

IKAROS 4. 11. 2008

Zpráva ze semináře Systémy pro zpřístupňování eVŠKP 2008

Již potřetí se v tradičním říjnovém termínu v budově rektorátu Vysokého učení technického v Brně (VUT) konal seminář Systémy pro zpřístupňování eVŠKP (7. října 2008). Letošní novinkou bylo tematické propojení se Seminářem ke zpřístupňování šedé literatury navazujícím na seminář následující den, této akci je věnována samostatná zpráva. Oba semináře byly navštíveny odborníky na zpřístupňování VŠKP a šedou literaturu z celé ČR.

Účastníky přivítala ředitelka Knihovny AMU a členka výkonného výboru Asociace knihoven vysokých škol ČR (AKVŠ) PhDr. Iva Horová. Sdělila, že se jedná již o třetí ročník akce, na které prezentuje svoji práci Odborná komise pro otázky elektronického zpřístupňování elektronických vysokoškolských kvalifikačních prací při AKVŠ (Komise pro eVŠKP), jejíž předsedkyní byla až do letoška právě dr. Horová. Novým předsedou se stal Ing. Jan Mach z Vysoké školy ekonomické v Praze. Nejdůležitější aktivitou Komise pro eVŠKP byla v tomto roce spolupráce na projektu Theses.cz. Na organizaci semináře se podílelo VUT a Ústřední knihovna VUT, Státní technická knihovna (STK) a AKVŠ. Finanční podpora pro konání akcí byla získána díky grantu Ministerstva kultury ČR. Dr. Horová poděkovala sponzorům seminářů, kterými byly společnosti Albertina icome Praha s.r.o. a MULTIDATA Praha s.r.o. Odborným garantem akce byla Komise pro eVŠKP.

Theses.CZ

Prvním odborným přednášejícím byl **RNDr. Miroslav Křipač, PhD.**, z Centra výpočetní techniky Fakulty informatiky Masarykovy univerzity, který informoval o **současném stavu projektu Theses.cz**. Dr. Křipač představil kolegy spolupracující na projektu – do realizačního týmu patří doc. Ing. Michal Brandejs, CSc., Ing. Jitka Brandejsová, Bc. Martin Stančík a Mgr. Jan Kasprzak. Vývoj systému Theses je ve své polovině podle časového harmonogramu, do projektu se zapojilo 17 veřejných vysokých škol, dvě soukromé vysoké školy a jedna zahraniční vysoká škola (tou je Ekonomická univerzita v Bratislavě).

Úroveň zapojení vysokých škol zachycuje graf 1:

Graf 1: Úroveň zapojení vysokých škol do projektu Theses.cz

Systém je i **nadále otevřený k zapojení** dalších vysokých škol.

Přednášející se dále věnoval mnohým **novinkám**:

- umožnění **vkládání prací samotnými studenty** pro školy, které nemají vlastní úložiště (funkcionalita lokálního registru);
- nové **vyhledávání v plných textech** zvládající lemmatizaci a distribuované vyhledávání a respektující přístupová práva konkrétního přihlášeného uživatele k systému;
- nový způsob **dotazování ve stylu vyhledávače Google**;
- nově **podporované formáty pro import metadat** theses.cz 1.0 a evskp.cz 1.1;
- **předávání metadat** třemi cestami (ručně https - přes webový formulář, automatizované vkládání systémem školy, OAI-PMH);

- výsledky **vyhledávání plagiátů** (byly nalezeny podobnosti mezi školami, většinou mezi pracemi z Masarykovy univerzity a nejen VŠKP; není cílem uspokojit média a veřejnost nálezy a exemplárními případy; naopak je snahou opisování odhalit co nejdříve a zamezit ostudě studenta).

Dr. Křípač vyhodnotil projekt jako unikátní v celosvětovém kontextu z hlediska systematického sběru eVŠKP pro potřeby kontroly možného plagiátorství a vysokého počtu zapojených vysokých škol.

