

# Přílohy ke Zprávě o realizaci Koncepte rozvoje Národní knihovny České republiky jako výzkumné organizace za r. 2010

Příloha 1: .....	2
NKCatalog v. poloprovoz - uživatelská příručka .....	2
Příloha 2: .....	8
Lokalizace stránky – implementace algoritmu .....	8
Příloha 3: .....	13
SYSTÉM PRAVIDEL PRO DIGITÁLNÍ ARCHIVACI (Digital preservation policies) .....	13
Příloha 4: .....	21
Open source nástroje pro dlouhodobou ochranu digitálních dat .....	21
Příloha 5: Analýza metodik pro sledování nákladů životního cyklu dokumentu .....	30
Příloha 6: Ukázky XML .....	39
Příloha 7: .....	44
Centrální znalostní báze RD – vypracoval Ing. Kocourek.....	44
Příloha 8: .....	49
Vstupní formulář průzkumu – vypracoval Ing. Kocourek .....	49
Příloha 9: .....	53
Depozitáře Národní knihovny České republiky 1943-2010 – vypracoval Dr. Šnýdr.....	53
Příloha 10: Příklad údajů o sledovaných klimatických podmínkách.....	55
Příloha 11:.....	57
Popis přístroje SurveNIR – vypracovala Ing. Marie Benešová, Ing. Vávrová .....	57
Příloha 12:.....	59
Tabulka naměřených údajů z přístroje SurveNIR – vypracovala Ing. Vávrová, Bc. Křečanová.....	59

# Příloha 1:

## NKCatalog v. poloprovoz - uživatelská příručka

7. prosince 2010

Jirka Hana, Barbora Hladká, Václav Novák

E-mail: [hladka@ufal.mff.cuni.cz](mailto:hladka@ufal.mff.cuni.cz)

Aplikace NKCatalog umožňuje uživateli vyhledat podobné segmenty textu v dokumentu uživatele a v dokumentech z předem stanoveného katalogu. Podobnost se hledá na úrovni různě dlouhých segmentů

textu. Aktuální verze aplikace pracuje s katalogem, který je částí projektu Manuscriptorium a obsahuje rukopisy v datových balíčcích Etno, NKCR, PARK, UJCAV. Jazyky dokumentů jsou čeština, latina a němčina. V rámci měření podobnosti jsou používány morfologické varianty pro příslušný jazyk, aby aplikace našla shodné segmenty i při odlišném pravopisu zápisu. Aplikace je implementována v české a

anglické lokalizaci.

### **Jak spustit aplikaci**

V internetovém prohlížeči zadat adresu <http://195.113.132.83:8080/match-web/>.

### **Hlavní obrazovka**

Hlavní obrazovka aplikace obsahuje tyto části: **prohlížení katalogu** (Zobrazit katalog), **správa dokumentu uživatele** - body 1. a 2., **správa porovnání** - body 3. a 4.

**Zobrazit katalog**

---

1. Nahrání dokumentu, který se má porovnat:

Document:

Zatím nebyl nahrán žádný dokument.

---

2. Analýza dokumentu:

Jazyk dokumentu:

Dokument nebyl zatím zanalyzován.

---

3. Porovnání dokumentu s dokumenty v katalogu:

Maximální počet výsledků:

---

4. Zobrazení výsledků:

Obrázek 1 Úvodní obrazovka

### Prohlížení katalogu

Tlačítkem Zobrazit katalog se zobrazí obsah katalogu. Pro každý dokument katalogu je zobrazen jeho název a údaje o rukopisu (autor, název, jazyk, počet slov). Políčka s nadpisy Id, Autor a Název umožňují vyhledávat dokumenty (rukopisy) z katalogu dle zadaných kritérií.

Zpět					
1 2					
Id	Autor	Název	Jazyk	Počet slov	
54_C_152_1543A_1T0W_TEI.xml		Artikulové sněmu království českého 1543	cs	8747	<input type="button" value="p"/>
54_C_152_1564__0KWI_TEI.xml		Artikulové sněmu království českého 1564	cs	2796	<input type="button" value="p"/>
54_C_152_1575A_1NRK_TEI.xml		Artikulové sněmu království českého 1575	cs	3402	<input type="button" value="p"/>
54_C_152_1582__1PTN_TEI.xml		Artikulové sněmu království českého 1582	cs	12933	<input type="button" value="p"/>
54_C_152_1627__29FL_TEI.xml		Artikulové sněmu království českého 1627	cs	12532	<input type="button" value="p"/>
A1.xml	Peregrinus de Opole	Sermones de tempore	latin	260	<input type="button" value="p"/>
A2.xml	Iohannes Neplach	Chronicon	latin	1116	<input type="button" value="p"/>
A2a.xml	Iohannes Neplach	Chronicon	latin	1116	<input type="button" value="p"/>
A2b.xml	Iohannes Neplach	Chronicon	latin	1116	<input type="button" value="p"/>
Marignola.xml	Iohannes de Marignolis	Chronicon Bohemorum	latin	74164	<input type="button" value="p"/>

Obrázek 2 Prohlížení katalogu

Obsah dokumentu se zobrazí kliknutím na symbol lupy na konci řádku.

The screenshot shows a web application interface. On the left is a table with columns 'Id' and 'Zpět'. The table lists several XML files. The right side is a window titled 'Zobrazit dokument' showing the content of a selected document. Below the document content is a table with columns 'Jazyk' and 'Počet slov'.

Id	Zpět
54_C_152_1543A_1T0W_TEI.xml	
54_C_152_1564__0KWJ_TEI.xml	
54_C_152_1575A_1NRK_TEI.xml	
54_C_152_1582__1PTN_TEI.xml	
54_C_152_1627__29FL_TEI.xml	
A1.xml	
A2.xml	
A2a.xml	
A2b.xml	
Manignola.xml	

**Zobrazit dokument**

A1.xml  
Peregrinus de Opole: Sermones de tempore

Peregrinus de Opole  
Sermones de tempore

1. Dominica prima adventus

Ecce Rex tuus veniet tibi etc. Cfr. Za 9, 9: Exsulta satis, filia Sion; juba, filia Jerusalem: ecce rex tuus veniet tibi justus, et salvator: ipse pauper, et ascendens super asinam et super pullum filium asinae; Mt 21, 5 Dicite filiae Sion: Ecce rex tuus veniet tibi mansuetus, sedens super asinam, et pullum filium subjugalis.

Item aliqui errabant sicut oves sine pastore et viam, que ad celum ducebat, invenire non poterant. Item aliqui erant captivi et qui eos redimeret, non habebant. Et ideo istis modis afflicti clamaverunt in celum, ut deus eis mitteret aliquem, qui eos ab excommunicatione absolveret et ad viam reduceret et qui eos de captivitate redimeret. Deus autem, qui est pius in omnes, qui invocant eum in veritate, audivit vocem lamentabilem eorum et respondit eis hodie in evangelio dicens: Popule meus, habete bene, quia Rex tuus veniet tibi, Mt 21, 5 Dicite filiae Sion: Ecce rex tuus veniet tibi mansuetus, sedens super asinam, et pullum filium subjugalis. ubi tria dicit: suam potentiam in hoc, quod dicit rex, qui scilicet est ita potens, quod te bene liberare potest; secundo sua propinquitas in hoc, quod dicit veniet, et ideo cum honore et sine mora et cum debita preparatione recipi debet; tercio eius utilitas cum dicit tibi, id est ad utilitatem tuam. Sed videamus hanc utilitatem. Certe in hoc notatur eius utilitas, quia venit tamquam pius mediator, tamquam bonus pastor, tamquam fidelis fideiussor, tamquam benignus medicus, tamquam fortis pugil, tamquam sapiens advocatus.

Hic notes dicere de Cronica Romanorum et cetera.

Jazyk	Počet slov
	8747
	2796
	3402
	12933
	12532
	260
	1116
	1116
	1116
	74164

Obrázek 3 Obsah dokumentu z katalogu

### Správa dokumentu uživatele

V bodu 1. se specifikuje dokument uživatele, jehož podobnost s dokumenty v katalogu bude měřena a vizualizována. Úložiště dokumentu se zadává přes tlačítko Procházet, následně se provede Nahrát.

Aplikace podporuje dokumenty těchto formátů: XML a txt (plain text).

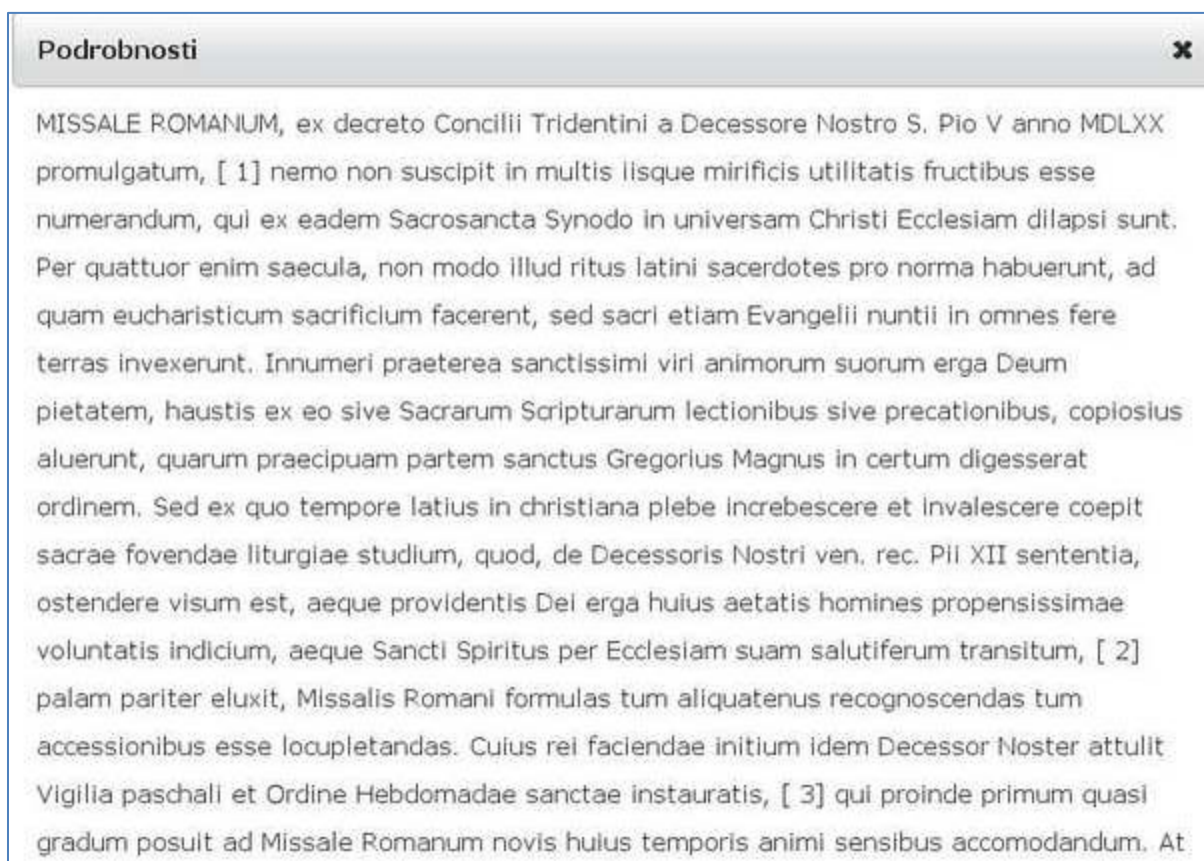
The screenshot shows a web application interface for document management. It is divided into four numbered steps:

- 1. Nahrání dokumentu, který se má porovnat:**  
Document:     
Dokument byl úspěšně nahrán. (userFile\_3969812981916925849)
- 2. Analýza dokumentu:**  
Jazyk dokumentu:     
Dokument nebyl zatím zanalyzován.
- 3. Porovnání dokumentu s dokumenty v katalogu:**  
Maximální počet výsledků:
- 4. Zobrazení výsledků:**

Obrázek 4 *Určení dokumentu uživatele*

V bodu 2. uživatel vybere jazyk jeho dokumentu a tlačítkem Analýza spustí analýzu dokumentu.

Následně je možné zobrazit obsah dokumentu tlačítkem Zobrazit můj dokument.



Obrázek 5 *Obsah dokumentu uživatele*

### **Správa pro porovnání**

Aplikace hledá podobné segmenty textu mezi dokumenty uživatele a dokumenty katalogu. Uživatel má

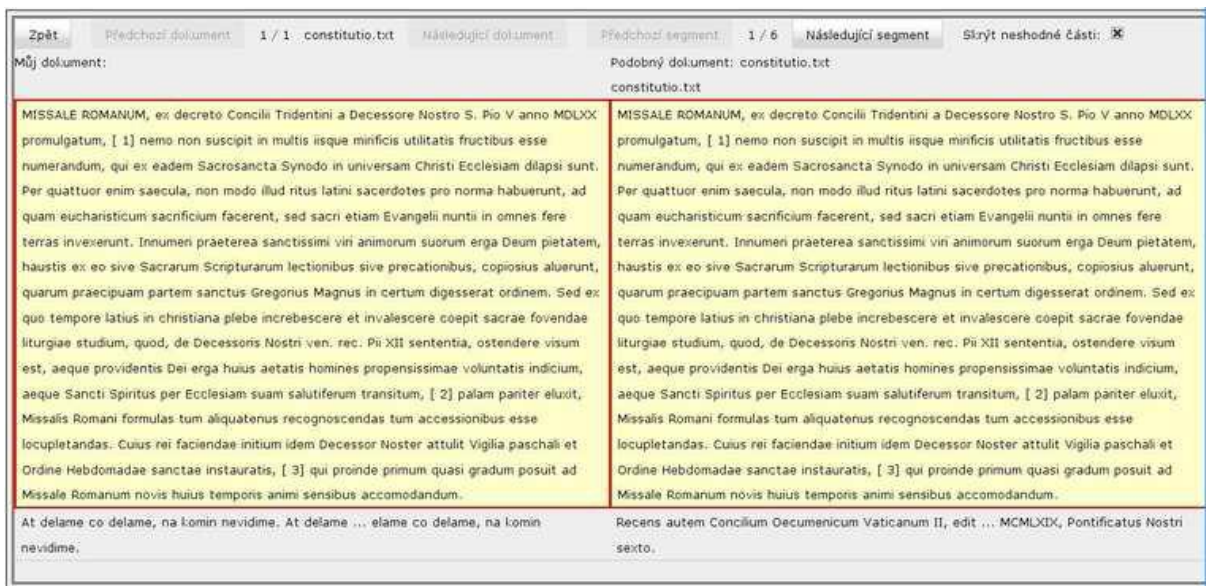
volnost v určení, jaký maximální počet shodných segmentů má být zobrazen. Nastavení tohoto počtu a spuštění porovnávací procedury (Porovnat) se provádí v bodu 3.

