem, který je součástí digitální knihovny Manuscriptorium a obsahuje plná znění rukopisných textů v datových balíčcích Etno, NKCR, PARK a UJCAV. Jazyky dokumentů jsou pro testovací účely omezeny na latinu, češtinu a němčinu. V rámci měření podobnosti jsou používány grafické varianty pro příslušný jazyk, aby aplikace našla shodné, resp. podobné segmenty i při více méně odlišné grafice konkrétního zápisu. Aplikace je implementována v české a anglické lokalizaci. Pro testování z vybraných IP adres je aplikace přístupná na adrese <http://195.113.132.83:8080/match-web/>.

V následujícím období bude porovnávání plných textů obohaceno o grafické vyhodnocení a zpřístupněno koncovým uživatelům.

### **Priorita 4: Diverzifikace obsahu – rozšíření o edice hudebních dokumentů**

Pokračovala práce na výzkumu zpřístupňování edic hudebních dokumentů (hudebního zápisu), s cílem ověřit možnost praktického zjednodušení značkování MEI úpravou některého standardního XML editoru a možnost jednoduché vizuální reprezentace v podobě obrazu. Bylo zjištěno, že tato cesta je neefektivní, protože náklady na ni by byly příliš vysoké. V rámci projektu CMME (Computerized mensural music notation): Dynamic Early music editions (<http://www.cmme.org/>) byl sice vytvořen editor pro hudební notaci, jeho použitelnost je však omezená, jak bylo zjištěno testováním v Seminárii dějin starší hudby při Ústavu hudební vědy Filozofické fakulty Univerzity Karlovy v Praze ([http://musicology.ff.cuni.cz/seminar\\_starsi.htm](http://musicology.ff.cuni.cz/seminar_starsi.htm)). Zároveň byl v tomto seminárii vytvořen základ pro editor, který může být s poměrně malými úpravami adaptován pro potřeby edic hudebních dokumentů (hudebního zápisu) určených pro Manuscriptorium.

V dalším období bude tento editor adaptován pro potřeby Manuscriptoria.

## **Priorita 5: Externí nástroje pro srovnávání obrazů**

Práce na této prioritě byly ve spolupráci s Centrem strojového vnímání katedry kybernetiky FEL ČVUT (<http://cmp.felk.cvut.cz/>) zahájeny výrazně dříve, než byl původní plán, protože souhlasí s dlouhodobým výzkumným zaměřením Centra, takže nebylo třeba žádného předběžného informačního průzkumu ani přípravných prací. Byly provedeny první kroky ve výzkumu porovnávání obrazů (dostupné z <http://cmp.felk.cvut.cz/~chaluv1/>). Bylo dosaženo prvních výsledků týkajících se rozlišení mezi stránkami obsahujícími a) písmo, b) obraz, c) hudební notaci, přičemž úspěšnost výsledků je 98,7%. Zatím však byl vytvořen jen základní algoritmus, nikoli nástroj.

V dalším období bude vytvořena první verze nástroje pro rozlišování obrazů.

## **Oblast 2: Dlouhodobá ochrany (LTP – Long-Term Preservation) digitálních dokumentů**

### **Priorita 1 : Plánování ochrany (2010-2013)**

Z oblasti 2 byly ze tří plánovaných priorit řešeny priorita 1 a 3, priorita 2 je plánována k řešení až od roku 2011. Prioritu 1 Plánování ochrany jsme rozdělili na 3 podčásti

- 1) strategie dlouhodobé ochrany digitálních dat;
- 2) nástroje na plánování dlouhodobé ochrany digitálních dat;
- 3) cost analysis – náklady na dlouhodobou ochranu jednotlivých typů digitálních dokumentů.

Toto rozdělení, jak se ukázalo, je plně funkční a bude zachováno i do roku 2011, během kterého by se výstupy postupně měly integrovat dle plánu tak, aby byly k užítku v nové oblasti, které se NK věnuje, tj. v dlouhodobé ochraně digitálních dat.

V rámci projektu NDK (Národní digitální knihovna) získá NK systém na dlouhodobou ochranu (dále LTP systém). Jeho provozování je podmíněno jasnou strategií ochrany dat, využitím relevantních nástrojů na plánování i vědomím nákladnosti ochrany u určitých dokumentů. Na základě jasné koncepce lze pak provádět rozhodnutí, která ve svém důsledku mohou přinést ušetření finančních prostředků, ochranu pro více důležitých dat, ale také např. vyloučení některých dat z dlouhodobé ochrany apod. Tj. výstupy řešení z roku 2010 a následně 2011 budou nutným předpokladem k fungování celého úložiště dat a vlastní dlouhodobé ochrany i po koncepční stránce (plánování celého životního cyklu digitálních dat), nejen po stránce technické.

Vzhledem k překotnému vývoji projektu NDK bude ovšem také nutné přehodnotit původní plány na jednotlivé priority, hlavně u priority 3, která se již nezdá být zcela aktuální – viz níže její popis.

Velmi pozitivní je, že u všech řešených priorit lze vycházet z podobných problémů řešených v zahraničí jak v paměťových institucích, tak v rámci návazných národních nebo evropských projektů.

### **Podčást 1 – strategie dlouhodobé ochrany digitálních dat**

Strategie dlouhodobé ochrany je velmi důležitá jak na institucionální úrovni, tak na úrovni národní. V prvním roce proběhl průzkum relevantních institucionálních strategií, ze kterých se bude dále vycházet. Teprve až poté bude vytvořen návrh na strategii širší (národní), která by měla být obecnější a mělo by se jí zabývat i Ministerstvo Kultury, např. v rámci nové koncepce knihovnictví do r. 2014, která je nyní připravována.

Strategie dlouhodobé ochrany digitálních dat NK musí stanovit měřitelné cíle a indikátory, musí určit explicitně odpovědnosti. Rozpracování zárodku strategií (dlouhodobé, střednědobé a projektové), které vznikají v NK od konce roku 2009 dosud, souvisí s tím, jaké konkrétní technologie budou využity při budování LTP systému v rámci projektu NDK. Obecná institucionální strategie musí být rozpracována do politik, které budou určovat způsob nakládání s jednotlivými typy digitálních dokumentů.



## Řešení v roce 2010

V rámci řešené koncepce vznikl rozklad základních znaků, které porovnávané strategie různých institucí mají. Z těchto bodů se bude vycházet při vytváření strategií pro NK.

Podle funkčního modelu normy OAIS, který v obecné rovině vymezuje základní funkce digitálního repozitáře, je plánování ochrany zajišťováno tzv. plánovací entitou (*preservation planning*). Plánovací entita představuje jednu z šesti základních funkčních entit repozitáře podle normy OAIS. Má za úkol poskytovat služby a funkce pro monitorování vnějšího kontextu repozitáře a dávat doporučení, která zajistí kontinuitu trvalého zpřístupňování informací bez ohledu na zastarávání počítačových technologií a další rizika spojená s plněním cílů digitální archivace. Mezi základní funkce plánovací entity patří vytváření doporučení pro institucionální systém pravidel a zavádění standardů nebo monitorování změn v technologiích a designované komunitě, která je klientem repozitáře.

### Plánování ochrany a systémy pravidel pro digitální archivaci

Podle řady odborných studií věnovaných problematice dlouhodobé ochrany digitálních dokumentů je systém pravidel pro digitální archivaci základním pilířem pro následně vykonávaná ochranná opatření. Poskytuje totiž instituci, která má za úkol tato opatření vykonávat, základní ukazatel pro směřování vlastních aktivit. Proto se také doporučuje, aby tento systém pravidel zůstal v rozumně obecné rovině. Například podle studie projektu ERPANET musí být všechny institucionální kroky podstupované při implementaci konkrétních ochranných opatření (jako jsou například formátová migrace, aktualizace metadat, emulace nebo vývoj metodik pro stanovování signifikantních vlastností archivovaných digitálních dokumentů) v souladu se systémem pravidel konkrétní instituce, aby byla zajištěna jejich koherence. Z vnějšího úhlu pohledu je (písemně zaznamenaný) systém pravidel znakem toho, že daná organizace přejímá odpovědnost za ochranu digitálních dokumentů. Systém pravidel pro digitální archivaci je podle praxe analyzovaných zahraničních paměťových institucí vždy zaznamenan v příslušných institucionálních dokumentech, které jsou volně přístupné na webových stránkách těchto institucí.

Podle normy OAIS je digitální repozitář vždy součástí širšího kontextu. Každý repozitář je vždy součástí nějaké širší instituce (knihovna, univerzita, vědecký ústav apod.). Tento širší institucionální kontext je v normě označen jako management (*management*), což je institucionální role, která reprezentuje instituci, jejíž je repozitář součástí, a která určuje celkovou strategii repozitáře. Management například podle normy OAIS schvaluje konkrétní povinnosti repozitáře a jeho dohody s producenty dat a designovanou komunitou, zajišťuje financování repozitáře, revizi jeho výkonu a pokroku v oblasti dlouhodobé ochrany nebo stanovuje a schvaluje cenovou politiku. Z toho vyplývá, že plánovací entita repozitáře nemůže vytvářet systém pravidel bez spolupráce s výše definovaným managementem, protože bez něj nelze zohlednit řadu otázek širšího kontextu (finanční otázky, vztahy s producenty dat a uživateli aj.).

