

Doktorand: Mgr. Jan H u t a ř

Název disertační práce: *Digitalizace, popis pomocí metadat a jejich formáty*

Praha, 2012, 397 s.

Oponent disertační práce: PhDr. Eva Bratková, Ph.D.

V Praze, dne 16. 5. 2012

Hodnocení oponenta

Předložená disertační práce Mgr. Jana Hutaře je věnována specifické dílčí otázce z oblasti masové digitalizace knihovních fondů, a to problematice tvorby komplexních metadat pro potřeby dlouhodobé archivace a ochrany digitalizovaných dat (LTP). Jde o závažný odborný i celospolečenský problém, kterému v ČR prozatím nebyla a stále není věnována náležitá pozornost. Jak autor zdůrazňuje v úvodní části, hrozba ztráty informací z dosavadního průběhu digitalizace knihovních fondů je proto veliká. Jako cíl disertační práce si autor vytknul provedení detailní analýzy, popisu a vyhodnocení vývoje a současného stavu tvorby metadat zejména ochranného charakteru (na příkladu aktivit Národní knihovny ČR, ale též v kontextu světového vývoje této problematiky) a zpracování národního standardu komplexního metadatového profilu pro potřeby digitalizace a dlouhodobé ochrany monografických a periodických publikací. Autor se po celou dobu výzkumu snažil, jak uvádí i v 1. kapitole na s. 11, změnit nepříznivý stav tvorby administrativních a technických metadat v rámci současné digitalizace v NK ČR. Takový záměr lze v rámci disertačního výzkumu zcela uvítat.

Výsledek disertační práce představuje jedinečný, původní a zásadní příspěvek k rozvoji oboru knihovní a informační vědy z pohledu obecně teoretického, metodologického/metodického, ale zejména praktického. Práce jako celek dává nahlédnout do zcela nového směru našeho oboru. Výrazně se mění paradigma procesů zpracovávání, uchovávání a zpřístupňování informací. Je potřeba ocenit, že disertační práce vznikala v průběhu 7-8 let na základě rozsáhlé výzkumné činnosti doktoranda, podílel se významně v celé řadě výzkumných úloh v ČR, řešitelsky vázaných především na NK ČR včetně dvou výzkumných záměrů z let 2004-2011 MK ČR (VZMK00002322101 řešitelů A. Knolla, Z. Uhlíře a J. Polišenského a VZMK00002322102 řešitelky B. Stoklasové). Podílel se také na několika zahraničních výzkumech v této oblasti. Odborná veřejnost se měla možnost seznámit s průběžnými výsledky práce doktoranda ve formě zveřejněných výzkumných zpráv a jejich příloh, pracovních analýz a metodických pokynů, příspěvků na konferencích a pracovních seminářích, článků v časopisech a také monografické práce vytvořené se zahraničními

spoluautory. Na zmíněném pracovišti byl též po celé období tvorby práce zaměstnán, v posledním období také ve vedoucí funkci relevantního odboru. Zkoumané problematice se věnoval i pedagogicky, zajišťoval odborné předměty magisterského studia v rámci ÚISK FF UK a vedl řadu bakalářských a diplomových prací k daným otázkám.

Obsah disertační práce je profesionálně fundovaný, reprezentuje nové výsledky aplikačního výzkumu, je předložen v kvalitní formě a vyznačuje se vhodnou logickou uspořádaností. Rozsah práce přesahuje úroveň běžně odevzdávaných disertačních prací na ÚISK (397 stran včetně příloh, v přepočtu na normostrany jde o cca 500 s.).

Z hlediska metodologického je práce postavena na vynikající průběžně prováděné bibliografické rešerši informačních zdrojů, pečlivém studiu a kritickém rozboru nosné literatury k tématu (zejména zahraniční), studiu, komparaci a vyhodnocení řady metadatových standardů a dalších souvisejících metodik v dané oblasti, průběžném získávání a vyhodnocování dalších odborných informací z provozů digitalizačních pracovišť a pracovišť, která provozují systémy pro dlouhodobou archivaci a ochranu digitálních dat v ČR a také v zahraničí. Praktickým původním přínosem je zpracování národního standardu - rozsáhlého komplexního metadatového profilu pro potřeby digitalizace a dlouhodobé archivace a ochrany periodických a knižních publikací národního charakteru. Disertace zahrnuje na konci seznam použitých zdrojů a literatury (s. 216-238), který čítá cca 240 titulů. Je připraven v souladu s mezinárodním standardem ISO 690:2010, zvolený harvardský citační styl je aplikován bez problémů.