V následující diskusi upozornila dr. Horová na fakt, že formát EVSKP-MS (vyvinutý Komisí pro eVŠKP) používá pro předávání záznamů zatím pouze VŠE. Daný formát se zdá perspektivní, cílem práce Komise eVŠKP je prosazení předávání metadat více školami v tomto formátu.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Kripac2.pdf>

Metadatový standard EVSKP-MS verze 1.1 pro popis VŠKP a standardy související

Na dr. Křípače navázala svým příspěvkem **PhDr. Eva Bratková** (Komise eVŠKP a Ústav informačních studií a knihovnictví FF UK v Praze). Prezentovala svou práci na metadatovém standardu EVSKP-MS 1.1 jako výsledek aktivity pracovní skupiny, zastřešující přípravu

- **národního standardizovaného formátu** pro metadatový popis VŠKP;
- **národního výměnného formátu** pro přenos metadatových záznamů VŠKP pro účely projektů Národního registru vysokoškolských kvalifikačních prací, Národního úložiště šedé literatury (NUŠL) a nadnárodních projektů NDLTD a DRIVER
- a **validátorů ověřujících korektní zápis** metadatových záznamů ve formátu XML.

Pracovní skupina vedená dr. Bratkovou jako **východisko** pro přípravu standardu zvolila následující **standardizační aktivity** a jejich výstupy:

- EVSKP-MS, verze 0.1 (2005) – předchozí, výchozí verze národního standardu;
- Dublin Core Metadata Element Set, verze 1.1 (2008);
- DCMII Metadata Terms (2008) – nový způsob vkládání popisných prvků a další možnosti umožněné zpřesněním syntaxe;
- ETD-MS, verze 1.00, revize 2 (2008);
- XMetaDiss, verze 1.3 (2006);
- TEF 2.0 : Métadonnées des theses électroniques française (2006).

Uvedené výchozí standardy zaznamenaly v minulých dvou letech řadu podstatných úprav a aktualizací, které následně převzal i národní standard:

- dvojtečková notace (podle Dublin Core);
- výhradně XML v kódování UTF-8;
- zrušení dalších úrovní ve jménech prvků;
- zpřesnění prvků a atributů (povinné a nepovinné; viz obr. 1);
- dc → dcterms (slovník termů nahradil kvalifikátory).

Obr. 1: Popisné metadatové prvky EVSKP-MS

Přednášející představila novinku v popisu fyzických osob a korporací. Metadatové soubory Pers.CZ a Corp.CZ slouží ke standardizaci popis osob a korporací v XML záznamech zachycujících dokumenty šedé literatury včetně VŠKP. Popis zahrnuje popisné údaje (jméno, příjmení, datum narození, tituly) a je připraven pro nasazení prozatím nespécifikovaného trvalého identifikátoru entit sloužícího k trvalé a globálně jednoznačné identifikaci osob.

Dr. Bratkovou **doplnil Ing. Mach**, který prezentoval **technické informace o změnách a úpravách metadatových standardů EVSKP-MS**. Shrnul nasazené prvky popisující metadatový záznam, soubory tvořící VŠKP a prvky řešící přenos metadat a těchto souborů do dalších registrů. Podstatné je mj. využití schémat pro zápis metadat podle jazyka Relax NG, což usnadňuje názornost, čitelnost i validaci pomocí dostupných nástrojů.

Prezentace ke stažení: <http://www.evskp.cz/dokumenty.php?ukol=3&sek=1&tsekce=2>

Netextové práce a standard MS-EVSKP

Dr. Horová informovala o specifikách netextových prací, které jsou na uměleckých vysokých školách součástí závěrečných absolventských zkoušek stejně jako obhajoba v případě textových VŠKP.

Problémy s popisem vznikají:

- kvůli **různé podstatě netextových prací** (scénář hry, interpretační výkon, projektová dokumentace, film, výtvarné dílo, ...), které jsou vždy součástí kvalifikačního úkolu společně s teoretickou textovou částí;
- kvůli **různým rolím stejných osob a korporací** podílejících se na vzniku, realizaci či provedení;
- nasazením **alternativních názvů**;
- množstvím v minulosti, současnosti a budoucnosti **využívaných technických formátů** originálu.

Řešení na AMU týkající se spolupráce s Národním registrem VŠKP spočívá v:

- předání jen záznamů textové části;
- dočasném upuštění od nabízení metadat přes OAI-PMH;
- importu do knihovního katalogu – platí pro samostatné záznamy;
- importu do školního úložiště v DSpace;
- iniciování dalšího rozvoje MS-EVSKP
- a v nahrazení záznamů v Národním registru po vyřešení problémů.

