

Open Repositories 2010

Madrid, Španělsko, 6. – 9.7.2010

Zpráva ze zahraniční cesty
Mgr. Lenka Němečková

Pátý ročník mezinárodní konference týkající se problematiky otevřeného přístupu a institucionálních repozitářů *Open Repositories 2010* pořádala španělská *Fundación Española para la Ciencia y la Tecnología, FECYT* (Španělská nadace pro vědu a techniku) počátkem července 2010 ve španělském Madridu.

Program byl tradičně rozdělen do dvou bloků – tzv. *General Sessions* (obecné prezentace) a *User Groups Sessions* (prezentace týkající se technických parametrů konkrétních softwarových systémů – *Dspace, Fedora, Eprints*). Pro účely rozvoje Ústřední knihovny ČVUT byly klíčové obecně zaměřené prezentace, které probíhaly první tři dny konference, tj. 6. – 8.7. 2010. Tento program byl rozložen do dvou dopoledních a dvou odpoledních bloků tak, že vždy probíhaly paralelně dvě sekce.

Program konference byl sestaven z prezentací projektů vytváření a provozu repozitářů jednotlivých institucí, resp. asociací na institucionální či národní úrovni. Bylo velmi pěkně patrné, že každá komunita je v celkovém rozvoji vytváření IR na odlišné úrovni a zároveň, na základě společenské a situace v dané komunitě, klade důraz na jiný aspekt Open Access repozitářů. Na základě stavu řešení problematiky budování institucionálního repozitáře na ČVUT jsme zvolily takový program, který měl potenciál přinést zajímavé informace pro naši univerzitu. Souhrnně je možné jej shrnout do následujících kategorií (nezávislé na členění bloků konference):

1. **Význam repozitářů pro vědecké pracovníky a hodnocení vědeckého výkonu,**
2. **Interoperabilita a integrace repozitáře do univerzitního informačního systému,**
3. **Uživatelská rozhraní, služby s přidanou hodnotou, hodnocení statistik využívání.**

Klíčový trend dalšího vývoje e-Science ve smyslu vytvoření platformy sociální sítě pro vědeckovýzkumné pracovníky představil v úvodní přednášce **David De Roure** z univerzity v Oxfordu ([Oxford e-Research Centre](http://www.oxford-research-centre.org)). Ve své prezentaci poukázal jednak na nedostatečný zájem o sdílení informací o vlastních výzkumných aktivitách mezi vědeckými pracovníky a současně na nevhodnost využití populárních sociálních sítí (Facebook, MySpace, LinkedIn) v prostředí VaV a prezentoval vlastní projekt kolaborativního prostředí pro VaV MyExperiment (<http://www.myexperiment.org>).

Význam institucionálních repozitářů pro VaV

Nejdiskutovanější tematikou byla bezesporu problematika jednoznačné identifikace autorů a na ně navázaná možnost zpracování citační analýzy. Německý projekt DOARC (<http://www.isn-oldenburg.de/projects/doarc2/> ; <http://doarc.projects.isn-oldenburg.de/>), který má představovat zastřešující platformu nad všemi institucionálními repozitáři a nabízet právě citační analýzy a další informace s přidanou hodnotou pro autory i uživatele dokumentů, zvolil využití volně dostupného systému AuthorClaim (<http://authorclaim.org>). Dalším zajímavým projektem je BibApp (<http://bibapp.org>) americké University of Illinois. Jedná se sice prakticky o registr publikační činnosti pro komunikaci a kolaboraci v rámci univerzity, ovšem zde mě zaujala funkce automatické kontroly časopisu (dle ISSN) oproti databázi SHERPA/RoMEO (<http://www.sherpa.ac.uk/romeo/>) s cílem zjistit možnost uveřejnění plného textu v repozitáři.

Prezentace systému INSPIRE (CERN) ukázala, že dle jejich výzkumů vědci v oboru High Energy Physics jednak téměř jednoznačně využívají repozitáře pro vyhledávání informací a jednak pokud své texty do repozitářů odešle, je tento text citován mnohem dříve a obdrží větší počet citací než pokud je publikován jen v tištěném časopise.

Obdobně jako projekt BibApp, i ve Velké Británii slouží institucionální repozitáře jako platforma pro evidenci výzkumné a publikační činnosti pro potřeby národního hodnocení VaV. Vzhledem k tomu, že pro nový systém hodnocení VaV musí autor prokázat vliv své publikace na rozvoj společnosti a svého oboru, shromažďuje repozitář také ohlasy z jiných zdrojů (např. Twitter, BBC, uskutečněné akce apod.). Právě úzká návaznost na hodnocení VaV je základní motivací autorů pro vkládání svých textů do institucionálních repozitářů.

