Oponentský posudek disertační práce

Autor disertační práce	PhDr. Mgr. Ladislav Cubr
Název disertační práce	Autenticita a digitální informace
Oponent disertační práce	Mgr. Jan Hutař, Ph.D.

Cíl práce a jeho naplnění

Cílem práce bylo popsat a analyzovat problematiku autenticity digitalizovaných knih. Na základě analýzy autor vypracoval návrh na postup posuzování a udržení autenticity digitalizovaných knih v reálném provozu instituce. Některé části návrhu byly již autorem částečně realizovány (URN:NBN resolver). Z textu je patrné, že autor na tématu pracoval několik let, průběžně výsledky publikoval a v rámci svého zaměstnání v Národní knihovně ČR také zaváděl do procesů digitalizace probíhajících v této instituci.

Volba informačních zdrojů

Autor využil množství dobře vybraných a k tématu relevantních zdrojů jak z českého, tak především světového prostředí. Použity byly jak odborné články, tak relevantní standardy a normy, na kterých do velké míry autor staví a precizně je interpretuje. Zdroje jsou korektně citovány.

Stylistická a formální úprava práce

Disertační práce je psána odborným stylem, za kterým můžeme tušit předchozí autorova sociálně vědní studia. Styl je v některých pasážích těžký a je často nutné se vracet k předchozímu textu pro odpovídající pochopení. Teoretické pasáže, které vysvětlují vztahy mezi entitami (model autenticity v částech 6.3.1 a 6.3.3), by si zasloužily grafické znázornění např. pomocí diagramu. Diagram by také napomohl k pochopení a přehledu vztahů např. při popisu "Preservation Description Information" a struktury informačních balíčků v kapitolách 2.2.2.2 a 2.2.2.3. Zamrzí překlepy a drobné chyby, tyto se ale nevyskytují v nadměrné míře. Některá slovní spojení používaná autorem se nezdají nejvhodnější, např. "požadavek autenticity" by měl spíše být "požadavek *na* autenticitu".

Struktura práce a členění do kapitol je logické a vhodné pro předkládané téma.

Shrnutí obsahu práce

Práce je členěna do deseti částí, včetně úvodu a závěru. Kapitoly 2 až 5 jsou přehledové a lze je označit za teoretickou část. Kapitoly následující, tedy 6 až 9, jsou vkladem autora do problematiky, obsahují analýzu předchozích částí (Kap. 6 Revize konceptů autenticity), autorem vytvořené specifikace (Kap. 7 Specifikace požadavků autenticity pro digitalizáty knih) a návrhy pro vlastní digitalizační praxi (Kap. 8 Návrh postupu pro udržení autenticity digitálizátů knih; Kap. 9 Návrh pravidel perzistentní identifikace digitalizátů knih).

Kapitola první je Úvod, kde autor shrnuje důvody, které jej vedly k výběru tématu, popsány jsou také cíle práce a její struktura.

Kapitola druhá popisuje konceptuální rámec řízení životního cyklu digitálních dokumentů. Je postavena na vysvětlení normy ISO 14721 (OAIS). Tato část je dle mého názoru zbytečně rozsáhlá. Norma je

dostupná a byla popsána již v jiných odborných pracích (disertačních) i v odborných článcích, a to i v českém jazyce. Autor disertace se aktivně účastnil oficiálního českého překladu normy ISO 14721, která byla publikována jako ČSN ISO 14721 v roce 2014. Přesto na str. 10 disertace uvádí, že nebude využívat oficiální terminologii a výrazy české normy. To považuji za nešťastné rozhodnutí, zvláště pokud se to týká i klíčových termínů normy (Administration přeloženo jako "administrace", oficiální překlad je ovšem "správa"). Autor takto bohužel nejde příkladem relevantní české komunitě digitální archivace. Příklady překladů pro ilustraci: termín *Designated Community* přeložen jako "cílová komunita", oficiální překlad používá termín "určená skupina"; termín *Producer* je přeložen jako "producent", v oficiálním překladu normy je ovšem překlad "tvůrce"; *Preservation Description Information* přeloženo jako "archivační informace", oficiální překlad je "informace o uchovávání", atp. Nejednoznačnost překladů může působit problémy, příkladem je část 2.2.4 a překlad typů digitálních migrací, který se od oficiálního překladu podstatně liší (*refreshment* přeloženo jako „renovace", oficiální překlad je „Obnova"; *replication* přeloženo jako „replikace", oficiální překlad je „kopírování").