Hlavní text disertační práce je uspořádán do 8 základních kapitol, přičemž osmá je věnována stručnému závěrečnému hodnocení. Poznamenala bych, že stránky s postliminarií (seznam zdrojů, výkladový slovník a příloha) by již neměly být číslovány (nejde o kapitoly hlavního textu). Struktura práce je v nosných částech (kap. 3-7) promyšlená a propracovaná a prozrazuje odbornou erudici autora. Titulky obsahu vystihují dílčí témata práce. Obsahy jednotlivých kapitol na sebe co do ústředního tématu logicky navazují. Za specifický bonus či přídavek k tématu lze ale považovat obsah kapitoly 6, který je věnován tématu souvisejícímu. Pravděpodobně technický problém zvoleného editoru způsobil, že některé podkapitoly (v tištěné i elektronické verzi) jsou, jak je vidět už z obsahu, zdvojnásobené nebo ztrojnásobené (3.2.1, 3.2.2, 3.2.3, 4.3.1, 4.8.1, 5.3.2.1), v důsledku čehož je několik málo odkazů v textu problematických, tj. ukazují na více částí najednou. Příloha práce (úplné znění

navrhovaných národních metadatových standardů) zaujímá v textu místo od strany 241- s tím, že se v ní vyskytuje dvakrát vnitřní číslování metadatových specifikací (78 s. a 74 s.).

První dvě kapitoly disertace tvoří nezbytný úvod do tématu. Autor v souladu s požadavky na disertační práce v první kapitole formuluje cíle práce, komentuje metody a okolnosti její tvorby a limity zpracování. Představuje také odborná východiska práce a vlastní výzkumnou a publikační činnost související s tématem disertace. Kapitola druhá pak zahrnuje přehled všech podstatných termínů, s nimiž se čtenář práce setkává v dalších kapitolách. S obsahem výkladů lze souhlasit, některé termíny jsou vymezeny ale možná hodně široce a pragmaticky, například „historické fondy“ na s. 18. Doporučovala bych zařadit spíše přesnější termín „historické knihovní fondy“ (je uveden i v databázi TDKIV). Na s. 22 je nekorektně uvedeno úplné jméno české organizace SKIP.

Nosnou či jádrovou část disertační práce tvoří kapitoly 3-7. K nejobsáhlejšími kapitolám patří kapitola 3 (48 s.). V jejím úvodu (3.1) je podán zasvěcený komentovaný přehled problematiky digitalizace v knihovnách, jejích důvodů a přínosů. Určitě lze souhlasit s názorem V. Readingové (autor na něj upozorňuje), že současná informační společnost stojí vedle digitalizace a online zpřístupňování informací také na dlouhodobé ochraně digitálních informací. Autor proto v části 3.2 přináší rozsáhlý (na cca 30 s.) kritický výklad otázek dlouhodobé ochrany digitálních dat včetně charakteristiky řady rizik, opatření, která mohou zabránit nevratným ztrátám informací. Autor osvětluje vývoj a proměny v náhledu a v přístupu k problematice dlouhodobé archivace a ochrany digitalizovaných dat v globálním rámci. Autor je hodně přesvědčivý v celé řadě argumentů. Text dokládá jeho vynikající znalost literatury (zejména zahraniční) k těmto otázkám, podstatné teze jsou konfrontovány s poznatky získanými z praxe. Třetí kapitolu uzavírá stručné představení referenčního modelu OAIS se zaměřením na popis balíčků informací a funkčního modelu OAIS. K textu této kapitoly nemám žádné poznámky.

