

# 10th International Symposium on Electronic Theses and Dissertations

13. - 16. června 2007, Uppsala, Švédsko

Zpráva ze zahraniční cesty  
ing. Jan Mach

Ve dnech 13. – 16. června 2007 se v Uppsala University konal 10. ročník konference na téma elektronických diplomových a kvalifikačních prací (zkráceně eVŠKP, angl. ETD), které jsem se zúčastnil jakožto zástupce Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ (dále jen Komise eVŠKP). Hlavní téma konference „Přidaná hodnota eVŠKP“ se dělilo na dva okruhy: „Integrace eVŠKP s procesy v institucích a praxe“ a „Služby s přidanou hodnotou“. Příspěvky se konkrétněji zaměřily na:

- integraci eVŠKP do vědy a výzkumu (mnoho institucí se zaměřuje na šedou literaturu obecně),
- integraci eVŠKP do lokálních služeb (zkušenosti z budování lokálních repositářů),
- integraci do národních a mezinárodních služeb (důležité v souvislosti se snahou zbudovat Národní registr VŠKP v ČR),
- zkušenosti, praxe z jednotlivých zemí,
- provázání eVŠKP na multimediální data, statistická data, tabulky apod.,
- dlouhodobé uchování dat (doporučena replikace dat do více repositářů),
- právní souvislosti (problematika plagiátorství, copyright).

Konference probíhala v několika sekcích, zaměřil jsem se na sekce související s trendy a se zkušenostmi s digitálními knihovnami na národní a mezinárodní úrovni, právní problematiku a standardizaci. Během konference jsem byl v kontaktu z ČR s Mgr. Evou Březinovou, UK a RNDr. Ing. Milanem Šormem, Ph.D., MZLU, který prezentoval poster se zkušenostmi s lokálním repositářem eVŠKP na MZLU. S mnoha přednášejícími jsem se během konference osobně setkal, většina byla velmi vstřícná ve výměně zkušeností s tvorbou repositářů, o problematiku eVŠKP v ČR se zajímali a získal jsem od nich kontakt pro případné další dotazy a spolupráci.

## 13. 6. 2007

Ve středu 13. 6. proběhla registrace účastníků, workshop skupiny NDLTD a večerní uvítací recepce v prostorách university.

## 14. 6. 2007

Konferenci zahájil ráno 14. 6. **Ulf Göranson**, ředitel Uppsala University Library, pokračoval ředitel NDLTD **Ed Fox**. Zdůraznil mimo jiné důležitost projektu LOCKSS (Lots of Copies Keep Stuff Safe) pro zpřístupňování a hlavně ochranu dat umožněnou replikací kopií na více míst, do více repositářů.

Dopoledne s plenárním příspěvkem vystoupil **Greg Crane: The dissertation in an age of data driven humanities** – obecněji zaměřenou přednáškou o významu digitálních prací a současných trendech, kdy práce je rozdělena do více nezávislých částí, kombinují se data, odděluje se obsah od grafické reprezentace, existuje dynamický obsah jako jsou např. různé Wiki, knihy „spolu mluví“. Specifikoval tři subjekty zájmu – digitální knihovnu,

čtenáře a knihy, subjekty si jsou vzájemně prospěšné a učí se od sebe navzájem, vyvíjí se v závislosti na druhých dvou subjektech (např. přizpůsobení se digitální knihovny chování čtenářů, vznik dokumentů ve Wiki systémech). V sekci Realising the Vision of ETDs jsem se zúčastnil přednášky **Gail McMillan: Toward the Future with Retrospective Hindsight**. Autor prezentoval, co by měl umět systém pro eVŠKP na příkladu programu ETD-db, který právě prochází redesignem. Druhá přednáška v sekci byla **Maurice Vanderfeesten, Gerard van Westrienen: A Portal for Doctoral e-Theses in Europe: Lessons learned from an interoperable demonstrator project for five European countries**. Ukázali velmi zajímavý ukázkový projekt ETD portálu agregujícího disertační práce pěti škol ([www.e-thesis.sharelab.cq2.org](http://www.e-thesis.sharelab.cq2.org)), který realizovalo sdružení SURFfoundation (Nizozemí), JISC (UK) a DIVA. Rozpočet za příspěví EU na projekt byl 20 000 €. Uváděli řešené problémy s metadaty, které jsme také řešili v rámci Komise eVŠKP a kde jsme dospěli k obdobným závěrům. Zmínili povinné prvky (český návrh jejich povinné prvky obsahuje), v metadatech pro dc:date doporučili kvůli možné nejednoznačnosti používat datum publikování, pro dc:type hodnoty Bachelor thesis, Master thesis, Doctoral thesis (dle DRIVER guidelines a E-Thesis Recommendations, <http://www.driver-support.eu>), do dc:contributor uvádět pouze vedoucího práce. Zmínili důležitost národních repositářů jakožto národních proxy serverů, které agregují lokální data a ty předávají jednotně do mezinárodní digitální knihovny. Dle následného osobního rozhovoru s Gerardem van Westrienenem jednání o následných projektech jsou vedena mezi skandinávskými zeměmi a větší anglicky mluvící pracovní skupina se zatím nepřipravuje. Jako metadatový standard doporučoval rozšířený Dublin Core, METS zvažovaný v Komisi eVŠKP nedoporučoval kvůli zbytečné složitosti pro potřeby registru. Pro účely trvalého identifikátoru jednoznačně doporučil Persistent Identifier [www.persistent-identifier.de](http://www.persistent-identifier.de).