Kapitola třetí popisuje jednotlivé složky digitalizované knihy (formáty, metadata apod.). Je dobrým přehledem problematiky, také ovšem již mnohokrát popsaným jinde. Popis standardů, teorie a jejich praktické využití v digitalizaci v této kapitole ukazuje velmi dobrý vhled autora do problematiky.

Kapitola čtvrtá se věnuje procesům a aktivitám vyskytujícím se v životním cyklu digitalizovaného dokumentu. V obou kapitolách (3 a 4) čerpá autor ze zkušeností z praxe Národní knihovny ČR, což je velmi cenné a pro práci a jejího čtenáře velmi přínosné.

Kapitola pátá přináší přehled a precizní výklad zdrojů, které se zabývají autenticitou digitálních dokumentů. Tato část je v českém prostředí ojedinělá a dovolil bych si tvrdit, že tématika autenticity digitálních dokumentů v knihovnách, vytvořená na základě dostupných pramenů, byla v českém prostředí autorem zpracována vůbec poprvé. Absence podobné analýzy a vyjasnění bylo ostatně i cílem celé disertační práce a popudem k jejímu napsání, jak autor popisuje v úvodu. Autor správně staví na konceptu autenticity pocházejícím z komunity archivářů a projektu InterPARES. V prostředí knihoven se do podobné míry autenticitou digitálních dokumentů zatím nikdo nezabýval. I přesto, že definice a dokumenty ze světa archivářů jsou podrobně rozebrány, unikla autorovi existence základní normy této komunity – ISO 15489 Records Management, která se autenticitou mj. zabývá. Norma ISO 14721 a standard PREMIS jsou analyzovány dostatečně. Části 5.5 a 5.6 jsou velmi pěknou sumarizací problematiky. Výstupem kapitoly jsou klíčové kategorie pro zkoumání autenticity, využité dále v kapitole následující.

V kapitole šesté autor reviduje a popisuje jednotlivé prvky popsané v kapitole páté, které posléze využije pro formulaci obecného rámce zkoumání autenticity digitálních knih. Zajímavou částí je část 6.2.3, která se věnuje otázce originálu a vztahu k digitalizované odvozenině. Dílčí závěr 6.2.13 je klíčový pro další části práce, včetně definice autenticity dle ISO 14721.

Obecný rámec ze šesté kapitoly je rozpracován v kapitole sedmé, kde autor specifikuje požadavky na autenticitu pro jednotlivé fáze existence digitálních dokumentů (knih) - produkce, archivace a zpřístupnění. Z textu této kapitoly je jasně vidět vhled autora do problematiky, teoretické i praktické znalosti. Závěr kapitoly je výborným shrnutím všech aspektů autenticity, které autor považuje za klíčové.

V kapitole osmé autor navrhuje obecná doporučení v podobě sad požadavků (P-0 a P-1) a dále doporučení pro jednotlivé fáze životního cyklu dokumentu (sady požadavků P-2, P-3 a P-4 pro

digitalizaci, archivaci a zpřístupnění/digitální knihovnu). Požadavky jsou dobře strukturovány a srozumitelně vysvětleny.

V kapitole deváté je rozpracován jeden z obecných požadavků z kapitoly osmé, konkrétně P-0.2 - Užití systému perzistentní identifikace digitalizátů knih. Je popsán systém na přidělování a správu identifikátorů URN:NBN, na jehož návrhu a poté nasazení do provozu v Národní knihovně ČR se autor přímo podílel.

Hlavní přínos a vklad autora je v analytické části, tedy kapitolách 6-9.

Hodnocení práce/Obsahové připomínky

Celkově hodnotím práci pozitivně, níže jsou zmíněny drobné nedostatky, které autorovi unikly.

