

Univerzita Karlova v Praze

Filozofická fakulta

Ústav informačních studií a knihovnictví

Informační věda – Informační studia a knihovnictví

Mgr. Andrea Fojtů

**Strategie, návrh, řízení a administrace rozsáhlých
digitálních knihoven a archivů**

**Strategy, design, management and administration of large
Digital Libraries and Archives**

Teze disertační práce

Vedoucí práce – doc. PhDr. Richard Papík, Ph.D.

2014

Klíčová slova: digitální dokument, digitální objekt, životní cyklus digitálních dokumentů, digitální knihovna, rozsáhlá digitální knihovna, digitální repozitář, digitální archiv, Referenční model OAIS, DELOS Referenční model digitálních knihoven, dlouhodobá ochrana, certifikace, audit, Best Practices

Anotace: Cílem disertační práce bylo sestavit obecně použitelnou strategii pro návrh, administraci a řízení rozsáhlých digitálních knihoven, repozitářů a archivů.

Abstrakt (CZ)

Předložená disertační práce „Strategie, návrh, řízení a administrace rozsáhlých digitálních knihoven a archivů“ se, jak vypovídá samotný název, soustřeďuje na přehled standardních a nejčastějších postupů či strategií při tvorbě a řízení digitálních knihoven, archivů, ale i repozitářů nikoli z technického, nýbrž z knihovnicko-informačního hlediska.

Teoretická část práce nabízí základní orientaci v komplexní problematice. Tu ulehčují nejrůznější typy modelů - od referenčních, přes objektové a procesní modely - až po ty funkční. Zachycují budoucí stav, ke kterému chceme dospět a představují abstraktní model reálného systému. Proto práce podrobně rozebírá dva nejdůležitější modely: ISO standard 14721:2003 Otevřený archivní informační systém a de facto standard DELOS Referenční model digitálních knihoven, jenž zastupují část návrhů a současně i možné strategické přístupy u digitálních knihoven, repozitářů a archivů.

Ve fázi návrhů je nutné myslet i na datový model, datový plán a respektovat pravidla informační architektury, jež se dotýkají i systémů pro organizaci znalostí (KOS) a do značné míry ovlivňují výslednou podobu uživatelského rozhraní, které má být dostatečně přívětivé, přehledné a mělo by se ním snadno pracovat.

To vše musí směřovat k důvěryhodným repozitářům a archivům, přičemž důležité je porozumět životnímu cyklu digitálních dat s využitím dostupných „open source“ či komerčních systémů pro řízení a administraci digitálních knihoven, repozitářů a archivů. Jen důvěryhodné systémy dokáží zabezpečit dlouhodobý přístup a ochranu svých digitálních objektů. Průkaznost důvěryhodnosti je postavena na auditu a následné certifikaci, čímž završují (nikoli ale nutně ukončují) práce na návrzích a administraci digitálních knihoven, repozitářů a archivů.

Praktická část disertační práce má podobu „Best Practices“, které popisují 10 základních okruhů pro rychlé zorientování se v dané oblasti. Je pro ně charakteristická maximální obsahová výtěžnost a co nejstručnější forma se zaměřením na praxi. Vychází z průzkumů a osobních rozhovorů se zástupci několika větších českých repozitářů, archivů a knihoven. Jejím cílem je poskytnout zejména managementu a netechnickým (i knihovnickým a informačním) pracovníkům digitálních repozitářů a archivů instrukce pro jejich budoucí či současný informační systém.

Keyterms: digital document, digital object, digital document life-cycle, digital library, Large Digital Library, VLDL, digital repository, digital archive, Reference model OAIS, The DELOS Digital Library Reference Model, digital preservation, long-term preservation, certification, audit, Best Practices

Anotation: The aim of the presented dissertation thesis lies in creating a general applicable strategy for design, administration and management large digital libraries, repositories and archives.

Abstract (EN)

The presented dissertation thesis “Strategy, design, management and administration of large Digital Repositories (Libraries) and Archives”, as the very name of it reveals, provides an overview of the standard and most common techniques, strategies in the design and management of digital libraries, repositories or archives from a librarianship’s and information science’s point of view.

The theoretical part presents with basic knowledge of this complex issue and various types of models, from the reference, object and process models to the functional ones. They represent the future and wanted state and stand for an abstract model of a real system. Therefore, this thesis deals with the two most important models, the ISO standard 14721:2003 Open Archival Information System and de facto standard DELOS Reference Model of Digital Libraries, which represent part of the design processes and strategic approaches to digital libraries, repositories and archives.

In the design phase, it is necessary to come up with a good data model, data plan and respect the rules of information architecture, including knowledge management with Knowledge Organization Systems (KOS) and to a large extent influence the final form of the user interface, which should be sufficiently friendly, clear and it should be easy to work with.

This should lead to trusted repositories and archives as well as being able to understand the digital data life cycle, while using available open source or commercial systems for the management and administration of digital libraries, repositories and archives. Only the trusted systems can secure ongoing access and long-term preservation of their digital assets. The sufficient evidence of trust can be taken from audit processes and subsequent certification, which complete (but not necessarily finish) works on the design and administration of digital libraries, repositories and archives.