Dr. Horová nastínila mnohé **otázky k rozhodování a řešení do budoucna** (způsob popisu netextových prací, propojování a identifikace jednotlivých komponent). Jako řešení nabízí **implementaci formátu METS**, který prostřednictvím navzájem propojených kontejnerů dokáže do jediného záznamu zahrnout popis více objektů včetně vzájemných vazeb.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Horova.ppt>

Strategie pro zvýšení viditelnosti vědeckých výstupů univerzity

Další část semináře po pauze na kávu moderoval Ing. Mach. **Ing. Vladimír Karen** ze společnosti Albertina icome Praha s.r.o. (AiP) prezentoval ve svém příspěvku databázi ProQuest Dissertations & Theses a její přínos pro zvýšení viditelnosti vědeckých výstupů univerzity. V systému ProQuest jsou zpřístupňovány produkty UMI založené v roce 1938 s cílem zachování kulturního dědictví a šíření textů vědecké úrovně.

Báze ProQuest Dissertations and Theses zahrnuje přes 2,5 mil. záznamů s retrospektivou od roku 1861. Roční přírůstek činí 80 000 záznamů a plných textů VŠKP z cca 700 škol z celého světa. Rozsáhlá retrospektiva je umožněna masivním nasazením technologie mikrofilmu od počátku existence databáze s udávanou životností 1 000 let.

Přestože je databáze komerční a placená, nabízí některé podstatné služby zdarma. Jsou jimi:

- archivace (online přístup k plným textům pocházejícím z instituce, ze které je k databázi přistupováno; záznamy prací ve formátu MARC)
- a usnadnění publikování a zpřístupnění (nástroj pro vytváření metadat ve formátech DTD a UMI-XML, kontrola výkonu autorského práva včetně vypořádání autorských poplatků; zařazení záznamů do dalších (tematických) databází společnosti ProQuest).

Databáze je dodávána ve **formě bibliografické** (PQDT Abstract and Index) i **plnotextové** ve dvou oborových modifikacích (PQDT Humanities & Social Sciences a STM).

Uložení plného textu je možné buď podle **tradičního modelu** (levnější možnost, text dostupný omezeně na základě licence), nebo podle **modelu Open Access** (dražší možnost, text dostupný online všem včetně indexování roboty a harvestery).

ProQuest Dissertations & Theses akceptuje multimediální (=netextové) přílohy, které rozšiřují a obohacují záznam (nyní je v systému obsaženo cca 2 000 prací s přílohami). Podporováno je několik desítek formátů příloh, jen pro některé formáty se předpokládá migrace do budoucích verzí formátů a jen některé formáty jsou dostupné bez aplikací třetích stran. Obsah je dostupný výhradně uživatelům plnotextových verzí databáze. Mezi formáty příloh figurují nejen obrázky, video, zvuky a prezentace, ale i zdrojové kódy programů a databáze.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Karen.ppt>

Trvalá identifikace fyzických osob a korporací v prostředí digitálních archivů: aktuální směry řešení v zahraničních systémech

Dr. Bratková ve svém dalším příspěvku informovala o současných směrech v identifikaci osob a korporací. Ve standardu EVSKP-MS jsou obsažené prvky dc:creator, dc:contributor, dc:publisher, thesis:grantor, popisující osoby a korporace. Možnosti jejich trvalé identifikace jsou nyní nedostatečné kvůli absenci standardizovaného identifikátoru. Budoucí trvalý identifikátor bude vkládán do prvku dc:identifier s cílem globálně jednoznačné a trvalé identifikace entit. Komise pro eVŠKP připravila standardy PersCZ a CorpCZ, využívající vnořené prvky pro zachycení data narození a titulů identifikované osoby. Příspěvek shrnuje jak tuzemské projekty předpokládající identifikaci osob (jednání s NK ČR na koordinovaném vytváření autoritních dat autorů VŠKP, situace v Registru informací o výsledcích), tak zahraniční zkušenosti s trvalou identifikací osob.

Současný vývoj vypadá následovně:

- Francouzská národní knihovna využívá identifikátor ARK pro fyzické entity, je sdílena infrastruktura resolverů s identifikátory pro knihy a jiné textové entity.
- OCLC WorldCat Identities zahrnuje veškeré fyzické osoby i korporace v bázi OCLC, pro vytvoření identifikátoru je využito identifikační číslo autoritního záznamu, což omezuje využití na lokální úrovni.
- OCLC VIAF (projekt propojující národní soubory autoritních záznamů do mezinárodních autoritních souborů) též užívá identifikační číslo autoritních záznamů.
- ISADN (IFLA) a INSAN (USA) jsou zatím rozpracovanými projekty.
- ISPI /ISNI (CISAC) je normou ISO pro veřejné autority.
- Researcher ID je komerční aktivita Thomson Reuters, producenta báze Web of Science, existují vazby na další systémy.
- Archiv RePec umožňuje trvalou identifikaci přispívajících autorů.
- Digitální archivy užívající systém EPrints 3 disponují vlastní autoritní službou.
- Digital Author Identification (DAI) je využíván v evropském portálu Narcis a je kompatibilní s URI.