- Str. 69 - tvrzení, že portugalská národní knihovna vyvíjí open source systém RODA není pravdou. Systém vznikl v roce 2006 s finanční podporou Centrální správy portugalských archivů, byl dále rozvíjen v evropských projektech SCAPE a E-ARK a do dnešního dne tento systém vyvíjí společnost KEEP Solutions.
- V disertační práci je několik míst, kde není jasné zda autorovo tvrzení platí obecně, nebo platí pro konkrétní příklad digitalizace v Národní knihovně ČR. Viz druhý odstavec na str. 70 (část 4.2.1), toto tvrzení platí pro NK ČR, ovšem zcela jistě ne obecně.
- Část 4.2.4 Metadata pro archivaci – při popisu formátu PREMIS autor uvádí, že komerční systémy Rosetta a Preservica užívají vlastní metadatový proprietární standard. Pravdou je, že Rosetta i Preservica mají vlastní schéma (Rosetta DNX a Preservica XIP), ta jsou ovšem založena na standardu PREMIS. Oba systémy mají mapování mezi DNX/XIP a PREMIS a odlišnosti oproti PREMIS standardu jsou minimální. Systém Preservica umí vyexportovat standardní PREMIS metadata. DNX schéma je navíc veřejně dokumentováno zde https://knowledge.exlibrisgroup.com/@api/deki/files/39700/Rosetta_AIP_Data_Model.pdf, tedy tvrzení o proprietárnosti není zcela na místě. Podobně na str. 169, část 7.8.5, se objevuje tvrzení, že "*(Preservica a Rosetta) užívají vlastní metadatová schémata ... . Problém však je, že jde o proprietární schémata. Je vhodnější trvat na užití standardu PREMIS a postupně aplikovat jeho novinky, než se spoléhat na komerční řešení svázaná s jednou firmou.*". Ve skutečnosti metadatová schémata obou systémů z PREMIS vychází a firmy tato schémata upravují tak aby byla s PREMIS kompatibilní, tj. vždy když je publikována nová verze schématu PREMIS.
- Část 4.3 Zpřístupňování - tvrzení, že hlavní formou zpřístupnění v současné praxi je prezentace digitalizované knihy v digitální knihovně. V této části a dalších částech práce autor zcela pomíjí koncept zpřístupnění přímo z digital assett management systému (DSpace např.) nebo ze systému na dlouhodobou ochranu (Rosetta, Preservica). Jmenované systémy mají zpřístupňovací vrstvu (modul), který využívá AIP balíčky, resp. jJejich části a vytváří diseminační balíček DIP na vyžádání. Nedrží tedy kopii dat pro zpřístupnění v jiném systému, zpřístupnění probíhá přímo z archivních dat, není zapotřebí mít oddělenou digitální knihovnu jak ji definuje autor na str. 71. Toto tvrzení se objevuje také v části 7.6.
- Část 7.7.2 Opravy balíčku SIP – v případě, že obsah SIP balíčku je poškozen již od původce, autor považuje za optimální řešení požádat původce o opravu a dodání nového SIP balíčku. V této části bych čekal konstatování, že tento přístup není vždy možný, např. u born-digital dokumentů,

které mohly být vytvořeny jako nevalidní již v prvopočátku. Digital preservation systémy jako Rosetta nebo Preservica, umožňují opravy dat během ingestu. Tyto jsou samozřejmě popsány a uchovány jako události v metadatech. V případě nahrazení (nevalidního) souboru za jiný (validní), se nevalidní uchová. Autenticita tedy nijak neutrpí. To samé platí pro opravy dat v balíčku AIP, část 7.7.3. Zmíněné systémy dovolují změny/editaci dat i metadat, vytvoří a uchovají metadata v databázi i v rámci AIP balíčku. S autorovým tvrzením, že je lepší si vyžádat nová data od producenta, uložit nový AIP a původní smazat, bych si tak dovolil nesouhlasit.

- Část 7.9.4 o nástrojích - Harvardská knihovna již nástroj JHOVE neudržuje a nevyvíjí, jak tvrdí autor. Oficiální sídlo a vývoj nástroje JHOVE zajišťuje od roku 2015 Open Preservation Foundation.

Otázky pro doktoranda

- Proč se doktorand rozhodl nevyužít terminologie oficiálního českého překladu normy ISO 14721, zvláště pokud byl spolu autorem českého překladu normy.

Závěr a doporučení

I přes výše uvedené připomínky hodnotím disertační práci kladně, autor prokázal vědomosti a předpoklady k tvořivé vědecké práci.

Disertační práci doporučuji k obhajobě a navrhuji klasifikaci - **prospěl**.

Wellington, 24.8.2017

Mgr. Jan Hutař, Ph.D.