The practical part of the thesis takes the form of Best Practices and describes 10 basic points for a quick reference in the field. It is typical for practices to be to the point and closely focused on real issues. Therefore they are based on surveys and interviews with representatives of several large Czech repositories. Its aim is to provide a non-technical management in particular employees and library and information professionals of mainly digital repositories and archives with key instructions for their future or current information system.

Předmluva

Jak ostatní informační systémy, tak i digitální repozitáře (knihovny) prošly během své poměrně krátké existence radikálním vývojem. Digitální knihovny nabízejí pomocí internetu přímý přístup k digitálním nebo zdigitalizovaným dokumentům, kdy jejich původní předloha byla převedena do digitální podoby. Díky digitálním knihovnám mají uživatelé v řadě případů přístup k ohroženým, historickým nebo unikátním dokumentům.

Z důvodu nejednoznačných definic a rozdílů v definicích jsou pro účely této disertační práce termíny digitální knihovna, digitální repozitář, digitální archiv, virtuální knihovna, elektronická knihovna a hybridní knihovna významově zaměnitelné a označují tak „organizaci, která může být virtuálního charakteru a jejímž účelem je komplexní sběr, správa a dlouhodobá ochrana bohatého digitálního obsahu“ [AMERICAN, 2007]. Větší digitální knihovny, repozitáře a knihovny jsou v této práci chápány a definovány jako instituce, které mohou se svou digitální knihovnou, repozitářem či archivem absolvovat alespoň jeden stupeň certifikace podle Evropského rámce pro certifikaci a audit digitálních repozitářů.

Před vlastním výběrem tématu disertační práce bylo důležité, aby výzkum s ním spojený byl užitečný a prospěšný pro praxi, aby měl poznatelný, resp. identifikovatelný předmět. Na počátku výzkumu byla stanovena tzv. pracovní hypotéza, která tvrdí, že: „Neexistuje ucelený přehled problematiky v českém jazyce, ideálně pro specifické české podmínky“, která byla dále přetavena do reálné hypotézy: „V ČR není kladen velký důraz na dodržování knihovních standardů při sestavování digitálních objektů (dokumentů), při návrzích a tvorbě digitálních knihoven, repozitářů a archivů. Možnou příčinou se zdá být i multidisciplinární záběr problematiky, rapidní technologický vývoj a příliš mnoho informací. Avšak standardy hrají vitální roli v efektivní a bezpečné správě, administraci, zpřístupňování a dlouhodobé ochraně obsahu digitálních knihoven, repozitářů a archivů.“ Tato hypotéza společně s definovanými tezemi (jež jsou součástí tohoto dokumentu), stanovily záběr, strukturu a samotný obsah předložené disertační práce. Navíc byla doplněna o množství obrazových materiálů, přehledných tabulek, příloh a poznámek pod čarou odkazujících na podrobnější informace či další doporučenou literaturu.

Zatímco v minulosti byly rozsáhlé digitální knihovny, repozitáře a archivy vnímány pouze jako „knihovny“ na národní či mezinárodní úrovni, dnes bývá rozsah dán především obsahem než jejich geografickým působením. Implementace moderních digitálních knihoven, repozitářů a archivů je v současné době více náročná než kdykoliv předtím a proto cílem této práce je analyzovat současný stav v oblasti problematiky (přehledový a vysvětlující profil) a v neposlední řadě sestavit obecně použitelnou strategii pro návrh, administraci a řízení rozsáhlých digitálních knihoven, repozitářů a archivů.

Seznam tezí

Teze 1

Digitální archiv/ knihovna by měla být chápána jako organizace, která uchovává, zpřístupňuje informace v primárně digitální (born-digital) či zdigitalizované podobě.

Teze 2

Architektura systémů pro digitální knihovny, repozitáře a archivy by měla vycházet z některého ze zavedených, ideálně standardních modelů.

Teze 3

Implementace informačních balíčků modelu OAIS je klíčová pro interoperabilitu.

Teze 4

Důležitější než samotná volba - komerční vs. „open source“ systémy - je dodržování standardů a cesta k důvěryhodným digitálním knihovnám, repozitářům a archivům

Teze 5

Digitální knihovny, repozitáře a archivy vznikají a fungují pro své uživatele, proto uživatelská rozhraní musí splňovat kritéria přístupnosti a použitelnosti.

Teze 6

Rozsáhlejší digitální knihovny, repozitáře a archivy představují komplexní záležitost, kterou ovlivňuje dobré projektové řízení jak na institucionální, tak funkcionální úrovni.

Teze 7

Funkcionalita plnotextového vyhledávání nenahradila systémy pořádání znalostí. V digitálních repozitářích mají svou nezastupitelnou funkci.

Teze 8

Nelze uchovávat vše, je nutné vybrat klíčové objekty a stanovit jejich signifikantní vlastnosti, přičemž proprietární formáty souborů způsobují značné komplikace při dlouhodobé ochraně digitálních objektů.

Teze 9

Pro efektivní chod digitálních knihoven, repozitářů a archivů je nevyhnutné pochopení životního cyklu digitálních dokumentů a dlouhodobé ochrany.