V bodu 4. jsou souhrnně prezentovány výsledky porovnávací procedury. Výpis zobrazuje sestupně dle kvantitativní míry podobnosti (tzv. skóre) ty dokumenty z katalogu, ve kterých byly nalezeny podobné úseky s úseky v dokumentu uživatele. Pokud uživatele zajímají pouze dokumenty, které překračují jistý





části se zobrazují pouze podobné (tj. žluté) segmenty.



Obrázek 8 Zobrazení výhradně podobných segmentů

## Příloha 2:

# Lokalizace stránky – implementace algoritmu

Cílem je lokalizovat na obrázku stránku knihy jak je to znázorněno níže na obrázku.



## Pozorování

- Na všech dodaných obrázcích mají okraje tmavou barvu
- Levou a pravou stranu lze rozlišit podle toho, zda je tmavý okraj nalevo nebo napravo v obrázku
- Některé stránky jsou oskenované dobře, není třeba stránku lokalizovat
- Všechny stránky jsou správně natočené (Vyjimečně stránky bývají natočené o 1.0 - 1.5°)
- Často se na oskenovaných obrázcích objevuje i část stránky sousední. Ta ve většině případů má větší výšku než stránka, kterou chceme lokalizovat. (Viz obrázek níže, kde vlevo sousední stránka sahá ke krajům obrázku, ale stránka, kterou chceme lokalizovat má nahoře a dole tmavé okraje)

## Popis algoritmu

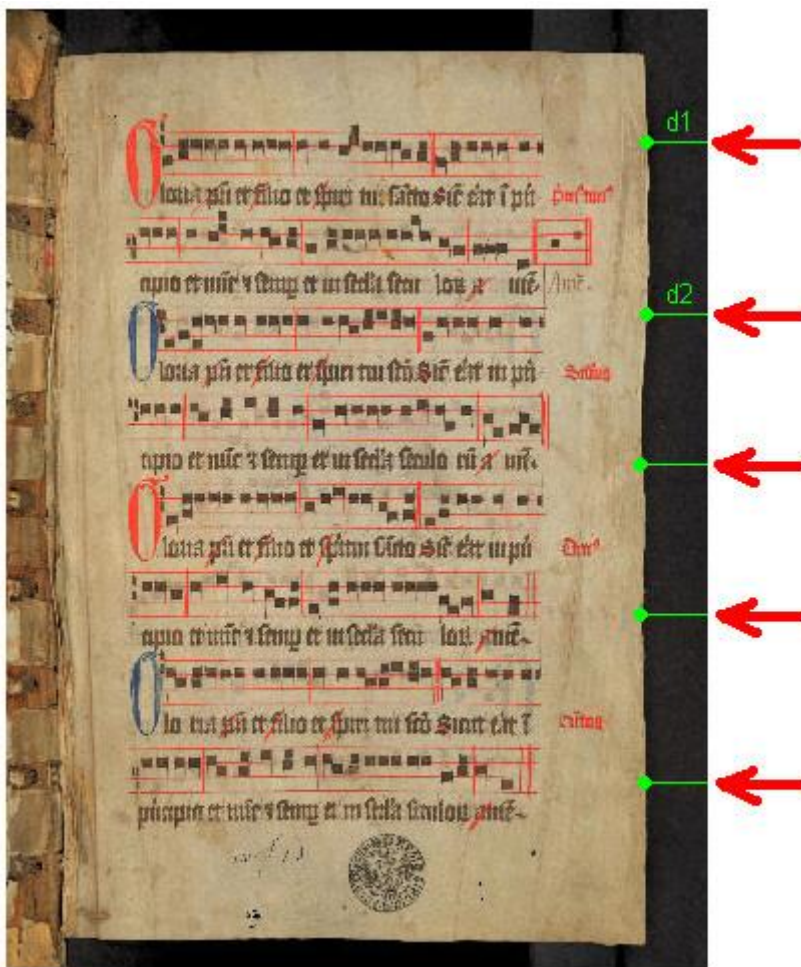
**VSTUP:** obrázek

**VÝSTUP:** dva body, které určují obdelník stránky

1) Porovnáním průměrného jasu úzkých oblastí (o šířce 5x) na krajích obrázku rozlišíme, jestli se jedná o stránku levou nebo pravou

2) Pokud je stránka pravá: Na pravé straně obrázku se zvolí několik rovnoměrně rozprostřených bodů, které se budou posouvat směrem doleva dokud narazí na světlé místo. Tím se určí vzdálenosti, které body urazily  $d_1$ ,  $d_2$ , ... Spočítá se jejich průměr  $d$ .





Totéž provedeme shora i zdola. Odříznou se oblasti znázorněny níže:



V těchto odříznutých oblastech provedeme podobný postup. Na levé straně zvolíme několik rovnoměrně rozprostřených bodů a posouváme je doprava dokud narazí na tmavé místo. Tím se určí vzdálenosti, které body urazily  $d_1$ ,  $d_2$ , ... Spočítá se jejich průměr  $d$ . Pokud některý z posouvajících se bodů se posune do prava příliš daleko, do průměru se nezapočítává.



3) Pokud je stránka levá: provedeme krok 2) s tím, že pravé zaměníme za levé a levé za pravé.

4) Dva body, které určují obdelník stránky vypočteme ze zjištěných průměrných vzdáleností od okrajů obrázku.

## Poznámka

Jestliže na stránce nejsou tmavé okraje, průměrné vzdálenosti od okrajů budou nulové. V takovém případě lokalizovat stránku není potřeba, stránka je na celém obrázku.

## Možná selhání algoritmu

- Jestliže okraje obrázku nemají tmavou barvu a přesto nepatří do stránky (**z pozorování vyloučeno**)
- Jestliže stránka je hodně natočená a nelze ji popsat obdelníkem (**z pozorování vyloučeno**)
- Jestliže stránky jsou z části utržené - může dojít k chybnému určení bodu obdelníku
- Jestliže je oskenovaná i část sousední stránky a má stejnou výšku jako stránka, kterou chceme lokalizovat

## Testování

Na 250 obrázcích z různých knih jsem označil, kde by se stránka měla správně lokalizovat.

### Způsob výpočtu chyby



- Zeleně je znázorněn obdelník, který určuje správnou lokalizaci stránky
- Červeně je znázorněn vypočtený obdelník

Zjistíme vzdálenosti vypočtených bodů obdelníku od bodů označených manuálně. Tyto vzdálenosti označíme  $e_1$  a  $e_2$ .

Chybu určíme jako průměr těchto dvou hodnot, tedy:  $(e_1 + e_2)/2$ .

Je-li chyba vyšší než 7 px, je lokalizace stránky označena jako chybná.

Testování jsem provedl na 250 obrázcích z 5 různých knih. Dosažená úspěšnost byla **94.8 %**. Průměrná chyba byla **4.46 px**.

### Několik náhodně vybraných výsledků

Vlevo je originální obrázek s vyznačeným obdelníkem správné lokalizace, a vpravo je stránka, kterou se podařilo lokalizovat algoritmem





A dalších několik zpráv na: <http://cmp.felk.cvut.cz/~chaluv1/>



## Příloha 3:

# SYSTÉM PRAVIDEL PRO DIGITÁLNÍ ARCHIVACI (Digital preservation policies)

PhDr. Ladislav Cubr

(Úvodní přehledová zpráva pro interní výzkumný záměr Národní knihovny ČR)

### *Úvodem*

Tato zpráva představuje vstupní přehled výsledku analýzy systému pravidel pro digitální archivaci (*digital preservation policies*) vybraných významných zahraničních paměťových institucí převážně anglosaské provenience.

Dlouhodobá ochrana digitálních dokumentů neboli digitální archivace (*digital preservation*) je komplexní činnost skládající se z řady specifických dílčích aktivit, které musejí být vykonávány ve vzájemné součinnosti. Jednou z těchto aktivit je plánování ochrany. 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. Plánovací entita 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 evaluace obsahu repozitáře, vytváření doporučení pro institucionální systém pravidel a zavádění standardů nebo monitorování změn v technologiích a designované komunitě, která je klientem repozitáře. Mezi další funkce plánovací entity patří například návrh šablon pro informační balíčky, vytváření plánů pro migrace, vývoj softwarových prototypů nebo testování plánů na implementaci migračních cílů.

### *Plánování ochrany a systémy pravidel pro digitální archivaci*

Zde se zaměříme na jeden z úkolů plánovací entity digitálního repozitáře, tedy vývoj systému pravidel pro digitální archivaci.<sup>1</sup> Podle řady odborných studií věnovaných problematice dlouhodobé ochrany digitálních dokumentů je systém pravidel pro digitální

---

<sup>1</sup> Anglický termín „policies“ překládáme jako „systém pravidel“. Jiné překlady, jako například „politiky“ nebo prostě „pravidla“ se nám zdají být méně vhodné, neboť nereflektují komplexitu termínu „policies“.

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 zaznamenán v příslušných institucionálních dokumentech, které jsou volně přístupné na webových stránkách těchto institucí.

V tomto místě je potřeba zpřesnění vzhledem k odkazu na funkční model OAIS. 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.).

Bližšími informacemi o tom, jakým způsobem má plánovací entita vyvíjet systém pravidel v součinnosti s managementem instituce, se norma OAIS nezmiňuje, proto je potřeba se inspirovat praxí významných zahraničních paměťových institucí, do jejichž dílčích povinností spadá také provozování digitálního repozitáře. Pro tento účel jsme provedli 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, které 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).

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 celoinstitucionálních 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. Uvedeme si je v několika bodech.

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

#### *Terminologie*

Většina pravidel obsahuje také vysvětlení základních pojmů. Tato část zdaleka není banální. Již jen samotná definice digitální archivace determinuje ochranná opatření instituce. Zatímco většina definic se shoduje v tom, že digitální archivace znamená nejen ochranu původních bitů, ale zejména jejich zpřístupnitelnosti (důraz na přístup k dokumentům navzdory změnám technologií, např. změnám formátů), ne vždy jsou shodné časové determinace (trvale / tak dlouho, jak to bude nutné / po vymezenou dobu). Z toho je zřejmé, že závazky se mohou v jednotlivých případech značně lišit. Podobně je potřeba vymežit, co se míní termínem digitální dokument, digitální objekt apod., tedy tím, co je předmětem digitální

archivace. Začlenění slovníčku pojmů tedy rozhodně není nadbytečnou činností, ale naopak pomáhá předejít řadě nedopatření.

### *Stakeholder*

Pro systém pravidel pro digitální archivaci je klíčové spojení s pojmem „stakeholder“. Stakeholder představuje takové jednotlivce nebo organizace (např. řešitele projektu nebo veřejnost), kteří se aktivně účastní nějakého projektu nebo „jejichž zájmy mohou být, ať již pozitivně nebo negativně, ovlivněny výsledky nebo dokončením projektu“ nebo kteří „mohou ovlivňovat projekt a jeho výstupy“ [Project Management Institute, 2004, s. 376].<sup>2</sup> Stakeholder je tedy de facto právě adresátem systému pravidel dané instituce. Jinými slovy, systém pravidel nějaké instituce je informací pro stakeholdera o tom, jakým způsobem, do jaké míry a za jakých podmínek instituce naplňuje nebo ovlivňuje jeho zájmy. Každá instituce musí vědět a jasně v dokumentové podobě vymežit (tj. uvést v systému pravidel), kdo je (případně není) jejím stakeholderem a komu poskytuje svoje služby, resp. kdo je na jejích aktivitách oprávněně zainteresován. Identifikace stakeholdera není jen nutností z hlediska legislativních konsekvencí, ale také pro potřeby finanční rozvahy nebo sociokulturního konsensu (záchrana ohroženého kulturního dědictví apod.).

### *Výběr dokumentů*

Většina systémů konkrétně vypisuje, jaké typy dokumentů se zavazuje dlouhodobě archivovat. Tyto soupisy často obsahují také informace o tom, jaké typy dokumentů instituce dlouhodobě archivovat nebude. Častou informací je také přiznání faktu, že všechny digitální dokumenty reálně archivovat je nemožný úkol.

Pokud pomineme legislativní nařízení, která v daných zemích určité instituce zavazují k archivaci jasně definovaných dokumentů (např. elektronický výtisk týkající se dokumentů na nosičích CD /DVD apod.), otázka výběru je dle většiny pravidel na institucích samotných. Obvykle se instituce zavazují archivovat digitalizované vlastní sbírky. Častý je též závazek

---

<sup>2</sup> Project Management Institute. 2004. A guide to the project management body

of knowledge. Third Edition. Newtown Township (USA) : Project Management

Institute, 2004. An American National Standard ANSI/PMI 99-001-2004.

ISBN 978-1-930699-45-8.

k archivaci webu, zpravidla na základě podrobněji stanovených kritérií. Některé univerzitní knihovny se – vedle archivace závěrečných vysokoškolských prací v digitální podobě – zavazují též k archivaci výukových materiálů, včetně těch multimediálních. Většina institucí naopak uvádí, že nearchivuje digitální publikace, ke kterým má pouze zakoupeny licence na vzdálené zpřístupňování prostřednictvím komerčních poskytovatelů dat. Kanadské knihovny a archivy zmiňují v otázce výběru budoucí možnost začlenění nových forem digitální komunikace. Někdy se setkáme s prohlášením, že u digitálních dokumentů, u nichž nelze dohledat vlastníka duševních práv, se instituce zřiká nutnosti jej archivovat.