V diskusi byl vznesen dotaz, zda bude systém identifikace osob využívat chystaného nástupce rodného čísla, který v současnosti připravuje Ministerstvo vnitra ČR.

VŠKP na švédské Universitě College of Boras

Mgr. Lucie Vyčítalová z Univerzity Pardubice představila zkušenosti ze studijního pobytu na University College of Boras, který absolvovala v letošním roce. Repozitářový systém **DSpace** je ve Švédsku nasazen na čtyřech univerzitách (Borås, Göteborg, Halmstad, Malmö), dále ve Stockholmském ústavu pro vzdělávání (Stockholm Institute of Education), užíván je metadatový standard SVEP, a to pro většinu dokumentů šedé literatury. Existuje **národní registr VŠKP** Upsatser.se (v anglické verzi Essays), přičemž není vyžadováno povinné odevzdávání VŠKP do národního systému. Disertace jsou odevzdávány a zpřístupňovány se souhlasem autora, dále jsou vydávány tiskem a opatřeny ISBN a identifikátorem handle. Univerzitní repozitář v systému DSpace nazvaný BADA zahrnuje kompletní univerzitní produkci odborných textů.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Vycitalova.ppt>

Zpřístupnění eVŠKP a thesis.cz v informačním systému IS/STAG

Po obědě představil **Ing. Tomáš Kotouč** ze Západočeské univerzity v Plzni možnosti kooperace IS/STAG s projektem Theses.cz. IS/STAG je studijní informační systém rozšířený na českých vysokých školách a nabízí mimo jiné komplexní workflow zpracování VŠKP od rezervace, přes zadání, recenzní posudky, obhajobu až po konečné zpracování a uložení. Pomocí webových služeb je v IS/STAG realizováno již několik desítek služeb (seznam studentů na předmětu, hromadný zápis hodnocení studentů na předmětu atd.). Pro spolupráci s dalšími systémy a projekty využívá IS/STAG webových služeb podle obecných standardů (pasivní pro export dat nebo aktivní pro import dat). Pro systém Theses.CZ umožňují webové služby z IS/STAG vyexportovat jak metadata, tak i plný text práce.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Kotouc.ppt>

Konverze metadat z formátu Marc21 do Dublin Core v prostředí univerzitního repozitáře

Následoval příspěvek **Bc. Ireny Baranayové** z Ústavu výpočetní techniky Univerzity Karlovy v Praze. Autorka prezentovala **Univerzitní repozitář** v systému DigiTool, jeho obsah a způsob metodického zajištění. Sběr prací je řízen decentralizovaně na úrovni jednotlivých fakult, VŠKP předává jedenáct fakult z celkových sedmnácti. Podporovány jsou formáty PDF, DOC, RTF, TeX a ODF (Open Document Format); v případě formátu PDF je sledována kvalita PDF souboru s důrazem na parametry indexování a podpory korektního kódování české diakritiky. Abstrakt VŠKP je součástí záznamu, pokud je autorem připraven. Vlastní workflow probíhá tak, že **záznamy tištěných VŠKP jsou primárně vytvářeny v Centrálním katalogu (AKS ALEPH)**, následně je **elektronická verze VŠKP vložena do repozitáře**, kde jsou generována technická metadata, index plného textu a náhled titulní strany. Bibliografický záznam ve formátu MARC21 je stažen z Centrálního katalogu pomocí protokolu Z39.50 (důvodem je budoucí potřeba synchronizace mezi systémy ALEPH a DigiTool pomocí OAI-PMH). Výsledkem je, že v repozitáři se nachází záznamy jak ve formátu MARC21, tak v Dublin Core (resp. EVSKP-MS), přičemž nejsou ošetřeny duplicity, vzniklé u dříve vložených prací ručním zpracováním. Po **konverzi** MARCovského záznamu na Dublin Core zahrnující vypracování vylepšeného mapování budou duplicity odstraněny a zůstanou záznamy pouze v EVSKP-MS. Nově vložené práce již mají bibliografický záznam pouze ve formátu MARC21, ze kterého je exportováno do cílového formátu DC obohaceného o pole EVSKP-MS.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Baranayova.ppt>