Teze 10

Posledním krokem před vypracováním konečné verze plánu dlouhodobé ochrany digitálních objektů je kontrola (audit) a případná certifikace předcházejících postupů a výsledků.

Strategie rozsáhlejších digitálních knihoven, repozitářů a archivů

Teze 1

Digitální archiv/ knihovna by měla být chápána jako organizace, která uchovává, zpřístupňuje informace v primárně digitální (born-digital) či digitalizované podobě.

Digitální knihovny, repozitáře a archivy nejsou jen pouhým úložištěm, naplněným digitálními daty, resp. digitálními objekty. Velmi důležitou roli zde ale hraje i lidský faktor (správci, management, uživatelé apod.), **postupy, procesy, finanční zabezpečení, komunity** atd. DELOS Referenční model digitálních knihoven definuje digitální knihovnu jako **“organizaci, která může být i virtuálního charakteru, jež přijímá, spravuje a dlouhodobě uchovává digitální obsah za účelem zpřístupňování uživatelským komunitám”** [CANDELA, 2007]. Jedná se v podstatě o informační systém, jehož součástí jsou podsystemy pro vytváření, zprostředkování a využívání informací.

Návrh a design rozsáhlejších digitálních knihoven, modely digitálních knihoven, repozitářů a archivů

Teze 2

Architektura systémů pro digitální knihovny, repozitáře a archivy by měla vycházet z některého ze zavedených, ideálně standardních modelů.

Pochopit komplexnost a záběr problematiky digitálních repozitářů, knihoven a archivů není snadné. Situaci však ulehčují nejrůznější typy modelů, od těch referenčních, přes objektové a procesní modely až po ty funkční. Na obecné úrovni se jedná o tzv. konceptuální modely (*Conceptual Model*), jež představují mentální model, zachycující koncepty (entity) a vztahy mezi nimi s cílem snadnějšího pochopení problému, resp. oblasti, kterou reprezentují. Typickým představitelem modelu digitálních knihoven tohoto typu je model Kahn-Wilenskeho architektury. Referenční modely (*Reference Model*) digitálních knihoven, repozitářů a archivů zachycují vždy budoucí, chtěný stav a představují abstraktní model reálného otevřeného systému. Příkladem je referenční model Otevřeného archivního informačního systému (*OAIS - Open Archival Information System*) a DELOS Referenční model digitálních knihoven.

Všechny tyto modely přistupují k digitálním knihovnám, repozitářům a archivům odlišným způsobem. Např. v popředí Kahn-Wilenskeho architektury, stojí digitální objekt s persistentními identifikátory, jež tvoří základ digitálních knihoven. Zajímavostí Kahn-Wilenskeho architektury je rozlišování mezi tzv. „master“ souborem (*Master File*), který můžeme považovat za archivní verzi a souborem pro zpřístupňování (*Access File*). Master soubory jsou obvykle (nikoliv nutně) výsledkem digitalizace velkého objemu (např. TIFF skeny) ve vysoké kvalitě. Pro uživatele se zobrazuje soubor pro zpřístupňování (např. JP2 nebo JPG).

Digitální objekty však nejsou jedinou podstatnou složkou digitálních knihoven, repozitářů či archivů. Významnou roli zde hrají lidé (specializovaný personál, uživatelé), zařízení, softwarové systémy, postupy, komunity apod. DELOS Digital Library Reference Model proto definuje třívrstvou konstrukci digitálních knihoven, kterou tvoří digitální knihovna na obecné úrovni, vybraný systém digitální knihovny a systém pro správu digitální knihovny (na technické úrovni). Tento model je průlomový především definováním prostředí digitálních knihoven, ve kterých působí a dlouhodobou ochranou (*Long-term Preservation*), která zásadním způsobem ovlivňuje dostupnost (*Availability*) digitálních objektů. Tuto myšlenku dále rozvíjí Referenční model Otevřeného archivního informačního systému (OAIS), který v rámci svého modelu definuje (zejména pro repozitáře a archivy) modul pro dlouhodobou ochranu (Plánování dlouhodobé ochrany – *Preservation Planning*). Velkou předností tohoto modelu je též vymezení terminologického základu, stanovení tzv. informačních balíčků a popis funkčních entit. **V současné době v odborných kruzích se téměř výlučnou normou pro digitální repozitáře a archivy stal právě OAIS model (viz Teze 3).**

Teze 3

Implementace informačních balíčků modelu OAIS je klíčová pro interoperabilitu.

Klíčovou podmínkou pro interoperabilitu mezi digitálními knihovnami, repozitáři a archivy navzájem nebo jinými informačními systémy je zavedení informačních balíčků SIP (Submission Information Package), DIP (Dissemination Information Package) a AIP (Archival Information Package). OAIS model nepreferuje žádný konkrétní způsob jejich implementace, ponechává volnou ruku tvůrcům.