### *Akvizice*

Řada institucí deklaruje, že je schopna archivovat dokumenty dodané pouze ve vybraných formátech. Knihovna Yaleovy Univerzity uvádí podrobný seznam akceptovaných a užívaných formátů. Australská národní knihovna a některé další instituce explicitně vyjadřují potřebu spojit systém pravidel pro digitální archivaci se systémem pravidel pro budování knihovního fondu. Objevuje se častý názor, že již při akvizici digitálních dokumentů je potřeba zvážit samotné perspektivy jejich možné ochrany (finanční náročnost, technická uskutečnitelnost archivace), resp. deponovat dokumenty pouze v takové podobě (neproprietární formát apod.), která nebude komplikovat budoucí ochranná opatření).

### *Finanční konsekvence*

Řada pravidel explicitně zmiňuje finanční náročnost závazku digitální archivace a skutečnost, že nejistota v kontinuitě financování může způsobit nevratné škody. Některá pravidla oznamují, že v případě nedostatku finančních zdrojů bude muset instituce prioritizovat ochranu pouze vybraných dokumentů a některé zmiňují i způsob výběru (ačkoliv jen obecně) při takovéto prioritizaci. Australská národní knihovna například deklaruje, že výběr těchto dokumentů bude proveden nejen na základě jejich relativní důležitosti (logický požadavek), ale také současné míry technické komplexity jejich archivace. Jinými slovy, systémy pravidel obsahují i oznámení o tom, že za určitých finančních okolností budou změněny cíle archivace, tedy že obecně deklarované cíle jsou pouze relativní, nikoliv absolutní.

Z hlediska finanční náročnosti digitální archivace také některé dokumenty zdůrazňují, že instituce se budou podílet na národních a mezinárodních výzkumných a kooperativních aktivitách, které pomohou snížit lokální náklady na digitální archivaci sdílením odborných znalostí a výstupu výzkumných a dalších projektů. Knihovnický svět je ostatně sdílením

svých výstupů, včetně standardů pro metadata, znám dobře, často se však v dokumentech objevuje důraz na rozšíření archivačních kooperací o další paměťové instituce (archivy, muzea).

### *Plán nástupnictví*

Pravidla pro digitální repozitář spravovaný konsorciem OCLC obsahují oznámení o tom, že v případě, že OCLC nebude moci dále provozovat tuto službu, vydají data depozitorům na vybraných datových nosičích a v aktuálně užívaném formátu. Je potřeba zdůraznit, že tento plán se vyskytuje ojediněle, patrně proto, že OCLC není přímo placena ze státních rozpočtů, ale z rozpočtů jednotlivých organizací.

### *Standardizace*

Obvyklou součástí institucionálních dokumentů pro digitální archivaci je seznamem užívaných nebo plánovaných standardů pro digitální archivaci, zejména pro metadata, digitální repozitáře a formáty.

### *Deklarované služby*

Řada institucí podrobně vypisuje seznam ochranných opatření, která se zavazuje vykonávat na archivovaných digitálních dokumentech. Některé instituce deklarují, že v současnosti deklarují jen určitá ochranná opatření. OCLC například deklaruje pouze bitovou ochranu (uchovávání dokumentů v původních verzích prostřednictvím duplikace) a v budoucnosti formátovou (formátové migrace).

### *Ochrana práv duševního vlastnictví*

Zahraníční pravidla berou vesměs zvláštní obezřetnost z hlediska toho, že archivující instituce musí zajistit nejen ochranu digitálních dokumentů, ale ochranu vztahu mezi digitálními objekty a vlastníky literárních nebo vědeckých, které tato data reprezentují. Vzhledem k snadnosti kopírování digitálních informací a jisté opožděnosti legislativy v této oblasti (která je typická téměř pro celý svět) mají vlastníci digitálních práv větší obavy o ochranu svých práv než kdykoli předtím a dokumenty námi analyzovaných institucí tento fakt velmi obezřetně reflektují v jasných prohlášeních v závazcích archivovaná data chránit i před neautorizovaným kopírováním. Ačkoliv to dokumenty explicitně nevyjadřují, je zde zřejmá snaha zmenšit obavy vlastníků copyrightu před internetovým pirátstvím a dalšími

nekalými praktikami ohrožujícími jejich práva. Mezi ochrannými opatřeními jsou často zmiňovány zabezpečovací a kryptografické mechanismy.

### *Kontextuální souvislosti*

Řada institucí uvádí svoje pravidla pro dlouhodobou ochranu digitálních dokumentů jako součástí širších pravidel. Australská národní knihovna například činí pravidla pro archivaci součástí širšího systému pravidel pro budování knihovního fondu. Australské národní muzeum má dokument, obsahující vzájemně provázaná pravidla jak pro digitalizaci, tak pro digitální archivaci (výběr z hlediska archivace vhodných metadat a formátů již při digitalizaci). Knihovna Yaleovy univerzity zdůrazňuje nutnost proaktivního přístup ke správě digitálních dat, neboť podle ní ochrana závisí na správných rozhodnutích v různých fázích digitálního životního cyklu dokumentu. Australská národní knihovna se v dokumentu také odvolává na nutnost explicitně zveřejnit požadavky na další výzkum v oblasti digitální archivace tak, aby se případní nezainteresovaní odborníci z australského vědeckého prostředí (případně i mezinárodního kontextu) dozvěděli blíže o problému a mohli se případně rozhodnout pomoci vlastními výzkumnými aktivitami.

### *Propagace*

Dokumenty národních knihoven (zejména australské) vyjadřují snahu o propagaci problematiky digitální archivace, včetně seznamování širšího uživatelského prostředí se všemi riziky, kterým je vystavena dlouhodobá ochrana digitálních dokumentů s tím, že tato rizika nejsou ještě dostatečně známa mezi širší veřejností, která cíle digitální archivace může považovat za snadněji dosažitelné, než je skutečnost, a která není plně seznámena s tím, že digitální svět není bezproblémový, pokud se na něj díváme optikou trvalé udržitelnosti.

### **Závěr**

Uvedené charakteristiky je třeba dále rozpracovat do systému pravidel pro digitální archivaci, který bude sepsán a schválen v kontextu Národní knihovny ČR a chystané Národní digitální knihovny. Je potřeba explicitně vyjádřit všechny 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.





## **Příloha 4:**

# **Open source nástroje pro dlouhodobou ochranu digitálních dat**

Mgr. Andrea Fojtu

**E-mail** • [andrea.fojtu@nkp.cz](mailto:andrea.fojtu@nkp.cz)

**Afiliace** • Odbor digitální ochrany (ODO)  
Národní knihovna ČR (<http://www.nkp.cz/>)

Praha, říjen - prosinec 2010

## Přehled open source nástrojů pro dlouhodobou ochranu

### **Fedora a její nadstavby:**

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

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

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

### **další nástroje:**

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

ARCHIVEMATICA (Unesco + Kanada) - <http://archivematica.org/>

### **nástroje pro plánování dlouhodobé ochrany:**

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

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

## HOPPLA (<http://www.ifs.tuwien.ac.at/dp/hoppla/>)

**Garant:** Department of Software Technology and Interactive Systems, Vienna University of Technology  
(<http://www.ifs.tuwien.ac.at/>)

**Instalace:** Windows XP, Linux, Mac (jednoduchý instalační postup)

**Požadavky:** Java Runtime Environment 1.6, doporučená instalace OpenOffice

**Poslední verze:** 2.0 (27.9.2010)

**Plány pro další verze:** vývoj automatického mechanismu pro řešení chyb, pokrytí různorodých formátů v malých institucích (ve větších je tendence k standardizaci), migrační nástroj PS2PDF

**Kontakt:** Michael Greifeneder, Stephan Strodl, Petar Petrov and Andreas Rauber {greifeneder, strodl, petrov, rauber}@ifs.tuwien.ac.at

**Implementace pro NK ČR:** nedoporučuje se z důvodu zaměření produktu na jinou cílovou uživatelskou skupinu

**Hoppla (Home and Office Painless Persistent Long-term Archiving)** představuje vhodné řešení pro širokou veřejnost, malé instituce a tzv. SOHO<sup>3</sup>. Open source software typu Fedora a DSpace jsou určeny pro velké instituce, které mají postačující finanční i lidské zdroje na dodatečné úpravy (z angl. customization). Proto největší výhodou Hoppla je fakt, že pro aplikaci řešení dlouhodobé ochrany digitálních dat není nutné sledovat nejnovější trendy v oblasti formátů, migračních nástrojů apod.

Funguje na principu antivirových programů s on-line balíčky updatů. Nezanedbatelný je i fakt, že je možné jej instalovat pro všechny nejrozšířenější operační systémy. Ve své metodologii dlouhodobé ochrany digitálních dat preferuje metodu migrací formátu v kombinaci se zálohováním (např. emulace, ochrana technologie<sup>4</sup>).

Systém je postaven na referenčním modelu OAIS a modulární architektuře. Umožněna je tak integrace nových či výměnu stávajících přídatných zásuvných modulů (plug-in) a externích služeb.

Metadata vztahující se k objektům ve sbírkách jsou důležité pro dodatečné vyhledávání a dlouhodobou ochranu. Hoppla sbírá dostupná metadata ze zdrojových systémů a doplňuje je o vyextrahovaná metadata z objektů (modul "Metadata Repository"). Aby systém mohl nabízet optimální ochranná pravidla a nástroje, je zaslán profil sbírky do webové služby s pravidelnými updaty (do modulu "Update service"). Uživatel má možnost vybrat jaká eventuálně citlivá data chce na server webové služby odeslat. Na základě přijatého profilu sbírky jsou vybrány migrační nástroje a

---

<sup>3</sup> Představuje akronym anglického Small Office /Home Office, tedy malé nebo domácí kanceláře.

<sup>4</sup> Podrobnější informace k metodám dlouhodobé ochrany např.: HENDLEY, T. Comparison of Methods & Costs of Digital Preservation. British Library Research and Innovation Report 106. 1998 [cit. 2010-10-10]. Dostupné i z WWW: <[http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html#\\_Toc422714272](http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html#_Toc422714272)>.

nejvhodnější postup. Po zmigrování dat systém provede kontroly migračních akcí a archivní část systému vytvoří z původně migrovaného objektu archivní verzi (tzv. verzování objektů). Modul uchovávání (“Storage Management”) spravuje mnohočetné zálohované objekty na různých médiích.

Systém má ale i ve své druhé verzi problémy, které musí nutně dořešit - především vyšší chybovost při migracích (pracuje se na vývoji automatického mechanismu pro řešení chyb) a podporu nejrůznějších formátů v malých institucích (ve větších institucích je tendence k standardizaci).

**Možnost implementace pro národní knihovny se nejeví jako příliš reálná, jelikož je Hoppla určena pro jinou cílovou skupinu a nepočítá s “náročnou” ochranou digitálních historických pramenů.**

**MOPSEUS** (<http://194.177.192.14/mopseus/>)

**Garant:** Digital Curation Unit (<http://www.dcu.gr>)

**Instalace:** Windows XP/ Vista (jednoduchý instalační postup )

**Požadavky:** Java Development Kit 6

**Poslední verze:** 1.0.3

**Plány pro další verze:** průvodce pracovních postupů (tzv. workflow wizards) pro snadnější definici, plánování a provedení aktivit pro správu objektů; uživatelsky přívětivější rozhraní; API jako náhrada za interní migrační nástroj (napojení na xterní služby typu PLATO)

**Kontakt:** Dimitris Gavrilis, Stavros Angelis {d.gavrilis, s.angelis}@dcu.gr, Christos Papatheodorou papatheodor@ionio.gr

**Implementace pro NK ČR:** nedoporučuje se z důvodu zaměření produktu na jinou cílovou uživatelskou skupinu a jeho “nemodulární” architekturu

**Mopseus** reprezentuje digitální repozitář s robustním datovým modelem (namapovaným na datový model PREMIS). Jeho strategie v oblasti dlouhodobé ochrany digitálních dat je namířena na menší a středně velké typy institucí s omezenými finančními prostředky.

Mopseus se při vývoji inspiroval referenčním modelem OAIS, ve kterém reprezentace informací tvoří digitální objekty a jsou nositelem klíčových informací pro správu a dlouhodobou ochranu digitálních sbírek. Tedy každý digitální objekt obsahuje datový stream a vazby.

Datové streamy nejsou součástí METS, pro popis objektů využívá pouze metadatové schéma nekvalifikovaného Dublin Core. Navíc nenabízí žádný interní migrační nástroj pro migraci (z důvodu omezených finančních prostředků na vývoj) institucionálních formátů souborů. Jeho základ tvoří Fedora a není nutné další programování a uživatelské úpravy.

**Výše uvedené nedostatky do podstatné míry ztěžují alespoň částečnou integraci Mopseus do systémů NK ČR. Cílovou skupinou jsou totiž malé až středně velké instituce.**

## **RODA** (<http://roda.di.uminho.pt/#home>) & **CRiB**

**Garant:** Portuguese National Archives (Directorate General of the Portuguese Archives), University of Minho (<http://antt.dgarq.gov.pt/>, <http://www.uminho.pt/>)

**Instalace:** GNU/Linux systém - ověřeno na Ubuntu 8.04 a Ubuntu 8.10

**Požadavky:** doporučená instalace OpenOffice

**Poslední verze:** 29.07.2009

**Plány pro další verze:** projekt byl ukončen (1.4.2006 - 31.1.2008)

**Kontakt:** José Carlos Ramalho [jcr@di.uminho.pt](mailto:jcr@di.uminho.pt), Miguel Ferreira [mferreira@dsi.uminho.pt](mailto:mferreira@dsi.uminho.pt)

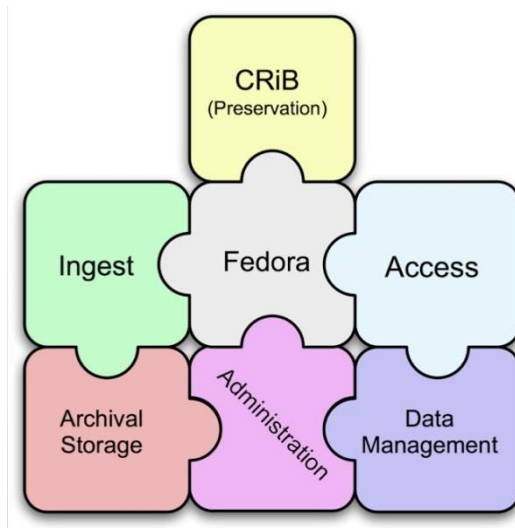
**Implementace pro NK ČR:** jde o technické řešení na úrovni národního archivu, nicméně díky servisně orientované architektuře je možné jej (především modul CRiB) integrovat do různých systémů NK ČR (např. do Transformačního modulu či Modulu plánování dlouhodobé ochrany budoucí NDK)

**RODA** – **Repositório de Objectos Digitais Autênticos** nebo-li Repozitář autentických digitálních objektů představuje technické řešení na úrovni národního archivu. Na rozdíl od ostatních hodnocených systémů, které se zaměřují pouze na některou z oblastí problematiky dlouhodobé ochrany, RODA je komplexní systém správy digitálních objektů - od vkládání až po jejich zpřístupnění.