Bližší informace o tom, jakým způsobem má plánovací entita vyvíjet systém pravidel v součinnosti s managementem instituce, norma OAIS neposkytuje, proto je *potřeba se inspirovat praxí významných zahraničních paměťových institucí, do jejichž dílčích povinností spadá také provozování digitálního repozitáře*. Pro tento účel byl proveden výzkum institucionálních systémů pravidel pro digitální archivaci publikovaných paměťovými institucemi (a v jednom případě celonárodním projektem) ze zemí, které jsou v oblasti vývoje digitální archivace nejdále, zejména tedy Austrálie, USA, Kanada, Nizozemsko a Německo, a identifikovali jsme několik hlavních referenčních institucí,

kteří jsou nejen nejvýznamnější ve svých zemích, ale také jsou relevantní z hlediska srovnání s Národní knihovnou ČR, a to jednak z národního hlediska (národní knihovny), ale také z hlediska předpokládané diverzity budované české Národní digitální knihovny. Z toho hlediska bylo potřeba všimnout si i univerzitních knihoven a dalších paměťových institucí (archivy, muzea, konsorcia).

### ***Charakteristické prvky zahraničních systémů pravidel pro digitální archivaci***

Všechny identifikované systémy pravidel pro digitální archivaci jsou systémy, které deklarují instituce jako celek, nikoliv pouze jejich oddělení zabývající se provozem nebo plánováním budování digitálního repozitáře. Systémy pravidel jsou zde deklarovány na nejvyšší úrovni instituce, protože zasahují do strategických celo-institucionálních otázek (zákonný mandát instituce, rozpočet, řízení lidských zdrojů, strategické směřování instituce apod.). Na základě analýzy vybraných institucionálních pravidel lze říci, že jde o pravidla, kterými dané instituce explicitně (v písemné / textové podobě) deklarují obecné směřování v oblasti digitální archivace. Pravidla obvykle obsahují informace o celkovém poslání instituce (mandát, zakládací listina apod.), kontextuálních souvislostech (vztahy se širším kontextem, zejména se stakeholdery aj.), finančních zdrojích instituce, řízení lidských zdrojů a ochraně duševních práv vázících se archivovaným digitálním dokumentům.

Systémy pravidel těchto vybraných institucí jsou různorodé, co se týče míry podrobnosti, oblastí, které identifikují, a rozsahu, který pokrývají pravidla. Některé dokumenty jsou velmi stručné, jiné velmi detailně propracované. Řada dokumentů se zaměřuje především na vybrané obecné oblasti. *Přesto lze ve většině systémů pravidel identifikovat některé společné charakteristiky, které můžeme považovat za základ pro vytvoření pravidel pro digitální archivaci pro Národní knihovnu ČR.*

V této části textu charakteristické body shodné při porovnávání systémů pravidel pouze vyjmenujeme, jejich podrobnější popis je v příloze.

- Terminologie
- Stakeholder
- Výběr dokumentů
- Akvizice
- Finanční konsekvence
- Plán nástupnictví
- Standardizace
- Deklarované služby
- Ochrana práv duševního vlastnictví
- Kontextuální souvislosti
- Propagace

### ***Návrh dalšího postupu prací na rok 2011***

Uvedené charakteristické body je třeba dále rozpracovat do systému pravidel pro digitální archivaci, který bude sepsán a schválen v kontextu Národní knihovny ČR a projektu NDK. Je potřeba explicitně vyjádřit všechna rizika, rozsah současně poskytovaných služeb, možnosti jejich rozšíření, plány pro případy omezení finančních prostředků na digitální archivaci, metodiku výběru dokumentů, využívané nebo zvažované standardy a řadu dalších aspektů, přičemž celý dokument musí být schválen širším vedením české národní knihovny.

V roce 2011 by měly být dokončeny práce na institucionálním systému pravidel (strategii), který bude postoupen vedení NK k případnému schválení. Podobně by měla být dokončena projektový systém

pravidel (strategie) dlouhodobé ochrany pro projekt NDK. Tento plán bude zahrnovat i ochranu externích dat přicházejících do NK z ostatních institucí apod.

## Podčást 2 – nástroje na plánování dlouhodobé ochrany digitálních dat

S otázkou strategií dlouhodobé ochrany digitálních dat souvisí vlastní plánování konkrétních postupů ochranných opatření. V této oblasti existuje několik externích nástrojů, jako např. PLATO nebo i open source systémy pro dlouhodobou ochranu, které mohou při plánování také výrazně pomoci. Komerční LTP systémy mají propracované moduly plánování ochrany, open source systémy v posledním roce začínají tuto službu, která jim dosud chyběla, také poskytovat. Je tedy možné použít open source systémy jako prostředek na testování a plánování, vedle komerčního systému, který NK získá v rámci projektu NDK. I tento komerční systém není všespásný, je potřeba si určité postupy migrací apod. předem promyslet, otestovat v určitém nástroji. Plánovací nástroje mohou být později včleněny do ostrého pracovního procesu projektu NDK i samostatné NK, kde by takto získané znalosti ovlivňovaly např. výběr formátů k digitalizaci, uložení apod.

### Řešení v roce 2010

V roce 2010 bylo cílem udělat vstupní průzkum dostupnosti nástrojů pro plánování dlouhodobé ochrany digitálních dat včetně open source systémů na dlouhodobou ochranu dat. Na základě tohoto průzkumu byly vybrány systémy pro další testování v roce 2011. V roce 2010 se prezentovalo několik nových open source nástrojů, což je velmi pozitivní. Rok předtím, tj. 2009, byl výběr omezen pouze na 2. Podařilo se nalézt a porovnat následující nástroje:

1) open source nástroje pro dlouhodobou ochranu dat

a) Fedora a její nadstavby:

RODA (Portugalsko, Portuguese National Archives) - <http://roda.di.uminho.pt/#home>

I když se jedná o technické řešení vytvořené pro národní archiv, díky servisně orientované architektuře je možné jej (především modul CRiB) integrovat do různých systémů NK ČR (např. do Transformačního modulu či Modulu plánování dlouhodobé ochrany budoucí Národní digitální knihovny).

CRIB (Portugalsko, University of Minho) - <http://crib.dsi.uminho.pt/>

Původně vznikl jako samostatný migrační nástroj, je integrován do systému RODA v rámci preservačního modulu.

MOPSEUS (Řecko, Digital Curation Unit) – <http://194.177.192.14/mopseus/>

Systém je určen pro menší a středně velké typy institucí, není velmi propracován. Metadatový popis je jen na úrovni Dublin Core a systém nenabízí žádný nástroj pro migraci. Výše uvedené nedostatky do podstatné míry ztěžují alespoň částečnou integraci Mopseus do systémů NK ČR.

b) další nástroje:

HOPPLA (Rakousko, Technische Uni Wien) - <http://www.ifs.tuwien.ac.at/dp/hoppla/>

Možnost implementace pro instituce typu národních knihoven se nejeví jako příliš reálná, jelikož je Hoppla určena pro jinou cílovou skupinu (domácí využití na domácích datech) a nepočítá s "náročnou" ochranou digitálních historických pramenů.

ARCHIVEMATICA (UNESCO + Kanada) - <http://archivemata.org/>

Implementace pro NK ČR se doporučuje téměř v úplném rozsahu - jednotlivé mikroslužby je možné připojit do systémů NK ČR. Zatím nejvyužitelnější součástí se zdá být "monitorPreservation". Archivemata uchovává originální formát všech importovaných dokumentů pro migraci a emulaci, přičemž normalizace souborů pro dlouhodobou ochranu je v otevřených formátech.

2) nástroje/metodiky pro plánování dlouhodobé ochrany

AIDA (Velká Británie, University of London + JISC) - <http://aida.jiscinvolve.org/wp>

Jedná se o metodologii, integrace AIDA je možná v případě její implementace do stávajících systémů či budoucích systémů. V kontextu NDK je možnost využití pro závěrečnou evaluaci projektu.

PLATO (Rakousko, Technische Uni Wien) - [www.ifs.tuwien.ac.at/dp/plato](http://www.ifs.tuwien.ac.at/dp/plato)

PLATO podporuje rozhodování v oblasti plánování dlouhodobé ochrany digitálních dat a výběru nejvhodnějšího formátu pro budoucí migrace. Tento nástroj vychází z referenčního modelu OAIS (konkrétně) jeho funkční entity Plánování dlouhodobé ochrany. Je o online softwarovou aplikaci (licencovanou jako CC-GNU LGPL) pro evaluaci potenciálních ochranných řešení a strategií. Výstupem je komplexní dokumentace s objektivním posouzením nejrozumnějších migračních alternativ dle specifických požadavků dané instituce. Proto se implementace nástroje do workflow dlouhodobé ochrany digitálních dat doporučuje v plném rozsahu.

Jednotlivé nástroje jsou dále podrobněji popsány v příloze.

Některé z výše uvedených nástrojů jsou vyvíjeny za účelem jejich integrace do stávajících systémů institucí či cílových skupin (např. CRIB, Hoppla). Jiné fungují jako první prototypy open source digitálních repozitářů odpovídajících referenčnímu modelu OAIS a plně podporujících životní cyklus dlouhodobé ochrany digitálních dat (např. Mopseus, Archivemata, RODA). Samostatnou skupinu nástrojů tvoří metodologie (online, offline) sledování, evaluace a hodnocení činnosti repozitáře, včetně plánování dlouhodobé ochrany dat.

Nejpoužitelnějším nástrojem na plánování dlouhodobé ochrany i nadále zůstává PLATO, které v současné době nemá žádný vhodný ekvivalent. PLATO je výstupem několikaletého evropského projektu PLANETS (<http://www.planets-project.eu/>).

Z open source systémů na dlouhodobou ochranu se zdá vhodná k testování zcela jistě Archivemata nebo systém RODA (CRIB).