Balíček představuje logický kontejner (princip podobný METS standardu), jež obsahuje vlastní informaci (Content Information), která má být dlouhodobě uchovávána a preservační informaci (Preservation Description Information), která formou metadat podporuje dlouhodobou ochranu. Obsah informace a preservační informace je spojen do logického celku, který se nazývá archivační informace (Packaging Information) [CUBR, 2012]:

- Dodavatelský balíček – SIP se v kontextu OAIS modelu dostává do digitálního archivu od producentů/tvůrců dat. Jeho forma je obvykle předem dojednána mezi producenty dat a přijímající stranou za archiv, nejčastěji formou smluv a dohod, které podrobně specifikují, jaké dodatečné (např. popisné) informace budou spolu s digitálními daty v balíčcích dodávat. SIP nemusí nutně představovat rovnici jeden digitální objekt = jeden SIP.
- Archivační balíček – AIP je vytvořen systémem po úspěšném přijetí a zpracování SIP. Systém ale nemění přijaté SIP, pouze k nim přidává doplňující archivační informace za účelem jejich dlouhodobé archivace.
- Uživatelský balíček – DIP reprezentuje uživateli požadované digitální dokumenty (na základě dotazu/ů). Nemusí nutně obsahovat veškeré archivační informace, které obsahuje AIP. Uživatele zpravidla nezajímají záznamy ochranných aktivit, provedených během jejich životního cyklu v digitálním repozitáři.

SIP, AIP a DIP nemusí být nezbytně rozdílné, to závisí na konkrétní implementaci modelu.

Velmi zjednodušeně lze konstatovat, že většina SIP metadat (popisných, technických a administrativních) může poskytnout producent, který do systému uploaduje své digitální objekty. Poté tyto balíčky mohou projít nástroji pro automatickou extrakci metadat (např. JHOVE, DROID, či New Zealand Metadata Extraction Tool) a validací (kontrola přítomnosti virů, kontrola autenticity a integrity apod.). Následně je balíček SIP přetransformován do archivního balíčku AIP. Další informace jsou přidány ve formě metadat (např. změny, které se na daném objektu prováděly, kdo a kdy je provedl). V případě potřeby je AIP změněn na DIP a zpřístupněn podle specifikace a přístupových práv.

Teze 4

Důležitější než samotná volba - komerční vs. „open source“ systémy - je dodržování standardů a cesta k důvěryhodným digitálním knihovnám, repozitářům a archivům

Důvěra uživatelů je jedním z klíčových aspektů knihoven, repozitářů a archivů jakéhokoliv typu a obsahu. Nejčastěji ji definujeme jako spolehlivost, víru či předpoklad, že osoba nebo organizace bude dodržovat rámec společných hodnot a představ. **Za důvěryhodný digitální repozitář možno označit repozitář, kterého posláním je poskytovat spolehlivý a dlouhodobý přístup k organizovaným digitálním zdrojům v cílové skupině - dnes i do budoucna.** Takový repozitář porozumí hrozbám i rizikům správy digitálních informací, přičemž důležité je vědět, že nejde ani tak o konkrétní řešení jako o koncepční, systematický přístup a dodržování standardních postupů a příslušných norem. **Systém sám o sobě nezajistí bezproblémový chod, pokud jsou opomíjeny ostatní problémy (personální, právní, technické, finanční, managerské, projektové apod.).**

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). 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 [FOJTŮ, 2010]. Za překotným vývojem „open source“ systémů a nástrojů dlouhodobé ochrany digitálních dokumentů stály do značné míry komerční systémy a jejich první „pokusy“ implementace ISO normy 14723:2003 do reálně fungujících informačních systémů. Mezi nejznámější komerční řešení patří Rosetta (Ex Libris) a SDB (Tessella).

Uživatelsky přívětivá rozhraní, uživatelsky orientovaný design

Teze 5

Digitální knihovny, repozitáře a archivy vznikají a fungují pro své uživatele, proto

uživatelská rozhraní musí splňovat kritéria přístupnosti a použitelnosti.

Uživatelská rozhraní do značné míry rozhodují, jak bude systém použitelný a tedy i úspěšný, proto je nutné věnovat jeho návrhu zvýšenou pozornost. **Přístupnost znamená takový stav, kdy daná věc neklade svým uživatelům žádné překážky** [ŠPINAR, 2004]. **Použitelnost lze v obecnějším chápání brát jako takový přístup k vývoji výrobku, který předpokládá přímou zpětnou vazbu uživatele během výrobního cyklu, s cílem snížení nákladů a naplnění potřeb uživatelů** [MAKULOVÁ, 2007].

Pro vytvoření digitální knihovny, repozitáře a archivu splňujících kritéria použitelnosti, je nutné se zaměřit zejména na tyto základní složky, vycházející mimo jiné i z informační architektury: obsah, funkčnost/funkce (navigace, vyhledávání) a rozhraní. Kling [1994] však zdůrazňuje nejen použitelnost rozhraní (*Interface Usability*), ale i tzv. organizační použitelnost (*Organizational Usability*). Organizační použitelnost určuje možnosti efektivního začlenění systémů do celkové struktury organizace. To vše ale není jednoduché podchytit, jelikož lidské, resp. uživatelské chování je složité a technologie pro digitální knihovny prochází neustálým a rychlým vývojem.