Jádro systému, obdobně jako u nástroje Mopseus (viz podrobnější popis výše), formuje Fedora Commons. Servisně orientovaná architektura, tedy jakýsi “balíčkový systém” interoperabilních služeb je možné využít v různých systémech. Proto má jeho integrace do systémů NK ČR reálný základ. RODA vychází z referenčního modelu OAIS, pro popisná metadata využívá schéma EAD. Digitální objekty jako soubor binárního obsahu, vztahů a reprezentací, jsou strukturovány v METS.

**CRiB** (Conversion and Recommendation of Digital Object Formats), který původně vznikl jako samostatný migrační nástroj, je integrován do RODA v rámci preservačního modulu. Zabezpečuje identifikaci formátů, doporučuje optimální verze pro migraci, provádí (distribuované) migrace objektů do nových, vhodnějších formátů a následně kontroluje jejich integritu. Po úspěšném vyhodnocení migračního procesu generuje ochranná metadata v PREMIS.

**I když se jedná o technické řešení vyvíjené na úrovni národního archivu, 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 NDK).**



**Obr. 1** Architektura systému RODA

## ARCHIVEMATICA (<http://archivematica.org/>)

**Garant:** Artefactual Systems (<http://artefactual.com/>), finanční garant - UNESCO Memory of the World Subcommittee on Technology

**Instalace:** Windows XP, Linux, Mac

**Požadavky:** Java Runtime Environment 1.6, doporučená instalace OpenOffice

**Poslední verze:** **0.6-alpha** nebo **0.6.2-alpha** (pouze pro vývojáře a technicky zdatné uživatele)

**Plány pro další verze:** ostré testování současné beta verze v partnerských institucích, které ukáže další směřování produktu

**Kontakt:** Peter Van Garderen [administrator@archivematica.org](mailto:administrator@archivematica.org)

**Implementace pro NK ČR:** doporučuje se implementace téměř v úplném rozsahu - jednotlivé nikroslužby je možné připojit do systémů NK ČR (např. monitorPreservation)

**Archivematica** je open source systém vyvíjený jako sada 24 “mikroslužeb” (z angl. microservices) postavených kolem souborového systému, rozdělených do 9 základních kategorií: 1. receiveSIP, 2. reviewSIP, 3. quarantineSIP, 4. appraiseSIP, 5. prepareAIP, 6. reviewAIP, 7. storeAIP, 8. provideDIP, 9. monitorPreservation. Celá sada pak vyhovuje referenčnímu modelu OAIS<sup>5</sup>.

<sup>5</sup> <http://archivematica.org/wiki/images/d/dc/Archivematica-architecture-7May2010-2.png>



Beta verze pro implementaci do partnerských institucí je naplánována na začátek roku 2011. Technicky zdatnější uživatelé (jde o virtual appliance -upravený Ubuntu Linux s open source nástroji) můžou systém testovat již dnes.

Cílem systému postavených na “mikroslužbách” je redukce technické komplexnosti a údržby, zjednodušení vývoje a možnost snadného nahrazení zastaralých prvků systému.

**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” - Archivematica uchovává originální formát všech importovaných dokumentů pro migraci a emulaci, přičemž normalizace souborů pro dlouhodobou ochranu je v otevřených formátech.**

**AIDA** (<http://aida.jiscinvolve.org/wp/>)

**Garant:** University of London Computer Centre, financováno z 1/07 JISC Capital Programme Call

**Instalace:** nejde o softwarovou aplikaci

**Požadavky:** nejsou známy žádné specifické požadavky, nicméně AIDA je cílen především na anglo-saské akademické instituce

**Poslední verze:** květen 2009

**Plány pro další verze:** projekt byl ukončen (1.10.2007 - 31.3.2009); v dalším samostatném projektu je snaha o vytvoření praktického nástroje, který pro jednotlivá kritéria bude navrhopvat praktická řešení

**Kontakt:** Ed Pinsent, Kevin Ashley, Patricia Sleeman {e.pinsent, k.ashley, p.sleeman}@ulcc.ac.uk

**Implementace pro NK ČR:** nedoporučuje se v celkovém rozsahu z důvodu zacílení nástroje na akademické instituce v anglo-saském prostředí; nejdůležitější pravidla pro dlouhodobou ochranu můžou být součástí open source systému nebo vlastního řešení

**AIDA** (The Assessing Institutional Digital Assets) představuje metodologii pro hodnocení činností souvisejících s repositářem, především kapacity, stav připravenosti a celkovou schopnost správy digitálních sbírek. Toto hodnocení však neprobíhá formou auditu (jak je tomu např. v případě nástroje DRAMBORA - <http://repositoryaudit.eu/>). Jde o evaluaci (z angl. “measurement”), proto „důkazy“ pro jednotlivé aktivity či naopak nečinnost nejsou nutné. V podstatě tedy špatný výsledek neexistuje.

Na evaluaci činnosti repositáře a procesů dlouhodobé ochrany se nepodílí jedna osoba, nýbrž všechny zodpovědné osoby (na úrovni oddělení i institucionální úrovni), které postupně vyplňují hodnotící tabulky. Pomocí nástroje se hodnotí 3 klíčové oblasti: organizace (z ang. organizational leg), technické zabezpečení (z angl. technical leg) a zdroje (z angl. resources leg). V každé této oblasti je

nutné zhodnotit 11 aspektů správy digitálních dat/sbírek (pro zdroje pouze 9) na 5 úrovních připravenosti<sup>6</sup>. Ty mají svůj stručný popis s příkladem, který vysvětluje a usnadňuje evaluaci. Ku příkladu prvek č. 6 (Sustainability of Funding), 2. stupeň na institucionální úrovni hodnotí kontinuitu finančních prostředků<sup>7</sup>.

Jelikož jde 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 ([www.ifs.tuwien.ac.at/dp/plato](http://www.ifs.tuwien.ac.at/dp/plato))

**Garant:** Department of Software Technology and Interactive Systems, Vienna University of Technology (<http://www.ifs.tuwien.ac.at/>)

**Instalace:** online nástroj

**Požadavky:** otevřené porty 8080 a 8443

**Poslední verze:** 3.0 (červenec 2010)

**Plány pro další verze:** nejsou známy

**Kontakt:** Christoph Becker, Hannes Kulovits, Michael Kraxner, Andreas Rauber {becker, kulovits, kraxner, rauber}@ifs.tuwien.ac.at

**Implementace pro NK ČR:** doporučuje se v celkovém rozsahu

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.

Celý proces evaluace je rozdělen na 3 fáze o 11 posloupných krocích:

1. **Definice požadavků** – pozůstává z definice a podrobnějšího popisu sbírky (obsahu dat), která je vybrána pro naplánování ochranných akcí. Tento krok je obdobný prvním kroků auditu DRAMBORA. Následně jsou vybrána konkrétní vzorová data, u kterých se podrobně požadavky na dlouhodobou ochranu. Autoři nástroje k tomuto účelu doporučují zejména open source nástroj FreeMind pro vytváření myšlenkových map.  
Základním hodnotícím kritériem je obvykle soubor nebo záznam sbírky.
2. **Hodnocení alternativ** – hodnotící kritéria z předchozí fáze jsou podkladem pro provádění experimentálních migračních aktivit.

---

<sup>6</sup> Stupeň připravenosti představuje úroveň splnění jednotlivých kritérií, přičemž úroveň 1 (z angl. stage) představuje nejnižší možný stupeň.

<sup>7</sup> „Funding comes in as short-term, one-off grants and ad-hoc awards.“

3. **Posouzení výsledků** - experimentální migrace jsou výstupem pro následnou analýzu a vyhodnocení nejlepšího formátu dané vzorové sbírky digitálních dat.

**Výstupem je komplexní dokumentace s objektivním posouzením nejruznějších migračních alternativ dle specifických požadavků dané instituce. Proto se implementace nástroje do workflow dlouhodobé ochrany digitálních dat doporučuje v plném rozsahu.**

## Závěr

V současné době vznikají open source nástroje pro dlouhodobou ochranu digitálních dat doslova jako “houby po dešti”. Některé 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, Archivematica, RODA). Samostatnou skupinu nástrojů tvoří metodologie (online, offline) sledování, evaluace a hodnocení činnosti repozitáře, včetně plánování dlouhodobé ochrany dat.

**Ne všechny níže hodnocené nástroje jsou však vhodné k integraci do systémů Národní knihovny ČR, potažmo Národní digitální knihovny. Důvodem je profilace na úzkou cílovou uživatelskou skupinu (Hoppla, RODA) nebo “nemodulární” architektura (Mopseus). Integrace metodologií typu AIDA je možná v případě dodatečného vývoje a implementace do stávajících systémů. V kontextu NDK je možnost využití AIDA pro závěrečnou evaluaci projektu. Nejpoužitelnějším nástrojem na plánování dlouhodobé ochrany i nadále zůstává PLATO, který v současné době nemá žádný vhodný ekvivalent. V následujícím roce by se mělo začít s návrhem použití tohoto systému 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 strategie dlouhodobé ochrany digitálních dat.**

# Příloha 5: Analýza metodik pro sledování nákladů životního cyklu dokumentu

PhDr. Bedřich Vychodil

---

Analýza v oblasti metodik pro sledování nákladů potřebných pro získávání, vytváření, dlouhodobé uchování a následné šíření je základním nástrojem pro možné úspory v celém procesu správy.

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 zabývajících 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.

Jak bylo uvedeno výše, problematikou sledování nákladů životního cyklu dokumentu se zabývá celá řada zahraničních projektů. Většina projektů využívá metodik teoretického charakteru a předkládá pouhá doporučení, která jsou často jen obecného charakteru. Některé projekty do svého vzniklého teoretického rámce implementují i vlastní empirická data. Takto prakticky ověřené metodiky pak lépe odpovídají praxi a jsou z pohledu praktického nasazení cennější.

Jako reprezentativní skupinu si představíme tři úspěšně zahraniční projekty, které byly již dříve veřejně prezentovány:

- **BRTF-SDPA**  
(Ensuring Long-Term Access to Digital Information)
- **DiCoMo**  
(An Algorithm Based Method to Estimate Digitization Costs in Digital Libraries)
- **LIFE I, II, III**  
(Life Cycle Information for E-Literature)

## **BRTF-SDPA** (Blue Ribbon Task Force on Sustainable Digital Preservation and Access)

Na projektu BRTF-SDPA, tedy trvale udržitelné ochraně a zpřístupnitelnosti digitálních dokumentů od roku 2007 spolupracovali významní hráči v oblasti uchování národního kulturního bohatství, jako: U.S. National Science Foundation (NSF), U.S. Library of Congress, U.K. Joint Information System Committee, Electronic Records Archives Program of the National Archives and Records Administration a Council on Library and Information Resources. Od roku vzniku bylo vydáno několik pracovních studií. 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 "ušitá" 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. Tyto kroky mají napomáhat ve zvládnutí celého procesu správy dokumentů. 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**. Tato studie nastiňuje problematiku tohoto komplexního procesu a ukazuje slabiny společně s neprobádanými oblastmi a předkládá, až provokativní otázky nad kterými se též zamýšlí:

- **Jaké digitální data bychom měli uchovávat?**  
(What digital information should we preserve?)
- **Kdo bude zodpovědný za uchování?**  
(Who will preserve it?)
- **Kdo to všechno zaplatí?**  
(Who will pay for it?)

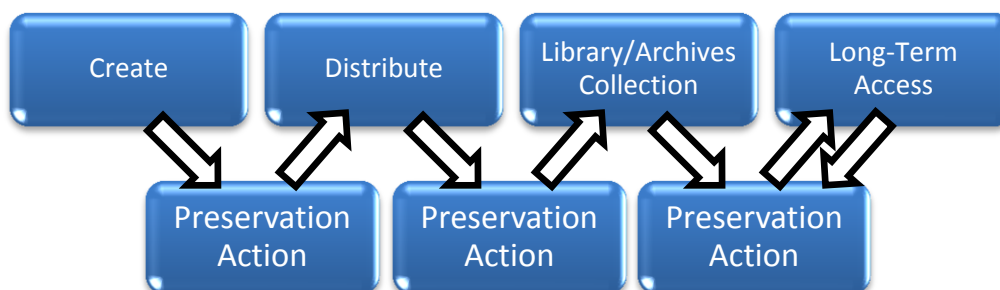
Jedná se jen velmi těžko zodpověditelné otázky v podobě obecného závěru. Zabírají velmi obsáhlé pole působnosti s velkou variabilitou v rámci prostředí, kde se daný projekt nachází. A v neposlední řadě geografické odlišnosti znemožňují vypracování obecně platného doporučení.

V projektu je též porovnávána tradiční dlouhodobá ochrana dokumentů s novým přístupem k ochraně digitálních dokumentů. Pro příklad uveďme alespoň toto porovnání, viz diagram níže:

### Tradiční dlouhodobá ochrana (Traditional Preservation)



### Digitální ochrana digitálních dokumentů (Digital Preservation Lifecycle)



Z rozsáhlé studie citujme následující text, který velmi přesně poukazuje na důležitost včasného a erudovaného přístupu:

*„Budoucí zpřístupňování cenných digitálních dokumentů závisí již na těch ochranných opatřeních, které vykonáváme v současnosti. Z dlouhodobého hlediska pak platí, že toto zpřístupňování závisí na soustavné a efektivní alokaci zdrojů určených pro dlouhodobou ochranu.“*

(volný překlad autora)

Z citovaného textu je patrné, že dlouhodobá ochrana digitálních dokumentů je odvislá od včasné aktivní účasti v celém životním cyklu digitálního dokumentu, za použití vhodných postupů a nástrojů.