### ***Návrh dalšího postupu prací na rok 2011***

V následujícím roce by se mělo začít s návrhem použití systému PLATO v NK ČR i v závislosti na tom, jaký LTP systém bude zvolen v rámci projektu NDK. Obecně lze ale PLATO využít např. i na stávající data bez LTP systému a na podporu tvorby krátkodobé (projektové) strategie dlouhodobé ochrany digitálních dat.

Otázkou zůstává, zda je žádoucí testovat i systém Archivemata. Vhodnější by bylo se asi v tomto roce řešení soustředit plně na nástroj PLATO, který má své uživatele, podporu a může významně přispět do procesu plánování dlouhodobé ochrany nejen Národní digitální knihovny, ale všech projektů NK.

### Podčást 3 – cost analysis - náklady na dlouhodobou ochranu jednotlivých typů digitálních dokumentů

Součástí každého procesu plánování (a výsledné metodiky jak pro NK i pro ostatní paměťové instituce) by měla být i analýza nákladnosti dlouhodobé ochrany jednotlivých typů datových objektů (jejich životního cyklu), tj. kolik to vlastně celé instituci bude stát, zda na to instituce vůbec má a bude mít prostředky i v budoucnu. Tento typ analýzy může významně ovlivnit strategie dlouhodobé ochrany. Např. v případě, že není dostatek financí, může se v rámci instituce dojít k rozhodnutí, že určité typy dat nebo výstupy určitého projektu budou z procesů dlouhodobé ochrany vyloučeny apod. K problematice existují projekty, velmi daleko jsou ve Velké Británii. Vzhledem k tomu, že digitalizace analogových dokumentů probíhá v masovém měřítku takřka na celém světě, je žádoucí vybrané projekty monitorovat a určité části implementovat i do českého prostředí. Každá z organizací zabývající se digitalizací má svá specifika a od nich se také odvíjí doporučení na základě vlastních analýz. Vzhledem k šíři celé problematiky a geografickým odlišnostem není možné přebírat doporučení a aplikovat je bez jakýchkoli změn. Tento proces vyžaduje jak znalost českého prostředí, tak znalost jiných, zejména zahraničních projektů a nacházet vhodná východiska pro optimalizaci procesů v našem prostředí.

#### Řešení v roce 2010

Na rok 2010 bylo plánováno udělat vstupní průzkum dostupnosti nástrojů a metodik a jejich základní porovnání. Z dostupných metodik byly porovnávány jen ty, které jsou svou propracovaností relevantní cílům NK. Na základě tohoto průzkumu byla vybrána metodika pro další návrhy implementace do prostředí NK v roce 2011. Do porovnání byly vybrány následující nástroje/metodiky:

#### BRTF-SDPA

(Ensuring Long-Term Access to Digital Information)

Jedná se o projekt, jehož výstupem je několik studií. Nelze říci, že by šlo o metodologii jako takovou. V roce 2010 byla publikována finální verze zprávy "Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information." Tato zpráva předkládá obecné zásady a činnosti na podporu dlouhodobé ekonomické udržitelnosti. Tyto obecné zásady by se ve zkratce daly popsat jako specifická doporučení "ušíťá" přímo na míru konkrétním scénářům, které jsou též podrobně analyzovány, dále uvádí doporučení pro nutnou agendu pro prioritní opatření a s tím spjaté další kroky, organizované podle daného způsobu a typu rozhodování v dané situaci. V následně vydané zprávě z roku 2010, pracovní skupina SDPA navrhla doporučení pro United States Office of Science and Technology Policy's nazvanou: „Grand Challenge: Sustainable Knowledge Infrastructure“. Tato studie se snaží definovat postup, který zaručí že, dnešní znalosti budou dostupné a použitelné i v budoucnosti.

Otázky a návody v těchto zprávách je velmi těžké přenést do podoby obecného závěru použitelného v prostředí ČR. Budoucí vývoj projektu počítá s vypracováním souboru ekonomicky životaschopných doporučení a sestavit přehled vývoje spolehlivých strategií pro uchování digitálních informací. I tak s dalším rozsáhlejším využitím v rámci NK nelze počítat.

## DiCoMo

(An Algorithm Based Method to Estimate Digitization Costs in Digital Libraries)

Projekt DiCoMo je odvozen od modelu CoCoMo, který je v praxi rutině nasazován pro odhady nákladů při vytváření softwaru. Model DiCoMo analogicky jako předešlý model CoCoMo vychází ze stejné premisy, a to že: „Čas jsou peníze“. Tím pádem vypočítáním časové náročnosti digitalizace nám dá velmi přesné vodítko k výsledným nákladům. Díky kvantitativnímu sběru dat v celém průběhu digitalizace bylo nashromážděno velké množství relevantních dat, které byly použity na vytvoření korekcí, aby výsledný výpočet nákladů, tedy potřebného času, co nejlépe odpovídal reálným, empiricky ověřeným hodnotám. Byly zohledněny veškeré známé faktory ovlivňující časovou náročnost procesu digitalizace. I tak je jasné, že předpokladem nasazení tohoto modelu je systematický sběr technických dat v čase, která následně slouží ke zpřesnění výpočtu potřebných nákladů.

Výstupem projektu je vzorec na výpočet finanční náročnosti procesu digitalizace, který by šlo použít za předpokladu, že máme v NK relevantní statistiky, neřeší ovšem problematiku dlouhodobé ochrany dat.

## LIFE I, II, III

(Life Cycle Information for E-Literature)

Projekt zabývající se odhadem nákladů na celý životní cyklus dokumentu vzniká ve spolupráci mezi University College London (UCL) a British Library. Projekt prošel za několik let svého vývoje již třemi fázemi. První fáze LIFE I byla dokončena v dubnu 2006, druhá fáze LIFE II v srpnu 2008 a třetí fáze LIFE III byla dokončena koncem minulého roku 2010. Z každé této fáze je na oficiální webové stránce tohoto projektu k dispozici výsledná analýza.

Ve studiích LIFE jsou velmi detailně rozpracovány tyto jednotlivé kroky správy digitálního dokumentu:

- § Digitalizace (Digitization)
- § Akvizice (Acquisition)
- § Deponování (Ingest)
- § Metadata (Metadata)
- § Zpřístupnění (Access)
- § Uložení (Storage)
- § Ochranná opatření (Preservation)

V nejbližší době (začátkem roku 2011) by měl být pro odbornou veřejnost dostupný nástroj umožňující predikci nákladů a výrazně zlepšit schopnost koordinace a plánování a kontrolovat dlouhodobé uchování digitálního dokumentu. Nástroje budou přístupné formou online web aplikace nebo offline verze postavené na excelové tabulce. Po kompletním dokončení fáze III se již počítá se čtvrtou fází tohoto projektu.

Z tohoto důvodu i díky velké propracovanosti a relevantnosti pro problémy NK je metodika projektu LIFE navržena pro další využití v NK.

Podrobnější popis projektů viz text v příloze.

### *Vyplývající doporučení z projektů využitelná v NK*

- Cíleně držet počet využívaných formátů pro dlouhodobé uchování a zpřístupnění na minimálním možném množství.
- Využívat doporučené standardy pro deponování a výrobu metadat, které byly ověřeny v praxi.
- Kooperace subjektů zabývajících se získáváním, vytvářením, dlouhodobým uchováním a šířením, společná výměna znalostí a zkušeností.

### *Návrh dalšího postupu prací na rok 2011*

V prostředí Národní knihovny České republiky by bylo možné implementovat následující metodiky vycházející z představených projektů:

- Za použití doporučení z projektu DiCoMo začít schraňovat detailních informace o provedených digitalizovaných zakázkách a provádět následnou evaluaci těchto cíleně nashromážděných dat v reálném procesu digitalizace. Následné vytvoření modifikátorů a koeficientů sloužících k objektivnímu propočtu nákladů procesu digitalizace.
- Pilotní nasazení offline, popřípadě online nástrojů projektu LIFE.
- Využívat postupy a doporučení, které již byly již prověřeny v praxi a navazují na aktivity týkající se dlouhodobého uchování a zpřístupnění v prostředí knihoven. Je možné vyjít z projektu BRTF-SDPA, který tyto postupy a doporučení předkládá.

## **Oblast 2 - priorita 3 : Vývoj podpůrných/návazných aplikací pro chod LTP systému (2010-2012)**

Během analýz se již v roce 2009 ukázalo, že systémy pro správu dlouhodobého úložiště (LTP) pracují s metadatami ve svém interním formátu. Pro příjem dokumentů do archivu (ingest) a pro distribuci nebo zpřístupňování dokumentů ven z archivu (delivery) bude vždy třeba metadata konvertovat. Konverzi pro zpřístupnění zajišťuje samotný LTP systém do běžných standardů jako např. METS apod. Lze také nastavit vlastní formáty, např. FOXML. Problematický je ovšem vstup dokumentů, kde je nutné vytvořit transformační aplikaci, která dokáže automatickou cestou převést data z metadatových formátů NK (DTD Kramerius) do interního formátu LTP systému. NK musí po té, co bude jasné, jaký z archivních systémů bude využit, vytvořit řadu aplikací, které budou tyto konverze zajišťovat. Především půjde o konverzi metadat z nového workflow digitalizace NDK projektu, dále o konverzi starých dat vzniklých v minulých letech a také o konverzi nebo vytváření metadat pro externí dokumenty.

Vytváření aplikace i mapování pro převod metadat a dat z nové digitalizace bude mít na starosti systémový integrátor projektu NDK a bude hrazeno z projektu samotného.