Rozhraní by ideálně mělo být jednotné a konzistentní napříč systémem konkrétní digitální knihovny, repozitáře či archivu. Uživatelsky orientovaný design (UCD - *User-centred design*) je design zaměřen na uživatele. Představuje moderní, rychle se rozvíjející filosofii designu, která má za to, že dobrý design musí sloužit především zájmu uživatelů, přičemž uživatelské požadavky jsou od začátku (životního cyklu tvorby uživatelského rozhraní) zaznamenávány a následně doladovány prostřednictvím různorodých metod (např. etnografických studií, testování prototypů apod.).

Řízení, projektové řízení rozsáhlejších digitálních knihoven, repozitářů a archivů

Teze 6

Rozsáhlejší digitální knihovny, repozitáře a archivy představují komplexní záležitost, kterou ovlivňuje dobré projektové řízení jak na institucionální, tak funkcionální úrovni.

K tomu, aby projekt mohl být úspěšně dokončen (a tím pádem i hlavní cíl naplněn či problém vyřešen) je nutné využít metod projektového řízení. V anglosaské literatuře se nejčastěji setkáme s termínem *Project management*. Do češtiny jej lze přeložit dvěma způsoby - jako řízení projektu, případně projektové řízení. **Projektové řízení je často chápáno jako filosofie přístupu k řízení projektu s jasně stanoveným cílem, který musí být dosažený v požadovaném čase, s určenými náklady a v požadované kvalitě. Projektový cíl se skládá ze tří základních podmínek, proto také někdy bývá označován jako trojpodmínka či trojimperativ. Jednotlivé podmínky - provedení (kvalita výstupu), čas a náklady - jsou na sobě závislé. Tedy změna jedné podmínky nutně vyvolá změnu dvou zbývajících. V ideálním případě jsou všechny tři roviny naplněny současně a nedochází ku příkladu k preferenci času na úkor kvality a nákladů.**

Během plánování a řešení projektů se nesmí opomenout ani možná rizika, resp. negativní externí faktory, které mohou ovlivnit postup či úspěch projektu. Ani ty, které nejsou příliš pravděpodobné, ale měly by být v maximálně možné míře pod kontrolou. **Řízení rizik (Risk Managment) patří k nejhůře chápaným činnostem v rámci projektů. Podcenění identifikace a potírání rizik však může ohrozit projekt v jakékoliv jeho fázi.**

Výběr nejvhodnější organizační formy projektového řízení ovlivňují zejména následující faktory: struktura projektu, jeho fáze a rozsah řešených problémů, způsob zapojení organizačních struktur účastníků projektů, míra institucionalizace subjektu projektového řízení. Dále právní, ekonomická a jiná omezení a formy. Systém legislativních, ekonomických, ale také "zvykových" norem vytváří specifické podmínky pro organizaci projektového řízení, potenciál spolupracovníků, míru ochoty a schopnost spolupráce účastníků projektu a v neposlední řadě úroveň informačních systémů účastníků projektu.

Řízení projektů je ale především o práci s lidmi a za její podcenění se draze platí! Spousta projektů je dnes svěřena lidem, kteří s řízením takto specifických aktivit nemají zkušenosti. Většinou bývají odborníky v oboru, skrze který bude získán výstup projektu. Ovšem odborník v oboru nemusí být vždy zdatným manažerem schopným uřídit tak nestabilní a nejistotou zatížený prvek, jakým je projekt obecně. Existuje však i opačný extrém, kdy se největší důraz klade nikoliv na lidskou složku, ale na používání nejrůznějších nástrojů a zpracování co nejvíce projektových metodik s cílem eliminovat rizika nesplnění termínů, kvality a nákladů, přičemž uniká samotná podstata řízení a řešení projektu.

Administrace digitálních knihoven, repozitářů a archivů a systémy pořádání znalostí

Teze 7

Funkcionalita plnotextového vyhledávání nenahradila systémy pořádání znalostí. V digitálních repozitářích mají svou nezastupitelnou funkci.

Informační architektura je pro digitální repozitáře, knihovny a archivy velmi důležitým oborem, jelikož sdílí stejný cíl – zaměřují se, kromě jiného, na vyhledatelnost informací, resp. digitálních objektů. Koncepty a principy informační architektury potřebují částečně odlišný přístup v aplikaci, jelikož digitální knihovny disponují nejen uživatelským rozhraním, ale i sbírkami, které mají efektivně uspořádat (obsah většinou není generovaný interně), zpřístupňovat a uchovávat [HOWLETT, 2011].

Systémy pořádání znalostí (KOS - Knowledge Organization System) mají svou dlouhou historii v klasickém knihovním prostředí, jsou běžnou součástí knihovních systémů a mají své neméně důležité postavení i v digitálních repozitářích. Pořádání či organizace znalostí (*Knowledge Organization*) je oblast lidského poznání, která se zabývá uspořádáním jednotek znalostí a objektů všech typů. Ty by měly odpovídat pojmům (teoriím) nebo pojmovým třídám tak, aby byly zachyceny znalosti o poznaném světě v uspořádané podobě spolu s umožněním rozšiřování těchto znalostí za účelem jejich využití [JONÁK, 1999]. **Jejich existence je důležitá**

pro uživatele, kteří díky nim dokážou snáze definovat své informační potřeby, informační požadavky a rychleji dohledat relevantní/pertinentní informace a znalosti v čase a místě, které potřebují. Abychom snáze pochopili organizační hledisko KOS, je vhodné se obeznámit s OAS referenčním modelem (ISO 14721:2003). Tento model nabízí kompletní funkční a informační popis digitálního repozitáře.