Poslední přednáška sekce **Peter Sefton: An integrated approach to preparing, publishing, presenting and preserving theses** zmiňovala zkušenosti z Austrálie, kde pro účely eVŠKP mají vlastní editor ICE s předdefinovanými styly, správou verzí řešenou přes SubVersion a generátorem HTML (souvisí s prací Komise eVŠKP na šablonách).

Odpoledne následovala panelová diskuse na téma **Usability: Perspectives from Research, Publishing, and Higher Education Institutions**. Použitelnost byla definována podle ISO 9241-11 (Ergonomické požadavky na kancelářské práce se zobrazovacími terminály – Část 11: Údaje o možnostech využití). Zajímavá byla prezentace **Joris van Rossuma** o testování použitelnosti uživatelského rozhraní SCOPUS, které probíhalo ve 27 kolech, v 5 institucích, po 25 užívatelích na každý test.

## 15. 6. 2007

V plenárním příspěvku **The Power of the Electronic Scientific Thesis** zdůraznil **Peter Murray-Rust** důležitost mladých lidí, kterým je nutno naslouchat. Nemají se dělat pouze texty, PDF, ale interaktivní aplikace, jako příklad uvedl DBPedia.org – wikipedii psanou za využití RDF, kterou je možné přes SQL dotazy prohledávat, projekt „Experimental Data Checker“ pro označování chemických dat v elektronickém textu a projekt BioClipse pro 3D modely chemikálií.

V dopolední sekci **Discovery and Access** v prezentaci **Sharon Reeves: User-Generated Metadata for ETDs: Added Value for Libraries** byly představeny zkušenosti z Kanady, kde v letech 1998-2002 využívali služeb Proquestu, v letech 2001-2002 vyvinuli celostátní program spravovaný v Library and Archives Canada (LAC), viz: [www.collectionscanada.ca/thesescanada](http://www.collectionscanada.ca/thesescanada). Původní náklady u Proquestu byly 29 CAD na jednu archivovanou práci, LAC za roky 2006-2007 tak ušetřil 95 000 CAD. Studenti zadávají metadata do archivu ručně. **Austin McLean** v příspěvku **ETDs and the Importance of**

**Discoverability** prezentoval Proquest, který obsahuje práce ze 700 institucí, získává 70 000 titulů ročně, celkem obsahuje 900 000 titulů eVŠKP. V sekci Implementation Solutions v přednášce **Uwe Müller a kol.: A Property Based Workflow System for E-Theses** prezentovali zkušenosti z Humbolt Universität Berlin ukazovali univerzitní systém pro sběr elektronických prací všeho druhu včetně např. konferenčních materiálů. Systém má propracované workflow, řeší např. i konverzi dat a připomínkování prací, obsahuje nástroje pro tvorbu, kontrolu a validaci dat, konverzi, extrakci metadat a elektronické podpisy. Požadavky na systém byly flexibilita, zpracování různých variant jednoho dokumentu, pre/post rozhodovací podmínky, manuální x poloautomatické x automatické akce (např. extrakce dat může proběhnout automaticky nebo v případě problému ručně). Využívají open source nástroje včetně systému Lucile pro indexování. V prezentaci autorů **Using Ajax to integrate institutional electronic theses and dissertation repository with corporate systems at University of São Paulo** prezentovali využití zkušeností NDLTD pro zapojení 39 knihoven ze 6 měst do repositáře eVŠKP. Provozují systém na bázi PHP, MySQL, AJAX (open source xajax pro PHP) a PEAR Sigma Library pro webové šablony. Systém obsahuje okolo 8000 prací.