Ovšem mapování ze stávajících metadatových formátů NK, konkrétně DTD monografie a DTD periodika z programu Kramerius, je odpovědnost a „výhradní“ znalost NK. Zároveň toto mapování bude nutné aplikovat i na data ostatních institucí, které dodávají do NK (resp. LTP systému) své výstupy, např. z projektu VISK7.

Podstatné je, aby bylo možné přenést stará data z digitalizace probíhající v NK do roku 2010 do nového systému (tj. změna struktury dat, metadat apod.). Jedná se o 8 milionů stran dokumentů různých generací skenování, s metadaty v několika generacích DTD popisu.

Řešení mělo tedy v roce 2010 dvě části – viz následující podčásti 1 a 2.

## **Podčást 1 – mapování DTD monografie a DTD periodika do interních formátů LTP systémů**

### **Řešení v roce 2010**

Vzhledem k blížícímu se výběrovému řízení na jednotlivé části systému NDK již od roku 2009 probíhaly analýzy komerčně dostupných LTP systémů. V letech 2008-2009 pracovníci NK viděli oba systémy, které připadají pro NK do úvahy v chodu v různých institucích, kde v některých proběhla dosti podrobná analýza funkcionalit. Ukázalo se ovšem, že nejlepší pochopení funkčnosti a návazností systémů bude možné získat pouze reálných testováním na skutečných datech, která má NK ve svém archivu. Zároveň jsme chtěli otestovat, jak je NK a její zaměstnanci připravena na provozování podobných systémů. Testovány byly systémy Rosetta od firmy Ex Libris (<http://www.exlibrisgroup.com/category/RosettaOverview>) a systém SDB od firmy Tessella (<http://www.digital-preservation.com/solution/safety-deposit-box>). Poznatky z testování byly využity do přípravy podkladů pro tendr NDK systému a budou využity i pro návrh nových metadatových a datových standardů pro knihovny v ČR a jejich projekty. Počítalo se i s tím, že mapování bude použito pro postupný převod archivních dat ještě před započítím NDK projektu, není to ovšem možné z toho důvodu, že stále nevíme, který z LTP systémů bude vybrán.

Pro testování LTP systémů, které probíhalo jako Proof of Concept (dále POC) od prosince 2009 do července 2010 *bylo nutné vytvořit alespoň základní mapování ze stávajících metadatových formátů do formátů LTP systémů.*

Cílem bylo na speciálním vzorku stávajících dat NK zkusit následující procesy – převod metadat do interního formátu LTP; vložení dat do LTP systému; procesy dlouhodobé ochrany. Důležité taky bylo posoudit vůbec vhodnost našich dat na uložení v LTP systému, tj. jak náročné bude mapování, zda bude možné v maximální míře, tj. zda nedojde ke ztrátě údajů apod.

V rámci POC vznikly 2 java aplikace na převod našich stávajících metadat do interních formátů obou systémů. Základ těchto aplikací i mapování bude využit v projektu NDK při převodu stávajících dat z archivu NK do LTP systému.

### **Mapování pro systém Rosetta**

#### **Formát DNX**

K mapování pro systém Rosetta nebyla k dispozici specifikace metadatového formátu DNX, který systém používá. Specifikace není veřejná. Mapování proto probíhalo ve vývojovém prostředí SDK (java aplikace), kde je celý DNX formát vyjádřen. Formát DNX je založen na formátu METS, jsou do něj ovšem přidána další pole a struktury firmou Ex Libris. Cílem bylo zjednodušení celkového data modelu. Data model systému Rosetta je záměrně celkem omezen, filosofii firmy je, že jde o LTP systém na archivaci dat a ne o content management systém.



Jako formát pro popisná metadata se pro správu používá Dublin Core. Dále jsou použity standardní součásti METS formátu, jako AmdSec (techMD, RightsMD, SourceMD, DigiProvMD), FileSec a další. Hlavní rozdílem je vložení DNX formátu do všech částí AmdSec, které namísto standardních formátů jako je PREMIS, MIX apod. obsahuje jejich zjednodušenou verzi, která z těchto formátů vychází a je vyjádřena v DNX. Tj. mapování muselo být vytvořeno pro popisná metadata do Dublin Core, pro technická a administrativní metadata obsažená v DTD periodika a monografie do formátu DNX a METS a to celé se muselo zabalit pomocí submission aplikace do jednoho záznamu XML.

### Datový model a mapování

Základní jednotkou je IE (intelektuální entita), která může mít sdílená metadata a metadata linkovaná zvenku (např. z katalogu knihovny). Intelektuální entitu si lze zvolit, může to být článek, číslo periodika, svazek monografie apod. Z toho pak vychází další mapování a funkcionalita systému. Ve všech známých implementacích systému Rosetta je základní intelektuální entitou číslo periodika a svazek monografie. Tato skutečnost vychází z procesu digitalizace, kde ve většině knihoven (Finsko, Norsko, Holandsko, Knihovna Kongresu, Austrálie) se digitalizuje na „úroveň“ čísel. Z toho vyplývá, že číslo/svazek musí obsahovat popisná metadata vyšší úrovně (IE), tak jak ji známe např. z Krameria – tj. ročník, titul.

Intelektuální entita záměrně v systému Rosetta nemůže obsahovat jiné IE (logické IE neexistují), pouze objekty tvořící onu základní entitu a jejich reprezentace. Tj. např. IE čísla obsahuje objekty stránek, která má každá např. tyto reprezentace – master obrázků v JPG; OCR soubor v TXT, náhled apod.

Z této premisy, muselo vycházet i mapování DTD periodika a monografie do formátu DNX. Tj. DTD monografie a periodika bylo nutné přemapovat tak, aby výstupem z vyvíjené submission aplikace byl DNX XML soubor popisující číslo časopisu a/nebo svazek monografie – popisná metadata, technická a administrativní a logická i fyzická strukturální mapa. Bylo nutné analyzovat strukturu obou DTD, která jsou vytvořena pro číslo zakázky, které v případě periodik je většinou ročník, který obsahuje vnitřní části (čísla, přílohy apod.). V případě monografií jednosvazkových je DTD pro jeden svazek, v případě vícesvazkových pro všechny svazky, tj. bylo nutné provést stejnou analýzu jako pro periodika a vytvořit výstup DNX tak, aby 1 DNX záznam se rovnal jednomu svazku monografie. Bylo tedy nutné popisná metadata vrchní úrovně (ročník, titul monografie) logicky a správně připojit k jednotlivým číslům/svazkům. DNX zároveň musí obsahovat popisná metadata, která byla v původním DTD výlučnou součástí popisu vnitřních částí (PeriodicalVolume, PeriodicalItem, MonographUnit, MonographPage atd.) jako např. pořadové číslo čísla periodika, datum vydání čísla, údaje o ročníku, číslo stránky (fyzické i logické), název kapitol apod. Typy jednotlivých entit z DTD (např. typ stránky – úvodní, normální strana, titulní strana, obsah aj.) byly namapovány přímo do DNX formátu.

Kromě polí z obou DTD jsme museli vzít do úvahy doprovodná administrativní a technická metadata, která jsou v současné době (od roku 2008) ke každému DTD vytvářena. Jedná se o metadata ve formátu PREMIS, MIX a METSRights v souborech [.amd] ve složkách [jpg\_amd], [txt\_amd]. Obsahují údaje o formátech, výstupy z PRONOMu, údaje o SW a HW digitalizace aj. Z těchto metadata náležitých ke každému DTD bylo nutno provést mapování polí, která jsou nutná a potřebná pro vstup do LTP systému. Na základě analýzy se ukázalo, že většina těchto polí není na vstupu potřeba, protože si je LTP systém vytvoří znovu sám (platí také pro systém SDB). Jedná se převážně o technická metadata, která na vstupu systém vytváří tak, aby na nich mohl později postavit další procesy, na

kteří je designován. Z údajů, které je nutné zachovat, jmenujme identifikátory (čísla zakázek aj.), původní názvy souborů (Rosetta soubory na vstupu přejmenovává), údaje o HW skeneru a použitém SW a nastavení skeneru a kontrolní součty, které je systém Rosetta schopen porovnat a případně zjistit poškození souboru způsobené přenosem dat. Zcela jsme v rámci POC vypustili údaje o autorských právech a oprávněních k přístupu. Údaje o oprávněních přístupu by neměly být uvedeny v metadatech, ale v nastavení systému LTP. Údaje o autorských právech budou řešeny ve finálním mapování, současný údaj obsažený v metadatech není relevantní.

Příklad výstupu ve formátu DNX viz příloha.

### *Mapování pro systém SDB (Safety Deposit Box)*

#### Formát XIP

K mapování byla k dispozici specifikace formátu XIP, která je dostupná pro zákazníky a testování u potenciačních zákazníků. Tj. mapování probíhalo klasicky spojením relevantních polí v obou formátech a následným vyjádřením těchto vazeb v java aplikaci (jako jeden krok ingest workflow). Převod dat pro systém SDB neprobíhá v samostatné aplikaci, ale jako jeden z kroků naspecifikovaného workflow – viz níže popis aplikace. Formát XIP zrcadlí jiný pohled na využívání LTP systému, jak ho vidí firma Tessella, která vidí LTP systém jako obohacený CMS systém o funkcionality dlouhodobé ochrany. Po analýzách se ukázalo, jde o formát (data model) nesmírně flexibilní, který umožňuje neomezenou integraci jakýchkoli metadatových schémat a neomezenou granularitu, včetně tvorby logických entit (sbírky). Otázkou zůstává, zda je to výhoda nebo v konečném důsledku nevýhoda. Podobně jako u formátu DNX vychází z běžných standardů (METS, PREMIS, MIX apod.), které jsou uvnitř formátu XIP upraveny/zjednodušeny.