Teze 8

Nelze uchovávat vše, je nutné vybrat klíčové objekty a stanovit jejich signifikantní vlastnosti, přičemž proprietární formáty souborů způsobují značné komplikace při dlouhodobé ochraně digitálních objektů.

Termín signifikantní vlastnosti v kontextu dlouhodobé ochrany digitálních dat (significant properties, significant characteristics, transformational information properties nebo také essence) byl poprvé zaveden v projektu CEDARS. Vyjadřuje důležité charakteristiky (komponenty) informačního objektu, které mají být uchovány tak, aby jeho použitelnost, přístupnost a srozumitelnost byla zabezpečena i do budoucna [WILSON, 2007]. **Základní požadavky na dlouhodobou ochranu je možné rozdělit do pěti základních okruhů, resp. kategorií: kontext (Context), obsah (Content), chování (Behavior), vzhled (Appearance) a struktura (Structure).** Z nich se poté „extrahují“ signifikantní vlastnosti.

Výběru signifikantních vlastností by však měl předcházet výběr samotných zdrojů (nebo jejich typů) k uchování, pro který se v anglo-amerických kruzích ustálil termín *Appraisal*. Do češtiny by se dal volně přeložit jako (kontrolovaný) výběr. Ten představuje proces evaluace aktivit, kterým se určuje jaké dokumenty a na jak dlouho mají být uchovány (při respektování organizačních pravidel a očekávání komunity). Základní stavební kameny kontrolovaného výběru představují kompletní analýzu dokumentů, hodnocení aktivit a jejich právních omezení, určení reálnosti uchování dokumentů a konečné rozhodnutí [FOJTŮ, 2011].

V posledních letech je čím dál tím víc dokumentů tvořeno primárně v digitální podobě, nebo vznikají jako výsledek digitalizačních snah. Nové druhy softwaru a hardwaru také ovlivňují vznik nových digitálních objektů. Tyto digitální dokumenty však vznikají v různých formátech na nejrůznějších (i vzájemně nekompatibilních) platformách a způsobují tak jejich přílišnou závislost na svých původcích/producentech.

Dlouhodobá ochrana digitálních dokumentů

Teze 9

Pro efektivní chod digitálních knihoven, repozitářů a archivů je nevyhnutné pochopení životního cyklu digitálních dokumentů a dlouhodobé ochrany.

Digitální dokumenty představují velmi komplexní a často diskutovanou problematiku, kterou možno lépe pochopit prostřednictvím modelu životního cyklu digitálních dokumentů. Ten

totiž zasahuje do organizační infrastruktury, administrace systému repozitáře a managementu rizik. Postupy správy digitálních objektů, technologický rámeček a zabezpečení dat musí být na rozumné úrovni a adekvátní úkolům a cílům repozitáře. Identifikovaná kritéria vychází především z právě vznikajícího referenčního modelu OAIS.

Podle DCC (*Digital Curation Centre*) modelu životního cyklu jsou klíčovým prvkem data, resp. digitální objekty či databáze (*Data Digital Objects Or Databases*). Data, jako libovolné informace v binární podobě, tvořící jednoduché až komplexní digitální objekty (textové a obrazové soubory společně s jejich identifikátory a metadaty). Pro tyto digitální objekty musí existovat dostatečný popis (informace ve formě reprezentace) pomocí popisných, technických, strukturálních, administrativních či ochranných metadat. Díky plánování a specifickému plánu dlouhodobé ochrany mají takto popsané objekty větší šanci na permanentní archivaci. Konkrétní preservační aktivity a nástroje se díky spolupráci v komunitě mohou vylepšovat a snadněji obměňovat. Tyto výše uvedené aktivity se prolínají celým životním cyklem a podporují je následující sekvenční aktivity: tvorba či příjem (*Create Or Receive*), výběr (*Appraise And Select*), import (*Ingest*), ochranné aktivity (*Preservation Action*), uložení (*Store*), zpřístupnění či využití (*Access, Use And Reuse*) a případná transformace (*Transformation*), která pomyslně uzavírá cyklus a eventuálně začíná celý proces nanovo [ŠVÁSTOVÁ, 2014].

Životní cyklus dlouhodobé ochrany by měl být zaveden již na počátku sestavování interního plánu, resp. politiky dlouhodobé ochrany, kdy je nutné vyhodnotit současnou situaci. Pokud repozitář ještě neexistuje, je možné vybrat vhodný nástroj a otestovat jej. Poté objektivně vyhodnotit auditem, externím nebo interním samo-auditem a sepsat patřičnou dokumentaci. Pokud se objeví případné nedostatky v politice, je nutné je upravit. Po určitém čase či zásadnější změně začíná hodnocení od začátku.