### Financování

BRTF-SDPA je financován z prostředků National Science Foundation a Andrew W. Mellon Foundation, ve spolupráci s Library of Congress, Joint Information Systems Committee of the United Kingdom, the Council on Library and Information Resources, a the National Archives and Records Administration.

## **Plány BRTF-SDPA do budoucna**

Projekt si klade za cíl provést analýzu dřívějších a stávajících modelů určených pro udržitelné uchovávání digitálních dokumentů a identifikování současných a především osvědčených postupů, které jsou využívány ve stávajících sbírkách, archivech a podobných subjekty. Dále vypracovat soubor ekonomicky životaschopných doporučení a sestavit přehled vývoje spolehlivých strategií pro uchovávání digitálních informací.

A nadále se zavazuje k angažování v oblasti ekonomické udržitelnosti digitálních informací a motivovat dalším k aktivitám a výzkumu v této oblasti.

### **Zdroj**

Kompletní výstup společně s publikacemi, které vznikly na základě tohoto projektu je dostupný na: <http://brtf.sdsc.edu>



## DiCoMo

Projekt DiCoMo, tedy odhad nákladů na digitalizaci dokumentů, je odvozený od modelu CoCoMo, který je v praxi rutině nasazován pro odhady nákladů při vytváření softwaru. Projekt je zaštitěn dvěma univerzitami ve Španělsku: CIO/DEMI, Miguel Hernández University in Spain a DLSI, University of Aliance in Spain.

Z modelu CoCoMo bylo vycházeno z důvodu jeho transparentnosti a snadné implementovatelnosti v reálném prostředí. Model byl upraven pro prostředí digitalizace a pomocí takto upraveného modelu nazvaného DiCoMo je možné velmi jednoduše a efektivně alokovat výdaje vznikající při digitalizaci.

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.

Základní premisa na počátku projektu byla taková, že časová náročnost (počet hodin potřebných na digitalizaci) je přímo úměrná počtu vyskenovaných stránek. Ta ovšem neodpovídala hodnotám, které byly naměřeny v reálném prostředí. Z toho důvodu byly definovány dva modifikátory, které slouží k zpřesnění výpočtu. Byly představeny tzv. modifikátory „a“ a „b“, pomocí níž se dá lépe kopírovat reálný exponenciální nárůst času. Dále byl definován koeficient „ $ea^f$ “, který zohledňuje další faktory ovlivňující časovou náročnost. Tento koeficient zahrnuje jak zkušenost obsluhy samé, tak kvalitativní ohodnocení vstupů a výstupů digitalizace, společně s ohodnocením použitého vybavení. Jen jako příklad uvedme výběr: Zohlednění zkušenost lidí, podobnost s předešlými projekty, cizí nebo starý jazyk, starý nebo poškozený papír, zastaralý font, speciální zacházení, vysoká kvalita výstupu, špatná technika, atd.). Dále byl zaveden modifikátor „S/O“. Praxe ukázala, že tento modifikátor není závislý na počtu stránek v daném projektu, jedná se tedy o fixní parametr. Z toho důvodu se přičítá pouze předem stanovené číselné vyjádření, které vzniká na základě dříve nashromážděných dat a zkušeností. Všechny tyto uvedené koeficienty a modifikátory jsou na základě zkušeností a empirických dat číselně ohodnoceny a mohou být dále korigovány v čase podle reálných výsledků nashromážděné v průběhu digitalizace.

**Z této metodiky vyplývá předpoklad, že nasazení tohoto modelu podmiňuje systematický sběr technických dat v čase, která následně slouží ke zpřesnění výpočtu potřebných nákladů.**

Vzorec na výpočet nákladů na digitalizaci dokumentu:

$$T = \sum_s \left( a \cdot P^b \cdot \Pi eaf_i + SIO \right)$$

**T** – Čas potřebný na digitalizaci dokumentu

**a** – Modifikátor

**P** – Počet digitalizovaných stran

**b** – Modifikátor

**eaf** – Koeficient

**SIO** – Modifikátor (Fixní čas potřebný na nastavení hardwarového zařízení a softwarových aplikací)

**Klíčovým faktorem pro relevantní odhad nákladů je identifikace hlavního/hlavních faktorů ovlivňujících časovou náročnost digitalizace.**

### **Zdroj**

Abstrakt k tomuto projektu je dostupný na:

<http://www.springerlink.com/content/5mxkw0fldcc4v70f/fulltext.pdf>

## Projekt LIFE

Projekt LIFE (Life Cycle Information for E-Literature), tedy odhad 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.

První fáze projektu LIFE I byla zaměřena na analýzu stěžejních sbírek UCL (University College London) a British Library, byly etablovány základní fáze cyklu dokumentů. Každá z těchto fází byla detailně prozkoumaná a následně byla finančně vyčíslena z důvodu určení celkové sumy potřebné na dlouhodobé uchování digitálních dokumentů v těchto zkoumaných sbírkách.

Ve Druhé fázi projektu LIFE II byly upřesněny výsledky z předešlého projektu, které byly dále rozvedeny. Byly přidány tři nové případové studie, které byly postaveny na předešlém projektu LIFE I.

V zatím poslední fázi projektu LIFE III byl rozšířen stávající „Generic Preservation Model“, který byl již definován v I. a posléze v II. fázi projektu. Byla též nově přidána a též analyzována fáze zabývající se digitalizací. Tím vznikla ucelená množina modelů, která pokrývá všechny fáze životního cyklu dokumentu. Tímto krokem vznikl rámec, který umožňuje generování odhadů s větší jistotou a především přesností.

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, uveďme si jejich výčet:

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

Celkový odhad nákladů se pak jednoduše provádí součtem těchto jednotlivých, předem definovaných kroků, viz matematické vyjádření níže:

$$L_T = D_T + A_q + I_T + M_T + AC_T + S_T + P_T$$

$L_T$  – Náklady v čase  
 $D_T$  – Digitalizace  
 $A_q$  – Akvizice  
 $I_T$  – Deponování  
 $M_T$  – Metadata  
 $AC_T$  – Zpřístupnění  
 $S_T$  – Uložení  
 $P_T$  – Ochranná opatření

V projekt LIFE I byla vyvinuta metodika na analýzu a výpočet nákladů na životní cyklus sbírky digitálních dokumentů, dále její reálné využití v procesu správy a bylo provedeno několik případových studií. Byl navrhnutý rámcový model životního cyklu pro odhad nákladů na uchování digitálních objektů. **Fáze projektu LIFE II nejen rozšiřuje fázi LIFE I a přináší další projektové studie, ale především dává návod, jak je možné tento postup převést do praxe i v jiných institucích.** Třetí fáze projektu přináší nástroje a metodiku predikce na efektivní mapování celého digitálního životního cyklu dokumentu včetně fáze digitalizace. Umožňuje vypočítávání nákladů na uchování digitálních dokumentů pro příštích 5, 10 až 20 let. **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.

Projekt LIFE je vypracován v kontextu nám blízkého pojetí Britské knihovny, společně s akademickou sférou. Dále projekt je v současné době ve své třetí fázi a oba subjekty již oznámily pokračování v tomto projektu i v budoucnu. V současné době je dokončovaná reálná aplikace na predikci nákladů. Jedná se tedy z pohledu představených projektů o nejslibnějšího kandidáta na úplnou implementaci v prostředí NKP.

## Financování

Projekt je financován z prostředků JISC (Joint Information Systems Committee).

## Zdroj

Kompletní výstup společně s publikacemi, které vznikly na základě tohoto projektu je dostupný na: <http://www.life.ac.uk>

## Výstup: Přednáška k tématu na AKM 2010

Problematika zabývající se analýzou metodik pro sledování nákladů životního cyklu dokumentu byla společně s nastíněním možných východisek prezentována v rámci každoroční 11. konference Archivy, knihovny, muzea v digitálním světě 2010 v příspěvku s názvem: **Úvod do problematiky (skrytých) nákladů na dlouhodobou archivaci**, tato prezentace je dostupná na oficiálních stránkách SKIP (Svaz knihovníků a informačních pracovníků České republiky), link: <http://skip.nkp.cz/KeStazeni/Archivy10/den2/Vychodil.pdf>.

Následná diskuze pléna odhalila nutnost detailnějšího výzkumu v této oblasti. Všichni zastoupení dotazovatelé se dovolovali nutnosti implementace nějaké formy monitoringu a zvýšení informovanosti v této oblasti. Vyplývá též, že kroky podle kterých se postupuje, jsou v naprosté většině nekonzistentní a nejsou empiricky ověřené. S tím je nedílně spjatá problematika užívaných pojmů, které nejsou dostatečně etablovány v prostředí archivů, knihoven a muzeí a jsou často vnímány nestandardně napříč celým institucionálním spektrem. Jako příklad uvádím jen některé, které byly přímo na konferenci diskutovány: Archivní kopie (Archival, Master Copy), Zpřístupňující kopie (Production Master Copy), Zástupná kopie (Surrogate Copy), Digitální reprezentace (Digital Representation), Autenticita digitálního objektu (Digital Authenticity of Digital object), a další.

## Doporučení

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.**
- Cíleně držet počet využívaných formátů pro dlouhodobé uchování a zpřístupnění na minimálním možném množství. Konkrétně formát JPEG a JPEG2000.
- 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.

## Příloha 6: Ukázky XML

1) Výstup DNX ze submission aplikace pro systém Rosetta (jen začátek s popisnými metadaty a ostatními metadaty relevantními pro 1. stránku dokumentu)

```
<?xml version="1.0" encoding="UTF-8" ?>
- <mets:mets xmlns:mets="http://www.loc.gov/METS/">
- <mets:dmdSec ID="ie-dmd">
- <mets:mdWrap MDTYPE="DC">
- <mets:xmlData>
- <dc:record xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<dc:identifier>d0e5fdf0-30ae-11de-8d73-000d606f5dc6</dc:identifier>
<dcterms:rightsHolder>ABA000</dcterms:rightsHolder>
<dc:source>54 G 002315/Č.1-15./Přív.13.</dc:source>
<dc:title>Silný Vašek &</dc:title>
<dcterms:alternative>Pohádka o kováři a krejčíkovi ; Chudá
  matička</dcterms:alternative>
<dc:creator>***Donator NF***</dc:creator>
<dcterms:Location>V Praze</dcterms:Location>
<dc:publisher>I.L. Kober</dc:publisher>
<dcterms:issued>1899</dcterms:issued>
<dc:format>21 cm</dc:format>
<dcterms:extent>16 s.</dcterms:extent>
<dcterms:isPartOf>Knihovna pohádek</dcterms:isPartOf>
<dc:language>cze</dc:language>
<dcterms:UDC>821.162.3-34</dcterms:UDC>
<dcterms:UDC>0:82-34</dcterms:UDC>
  </dc:record>
  </mets:xmlData>
  </mets:mdWrap>
  </mets:dmdSec>
- <mets:amdSec ID="rep1-amd">
- <mets:techMD ID="rep1-amd-tech">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
- <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
- <section id="generalRepCharacteristics">
- <record>
  <key id="preservationType">PRESERVATION_MASTER</key>
  <key id="usageType">VIEW</key>
  <key id="RevisionNumber">1</key>
  <key id="DigitalOriginal">>true</key>
  </record>
  </section>
  </dnx>
  </mets:xmlData>
  </mets:mdWrap>
  </mets:techMD>
```

```

- <mets:rightsMD ID="rep1-amd-rights">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx" />
  </mets:xmlData>
  </mets:mdWrap>
  </mets:rightsMD>
- <mets:sourceMD ID="rep1-amd-source">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx" />
  </mets:xmlData>
  </mets:mdWrap>
  </mets:sourceMD>
- <mets:digiprovMD ID="rep1-amd-digiprov">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx" />
  </mets:xmlData>
  </mets:mdWrap>
  </mets:digiprovMD>
  </mets:amdSec>
- <mets:amdSec ID="fid1-1-amd">
- <mets:techMD ID="fid1-1-amd-tech">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
- <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
- <section id="generalFileCharacteristics">
- <record>
  <key id="label">[a] - TitlePage</key>
  <key id="note" />
  <key id="fileCreationDate" />
  <key id="fileModificationDate" />
  <key id="FileEntityType" />
  <key id="compositionLevel" />
  <key id="fileLocationType" />
  <key id="fileLocation" />
  <key id="fileOriginalName" />
  <key id="fileOriginalPath">C:\rosetta-
    develop\rosetta\packages\35945\35945\jpg\3594500001.jpg</key>
  <key id="fileOriginalID" />
  <key id="fileExtension" />
  <key id="fileMIMETYPE" />
  <key id="fileSizeBytes">0</key>
  <key id="storageID" />
  <key id="streamRefId" />
  <key id="formatLibraryId" />
  <key id="riskLibraryIdentifiers" />
  </record>
  </section>
- <section id="creatingApplication">
- <record>
  <key id="creatingApplicationName">Sirius</key>