#### Datový model a mapování

XIP je dobře zdokumentován a plně popsán XSD schématem, což umožňuje použít standardní Java nástroj (např. JAXB) pro jeho zpracování. Důsledkem komplexnosti a flexibility datového modelu je náročnější implementace. Hlavně dobře rozmyslet a domyslet všechny důsledky určitého rozhodnutí o struktuře a použitém datovém modelu.

Z hlediska uživatele/správce archivu je nejdůležitější jednotkou datového/logického modelu SDB tzv. Deliverable unit (dále DU). DU obsahuje manifestace, přes které je spojena s digitálními soubory. DU mohou být navzájem spojeny pomocí vazeb „is related to“ nebo „is part of“. DU podle definice odpovídá v modelu OAIS tzv. Content information, tj. tomu, co lze komunitě uživatelů předložit jako DIP. Jednotlivé DU jsou spojeny s Collections, což už jsou zcela abstraktní entity, sbírky.

Collections a DU lze libovolně vkládat do sebe, tedy collection může obsahovat jakékoli další collections a ty zase jakékoliv DU, obsahující další DU. DU ovšem nemůže obsahovat žádnou collection. Na každé úrovni je možné připojit metadata podle potřeby, včetně metadata ve specifickém formátu, tzv. generic metadata. Všechna schémata metadat jsou v systému zaregistrována a provádí se validace. Mezi entitami collection a DU mohou existovat další vztahy, a DU a collections, mohou mít další části označené jako compoments.

Z formátu XIP jsme použili následující části:

- § Collections (collection, department – identifikace oddělení odpovědného za data - and series - RecordSeriesRef)

- § Aggregation (Accumulation, Accession)
- § IngestedFileSets
- § DeliverableUnits (identifikace DU, jejich manifestací a souvisejících DU)
- § Files (veškeré informace o vkládaných souborech – identifikátory, příslušnost k FilesSetu, rozlišení souboru/složky, generická/specifická metadata podle našeho schématu (technická metadata, metadata událostí, informace o vložených bitstreamech aj.)

Ostatní části formátu XIP jsou využívány až při archivních procesech a nejsou pro převod metadat do interního formátu určeného ke vstupu do systému relevantní.

Strukturu jsme se rozhodli udělat stejnou, jako v případě systému Rosetta, tj. hlavní IE číslo/svazek. Tyto entity byly stanoveny jako tzv. deliverable units. V případě SDB je možné mít logické IE jako collections, čehož jsme využili pro titul a ročník.

Popisná metadata:

- na úrovni collections a deliverable unit - některá metadata jsou namapována přímo do struktury XIPu, jinak je pro každý titul vytvořeno na příslušné úrovni DC (monographTitle, monographChapter, periodicalTitle) kam jsou naše metadata z DTD kompletně přemapována, ostatní popisná metadata (číslo výtisku, číslování stránek apod., názvy kapitol apod.) jsou namapována přímo do elementů v XIPu v rámci elementů Deliverable Unit.

Technická metadata:

- veškerá technická metadata jsme namapovali na úroveň files, kde jsou částečně ve struktuře XIPu a vedle toho ještě ve vloženém MIXu, případně PREMISobject formátu, které byly použity v té podobě, ve které jsou připojeny ke stávajícím DTD. Byly použity opět jen relevantní části.

Příklad výstupu ve formátu XIP viz příloha.

## Podčást 2 – tvorba aplikací pro převod starých dat do interních formátů LTP systému

### Řešení v roce 2010

Na základě vytvořeného mapování vznikly v průběhu ledna až července 2010 dvě aplikace, schopné automaticky vzít stávající archivní data a metadata, aplikovat pravidla mapování, vytvořit balíček pro LTP systém v jeho vnitřním formátu včetně dat, zavolat ingest modul LTP systému a poslat balíček dat i metadat do systému samotného.

Aplikace vznikaly ve spolupráci s Ing. Přemyslem Špičkou z firmy Datacons. Bylo nutné dělat časté schůzky ohledně správného mapování a funkcionalit aplikací. Velmi důležité byly i přímé konzultace s oběma firmami (ExLibris, Tessella). Každá z firem poskytla úvodní školení a 2-3 denní workshop na vytvoření aplikací, resp. návod, jak aplikace vytvořit a zapojit ke konkrétnímu LTP systému.

Důvod, proč vznikly aplikace dvě, je nasnadě, každý LTP systém používá jiný formát metadat a jinou filosofii uložení i vstupu do systému. Náročné bylo rozmyslet jak vyjádřit strukturu dokumentů v těchto formátech a podle toho aplikace připravit.

### ***Vstupní aplikace (submission application) pro LTP systém Rosetta***

V rámci „Proof of concept“ (POC) byla vyvinuta aplikace pro konverzi metadat digitalizovaných monografií a periodik ve verzi 2.1 (<http://digit.nkp.cz/techstandards.html>) do formátu podporovaného systémem Rosetta dodávaného firmou Exlibris.

Aplikace je implementována v jazyce Java, k vytvoření xml souboru (a celého SIP balíčku) pro Rosettu využívá vývojový nástroj DPS-SDK od firmy Exlibris. DPS-SDK poskytuje reprezentaci xml v jazyce Java a umožňuje serializaci této reprezentace do xml. Dále je DPS-SDK využito k volání webových služeb systému Rosetta. DPS-SDK usnadnilo a zrychlilo vývoj aplikace, která vytváří xml soubor (a celý SIP balíček) pro Rosettu.

Submission application pracuje ve dvou krocích, v prvním vytvořila z našeho DTD interní formát pro Rosettu (DNX) a data přebalila do struktury pro Rosettu a vystavila na NFS Rosetty. Ve druhém volala aplikace deposit API Rosetty, spustila material flow na účet nějakého producenta pro data z určité lokace.

Aplikace mapovala metadata jak pro periodika, tak pro monografie. Ukládání OCR souborů proběhlo v Rosettě formou další reprezentace IE, vedle obrazových souborů je tak ještě OCR v TXT.

### ***Rozšíření vstupního workflow LTP systému SDB***

V rámci POC bylo vyvinuto rozšíření aplikace SDB (Secure Deposit Box) od firmy Tessella umožňující ingest monografií a periodik popsaných XSD ve verzi 2.1.

SDB je Java serverová aplikace, která se skládá ze dvou základních částí. Z uživatelského rozhraní SDB Web Application a výpočetní části JobQueue. Komunikace mezi těmito částmi probíhá pomocí webových služeb a každá z nich může být instalována na samostatný server, případně serverový klastr.

Funkčnost je v SDB realizována pomocí Drools flow, které se skládá z jednotlivých kroků Workflow Steps. Jedním z cílů POC bylo vytvořit nové Workflow pro ingest monografií a periodik. Rozšíření je realizováno Jar (Java Archive) souborem, který se nahrává na příslušný server a je následně odkazován v odpovídajícím Workflow, které toto rozšíření využívá. Drools flow je tedy abstraktní popis procesu, který je realizován Java kódem. Kód je s Workflow provázán pomocí jmenné konvence. Tento model je velmi flexibilní a umožňuje snadné, transparentní a plnohodnotné rozšíření. Workflow lze spouštět manuálně, časovačem, na základě změny file systému nebo vzdáleně přes webové služby.

### ***Návrh dalšího postupu prací na roky 2011-2012***

Prvotní mapování a vývoj aplikace v roce 2010 bylo cílené na testování LTP systémů a přípravu základů mapování pro finální převod stávajících dat do LTP systémů.

*Původně plánovaný vývoj aplikace pro převod starých dat bude financován z projektu NDK v jehož rámci bude naplněn původní plán Priority 3 Oblasti 2. Bude samozřejmě využít buď jeden, nebo druhý základ aplikací popsaných výše, včetně základu mapování vzniklého řešení v roce 2010.*

Z výše uvedených důvodů bychom finance plánované pro Prioritu 3 na roky 2011-2012 rádi využili jiným relevantním způsobem.

*Navrhujeme zacílit Oblast 2 priority 3 na pokračování výzkumu v oblasti jednoznačných identifikátorů digitálních objektů.*

Problematikou, která je s velkými objemy digitálních dat nedílně spojena, je jejich jednoznačná identifikace - ideálně globálně jednoznačným a persistentním identifikátorem. V digitálním světě je z mnoha důvodů (vyhledávání, citace, manipulace s daty, sdružování, dlouhodobá ochrana aj.) velmi důležité být schopen jednoznačně označit a posléze identifikovat digitální objekt nebo logické kombinace objektů. Zvláště pokud instituce uchovává miliony digitálních dokumentů. Životně důležité jsou identifikátory pro různé agregátory dat, nebo služby vyhledávání z více zdrojů dat/metadat apod.

V roce 2010 vznikla v rámci končícího Výzkumného záměru MK00002322102 (<http://www.isvav.cz/researchPlanDetail.do?rowId=MK00002322102>) v NK koncepcí pilotního využití identifikátoru URN:NBN pro data uložená v archivu NK. Na základě této koncepce vznikl pilotní provoz tzv. resolveru URN:NBN, který bude procházet během 1Q 2011 testováním. Jedná se pouze o pilotní provoz se základní funkcionalitou. URN:NBN je v současné době nejrozšířenějším systémem v Evropě na úrovni národních institucí (archiv, knihovna) a je využíván např. v Norsku, Finsku, Švédsku, Maďarsku, Německu, Rakousku, Švýcarsku atp. Systém a vyvinutý SW pro správu, přidělování a resolvování identifikátorů by měl být základem národní služby pro URN:NBN, jak je tomu obvyklé v okolních státech (Německo).