Odpolední sekci **Next Generation ETDs** zahájil přednáškou **David Lawrence, Peter Berkesand: Making the most of Electronically Published Theses: Combining Text with Video, Sound and Data**. Zmínili možnost ukládat multimediální data do PDF. Jako příklad multimediálních dat uvedli video, zvuk, Excel tabulky, simulace (grafy, 3D modely), synchronizované video a prezentaci. Tím lze vytvářet interaktivní materiály oproti tištěné verzi. U těchto prací vznikají zde složitější právní vztahy, kdy např. kvůli copyrightu je potřeba získat svolení od lidí na videu. **Joan F. Cheverie, Austin McLean: Social Science Data and ETDs: Opportunities and Challenges** uvedli hodně otázek, které řeší v souvislosti s ETD – jaká data/formáty povolit, kdo o tom rozhoduje, jaké použít politiky přístupu k datům a k vyhledávání, jaká použít metadata, jak řešit otázky migrace a dlouhodobé ochrany dat (zde navrhuje využití Proquestu). Zdůrazňují důležitost využití standardů při vývoji. **Edward A. Fox** ve svém příspěvku **Using Concept Maps in NDLTD as a Cross-Language Summarization Tool for Computing-Related ETDs** uvedl zkušenosti s automatickou tvorbou abstraktů a myšlenkových map pro automatizované náhledy obsahu prací. Porovnával využitelnost obsahů práce vs. automaticky generovaných náhledů Relex (myšlenkové mapy). Problematika automatizovaného překladu byla ukázána na příkladu dvou korpusů dat – anglického a španělského. Závěr pokusu NDLTD je, že automatické překlady jsou použitelné a dokáží automaticky vytvářet přínosné myšlenkové mapy. Do budoucna se chtějí zaměřit na export obrázků a grafů do myšlenkových map a jejich provázání do dokumentu.

V panelové diskusi **Added Values and E-Science** bylo zmíněno velké množství dat dostupných v medicíně a chemii (př. PubChem, PubMed), důležitá je jejich archivace, analýza, uživatelský přístup k datům a data mining. Bylo poukázáno na problematiku osobních dat v datech používaných v kvalifikačních pracích. Data, se kterými se provádí výzkum, je třeba uchovat pro potřeby navazujícího výzkumu, ale v případě jejich nutné anonymizace může být problém navázat dalším výzkumem – nejdou pro výpočty vybrat shodné vzorky.

## 16. 6. 2007

Odpolední sekci **Policy Issues** zahájili **Niamh Brennan, Austin McLean: Plagiarism and ETDs: Confronting the Reality** na téma problematiky plagiátorství. Otázkou dle nich je, zda počet plagiátů roste, nebo je pouze jednodušší je najít. Řešením dle nich je software na automatickou detekci plagiátů a nutnost autory podepsat Academic code of Conduct.

**Holly Mercer** v **Open Access and Responsible Research: Preparing Future Faculty at the University of Kansas** ukázala zkušenosti z Kansas University, kde vybírají elektronicky články, drafty, prezentace, preprinty a eVŠKP ([www.graduate.ku.edu/~etd](http://www.graduate.ku.edu/~etd)). Proquest plně nevyhovuje, neboť k pracím není volný přístup, proto využívají i vlastní nástroj v DSpace. Poplatky za zveřejnění v Proquestu jsou univerzitou částečně dotovány. Studenti musí souhlasit s open access, používá se Creative Common Licence. Prezentace **Wendy White: Opening access and closing risk: delivering the mandate for e-theses deposit** zdůraznila význam získání oprávnění pro manipulaci s daty na základě analýzy workflow, potřebu sofistikovanějších licenčních smluv (Creative Commons) a nutnost politik uchování dokumentů. Zmínila dále projekty CLADDIER a KULTUR.

**Příští konference ETD se budou konat v roce 2008 v The Robert Gordon University v Aberdeenu (se zaměřením na multimédia a dlouhodobou ochranu) a v roce 2009 v Pittsburghu.**