Přenos VŠKP pomocí protokolu OAI-PMH

Ing. Mach ve svém dalším příspěvku informoval o vlastním vývoji **exportu metadat ve formátu EVSKP-MS prostřednictvím protokolu OAI-PMH**. Harvester na straně NR VŠKP (projekt Theses.CZ) získává metadata z repozitáře na straně VŠE. Záznamy jsou opatřeny unikátním identifikátorem oai:vse.cz:vskp, jsou ve formátu Dublin Core a EVSKP-MS. Implementace VŠE plně respektuje standard OAI-PMH s výjimkou zahrnutí implementace funkcionality Resumption Token. Funkcionalitu OAI-PMH lze otestovat prostřednictvím nástroje Repository Explorer.

Prezentace ke stažení: <http://www.evskp.cz/dokumenty.php?ukol=3&sek=1&tsekce=2>

IS.MUNI.CZ a služba pro odhalování plagiátů

Po přestávce na kávu pokračoval seminář sekcí na téma plagiátorství.

V příspěvku **Bc. Martina Stančíka** byla představena **koncepce kontroly plagiátorství na Masarykově univerzitě**. Vyhledávání plagiátů VŠKP bylo do Informačního systému Masarykovy univerzity (IS MUNI) zařazeno v roce 2006. Funguje na principu vyhledávání podobností, přičemž je zachována přístupnost či nepřístupnost plného textu koncovému uživateli podle nastavení přístupové politiky. Pokud dokument není označen jako přístupný, je možno přímo kontaktovat správce práv dokumentu. **Workflow kontroly** je rozděleno na **celouniverzitní** (studijní oddělení rektorátu) a **fakultní** úroveň (vedoucí práce a studijní proděkan). Kompetenci k řešení podezření na plagiarismus mají jednotlivé fakulty. Přístup k odevzdávání a archivaci tištěných VŠKP se liší podle jednotlivých fakult, přičemž většina fakult vyžaduje odevzdání a nevyžaduje archivaci tištěné formy (viz obr. 2):

Obr. 2: Odevzdávání a archivace papírových závěrečných prací

Ze zkušeností Masarykovy univerzity vyplývá, že mezi pracemi z období, kdy je protiplagiátorský systém v provozu (2007-2008), je výrazně méně opsaných prací než v roce 2006.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Stancik.ppt>

Odevzdej.cz

Dr. Křípač představil příbuzný projekt systému na odhalování plagiátů nazvaný Odevzdej.cz. Jeho cílem je **rozšířit protiplagiátorské funkce na seminární práce, referáty, eseje a další typy prací** odevzdávaných studenty ve vysokoškolském prostředí. Projekt se nachází ve **fázi žádosti** o přidělení financí pro Centralizovaný rozvojový projekt MŠMT pro rok 2009. Projektovaný systém má podporovat různé způsoby vkládání souborů (individuálně, automatizovaně, ...). Podstatným rysem projektu je **diferencovaný přístup** jednotlivých účastnických škol. Chystá se odlišný mechanismus vkládání i odlišný rozsah účasti jednotlivých institucí. Oproti systému Theses.CZ systém **nepředpokládá zveřejňování** metadat seminárních prací ani plných textů. O projekt projevil zájem deset vysokých škol z ČR.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Kripac1.pdf>

Definice plagiátorství v pojetí VŠE

Na prezentaci dr. Křípače navázal **Ing. Petr Bartoň** z VŠE s přednáškou popisující **přístup VŠE k problému plagiátorství**. Aktivita v oblasti boje proti plagiátům reagují na výzvu děkana Národohospodářské fakulty VŠE (NH VŠE) doc. Jiřího Schwarze, slibující odměnu za každý odhalený plagiát. Smyslem této výzvy byl požadavek na změnu vztahu akademické veřejnosti k problému plagiátorství a zejména podpora „nadrámkové aktivity“ doktorandů a odborných pracovníků NH VŠE. Přednášející shrnul **důvody rozšíření** plagiátorství, nastínil typický **studentský přístup** k problému a definoval na VŠE používaná pravidla pro vyhodnocování plagiátorství v případě pochybností. Závěr přednášky věnoval technikám používaným pro odhalování plagiátorů.

Prezentace ke stažení: <http://www.evskp.cz/Seminar3/seminar3-Barton.pps>

Na odborný program navazoval společenský večer v prostorách Rektorátu VUT.

Zdroj: <<http://www.ikaros.cz/node/5087>>