Cílem v roce 2010 bylo připravit prostředí a základní SW aplikaci pro pilotní test systému pro využití URN:NBN v NK ČR. Aplikace umožňuje přiřazování globálně jedinečného identifikátoru odpovídajícího pravidlům URN:NBN, dále jeho správu (administrátorský modul), vyhledávání dle identifikátoru. V pilotním provozu je logika pilotního řešení zatím omezena na dokumenty, které jsou registrovány v jednom balíčku jako zakázka v systému RD.CZ, a to pouze ty z nich, které jsou již reálně zdigitalizovány a budou trvale archivovány v digitálním repozitáři Národní knihovny ČR.

Zároveň umožňuje rovněž vyhledávat identifikované digitální dokumenty, a to nejen podle identifikátoru URN:NBN podle kterého vyhledává primárně, ale také podle dalších užívaných identifikátorů, a sice ISSN, ISBN a čČNB (číslo České národní bibliografie). Po vyhledání dostane uživatel relevantní metadata k vyhledávanému dokumentu, vidí, zda a kým byl zdigitalizován a kde je zpřístupněn, včetně URL linku do konkrétní digitální knihovny.

*Již nyní víme, že některé premisy koncepce nebyly kompletní a celá koncepce i aplikace budou potřebovat další výzkum a vývoj tak, aby poskytovala další služby žádané nejen interně v NK a v projektu NDK, ale i v ostatních paměťových institucích, které o využívání resolveru také velmi stojí pro svá digitální data.*

Návrh dalšího vývoje resolveru URN:NBN pro rok 2011

- návrh optimalizace životního cyklu digitálních dokumentů s ohledem na identifikaci pomocí URN:NBN

- analýza provázanosti aplikace resolver s workflow digitalizace, Alephem NKC, RD.cz a novými systémy, které vzniknou v rámci projektu NDK
  - následné úpravy aplikace resolver
- netechnické věci související se strategií přidělování a odvíjející se od aktuálního workflow digitalizace je nutné ještě dořešit
- rozšíření současné funkcionality pilotního provozu Resolveru URN:NBN
  - možnost získání dat z ostatních aplikací (např. URL dokumentu z digitálních knihoven, URN:NBN přidělené ve workflow digitalizace apod.) – možná přes XML API nebo přes webservice pro volání služeb resolveru
  - hromadné exporty (vystavení) identifikátorů pro další aplikace (např. katalog NKC apod.)

## Oblast č. 5: Vývoj nových metod konzervace novodobých knihovních dokumentů

### *Navržené úkoly a vyhodnocení jejich plnění - rok 2010:*

1. Rešerše odborné literatury
2. Vývoj metodiky průzkumu
3. Typologie skladovacích prostor NK ČR
4. Analýza požadavků na funkcionalitu znalostní báze
5. Zahájení průzkumu

#### **Ad 1.) Rešerše odborné literatury**

Byla provedena rešerše odborné české i mezinárodní literatury. Novodobý (bohemikální) knihovní fond (dále NF) zahrnuje dokumenty vzniklé v rozmezí více než dvou set let, počínaje rokem 1801. Zachycuje vývoj české kultury a národní svěbytnosti, má neocenitelný historický, umělecký, společenský význam a zásadní informační hodnotu. NF je ohrožen účinky degradačních faktorů vnitřních (rozpad dřevitého papíru) a vnějších (např. nevhodné uložení v minulosti). Sbírký opatrované v tzv. paměťových institucích (knihovny s konzervační povinností, archivy, muzea) jsou proto v alarmujícím stavu. Řada dokumentů je poškozena, v přibližně 30 % případů kriticky. Bez rychlého řešení reálně hrozí, že tištěné kulturní dědictví bude nenávratně ztraceno. U nás ani v zahraničí nebyl dosud proveden zásadní průzkum cílený speciálně na NF a nebyla řešena konzervace a restaurování novodobých dokumentů. Pro NF nejsou vytvořeny a rozpracovány komplexní restaurátorské postupy – restaurování je vztahováno z drtivé většiny na historické dokumenty, tj. rukopisy a tisky vyrobené do r. 1800. Není nám známo, že by se někdo v České republice nebo v zahraničí systematicky věnoval této problematice. Pro restaurátorské zásahy na NF nejsou dostatečně vytýčeny etické a estetické normy. Metody a technologie užívané pro historické sbírky lze na dokumenty NF aplikovat pouze částečně, vzhledem k odlišné chemické podstatě, různorodosti materiálů i provedení a k velkému rozsahu NF.

Postupy záchrany u NF se sestávají z optimalizace úložných podmínek (vhodné mikroklimatické prostředí, prevence proti biologickému napadení, pravidelná očista) a zásahů konzervačního charakteru: knihvazačské práce, dezinfekce, případně odkyselení – u nás ojediněle, a ukládání do ochranných obalů.

#### **Ad 2.) Vývoj metodiky průzkumu**

Změnou legislativy po r. 1989 byl podstatně redukován počet povinných výtisků uchovávaných knihovnamí. Z původní 22 exemplářů se v současné době uchovávají 4 povinné výtisky (2x v NK ČR, 1x v MZK a VKOL) a regionální povinný výtisk v příslušné regionální knihovně. Zatímco původně knihovny spravující konzervační sbírky budované získanými povinnými výtisky, nebyly vázány povinností tyto sbírky trvale uchovávat, je nyní tato povinnost daná zákonem.

Redukce povinných výtisků přináší samozřejmě vyšší rizika pro trvalé dochování konzervačních sbírek v maximální úplnosti. To je navíc akcentováno nízkou mírou využití mikrofilmu

jako nástroje trvalého dochování ohrožených dokumentů v minulých desetiletích v ČR. Digitalizace, která by měla tuto roli převzít, je v počátcích a významný přínos lze očekávat v horizontu nejméně deseti let. Dalším faktorem negativně ovlivňujícím míru a stav dochování konzervačních sbírek je skutečnost, že tři ze čtyř povinných výtisků se využívají v běžných knihovních službách, v jednom případě dokonce absenčních. To samozřejmě způsobuje vyšší míru ztrát a poškození konzervačních exemplářů. I když lze očekávat, že v příštích desetiletích se bude výrazně transformovat role knihoven díky uplatnění elektronických médií, zůstane význam knih, uchovaných v našich konzervačních sbírkách jako artefaktů kulturního dědictví i jako nástroje autentifikace obsahu převedeného do elektronické podoby, stejně důležitý, příp. se ještě zvýší.

Zákonný požadavek na trvalé uchování konzervačních sbírek klade daleko vyšší nároky na příslušné knihovny a jejich aktivity v oblasti ochrany a správy těchto dokumentů, než tomu bylo v minulosti. Účinná péče cílená na trvalé uchování musí vycházet z podrobného zmapování stavu dochování konzervačních dokumentů. Proto bude stěžejním bodem našich aktivit nejprve podrobný průzkum fondů, jehož výsledkem bude zjištění opravdového a skutečného fyzického stavu jednotlivých exemplářů. Z výsledků takového podrobného průzkumu pak bude možné vyhodnocení typů poškození, diagnostika jejich příčin a návrhy možných řešení (opravy, konzervátorské a restaurátorské zásahy na novodobém fondu, uložení do ochranných obalů, reformátování).

Národní kulturní dědictví z oblasti knižní kultury obsahuje cca 1,5 mil. dokumentů různých typů. Toto množství bude vyžadovat uplatnění některých hromadných technologií (např. hromadné odkyselování) a koordinaci konzervačních aktivit uvedených knihoven. Informace z průzkumu by měly být dostupné odborníkům v různých institucích, proto bude součástí řešení vytvoření znalostní báze shromažďující co nejvíce informací o našich konzervačních sbírkách. Tyto informace by měly být dostupné také prostřednictvím katalogů knihoven, což umožní využití některých zavedených identifikátorů (např. číslo ČNB). Postupy a nástroje vyvinuté v NK ČR budou moci uplatňovat i další konzervační knihovny.

Průzkum novodobých knihovních fondů 19. a 20. století vznikl na základě nutnosti zjištění faktického fyzického stavu novodobých fondů v NK ČR. Vzhledem k tomu, že běžné formuláře a postupy pro restaurátorské průzkumy historických fondů není úplně možné aplikovat na tento typ fondů a z důvodu velké časové náročnosti těchto průzkumů, bylo nutné, aby vznikla přehledná a srozumitelná aplikace, do které se zjištěné údaje zaznamenávají.

Z aplikace bude možné získat seznamy knih, které mají různé typy poškození, nebo např. mají hodnotu pH nižší než 7; nevhodné ochranné obaly; mají mechanické, biologické či chemické poškození; které mají koženou, plátěnou, papírovou vazbu a v jakém je stavu; které již z důvodu fyzického stavu, nejsou vhodné pro výpůjčku, apod.

Výsledkem tohoto průzkumu budou přesné a statisticky zpracovatelné údaje o tom, které knihy:

- a) potřebují nutný restaurátorský či konzervátorský zásah (jsou mechanicky, biologicky či chemicky poškozené)
- b) jaký typ papíru obsahují a jaké je jeho pH
- c) potřebují ochranný obal
- d) mají přílohu, a jakou
- e) přesné hodnoty, v jakém uložení se knihy nacházejí
- f) doporučujeme reformátovat



Vývoj a upřesnění softwarových nástrojů pro vytvoření elektronického formuláře a databáze pro průzkum novodobých fondů bude pokračovat i v roce 2011 v závislosti na zjištěných nejasnostech a nefunkčnostech prováděním zkušební testování formuláře. Ukázka elektronického formuláře a jednotlivé údaje potřebné pro průzkum fondů jsou uvedeny v Příloze č. 1 a č. 2.