Stejně metody práce uplatnil autor také v rámci zpracování další čtvrté kapitoly, nazvané „Úvod do metadat“ (cca 40 s.). Na základě hlubokých znalostí odborné literatury a dalších zdrojů i na základě vlastních dlouholetých zkušeností autor i v tomto případě podává kvalitně připravený přehled definic termínů „metadata“, „metadatové schéma“ a „metadatový profil“, dále typologii metadat a charakteristiku jejich jednotlivých typů. Koncentruje se zejména na administrativní metadata (technická, ochranná a právní). Věnuje se také procesům tvorby a ukládání metadat. Na s. 86 je uvedeno pravidlo, že se digitalizuje zásadně pouze takový

dokument, který má z dřívější doby zpracován bibliografický záznam. Mohl by autor u obhajoby doplnit informaci, jaká je praxe NK ČR v případě, že národní digitalizované historické dokumenty takový záznam prozatím nemají (u nezpracovaných fondů)? Vynechávají dočasně se z digitalizace? V rámci popisu vývoje metadat autor na s. 103 zmiňuje také oblast muzejnictví a staré metadatové schéma CIMI. Zeptala bych se, proč není zmíněn nový referenční model CRM (ICOM/CIDOC)? V návaznosti na část 3.3 podává autor v části 4.9 podrobnější výklad informačního modelu OAIS s osvětlením vztahu tohoto modelu a metadat.

Za velmi přínosnou část pokládám text kapitoly páté. Je v ní předložen výsledek kritické analýzy složité implementace a vývoje používání metadat v NK ČR. Autor maximálně těží ze znalostí získaných na svém mateřském pracovišti. Celá kapitola je vhodně strukturována, na druhé úrovni do dvou časových období (léta 1996-2003 a 2004-2011). Vnitřně jsou pak odděleny metadatové popisy historických dokumentů a dokumentů novodobých. Čtenáře práce určitě zaujmou popisy „dramatických“ konverzních procesů. Autor v této kapitole představuje také vlastní práce na specifikaci metadat (jsou uvedeny také v závěrečném seznamu referencí), které pro tuto knihovnu připravil v letech 2003-2010. Oceňuji i kritiku vlastních omylů autora (např. na s. 149 zmiňovaný návrh na zavedení identifikátoru SICI, který se nezačal nikdy používat). V části 5.3.2.1 (v prvním výskytu!) autor zmiňuje otázku začlenění identifikátoru UUID do syntaxe identifikátoru Handle v rámci lokální identifikace dokumentů v systému Kramerius. Mohl by autor v rámci obhajoby osvětlit, proč a jak k tomu došlo? Lze v tomto případě ještě hovořit o identifikátoru Handle?

Jak jsem již uvedla výše, za část „navíc“ v rámci disertačního tématu bych pokládala obsah kapitoly 6, nazvané „Vývoj konceptu správy a uložení digitálních dat“ (se zaměřením na NK ČR). Jde o související problematiku, kterou autor velmi důvěrně zná a zřejmě proto ji do textu zařadil. I v tomto případě podává s velkou erudicí výklad vývoje ukládání dat na různé nosiče. Předkládá výborné postřehy k pozitivům i negativům tohoto procesu. Upozorňuje na problémy testování datového úložiště NK ČR aj. Na s. 160 vyslovuje obavu že NK ČR v rámci nové koncepce tvorby digitálního repozitáře, bohužel, asi nepřeskočí etapu DAM, jak se původně plánovalo. Výborný přehled systémů pro dlouhodobou archivaci a ochranu (LTP) je podán v části 6.3.3. Oceňuji určitě, jakož i řada dalších odborníků v ČR, představení konceptu důvěryhodného repozitáře a jeho certifikaci včetně výsledků prováděných testů vybraných systémů v ČR. Autor v této části rekapituluje a vyhodnocuje informace dříve

publikované v odborném tisku. V části 6.4.4 je také odkaz na monografickou práci k metodice PLATTER, jíž je spoluautorem (s kolektivem renomovaných zahraničních odborníků).