Cílem dlouhodobé ochrany není uchovávaní veškerého obsahu, nýbrž stanovení pravidel a podmínek za jakých dokumenty budou uchovávány. Slouží k tomu Plán dlouhodobé ochrany (*Preservation Plan*). Definuje ochranné akce (označovány jako akční plán dlouhodobé ochrany), které má zodpovědná instituce provádět, aby dokázala úspěšně potírat rizika, která se v průběhu dlouhodobé ochrany digitálních dokumentů (potažmo sbírek) můžou vyskytnout. Vymezuje též politiku provozu, právní závazky, organizační a technická omezení v kontextu uživatelských nároků a potřeb.

Certifikace a audit digitálních knihoven, repozitářů a archivů

Teze 10

Posledním krokem před vypracováním konečné verze plánu dlouhodobé ochrany digitálních objektů je kontrola (audit) a případná certifikace předcházejících postupů a výsledků.

Posledním krokem před vypracováním konečné verze plánu dlouhodobé ochrany digitálních dokumentů je kontrola (audit) předcházejících postupů a výsledků. Kontrola

nejčastěji probíhá pomocí ověřených auditních a certifikačních nástrojů. Mezi nejznámější metriky, certifikace a auditní nástroje patří:

- Preserving Digital Information – 1996
- Trustworthy Digital Repositories - 2002
- Catalogue of Criteria for Trusted Digital Repositories (nestor Catalogue) – 2006, 2009 (později schváleno jako německý standard DIN 31644 Criteria for trustworthy digital archives)
- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) - DCC and DPE - 2007
- Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) – 2007
- Data Seal of Approval (DSA) – 2008.

Výběr konkrétního auditního nástroje a typu certifikace je podmíněn personálním obsazením, materiálovým vybavením a v neposlední řadě finančními zdroji. Všechny tyto metodiky, auditní a certifikační nástroje slouží především pro instituce uchovávající kulturní dědictví – knihovny, muzea a archivy – a hodí se jak v přípravných (pro kontrolu návrhů), tak post-implementačních fázích běžného chodu repozitáře (pro kontrolu implementace). Poslouží však i komerčním a nekomerčním poskytovatelům služeb, vývojářům software a dodavatelům třetích stran pro případnou spolupráci s kulturními institucemi při návrhu, vývoji a provozu technologií pro digitální repozitáře. V neposlední řadě prokáže službu všem zájmovým skupinám důvěryhodnosti repozitáře:

- uživatelům repozitáře, kteří potřebují důvěryhodný zdroj informací dnes i do budoucna
- producentům dat a poskytovatelům obsahu, pro které certifikace repozitáře představuje záruku správně zvoleného partnera
- finančním institucím a grantovým agenturám či jiným organizacím, pro které důvěryhodnost repozitáře znamená garanci správně investovaných prostředků
- samotným digitálními repozitářům, které auditem a certifikací prokáží své odpodstatnění.

Cílem auditu je kritické zhodnocení jak pozitivních, tak negativních aspektů funkčnosti digitálních knihoven, repozitářů a archivů, nikoliv pouze samoučelné hledání chyb. DRAMBORA interactive představuje metodologii „sebe-hodnocení“ (*Self-Assessment*) repozitáře, které je (v kontextu projektu DRAMBORA) vnímáno jako interní audit. Na rozdíl od jiných auditních a certifikačních metod je více zaměřen na kvantitativní než kvalitativní hodnocení. Zastupuje metodologický přístup k hodnocení důvěryhodnosti repozitáře. DRAMBORA chápe digitální ochranu (*Digital Curation*) jako aktivitu managementu rizik, v rámci kterého je nutné podchytit nejistoty a hrozby, které ohrožují autenticitu, integritu, pochopení digitálních objektů a jejich proměnu na rizika, které je možné zachytit a do budoucna odstranit.

Závěr

Disertační práce prokázala platnost předložených tezí. Byl zanalyzován současný stav, návrhy, administrace a řízení v oblasti problematiky digitálních knihoven, repozitářů a archivů. Disertační práce má především přehledový a vysvětlující charakter problematiky a zpřístupňuje nejdůležitější poznatky v této multidisciplinární problematice.

Důraz je kladen na důležitost výběru signifikantních vlastností dig. objektů pro efektivní dlouhodobou ochranu. Není totiž možné dlouhodobě uchovávat vše. Je nutný výběr dokumentů, včetně jejich signifikantních vlastností. Praktická ukázka (sestavena autorkou) možných stanovených klíčových vlastností objektů je dostupná formou myšlenkové mapy. Neméně důležitý je i souhrn nejdůležitějších standardů a de facto standardů podle problematických okruhů. Dále šlo o nasměrování pozornosti na uživatelská rozhraní z pohledu uživatelsky orientovaného designu a na požadavky uživatelů na digitální knihovny, repozitáře a archivy (včetně získávání zpětné vazby).