### Ad 3.) Typologie skladovacích prostor NK ČR

Do konce roku 2010 byla zpracována typologie skladovacích prostor, resp. depozitářů NK ČR. V dalších letech bude provedena tzv. pasportizace signatur (kde a kdy byla signatura uložena) a popsáno, kde a v jakém časovém období byly jednotlivé fondy uloženy s ohledem na klimatické podmínky. Podle těchto údajů pak bude možné vysvětlit určitý typ poškození u jednotlivých fondů.

Pan Dr. Šnýdr vytvořil seznam všech depozitářů Národní knihovny ČR v letech 1943-2010, tyto údaje jsou uvedeny v Příloze č. 3. V Příloze č. 4 jsou pro názorný příklad uvedeny klimatické parametry v depozitáři v Neratovicích, které sledují pracovníci Oddělení péče o novodobé fondy a upozorňují na případné výkyvy nebo nejasnosti.

V dalších letech projektu bude popsáno, jak má vypadat ideální depozitář pro vhodné uložení novodobých fondů s ohledem na jejich složení.

### Ad 4.) Analýza požadavků na funkcionalitu znalostní báze

V říjnu 2010 byl zahájen testovací průzkum a z praktického testování vzešly požadavky na úpravy a doplnění informací. Tyto požadavky byly postupně zpracovány do elektronického formuláře a doplněny. V současné době je již připravena verze formuláře, která plně vyhovuje požadavkům průzkumu a funguje v praxi.

Byly zvoleny tyto požadavky na funkcionalitu znalostní báze a elektronického formuláře:

- možnost stáhnout pomocí zadání signatury bibliografické záznamy včetně identifikátorů a údajů o jednotce (autor, název, místo a rok vydání, ČNB, pole 001, SYSNO, ISSN, ISBN, čarový kód, odkaz do digitální knihovny, aj.),
- stahování bude probíhat v dávkách se záznamy seřazenými podle signatury (jako místní seznam),
- údaje o průzkumu budou shromažďovány v elektronických formulářích a budou částečně vygenerovány, částečně vytvářeny (většina volbou z dané nabídky, minimálně zapisováním, popisováním),
- vytvořená data budou uložena v relační databázi a budou zálohována (archivována),
- vizualizace dat by měla umožnit přechíst všechny údaje k danému dokumentu, vyexportovat dávku údajů k dílčí sbírce (např. signatura nebo část, všechny bibliofilie, všechny knihy k jednomu roku, všechny knihy s definovanou hodnotou pH, všechny knihy doporučené k uložení do ochranného obalu apod.) a pak číst údaje k jednotlivým dokumentům,
- každý prováděný průzkum musí být jednoznačně identifikován – názvem, předmětem a časem, časový údaj by se měl generovat na každém formuláři, stejně jako zvolený identifikátor daného objektu.

V databázi budou uloženy následující typy dat:

- výsledky měření (minulých i průběžných), např. ze systému Hanvell a to z různých depozitářů (Klementinum, CDH, RF v Neratovicích), vlákninového složení, hodnoty pH, aj.,
- výsledky průzkumů fondů,
- obrazová fotodokumentace ve formátu JPG, TIFF, BMP,
- odborné statě ve formátu Word, Excel, PDF, RTF, HTML,
- možnost rozšíření na další obrazové nebo textové formáty,
- registr současných i minulých depozitářů s popisem klimatických parametrů v depozitářích,
- průběh sušení nebo ošetření nebo konzervátorských či restaurátorských zásahů.

## Ad 5.) Zahájení průzkumu

V říjnu 2010 byl předložen elektronický formulář pro databázi a průzkum novodobých fondů. Byli přijati 4 pracovníci na provádění průzkumu v Klementinu a 1 pracovník na získávání dalších informací pro databázi – určování vlákninového složení papírů novodobých fondů. Za období říjen až prosinec 2010 bylo prozkoumáno asi 2400 knih a vyplněny jednotlivé formuláře pro knihy. Z těchto výsledků pak bude v roce 2011 možné statisticky zpracovat stav prozkoumané části fondů a zmapovat jednotlivé typy poškození a navrhnout metody konzervace a restaurování a vytvořit cenovou kalkulaci pro jednotlivé konzervátorské a restaurátorské zásahy.

Pro objektivní vyhodnocení výsledků průzkumu bude nutné provést řadu materiálových analýz a analýz zaměřených na zjištění stupně degradace (kromě stupně rozpadu papíru také, např. analýzy použitého textilu, lepidel a barevných součástí knih). K tomu bude nutné využít nové nedestruktivní metody. Díky takto získaným poznatkům bude možné vyvinout materiály a speciální postupy vhodné pro konzervaci a restaurování novodobých dokumentů; tyto aktivity byly v našich podmínkách dlouhodobě podceňovány.

Samotný průzkum fondů byl započat v říjnu roku 2010 po vytvoření elektronického formuláře pro každý exemplář (dále jen „formulář“) a databáze těchto formulářů. Příklad první verze formuláře je uveden v příloze. Ve formuláři jsou shrnuty jak informace bibliografické, tak především informace o fyzickém stavu knihy, dokumentu, fotodokumentace apod. Je třeba uvést, že průzkum provádějí zaškolení pracovníci se vzděláním v oblasti chemické technologie restaurování památek a restaurování papírových materiálů. Byl vypracován metodický manuál pro zaškolení pracovníků pro průzkum, kde jsou podrobně popsány jednotlivé pojmy, vlastnosti papíru, jejich identifikace a vše je doplněno o názorné fotografie. Z toho také vyplývá, že průzkum je prováděn velmi podrobně a profesionálně.

Dalším přínosem je rozšíření formuláře o údaje o chemickém složení, fyzikálních, optických a chemických vlastnostech materiálů. Zatím bylo u několika exemplářů např. provedeno vlákninové složení původních surovin pro výrobu papíru.

Rovněž byl za pomoci finanční podpory z Norských fondů (o grant v rámci 3. výzvy finančních mechanismů EHP/Norsko), využít k analýzám přístroj SurveNIR (bližší informace o metodě jsou uvedeny v Příloze č. 5) - je to nová metoda nedestruktivního průzkumu papírových dokumentů, kdy jsou pomocí naměření spekter v blízké infračervené oblasti světelného spektra, využitím chemometrických metod a srovnáním s databází standardů papírů určeny vlastnosti měřeného

papíru. Byly takto proměřeny některé exempláře v rámci průzkumu, tyto informace jsou uvedeny v Příloze č. 6. Tyto informace budou vloženy do elektronického formuláře pro každý exemplář a budou velmi cenné při vyhodnocování výsledků průzkumu a určování typů poškození, diagnostice jejich příčin a návrhy možných řešení a postupů konzervace a restaurování.

### Závěr:

Z průběžného zjištění stavu jednotlivých knih a dokumentů vyplývá, že jsou papíry velmi poškozené a především jejich hodnota pH je velmi nízká (papíry jsou velmi kyselé). Záchrana takto poškozených a kyselých papírů je velmi finančně i technologicky náročná, ale bez její realizace hrozí, že se tyto dokumenty během následujících několika desítek let stanou zcela nečitelnými a dojde k jejich rozpadu. Žloutnutí a křehnutí papíru a posléze jeho rozpad se stává klíčovým problémem knihoven, jejichž úkolem je uchování tohoto materiálu coby nejrozšířenějšího nosiče informací pro další generace čtenářů a badatelů. K nejvíce ohroženému dnes patří především papír vyráběný od 2. poloviny 19. století ze dřeva s použitím kyselého pryskyřičného klížení. Takto vyrobený papír vykazuje kyselou reakci a tím i nízkou odolnost proti přirozenému stárnutí. Jeho životnost je závislá na urychleném konzervačním zásahu, při kterém je provedena neutralizace kyselin a současně zavedena zásoba alkalické látky chránící papírová vlákna před dalšími kyselinami vznikajícími v budoucnu – tzv. alkalická rezerva. Aby tento konzervační zásah byl smysluplný a prodloužil užitnou životnost papíru, musí být proveden v době, kdy papír ještě vykazuje dobré mechanické vlastnosti.

O hromadném odkyselování archiválií se v českém archivnictví mluví již od začátku osmdesátých let minulého století. Bylo by potřeba vybudovat odkyselovací pracoviště, např. po vzoru polských kolegů z Národní knihovny ve Varšavě a Jagelonské university v Krakově. Budování sítě pracovišť hromadného odkyselování v Polsku bylo součástí rozsáhlého osmiletého vládního programu Kwasny Papier, připravovaného od roku 1998 a zahájeného v roce 2000. Podobný program, zkoumající stav poškození knihovnických sbírek kyselostí papíru a hledající optimální způsob odkyselení, je v současné době realizován i ve Slovenské republice pod názvem Kniha.

Situace v ostatních státních knihovnách je zřejmě obdobná, ale tam průzkumy zatím provedeny nebyly. Záchrana archivního dědictví je tedy více než aktuální. Proto by měl náš formulář také sloužit jako metodická pomůcka pro všechny státní knihovny, které by chtěly započít průzkum svých fondů – měli bychom fungovat jako metodické centrum pro knihovny v České republice. Pracovníci Oddělení péče o novodobé fondy mohou proškolit pracovníky ostatních knihoven, které budou průzkumy fondů provádět.