```



```

<key id="creatingApplicationVersion">2.5</key>
<key id="dateCreatedByApplication" />
<key id="creatingApplicationExtension" />
  </record>
</section>
- <section id="fileFixity">
- <record>
  <key id="agent" />
  <key id="fixityType">MD5</key>
  <key id="fixityValue">30b9acadba8b755934ecd320791c0910</key>
  </record>
  </section>
  </dnx>
  </mets:xmlData>
  </mets:mdWrap>
  </mets:techMD>
- <mets:rightsMD ID="fid1-1-amd-rights">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx" />
  </mets:xmlData>
  </mets:mdWrap>
  </mets:rightsMD>
- <mets:sourceMD ID="fid1-1-amd-source">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx" />
  </mets:xmlData>
  </mets:mdWrap>
  </mets:sourceMD>
- <mets:digiprovMD ID="fid1-1-amd-digiprov">
- <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="dnx">
- <mets:xmlData>
  <dnx xmlns="http://www.exlibrisgroup.com/dps/dnx">
- <section id="event">
- <record>
  <key id="eventIdentifierType">NKP Identifiers</key>
  <key id="eventIdentifierValue">a7c3f242-bac2-41fb-99e7-1f7fea4d32d5</key>
  <key id="eventType">Image Creation</key>
  <key id="eventDescription" />
  <key id="eventDateTime" />
  <key id="eventOutcome1">SUCCESS</key>
  <key id="eventOutcomeDetail1" />
  <key id="eventOutcomeDetailExtension1" />
  <key id="eventOutcome2" />
  <key id="eventOutcomeDetail2" />
  <key id="eventOutcomeDetailExtension2" />
  <key id="eventOutcome3" />
  <key id="eventOutcomeDetail3" />
  <key id="eventOutcomeDetailExtension3" />
  <key id="linkingAgentIdentifierXMLID1" />
  <key id="linkingAgentIdentifierType1">Scanning device</key>

```

```

<key id="linkingAgentIdentifierValue1">HYBRID TECH 40li,PROSERV
  Datentechnik,40li,0(11811.02x11811.02)</key>
<key id="linkingAgentRole1" />
<key id="linkingAgentIdentifierXMLID2" />
<key id="linkingAgentIdentifierType2">Scanning software</key>
<key id="linkingAgentIdentifierValue2">PROView 6.10</key>
<key id="linkingAgentRole2" />
<key id="linkingAgentIdentifierXMLID3" />
<key id="linkingAgentIdentifierType3" />
<key id="linkingAgentIdentifierValue3" />
<key id="linkingAgentRole3" />
  </record>
</section>
</dnx>
</mets:xmlData>
</mets:mdWrap>
</mets:digiprovMD>
</mets:amdSec>
- <mets:amdSec ID="fid2-1-amd">

```

## 2) Ukázka XML ve formátu XIP – část popisných metadat v tagu collections

```

- <Collections>
- <Collection status="new">
- <CollectionRef>uuid0</CollectionRef>
  <CollectionTypeRef>1</CollectionTypeRef>
  <CollectionCode>PERIOD</CollectionCode>
  <Title>periodikum</Title>
  <ClosureStatus>O</ClosureStatus>
  <ClosureType>I</ClosureType>
  </Collection>
- <Collection status="new">
- <CollectionRef>uuid1</CollectionRef>
  <CollectionTypeRef>1</CollectionTypeRef>
  <ParentRef>uuid0</ParentRef>
  <CollectionCode>PERIODICAL</CollectionCode>
  <Title>Našinec</Title>
  <ClosureStatus>O</ClosureStatus>
  <ClosureType>I</ClosureType>
- <Metadata
  schemaURI="http://dublincore.org/schemas/xmls/qdc/2008/02/11/dcterms.x
  sd">
- <dc:record
  xsi:schemaLocation="http://dublincore.org/schemas/xmls/qdc/2008/02/11/dct
  erms.xsd" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/">
  <dcterms:rightsHolder>OLA001</dcterms:rightsHolder>
  <dc:source>IV 53.435</dc:source>
  <dc:type>periodikum</dc:type>
  <dc:title>Našinec</dc:title>
  <dc:publisher>Josefina Černochová</dc:publisher>

```

```

<dcterms:issued>1869 - 1884</dcterms:issued>
<dc:publisher>Alexandr Jiříček</dc:publisher>
<dcterms:issued>1885 - 1904</dcterms:issued>
<dc:publisher>Vydavatelské družstvo Našinec</dc:publisher>
<dcterms:issued>1905 - 1920</dcterms:issued>
<dc:publisher>Lidové závody tiskařské a nakladatelské</dc:publisher>
<dcterms:issued>1921 - 1941</dcterms:issued>
<dc:language>cze</dc:language>
<dcterms:bibliographicCitation>68</dcterms:bibliographicCitation>
<dcterms:bibliographicCitation>1932</dcterms:bibliographicCitation>
<dc:description>nekvalitní a poškozená předloha; č. 296 není k
dispozici;</dc:description>
</dc:record>
</Metadata>
</Collection>

```

### 3) Ukázka Deliverable Unit čísla periodika a stránky čísla ve formátu XIP

```

= <DeliverableUnit status="new">
<DeliverableUnitRef>uuid5754</DeliverableUnitRef>
<CollectionRef>uuid1</CollectionRef>
<AccessionRef>uuid3</AccessionRef>
<AccumulationRef>uuid2</AccumulationRef>
<ParentRef>uuid5718</ParentRef>
<DigitalSurrogate>true</DigitalSurrogate>
<CatalogueReference>7</CatalogueReference>
<ScopeAndContent>NormalPage</ScopeAndContent>
<ClosureStatus>O</ClosureStatus>
<ClosureType>I</ClosureType>
<CoverageFrom>2000-01-31T00:00:00.000+01:00</CoverageFrom>
<CoverageTo>2000-01-31T00:01:00.000+01:00</CoverageTo>
<Title>8</Title>
<TypeRef>1</TypeRef>
</DeliverableUnit>
= <DeliverableUnit status="new">
<DeliverableUnitRef>uuid5759</DeliverableUnitRef>
<CollectionRef>uuid1</CollectionRef>
<AccessionRef>uuid3</AccessionRef>
<AccumulationRef>uuid2</AccumulationRef>
<ParentRef>uuid5</ParentRef>
<DigitalSurrogate>true</DigitalSurrogate>
<CatalogueReference>227</CatalogueReference>
<ScopeAndContent>PeriodicalIssue</ScopeAndContent>
<ClosureStatus>O</ClosureStatus>
<ClosureType>I</ClosureType>
<CoverageFrom>2000-01-31T00:00:00.000+01:00</CoverageFrom>
<CoverageTo>2000-01-31T00:01:00.000+01:00</CoverageTo>
<Title>4.10.1932 Našinec</Title>
<Language />
<TypeRef>1</TypeRef>
</DeliverableUnit>

```

# Příloha 7:

## Centrální znalostní báze RD – vypracoval Ing. Kocourek

Ochrana novodobých dokumentů

### Aplikace pro průzkum fondů

Záměrem je vytvořit prostředí pro evidenci a sdílení odborných dokumentů týkajících se ochrany novodobých dokumentů.

Archiv článků a zpráv dynamicky zařazovaných do řešených témat na základě obsahu. Dokumenty a zprávy mohou výhledově obsahovat diskuze a hodnocení uživatelů. K řešeným tématům může být sestavován seznam odborníků a osob, které se problematikou zabývají.

Uživatelé služby budou v první řadě pracovníci ochrany fondů, potom pracovníci institucí spolupracujících na některých projektech, případně další odborníci z tuzemska i zahraničí, podle situace. Výhledově se dá předpokládat přístup neautorizovaných osob a případně implementace hodnocení a komentování.

### APLIKACE PRO PRŮZKUM FONDŮ

Aplikace pro zjišťování stavu dokumentů fondů pracovníky bude založena na platformě RIII, kterou využívá Registr digitalizace a bude pro provoz používat také jeho technologické prostředí.

#### Komponenty řešení

Úložiště – RDBMS ORACLE.

Prezentační rozhraní – vyhledávací systém FAST ESP.

#### Datová struktura aplikace:

Přehled tabulek obsahuje návrh datové struktury dle požadavků pracovníků NK.

Identifikace exempláře

Název pole	Typ/Rozsah	Poznámka/kod
ČNB	Text/ 20 znaků	Přebírají se data
Pole 001	Text/ 15 znaků	Přebírají se data
SYSNO	Text/ 15 znaků	Přebírají se data
Signature	Text/ 20 znaků	Přebírají se data / zapis
Čárový kód	Alfanumerické/ 10 znaků	Přebírají se data
Název	Text/ 100 znaků	Přebírají se data / zapis
Autor	Text/ 50 znaků	Přebírají se data / zapis
Místo vydání	Text/ 50 znaků	Přebírají se data / zapis
Rok vydání	Alfanumerické/ 4 znaky	Přebírají se data / zapis
Vzácnost NKC	Text/15znaků	Přebírají se data

Vzácnost SKC	Text/15znaků	Přebírají se data
--------------	--------------	-------------------

Šedou barvou označená pole jsou přebírána z katalogu

#### Skladové podmínky

Název pole	Typ/Rozsah	Poznámka/kod
Kde (místnost/sklad)	Text/15 znaků	Nastavitelné administrátorem
Teplota v °C	Alfanumerické/2 znaky	
Vlhkost v %	Alfanumerické/2 znaky	

#### Údaje o zpracování dokumentu

Název pole	Typ/Rozsah	Poznámka/kod
Digitalizace	A/N checkbox	
Mikrofilm	A/N checkbox	
Záznam v katalogu	A/N checkbox	
Záznam jednotky (exempláře)	A/N checkbox	
Přílohy	A/N checkbox	Aktivuje pole přílohy
ID průzkumu	Generované číslo	

#### Fyzický popis exempláře

Název pole	Typ/Rozsah	Poznámka/kod
Typ fondu	RadioButton 3	Monografie / Periodika / Hudebniny
Typ vazby	RadioButton 5	Pevná / Polotuhá / Brožovaná / Brožura / Desky s kapsou
Druh vazby	RadioButton 6	Celopapírová / Celopátěná / Celokožená / Poloplátěná / polokožená / Saténová /
Neúplnost exempláře	Text/200 znaků	
Poznámky	Text/200 znaků	

Ochranný obal	RadioButton 3	Krabice / Desky / Není
Materiál	RadioButton 3	Archivní lepenka / Původní / nevhodné
pH	2 znaky	

Typ papíru	RadioButton 5	Ruční / Dřevitý / Křídový / Novinový / Jiný
Zabarvení papíru	RadioButton 4	Světlý / Tmavý / Barevný / Jiný

Písmo	RadioButton 4	Dobré/ nečitelné/ zdegradované/ jiné
-------	---------------	--------------------------------------

Typ tisku	RadioButton 3	Knihtisk / Rukopis / Jiný
-----------	---------------	---------------------------

#### Přílohy

Název pole	Typ/Rozsah	Poznámka/kod
Typ papíru	RadioButton 5	Stejný jako KB / Křídový / Ruční / Dřevitý / Jiný
Typ přílohy	RadioButton 7	CD / Gramofonová deska / Disketa / Mapa / Střih / Rys / Jiná
Typ tisku	RadioButton 7	Grafika / Hlubitisk / Plochý tisk / Střih / Mapa / Rysy / Jiný
Barva tisku	RadioButton 3	Černá / jednobarevná / Jiná

Poznámka	Text/200 znaků	Vepsání textu
----------	----------------	---------------

#### Mechanické poškození dokumentu

Název pole	Typ/Rozsah	Poznámka/kod
Desky	1 z 5	Výběr bodového hodnocení
Hřbetník	1 z 5	
Kapitálek	1 z 5	
Záložková stužka	1 z 3	

#### Poškození knižní blok

Předsádka	1 z 5	
Vazba	1 z 5	

#### Poškození papíru

##### Mechanické

Název pole	Typ/Rozsah	Poznámka/kod
Stav papíru	M z N	Chybějící část (utržený, ohořelý)/ chybějící celá stránka (možnost vepsat jaká) /Přehyby, překlady/ Slepené listy / Činnost hlodavců, hmyzu / Křehkost/ lepící pásy/Není

##### Biologické

Stav papíru	M z N	Hmyz / Hlodavci/Plísně, bakterie
-------------	-------	----------------------------------

##### Chemické

Stav papíru	M z N	Voda/ Jiná tekutina/ Mastnota/ Prach/ Tepelné poškození (oheň)/ Barevné
-------------	-------	--

		skvrny (Pokreslené) / Nečitelný text/ Zdegradovaný (křehký)
--	--	--

Poznámka	Text/200 znaků	Vepsání textu
----------	----------------	---------------

#### Obálka

Na deskách	1 z 5	Výběr bodového hodnocení
Vevázaná	1 z 5	
Není		

Poznámka	Text/200 znaků	Vepsání textu
----------	----------------	---------------

#### Měření :

##### Obálka

Název pole	Typ/Rozsah	Poznámka/kod
Způsob	RadioButton	Elektroda / Tužka
Kde	Text / 15 znaků	
pH	2 znaky	

##### Knižní blok

Způsob	RadioButton	Elektroda / Tužka
Kde	Text / 15 znaků	
Strana	Text / 4 znaky	Zahrnout do kde ?
pH	2 znaky	

##### Příloha

Způsob	RadioButton	Sonda / Tužka
Kde	Text / 15 znaků	
Strana	Text / 4 znaky	Zahrnout do kde ?
pH	2 znaky	

Poznámky	Text/200 znaků	Vepsání textu
----------	----------------	---------------

#### Připojený soubor

Název pole	Typ/Rozsah	Poznámka/kod
Název	Txt	
File	Img	
Velikost	kB	



Typ souboru	(mime type)	
Thumbnail	Img	

Fotodokumentace bude pocházet z digitálního fotoaparátu. Soubory se budou muset přejmenovat a importovat nebo vložit do aplikace.

# **Příloha 8:**

## **Vstupní formulář průzkumu – vypracoval Ing. Kocourek**

Formulář aplikace RIII, který je využíván pro zápis informací z průzkumu fondu.