V průběhu zpracování práce byla hlavní hypotéza potvrzena na základě dotazníkového šetření (mezi uživateli/studenty i institucemi), studia dostupné odborné literatury a komunikace s odborníky nejen na konferencích. Výzkum (především dotazníkové šetření a interview) v této disertační práci ukázal, že standardy pro dlouhodobou ochranu nejsou téměř vůbec dodržovány (potvrdila se tak i pracovní hypotéza), což je velmi zneklidňující. Obdobně je to i se standardy v oblasti návrhu architektury digitálních knihoven a repozitářů, sestavování digitálních objektů, jako i při návrzích uživatelského rozhraní a certifikace procesů uvnitř knihoven a archivů. Knihovny rozlišují pojmy digitální knihovna a digitální repozitář, kdy druhý zmíněný je buď přesný protějšek kamenné knihovny či archivu nebo má na rozdíl od digitální knihovny nastavené procesy pro dlouhodobou ochranu a jim odpovídající nástroje.

Pro knihovny a archivy, jež ještě nemají implementovaný repozitář s prvky dlouhodobé ochrany byly autorkou sestaveny následující „Best Practices“, které prokrývají problematické body z dotazníku a představují výzkumnou součást práce. Slouží k rychlé orientaci v komplexní problematice.

Jde tedy o pomůcku, návrh či doporučení v podobě „Best Practices“ pro obecné strategie, plány, management a administraci v českých (případně slovenských) digitálních knihovnách, repozitářích a archivech.

I když je zřejmé, že díky rapidnímu vývoji technologií a získávání nových poznatků v multidisciplinární problematice repozitářů může dojít k částečnému posunu aktuálnosti této disertační práce. Přesto **je možné konstatovat, že předložená disertační práce odpovídá na otázky a plní cíle, které stály na počátku jejího vzniku a ukazuje tak možné směřování do budoucna.**

Přehled citovaných informačních zdrojů pro teze disertační práce

Seznam citovaných zdrojů je řazen abecedně dle prvního údaje v záznamu (záhlaví), seznam není číslován. Pro lepší přehlednost je v seznamu literatury každé záhlaví zvýrazněno. Jednotlivé záznamy jsou v souladu s pravidly uvedenými v normě ČSN ISO 690 (01 0197) platné od 1. dubna 2011. Výjimkou je uvádění roku vydání, který je uveden oproti normě dvakrát, jednou za záhlavím a jednou za údaji o vydavateli a to z důvodu rychlejší orientace v seznamu.

AMERICAN LIBRARY ASSOCIATION. 2007. Fostering media diversity in libraries strategies and actions [online]. American Library Association, June 2007 [cit. 2009-09-05]. Dostupné z: http://www.ala.org/ala/aboutala/offices/oif/ifissues/fostering_media_div1.pdf

CANDELA, L. et al. 2007a. Setting the Foundations of Digital Libraries. *DLib Magazine* [online]. Vol 13, no 3/4. [cit. 2012-11-08]. ISSN 1082-9873. Dostupné z: <http://www.dlib.org/dlib/march07/castelli/03castelli.html>

CUBR, Ladislav. 2010. *Dlouhodobá ochrana digitálních dokumentů*. 1. vyd. Praha: Národní knihovna České republiky, 2010. 154 s. ISBN 978-80-7050-588-5.

FOJTŮ, Andrea. 2010. *Open source nástroje pro dlouhodobou ochranu digitálních dat: Podklad pro závěrečnou zprávu grantu Výzkum a vývoj*. Praha: Národní knihovna ČR, prosinec 2010.

FOJTŮ, Andrea. 2011. *Plánování dlouhodobé ochrany pomocí nástroje PLATO Preservation Planning Tool: Podklad pro závěrečnou zprávu grantu Výzkum a vývoj*. Praha: Národní knihovna ČR, září 2011.

HOWLETT, Allison. 2011. *Information Architecture and Digital Libraries (Part 1)* [online]. June 11, 2011 [cit. 2013-08-26]. Dostupné z: <http://acrystelle.com/2011/06/11/information-architecture-and-digital-libraries-part-1/>

JONÁK, Z., PÁLKA, P. a SCHWARZ, J. 1999. Potřeby a možnosti mezinárodní spolupráce v oblasti pořádání znalostí. *Ikaros* [online]. 1999, roč. 3, č. 1. [cit. 2009-09-05]. URN-NBN:cz-ik255. ISSN 1212-5075. Dostupné z: <http://www.ikaros.cz/node/255>

KLING, Rob a ELLIOTT, Margaret. 1994. Digital library design for organizational usability. In: *ACM SIGOIS Bulletin* [online]. 1994 [cit. 2014-04-21]. DOI: 10.1145/192611.192746. Dostupné z: <http://dl.acm.org/citation.cfm?id=192746>

ŠPINAR, D. 2004. *Tvoříme přístupné webové stránky*. 1. vyd. Brno : Zoner Press, 2004. 360s. ISBN 80-86815-11-0.

ŠVÁSTOVÁ, Pavla. 2014. *Digital Curation* [online prezentace]. Brno : Moravská zemská knihovna, 2014 [cit. 2014-21-05]. Dostupné z: <http://www.ics.muni.cz/mba/eiz/eiz11c-6.pdf>