### Upřesnění úkolů na rok 2011

- Doplnění rešerše o nové poznatky – zejména publikované informace koncem roku 2010 a v roce 2011
- Vývoj aplikace a databáze pro znalostní bázi – úpravy formuláře a databáze vyplývající z testovacího průzkumu v roce 2010
- Průzkum novodobých fondů a vyhodnocování získaných poznatků – bude pokračovat a výsledky budou průběžně vyhodnocovány
- Analýza problematiky plyných polutantů a prachových částic – metody stanovení VOC látek a prachových částic uvolňujících se z knih a dalších materiálů

v depozitářích. S partnery (např. Ústavem chemie a fyziky aerosolů, AV ČR) budou sledovány koncentrace částic a polutantů v depozitářích. Podle těchto parametrů pak mohou být zvolena preventivní opatření při dlouhodobém uložení fondů.

- Průzkum metod odkyselování – pro získání výsledků dosažené těmito metodami budou zvoleny materiály charakterizující novodobé fondy a tyto budou odkyseleny ve světě dostupnými odkyselovacími metodami (hromadnými i individuálními) a u těchto vzorků budou proměřeny jejich vlastnosti (chemické, optické, mechanické) a také budou výsledky porovnány s námi vyvinutou metodou odkyselováním ve vakuových balíčcích.

# Návrh výsledků výzkumu a vývoje k nahlášení do edatabáze RIV

## Publikace

### Výsledek č. 1

FOJTŮ, Andrea. Open source nástroje pro dlouhodobou ochranu digitálních dokumentů. *Archivy, knihovny, muzea v digitálním světě. Praha, NA ČR, 2.12.2010.* [online]. [cit. 2010-12-04]. Dostupné z WWW: <<http://skip.nkp.cz/KeStazeni/Archivy10/den2/Fojtu.pdf>>.

### Výsledek č. 2

VYCHODIL, Bedřich. Úvod do problematiky (skrytých) nákladů na dlouhodobou archivaci. *Archivy, knihovny, muzea v digitálním světě. Praha, NA ČR, 2.12.2010.* [online]. [cit. 2010-12-04]. Dostupné z WWW: <<http://skip.nkp.cz/KeStazeni/Archivy10/den2/Vychodil.pdf>>.

### Výsledek č. 3

OHLÍDALOVÁ, M., VÁVROVÁ, P. Monitorování vlivu světla na sbírkové předměty v expozicích – „světelný životopis“. *MUEUM – Muzejní a vlastivědná práce*, 48, (2), 2010, s. 74 – 82

### Výsledek č. 4

R04: AF

R05: D

R06: Plurality of Similarities

R07: eng

R09: 2010

R10: 1

R11: 1

R13: written and documentary heritage – transcultural research

RX1: U

R27: 978-80-7050-587-8

R29: Praha

R30: REDISCOVER Final Conference Proceedings, Prague, 15 September 2010

R33: 15

R34: Národní knihovna České republiky

R35: 15.09.2010

R55: EUR

R42: Transcultural comparison of Central European medieval and early modern chronicles and other historical works on the base of material that is aggregated in the Manuscriptorium digital library

A02: Klimek

A03: Tomáš

A04: A

A08: OCR

A05: 7904121236

A06: G

N01: Z

## Výsledek č. 5

R04: AF  
R05: D  
R06: Central European University Literature in the Late Middle Ages  
R07: eng  
R09: 2010  
R10: 1  
R11: 1  
R13: written and documentary heritage – transcultural research  
RX1: U  
R27: 978-80-7050-587-8  
R29: Praha  
R30: REDISCOVER Final Conference Proceedings, Prague, 15 September 2010  
R33: 11  
R34: Národní knihovna České republiky  
R35: 15.09.2010  
R55: EUR  
R42: Transcultural comparison of Central European university literature on the base of material that is aggregated in the Manuscriptorium digital library  
A02: Marek  
A03: Jindřich  
A04: A  
A08: OCR  
A05: 7811280895  
A06: G  
N01: Z

## Výsledek č. 6

R04: AF  
R05: D  
R06: The Uniformity and Variability of the Book Culture of Women's medieval Order Institutions in Central European Lands  
R07: eng  
R09: 2010  
R10: 1  
R11: 1  
R13: written and documentary heritage – transcultural research  
RX1: U  
R27: 978-80-7050-587-8  
R29: Praha  
R30: REDISCOVER Final Conference Proceedings, Prague, 15 September 2010  
R33: 20  
R34: Národní knihovna České republiky  
R35: 15.09.2010  
R55: EUR  
R42: Transcultural comparison of Central European Book Culture of Women's medieval Order Institutions on the base of material that is aggregated in the Manuscriptorium digital library  
A02: Modráková  
A03: Renáta

A04: A  
A08: OCR  
A05: 7856021679  
A06: G  
N01: Z

#### Výsledek č. 7

R04: AF  
R05: D  
R06: Central European Preaching in the High and Late Middle Ages and Its Polymorphic Unity  
R07: eng  
R09: 2010  
R10: 1  
R11: 1  
R13: written and documentary heritage – transcultural research  
RX1: U  
R27: 978-80-7050-587-8  
R29: Praha  
R30: REDISCOVER Final Conference Proceedings, Prague, 15 September 2010  
R33: 20  
R34: Národní knihovna České republiky  
R35: 15.09.2010  
R55: EUR  
R42: Transcultural comparison of Central European preaching manuscripts and preaching literature on the base of material that is aggregated in the Manuscriptorium digital library  
A02: Uhlíř  
A03: Zdeněk  
A04: A  
A08: OCR  
A05: 5609140416  
A06: G  
N01: Z

#### Výsledek č. 8

R04: AF  
R05: D  
R06: The Central European Cultural Region in the Middle Ages and Early Modern Region from the Czech Perspective  
R07: eng  
R09: 2010  
R10: 1  
R11: 1  
R13: written and documentary heritage – transcultural research  
RX1: U  
R27: 978-80-7050-584-7  
R29: Praha  
R30: Four Versions of One Culture: A Synthesis of the Dispersed Content of Central European Literature

R33: 25  
R34: Národní knihovna České republiky  
R35: 15.09.2010  
R55: EUR  
R42: Transcultural comparison of national cultures in Central Europe on the base of material that is aggregated in the Manuscriptorium digital library  
A02: Uhlíř  
A03: Zdeněk  
A04: A  
A08: OCR  
A05: 5609140416  
A06: G  
N01: Z

## Aplikované výsledky

### Výsledek č. 1

R04: AF  
R05: Z  
R06: NKCatalog  
R07: cze  
R08: NKCatalog  
R09: 2010  
R11: 1  
RX1: U  
R72: A  
R73: A  
R59: NKCatalog (Build: 3.1.1)  
R36: dostupný z URL: <http://195.113.132.83:8080/match-web/>  
R37: Webová aplikace NKCatalog umožňuje uživateli vyhledat podobné segmenty textu mezi dokumentem uživatele a dokumenty v předem stanoveném repozitáři. Podobnost se hledá na úrovni různě dlouhých segmentů textu na základě statistického vyhodnocení vektorové analýzy. Aktuální verze aplikace, jejíž poloprovoz byl zpřístupněn pro soustavné testování, pracuje s repozitářem, který je součástí digitální knihovny Manuscriptorium a obsahuje plná znění rukopisných textů v datových balíčcích Etno, NKCR, PARK a UJCAV. Jazyky dokumentů jsou pro testovací účely omezeny na latinu, češtinu a němčinu. V rámci měření podobnosti jsou používány grafické varianty pro příslušný jazyk, aby aplikace našla shodné, resp. podobné segmenty i při více méně odlišné grafice konkrétního zápisu. Aplikace je implementována v české a anglické lokalizaci.  
R39: Národní knihovna České republiky  
R40: 00023221  
R41: CZ  
A02: Uhlíř  
A03: Zdeněk  
A04: A  
A08: OCR  
A05: 5609140416  
A06: G  
N01: Z



## Použití finančních prostředků

Poř.č. řádku	Název položky	účet	schválený rozpočet projektu	úprava rozpočtu projektu	úprava rozpočtu projektu	úpravený rozpočet projektu
			1	2	3	4
1	Spotřeba materiálu	(501)	70			70
2	Spotřeba energie	(502)				0
3	Spotřeba ostatních neskladovatelných dodávek	(503)				0
4	Prodané zboží	(504)				0
5	Opravy a udržování	(511)				0
6	Cestovné	(512)	100			100
7	Náklady na reprezentaci	(513)				0
8	Ostatní služby	(518)	510			510
9	Mzdové náklady x) (ř.10+ř.11)	(521)	974			974
10	v tom: platy zaměstnanců x) (z AE k účtu 521)		120			120
11	ostatní osobní náklady (z AE k účtu 521)		854			854
12	Zákonné sociální pojištění	(524)	315			315
13	Jiné sociální pojištění	(525)				0
14	Zákonné sociální náklady - FKSP	(527)	2			2
15	Jiné sociální náklady	(528)				0
16	Ostatní náklady z činnosti	(549)				0
17	Odpisy dlouhodobého nehmotného a hmotného majetku	(551)				0
18	Ostatní finanční náklady	(569)				0
19	Náklady PO - účtová třída 5 celkem		1 971	0	0	1 971
20	Příspěvky a dotace na provoz x)	(671)				0
21	Hospodářský výsledek (ř.20 -ř. 19)		-1 971	0	0	-1 971