Tabulka		Formulář	
<b>Identifikace exempláře</b>			
Číslo záznamu:			
Em000000001			
<b>Hlavní identifikační znaky exempláře</b>			
Název:			
Diamantový důl			
ČNB:	Autor:	<b>Údaje o zpracování</b>	
CNB123456789	Jan Mochl	Digitalizace:	<input checked="" type="checkbox"/>
Pole 001:	Rok vydání:	Mikrofilm:	<input type="checkbox"/>
987456321	2010	Záznam v katalogu:	<input type="checkbox"/>
SysNo:	Místo vydání:	Záznam jednotky (exempláře):	<input checked="" type="checkbox"/>
S5482515			
Čárový kód:	Vzácnost NKC:		
95124659785	vzácné		
Signatura:	Vzácnost SKC:		
SIG98547	nevzácné		
<b>Stávající skladovací podmínky</b>			
Uložení:	Místo uložení:	Teplota v °C:	Vlhkost v %:
	testSklad01	18	59
Typ fondu:	Typ vazby:	Druh vazby:	Ochranný obal:
<input checked="" type="radio"/> Hudebniny	<input type="radio"/> Brožovaná	<input type="radio"/> Celokožená	<input type="radio"/> Desky
<input type="radio"/> Monografie	<input type="radio"/> Brožura	<input checked="" type="radio"/> Celopapírová	<input checked="" type="radio"/> Krabice
<input type="radio"/> Periodika	<input checked="" type="radio"/> Desky s kapsou	<input type="radio"/> Celoplátěná	<input type="radio"/> Není
	<input type="radio"/> Pevná	<input type="radio"/> Jiná	
	<input type="radio"/> Polotuhá	<input type="radio"/> Polokožená	
		<input type="radio"/> Poloplátěná	
			Materiál:
			<input type="radio"/> Archivní lepenka
			<input checked="" type="radio"/> Nevhodné
			<input type="radio"/> Původní
			pH: 9
Neúplnost exempláře:			
Exemplář je úplný			
Poznámky:			
první poznámka			
<b>Stav exempláře</b>			
Typ papíru	Zabarvení papíru	Písmo:	Typ tisku:
<input type="radio"/> Dřevitý	<input checked="" type="radio"/> Barevný	<input type="radio"/> 1	<input type="radio"/> Jiný
<input type="radio"/> Jiný	<input type="radio"/> Jiný	<input checked="" type="radio"/> 2	<input checked="" type="radio"/> Knihtisk
<input checked="" type="radio"/> Křídový	<input type="radio"/> Světlý	<input type="radio"/> 3	<input type="radio"/> Rukopis
<input type="radio"/> Novinový	<input type="radio"/> Tmavý		
<input type="radio"/> Ruční			
<b>Mechanické poškození dokumentu</b>			
Desky:	Hřbetník:	Kapitálek:	Záložková stužka:
<input type="radio"/> 1	<input type="radio"/> 1	<input type="radio"/> 1	<input type="radio"/> 1
<input checked="" type="radio"/> 2	<input type="radio"/> 2	<input checked="" type="radio"/> 2	<input type="radio"/> 2
<input type="radio"/> 3	<input checked="" type="radio"/> 3	<input type="radio"/> 3	<input checked="" type="radio"/> 3
<input type="radio"/> 4	<input type="radio"/> 4	<input type="radio"/> 4	
<input type="radio"/> 5	<input type="radio"/> 5	<input type="radio"/> 5	
<b>Poškození knižní blok</b>			
Předsádka:	Vazba:	<b>Obálka</b>	
<input type="radio"/> 1	<input type="radio"/> 1	Obálka:	Hodnota:
<input checked="" type="radio"/> 2	<input type="radio"/> 2	<input checked="" type="radio"/> Na deskách	<input type="radio"/> 1
<input type="radio"/> 3	<input type="radio"/> 3	<input type="radio"/> Není	<input type="radio"/> 2
<input type="radio"/> 4	<input checked="" type="radio"/> 4	<input type="radio"/> Vevázaná	<input checked="" type="radio"/> 3
<input type="radio"/> 5	<input type="radio"/> 5		<input type="radio"/> 4
			<input type="radio"/> 5
Poznámka (mechanické poškození dokumentu):			
druhá poznámka			

**Měření**

**Obálka**

Způsob (obálka): Kde (obálka): pH (obálka):

Elektroda    nahoře    2

Tužka

**Knižní blok**

Způsob (k.blok): Kde (k.blok): Strana (k.blok): pH (k.blok):

Elektroda    nahoře    78    7

Tužka

Poznámky (měření):

pátá poznámka

**Přílohy**

Typ přílohy:      Typ papíru:      Typ tisku:      Barva tisku:

CD                       Dřevitý                       Grafika                       Černá

Disketa                       Jiný                       Hlubitisk                       Jednobarevná

Gramofonová deska       Křídový                       Jiný                       Jiná

Jiný                       Ruční                       Mapa                       Plochý tisk

Mapa                       Stejný jako KB                       Rys

Rys                       Střih

Střih

Způsob:

Elektroda     Tužka

Kde:                      Strana:                      pH:

Poznámka:

**Poškození papíru**

**Mechanické**

Chybějící stránka:  Strana: 78

Činnost hlodavců, hmyzu:

Chybějící část:

Křehkost:

Lepicí pásky:

Přehyby, překlady:

Slepené listy:

Není:

**Biologické**

Hlodavci:

Hmyz:

Plísně, bakterie:

Není:

**Chemické**

Zdegradovaný(křehký):

Barevné skvrny (pokreslené):

Jiná tekutina:

Mastnota:

Nečitelný text:

Prach:

Tepelné poškození (oheň):

Voda:

Není:

Poznámka (Poškození papíru):

třetí poznámka

**Poškození dokumentu**

**Biologické**

Hlodavci:

Hmyz:

Není:

Plísně, bakterie:

Zvířecí exkrementy:

**Mechanické**

Chybějící část(utržený ohořelý):

Činnost hlodavců, hmyzu:

Křehkost:

Není:

Přehyby, překlady:

Slepené listy:

**Chemické**

Barevné skvrny (pokreslené):

Jiná tekutina:

Mastnota:

Není:

Prach:

Tepelné poškození (oheň):

Voda:

**Poškození příloh:**

1

2

3

4

5

Poznámka (poškození):

čtvrtá poznámka

**Přílohy**

Typ přílohy:  CD  Disketa  Gramofonová deska  Jiný  Mapa  Rys  Střih

Typ papíru:  Dřevitý  Jiný  Křídový  Ruční  Stejný jako KB

Typ tisku:  Grafika  Hlubitisk  Jiný  Mapa  Plochý tisk  Rys  Střih

Barva tisku:  Černá  Jednobarevná  Jiná

Způsob:  Elektroda  Tužka

Kde: \_\_\_\_\_ Strana: \_\_\_\_\_ pH: \_\_\_\_\_

Poznámka: \_\_\_\_\_

---

Typ přílohy:  CD  Disketa  Gramofonová deska  Jiný  Mapa  Rys  Střih

Typ papíru:  Dřevitý  Jiný  Křídový  Ruční  Stejný jako KB

Typ tisku:  Grafika  Hlubitisk  Jiný  Mapa  Plochý tisk  Rys  Střih

Barva tisku:  Černá  Jednobarevná  Jiná

Způsob:  Elektroda  Tužka

Formulář obsahuje některé odlišnosti v uspořádání a detailech popisu oproti popisu datové struktury v zadání projektu, které vyplynuly z testování a požadavků na optimalizaci navržené datové struktury aplikace pro průzkum.

### Vyhledávání ve Znalostní bázi

Pro zpřístupnění pomocí vyhledávací technologie bude využito systému ESP. Ten bude indexovat data uložená v relační databázi a umožní prostřednictvím extrakce entit využívat Navigace na základě Tématických kategorií, Typů dokumentů, Časového určení, Autorů dokumentů, případě výhledově osob, které dokumenty komentují.

Vizuálně bude vyhledávací rozhraní uzpůsobenou Registru digitalizace.

Nasazení uživatelského rozhraní založeného na vyhledávání se předpokládá po odladění datové struktury a jejího ověření v praxi při průzkumu první části fondu.

# Příloha 9:

## Depozitáře Národní knihovny České republiky 1943-2010 – vypracoval Dr. Šnýdr

Do začátku 2. světové války nepotřebovala tehdejší Národní a universitní knihovna depozitáře mimo Klementinum k uložení knižních fondů.

V souvislosti s vývojem válečné situace se v roce 1943 přemístilo cca 500.000 svazků mimo Prahu takto:

- Karlštejn – 10.284 svazků – rukopisy a nejvzácnější staré tisky – uloženy byly v celém purkrabství a v Křížové kapli. Využity byly i truhlice, v nichž byl kdysi uložen korunní archiv.
- Zlatá Koruna – 399.716 svazků, včetně Národního konzervačního fondu
- Pohled u Havlíčkova Brodu - 82.101 svazků
- Horažďovice – 22.620 svazků

V období 1945-1952 se knihovna rozrostla nejen o nové přírůstky, především sovětské literatury, ale i západní produkce z doby nesvobody, ale i o obrovské množství literatury ze sozů a likvidace německých a starých a církevních institucí. Byly proto hledány nové prostory pro uskladnění těchto svazků:

- Houska – od cca 1955 do začátku 70. let, byla zde signatura 52, část Akademická knihovny a rezervních fondů. Prostory byly zcela nevyhovující. Dřevěné regály, bez klimatizace a topení. V zimním období se nedalo z tohoto objektu expedovat knihy pro čtenáře. Ve výjimečných případech bylo nutno k silnici přivážet knihy na sáních.
- Praha – Klášterní knihovna benediktnů, Praha 6, Markétská - 25.000 svazků
- Praha - Klášterní knihovna dominikánů, Praha 1, Husova 8 - 9.622 svazků
- Praha - Klášterní knihovna Augustiánů 1, Josefská 8 - 18.729 svazků
- Praha - Klášterní knihovna křižovníků, Praha 1, Husova 8 a Praha 2, Voršilská 1 - 41.471 svazků
- Praha - Klášterní knihovna františkánů, Praha 1, Jungmannovo nám 18 – 9.559 svazků
- Zámek Postoloprty – od poloviny 60. let do poloviny 90. let: Roudnická Lobkovická knihovna, Kinského knihovna, Akademická knihovna + Rezervní fondy + Knihovna Ústavu TGM, Odd. 20, 23, D, Pa, Fondy byly uloženy na provizorních dřevěných regálech i na chodbách. Na začátku 90. let se tyto fondy (především Roudnická Lobkovická knihovna) staly terčem organizovaných zlodějských akcí. Jen posledně jmenovaná knihovna podle soudních znaleců přišla o staré a vzácné tisky v hodnotě 14 milionů Kč. Poté byla do doby předání zpět v restituci přemístěna do depozitáře Most (viz dále).
- Buštěhrad – na počátku 70. let byl depozitář Houska vyměněn se Spolanou Neratovice objekt v Neratovicích. Signatura 52 byla z Housky přemístěna do Buštěhradu, kde byly jen o něco lepší skladovací podmínky než na Housce. Přes veškerou snahu ve skladištích hnízili ptáci a hlodavci. I zde byly jen provizorní dřevěné regály.

- Most - ze všech depozitářů NK zde byly pro uložení knihovních fondů nejlepší podmínky. Budova byla uvnitř střeženého areálu GŘ Mosteckých dolů. Byla klimatizována a s regulovanou teplotou. Od roku 1991 do poloviny 90. let zde byly fondy převezené z depozitáře Postoloprty:
- Roudnická Lobkovická knihovna do protokolárního předání Lobkovicům v rámci restituce, Kinského knihovna, Odd. 20, 23
- Nučice - v provozu od počátku 70. let do poloviny 90. let. Jednalo se celkem o 3 plechové haly s kovovou jednopatrovou skladištní konstrukcí. V každé hale bylo umístěno cca 200 – 250 tisíc knih. Do roku 1992 byl depozitář využíván pro skladování zvláštních fondů (vyřazená „nevhodná literatura“ z veřejných knihoven.) Do konečného opuštění depozitáře byly 3 haly využívány takto:
  - 1. hala – 244000 svazků – signatury 4, 6, 7, 17, 19, 24, G, Sp
  - 2. hala – rezervní fondy, rezervní fondy Slovanské knihovny, rezervní fondy a odepsaná literatura Ústřední ekonomické knihovny
  - 3. hala - signatury F, G, K, Diss, 74, Kand

Byly připravovány další depozitáře: Brnky, Tmář, Filipov, kostel sv. Michala v Praze.

#### Depozitář Neratovice

V současné době jediný mimopražský depozitář NK ČR. Jedná se o zděný objekt na okraji Neratovic. Od 70. let sloužil pro uložení Národního konzervačního fondu a sign. 73 a části rezervních fondů. V současnosti je zde uloženo cca 600.000 svazků rezervních fondů. Skladovací prostory jsou zde bez oken a vnitřní klima ovlivňuje venkovní prostředí (teplota, relativní vlhkost). Celoroční hodnoty klimatu v prostorách se u teploty každý rok pohybují v rozmezí 14-22 °C a u relativní vlhkosti v rozmezí 30-50% a k jejich změnám dochází postupně. V roce 2002 byl depozitář postižen vniknutím vody z povodně a došlo k poškození fondů uložených v přízemí objektu. Depozitní prostory jsou vybaveny systémem pro přenos dat klimatických podmínek teplotními čidly a vlhkoměry. V objektu chybí účinná klimatizace a účinné vzduchové filtry. Do objektu vniká prach a další degradační činitele z venkovního prostředí. Dochází ke znečišťování fondu i k degradaci papírových dokumentů. (Údaje ze října 2008). V depozitáři se pravidelně provádí vizuální kontrola stavu fondu, pravidelně se odebírají stěry pro kontrolu mikrobiologického stavu fondu, poslední dílčí mechanická očista fondu byla provedena při likvidaci následků povodní v roce 2003. Objekt je vybaven EZS a EPS.

#### Centrální depozitář Hostivař

Vlastní plocha skladovacích hal v CDH je cca 4100 m<sup>2</sup> a tři patra regálových konstrukcí je cca 12300 m<sup>2</sup> na cca 4200000 knižních svazků. Část skladovacího prostoru je využívána pro Archiv NK ČR a pro Pekařovu knihovnu University Karlovy, Akademickou knihovnu, rezervní fondy a částečně zpracované výroční zprávy vysokých a středních škol. Dále jsou zde periodika Konzervačního fondu 19. století a Poškozený fond UKF. V CDH jsou uloženy veškeré fondy Národního konzervačního fondu a převážná většina cizojazyčného fondu Universálního knihovního fondu.



# Příloha 10: Příklad údajů o sledovaných klimatických podmínkách

Vypracoval – Ing. Franci

Příklad údajů o sledovaných klimatických podmínkách, které budou vkládány do databáze a elektronických formulářů dokumentů. Zde konkrétně klimatické parametry z depozitáře v Neratovicích.