Poslední nosná kapitola sedmá představuje komentář k návrhu aplikačního metadatového profilu pro digitalizaci v projektu Národní digitální knihovny. Návrhy jsou předloženy v příloze práce (celkem cca 150 s.). Jde o hlavní vlastní přínos doktoranda. Autor na základě rozsáhlé analýzy všech dostupných zdrojů připravil návrh metadatového profilu (zvláště pro periodické a monografické publikace), který je postaven na známém standardu (struktuře) METS (též s využitím elementů specifikace metadat datového modelu PREMIS). Inspirativním modelem se stala specifikace Norské národní knihovny a částečně také specifikace nizozemské Královské knihovny. Jak autor uvádí v textu práce, nejde o metadata pro potřeby momentálního dlouhodobého uložení, ale o metadata, která budou vznikat v průběhu digitalizace. Cílem je vytvoření dostatečné množiny zejména administrativních metadat (ochranných, a technických) pro potřeby budoucího využití v digitálním repozitáři (v souladu s koncepcí modelu OAIS). Jde o velmi zásadní práci doktoranda pro odbornou praxi, kterou je potřebné náležitě ocenit i v rámci hodnocení disertační práce. Specifikace připojená v příloze je vystavena také na webu NK ČR jako oficiální specifikace metadat pro dotační projekt „Vytvoření Národní digitální knihovny“. Téměř „heroické“ úsilí doktoranda se tak obhajobou této práce završuje. Jaké perspektivy tohoto standardu vidí autor v dalším vývoji informačních systémů NK ČR? Mohl by to rozvést v rámci obhajoby? K přiloženým specifikacím nemám žádné podstatné připomínky, jenom formální připomínku: v obou částech je v podkapitolách 8.4.1.1. chybně uvedena syntax elementu specifikace Dublin Core. Namísto „dc.title“ má být zapsáno správně „dc:title“.

Disertační práce je ukončena v kapitole 8 stručným hodnotícím závěrem.

Disertační práce má kromě vynikajícího obsahu také pěknou formální úpravu. Stran gramatických chyb má text jisté nedostatky. Uvedu některé základní, ostatní jsou poznamenány tužkou v jednom papírovém exempláři. V textu se vyskytuje vícekrát chyba nekorektně oddělné věty v rámci souvětí (například na s. 13, odst. 4, první věta, chybí čárka za řetězcem „2006-2009“, nebo na s. 19, odst. 2, poslední věta, chybí čárka za slovem „formátu“ aj.). Jména některých geografických entit nejsou uvedena korektně (určitě jde o „Nizozemsko“, které je uvedeno ve většině případů nesprávně jako „Nizozemí“, jenom jednou je správně, a to na s. 210; na s. 13 by měl být uveden v českém textu korektní tvar „v Lublani“ namísto „v Ljublani“). Jména některých známých knihoven jsou uváděna

v různých tvarech, některé jsou i nekorektní (vyskytují se například 4 formy: Německá národní knihovna – německá národní knihovna - Národní knihovna Německa – národní knihovna Německa; viz též jméno Australské národní knihovny ve více tvarech a další jména), doporučit lze jednotný tvar (viz první tvar). Zápis procent pomocí symbolu je v několika případech nekorektní, jde-li o podstatné jméno (například na s. 30 je nesprávně zapsáno „37%“, správně má být s mezerou 37 %). Doporučit lze také jednotné zapisování číslovek řádu tisíců, autor užívá více modelů (například na s. 31 „999 960 Euro“ nebo „2000 stran“, na s. 155 „31.000 kusů“ aj.). Ve velkém počtu užívá autor nejednotně řadu sousloví (například záznam METS – METS záznam, standard PREMIS – PREMIS standard, schéma XML – XML schéma a řada dalších), korektní v českém textu je první varianta, nikoliv varianta s anglicistním pořadím slov. Dvakrát se vyskytuje překlep u slova „rozhranní“. Na s. 74 se v poznámkách vyskytuje omylem dvakrát stejné URL.

Závěrem konstatuji, že rozsáhlá disertační práce Mgr. Jana Hutaře je z hlediska obsahu mimořádně kvalitním dílem věnovaným palčivé otázce dlouhodobé archivace a ochrany digitálních dat se zaměřením na specifická metadata ochranné a administrativní povahy. Kromě přínosných analytických textů v dílčích kapitolách je velkým vkladem doktoranda návrh rozsáhlého národního metadatového standardu. Výše zmíněné gramatické prohřešky nijak nesnižují kvalitu díla a jsou spíše zmíněny jako pomoc autorovi pro případ další práce nad tímto textem v příštím období. Doktorand naplnil vrchovatě požadavky kladené na disertační práce. Prokázal řadou vlastních publikací i touto závěrečnou prací doktorského studia, že je uznávaným expertem v ČR i v zahraničí na danou problematiku. Práci doporučuji k obhajobě v plném rozsahu.

Eva Bratková