Květen 2010

Prostor	Teplota (min)	Teplota (max)	Teplota (průměr)	Kolísání teploty (průměr)	Relativní vlhkost (min)	Relativní vlhkost (max)	Relativní vlhkost (průměr)	Kolísání relativní vlhkosti (průměr)
P-1	15,4°C	22,3°C	17,4°C	0,3°C/24 h	33,00%	44,00%	37,60%	2,0%/24 h
P-2	14,5°C	18,7°C	16,0°C	0,3°C/24 h	39,00%	46,00%	41,70%	1,6%/24 h
I-1	15,5°C	23,9°C	17,7°C	0,4°C/24 h	31,00%	42,00%	36,40%	1,8%/24 h
I-2	14,7°C	20,5°C	16,7°C	0,4°C/24 h	36,00%	44,00%	40,60%	1,5%/24 h
II-1	15,6°C	24,6°C	18,1°C	0,7°C/24 h	30,00%	42,00%	36,60%	2,3%/24 h
II-2	14,9°C	21,6°C	17,2°C	0,6°C/24 h	34,00%	42,00%	38,60%	1,6%/24 h

Červen 2010

P-1	16,2°C	20,4°C	18,6°C	0,4°C/24 h	41,00%	51,00%	44,30%	2,5%/24 h
P-2	15,1°C	19,7°C	17,8°C	0,3°C/24 h	41,00%	51,00%	46,50%	2,5%/24 h
I-1	16,3°C	22,4°C	19,6°C	0,5°C/24 h	40,00%	47,00%	43,00%	2,1%/24 h
I-2	15,4°C	21,7°C	19,0°C	0,5°C/24 h	42,00%	48,00%	44,90%	1,9%/24 h
II-1	16,4°C	24,3°C	20,4°C	1,1°C/24 h	40,00%	51,00%	42,70%	3,1%/24 h
II-2	15,7°C	23,7°C	20,0°C	1,0°C/24 h	39,00%	49,00%	42,10%	2,8%/24 h

Červenec 2010

P-1	20,1°C	23,7°C	22,0°C	0,4°C/24 h	40,00%	53,00%	44,70%	3,3%/24 h
P-2	19,7°C	24,0°C	22,0°C	0,4°C/24 h	39,00%	53,00%	43,70%	2,0%/24 h
I-1	20,9°C	25,4°C	23,1°C	0,5°C/24 h	40,00%	48,00%	44,20%	2,6%/24 h
I-2	21,0°C	25,2°C	22,9°C	0,5°C/24 h	41,00%	49,00%	44,00%	2,1%/24 h
II-1	21,0°C	27,3°C	24,0°C	1,1°C/24 h	39,00%	47,00%	42,80%	2,2%/24 h

II-2	21,1°C	27,0°C	24,0°C	1,0°C/24 h	37,00%	44,00%	41,10%	2,3%/24 h
------	--------	--------	--------	------------	--------	--------	--------	-----------

Srpen 2010

P-1	19,6°C	22,3°C	21,1°C	0,5°C/24 h	42,00%	50,00%	46,00%	2,6%/24 h
P-2	19,4°C	22,5°C	21,4°C	0,4°C/24 h	38,00%	47,00%	42,20%	1,7%/24 h
I-1	19,6°C	23,2°C	21,5°C	0,5°C/24 h	43,00%	50,00%	45,70%	2,3%/24 h
I-2	19,6°C	23,1°C	21,6°C	0,5°C/24 h	40,00%	46,00%	42,80%	2,1%/24 h
II-1	19,4°C	24,4°C	21,9°C	1,0°C/24 h	41,00%	57,00%	46,10%	4,5%/24 h
II-2	19,6°C	24,5°C	22,0°C	0,8°C/24 h	39,00%	53,00%	41,80%	3,0%/24 h

Září

2010

P-1	15,6°C	19,6°C	17,4°C	0,4°C/24 h	41,00%	55,00%	46,80%	1,8%/24 h
P-2	15,1°C	19,4°C	17,1°C	0,3°C/24 h	39,00%	48,00%	44,60%	1,7%/24 h
I-1	15,6°C	19,5°C	17,5°C	0,4°C/24 h	40,00%	52,00%	46,60%	2,2%/24 h
I-2	15,4°C	19,6°C	17,1°C	0,3°C/24 h	41,00%	49,00%	45,00%	1,8%/24 h
II-1	15,5°C	19,6°C	17,5°C	0,6°C/24 h	43,00%	55,00%	47,40%	2,3%/24 h
II-2	15,6°C	19,6°C	17,4°C	0,5°C/24 h	40,00%	48,00%	43,80%	1,8%/24 h

Říjen 2010

P-1	17,7°C	21,1°C	19,0°C	0,3°C/24 h	34,00%	52,00%	42,50%	2,2%/24 h
P-2	13,1°C	17,6°C	15,3°C	0,4°C/24 h	43,00%	51,00%	47,20%	2,5%/24 h
I-1	19,1°C	22,4°C	20,4°C	0,4°C/24 h	31,00%	48,00%	38,70%	1,7%/24 h
I-2	13,1°C	18,4°C	15,6°C	0,5°C/24 h	44,00%	50,00%	47,10%	2,5%/24 h
II-1	19,1°C	22,7°C	20,5°C	0,5°C/24 h	31,00%	47,00%	38,70%	1,6%/24 h
II-2	13,5°C	19,2°C	16,2°C	0,7°C/24 h	42,00%	49,00%	45,10%	2,2%/24 h

## **Příloha 11:**

# **Popis přístroje SurveNIR – vypracovala Ing. Marie Benešová, Ing. Vávrová**

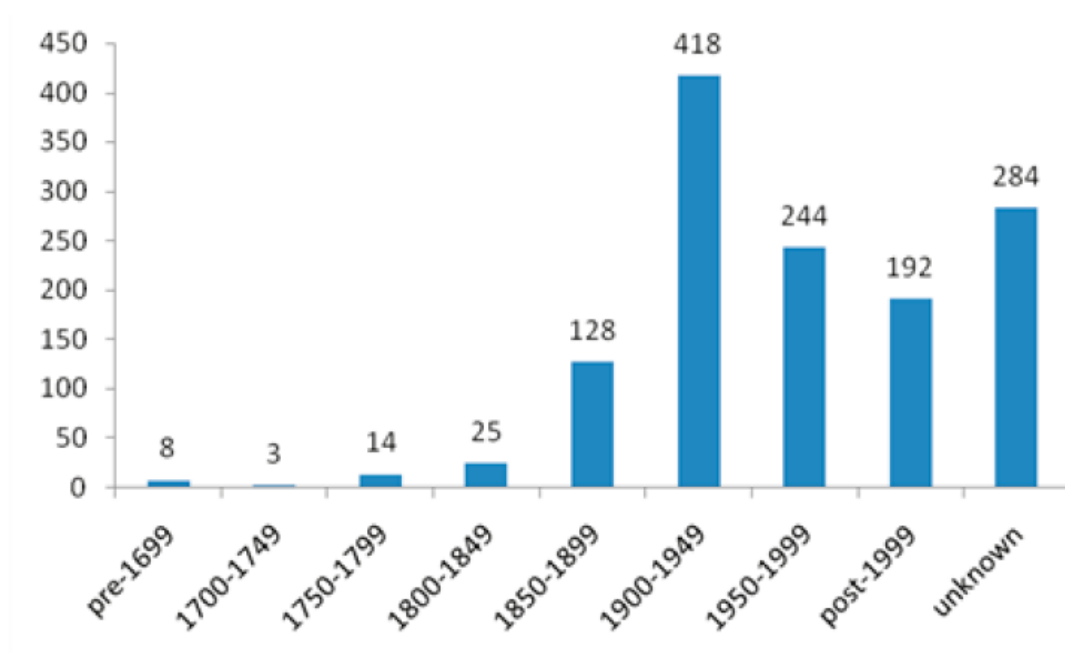
SurveNIR je spektrometrická metoda používaná pro vyhodnocování stavu papíru. Je to metoda vycházející z porovnání naměřených infračervených spekter s obsáhlou knihovnou spekter známých vzorků papírů. Známé papíry zde znamenají historické papíry, které byly naměřeny pomocí přístroje SurveNIR a poté byly analyticky vyhodnoceny hodnota pH studeného výluhu, polymerační stupeň, molekulová hmotnost a obsah ligninu, proteinů a pryskyřic. Dále byly měřeny hodnoty pevnosti v tahu, pevnosti v tahu po ohybu a přítomnost optických zjasňujících prostředků. Jedná se o zhruba 1316 listů z historických objektů převážně z 18. Století (zastoupení materiálů je uveden v tabulce níže). Výzkum se nezaměřil na neevropské, transparentní a jiné papíry se speciálním účelem. Dále nebyly vyhodnocovány papíry poškozené ohněm, vodou a plísněmi. Software automaticky tyto hodnoty vyhodnocuje pomocí již zmíněné knihovny spekter.

Tento přístroj byl sestaven pro orientační zhodnocení stavu celých fondů knihoven a archivů. Při provádění takovýchto průzkumů lze ke každému měřenému objektu zadat popis fyzického stavu podle kategorií, které si sám uživatel zvolí. Součástí softwaru je také statistické vyhodnocení naměřených výsledků tzn. stavu papírové podložky spolu s vyhodnocením celkového fyzického stavu fondu.

Přístroj SurveNIR využívá vlnové délky v rozmezí 1100 – 2500 nm ( $9100 - 4000 \text{ cm}^{-1}$ ).

Vyhodnocení chemických vlastností papírové podložky objektu je prováděno pomocí Partial Least Squares (PLS). Tato metoda závisí na kvalitě získaných spekter a na dostatečném množství naměřených reprezentativních vzorků v knihovně. Podle nich jsou poté predikovány chemické vlastnosti neznámého vzorku.

Spektrum reflektance blízkého infračerveného záření bylo měřeno z kruhové plochy vzorků o průměru 5 mm. Spektrum pro vyhodnocení bylo průměrováno z 20 měření. Pro vyhodnocení byla použita oblast v rozsahu 1100 – 2500 nm. Pro měření byl konkrétně použit přístroj SurveNIR prototyp (Lichtblau e.k. Lipsko, Německo) propojený s počítačem pomocí softwaru SurveNIR 2.0.



Tabulka: Zastoupení analyzovaných papírů v knihovně standardů systému SurvenIR podle stáří.

## Příloha 12:

# Tabulka naměřených údajů z přístroje SurveNIR – vypracovala Ing. Vávrová, Bc. Křečanová

V tabulce jsou uvedeny naměřené vlastnosti papírů novodobých fondů pomocí systému SurveNIR.

Signatura	stranka	Mass (g)	Paper Type	pH	Degree of Polymerisation	Tensile Strength	Tensile Strength Folded	Lignin Content	Protein Content	Rosin Content	Optical Brightener
54 K 26490	19		Ground Wood Paper	N/A		24	7	211	0,1		NO
54 K 26490	85	261	Ground Wood Paper	3		28	10	215	0		NO
54 K 26490	271	261	Ground Wood Paper	N/A		20	N/A	202	0	3	NO
K 4975	69	280	Bleached Pulp Paper	5	N/A	33	0	51	0		NO
K 4975	69	280	Bleached Pulp Paper	4	N/A	37	4	48	0		NO
K 4975	67	280	Bleached Pulp Paper	4	N/A	33	1	46	0		NO
K 4975	50-51	280	Bleached Pulp Paper	4	N/A	8	N/A	49	0,2		NO
K 4975	146-147	280	Bleached Pulp Paper	4	N/A	10	N/A	53	0,2		NO
B 10525	5	164	Bleached Pulp Paper	6	500	36	16	26	0		NO
A 207	obrazek 2	207	Rag Paper	6	2 400	29	50	27	0,1		NO
A 207	za obrazkem 2	207	Rag Paper	6	2 700	22	50	32	0,1		NO
54 H 790	223	778	Ground Wood Paper	N/A		17	0	166	0,1		NO
54 H 790	347	778	Ground Wood Paper	N/A		25	3	174	0,1		NO
54 H 790	287	778	Ground Wood Paper	N/A		21	0	166	0,1		NO
54 H 790	43	778	Ground Wood Paper	N/A		23	5	161	0,1		NO
54 K 22969	77	129	Ground Wood Paper	N/A		21	0	137	0		NO
54 K 22969	84	129	Ground Wood Paper	N/A		21	1	133	0		NO
54 K 94743	23	40	Ground Wood Paper	N/A		20	4	132	0,3		NO
54 K 94743	28	40	Ground Wood Paper	N/A		20	4	136	0,2		NO
54 H 304177	17	38	Rag Paper	6	1 700	35	37	18	0,5		NO
54 K 94546	25	176	Rag Paper	3	700	14	35	30	0		NO
54 H 305525	15	191	Ground Wood Paper	N/A		24	4	228	0		NO
54 H 305525	29	191	Ground Wood Paper	3		24	2	256	0,2		NO
C 12425	105	569	Bleached	6	900	46	25	35	0		NO

			Pulp Paper								
O 8180	117	757	Bleached Pulp Paper	6	600	36	17	33	0		NO
54 K 80486	81	158	Rag Paper	4	800	31	36	19	3		NO
54 K 80485	23	103	Rag Paper	5	600	20	39	38	0,7		NO
54 K 28974	27	70	Ground Wood Paper	N/A		27	15	188	0,1		NO
54 K 28974	35	70	Ground Wood Paper	N/A		31	16	210	0,2		NO
54 K 73046	15	65	Ground Wood Paper	N/A		26	4	287	0		NO
54 K 73046	5 (Obsah)	65	Ground Wood Paper	N/A		18	0	290	0		NO
54 K 93124	15	20	Bleached Pulp Paper	4	N/A	37	6	58	0,1		NO
54 K 93124	9	20	Bleached Pulp Paper	4	N/A	35	3	57	0		NO
54 H 298601	67	273	Rag Paper	5	1 600	28	36	26	0,3		NO
K 7471	51	242	Ground Wood Paper	4		36	20	294	0,1		NO
54 H 303249	141	356	Bleached Pulp Paper	4	N/A	53	18	53	0		NO
54 H 303249	83	365									