V Nizozemí existuje jako národní agregátor všech institucionálních repozitářů volně dostupný portál NARCIS (www.narcis.nl), který zastřešuje všech sedmáct univerzitních repozitářů a mimo jiné spravuje systém jednoznačných identifikátorů autorů – tzv. Digital author identifier (DAI).

Na Cornell university existují dvě aktivity – repozitář ArXiv (<http://arxiv.org>), který systém autorských ID provozuje od roku 1999 a nezávislý otevřený registr identifikátorů původců odborných textů ORCID (<http://orcid.securesites.net/>), který by měl uchovávat a znázorňovat vztahy mezi osobami a díly, včetně citačních ohlasů. Takovýto identifikátor původce díla by měl být trvalý, jednoznačný, oborově, geograficky a institucionálně nezávislý (univerzální) a měl by umět definovat více rolí než jen autorské.

V Japonsku například existují identifikátory přidělené vládou, resp. vládními institucemi. V plánu je vytvoření Name Resolveru, které každému výzkumnému pracovníkovi přidělí unikátní ID na základě již získaných identifikátorů, příp. zcela nové) a propojí je na obdobné webové stránky i záznamy publikací těchto autorů.

Interoperabilita a integrace repozitáře do univerzitního informačního systému

Jak bylo zmíněno výše, řada institucionálních repozitářů vznikla zejména kvůli absenci jakéhokoli systému pro sběr a uchování publikační činnosti univerzity – pro potřeby hodnocení VaV. Zde je patrný rozdíl oproti České republice, kde v důsledku aktuální metodiky hodnocení VaV vysoké školy obdobné systémy provozují – ovšem nejsou provozovány na úrovni univerzitní knihovny (jak je tomu často v zahraničí) ani ze své podstaty nejsou budovány jako plnotextové a nemají žádné výraznější služby s přidanou hodnotou.

Velkým problémem integrace je využívání různorodých metadatových formátů v jednotlivých systémech. To se týká sklizně dat, uložení uchování a propojení s dalšími systémy.

V prezentaci odborníků z Oxfordu také zazněla jedna klíčová skutečnost, že samotní uživatelé se stále agregace systémů obávají – a to i navzdory skutečnosti, že všechny uvedené údaje jsou veřejně přístupné. Přesto integraci informací o publikační a citační aktivitě akademických a vědeckovýzkumných pracovníků spolu s registrem zaměstnanců univerzity považují za významný krok.

Integrace ovšem závisí zejména na implementaci jednotných metadatových standardů. Zástupci z University of North Carolina at Chapel Hill představili svůj projekt *Carolina Digital Repository* (<http://cdr.lib.unc.edu>) - integrace publikační činnosti a dalších digitálních textů (speciálních sbírek, diplomových prací, konferenčních příspěvků a časopiseckých článků). Největší výzvou byla implementace metadatových standardů pro dlouhodobé uchování dokumentů (OAIS, METS, PREMIS, MODS) do všech dílčích repozitářů, vytvoření hierarchie v rámci repozitáře a shoda mezi knihovníky a programátory v oblasti terminologie.

Německý projekt *WissGrid* (www.wissgrid.de), resp. německá národní iniciativa vytváření sítí repozitářů *D-Grid* představují další z řady iniciativ vytvoření zastřešující platformy nad individuálními institucionálními či předmětovými repozitáři. Vytváření obdobných zastřešujících platform nad individuálními repozitáři je zjevně dalším trendem rozvoje a integrace individuálních repozitářů. Z toho důvodu je při budování repozitářů nezbytné pokrýt jednak maximální možný rozsah publikační

činnosti dané komunity a zároveň implementovat globálně využívané standardy tak, aby byla umožněna integrace a kooperace s dalšími domácími i zahraničními systémy.

Uživatelská rozhraní, služby s přidanou hodnotou a hodnocení statistik využívanosti

Důraz na vytváření rozhraní s přidanou hodnotou – s funkcemi pro osobní klasifikaci, ukládání, sdílení a komunikaci výsledků VaV – byl prezentován již v úvodní přednášce Davida De Roure. Od cca roku 2007, kdy se o využití prvků Web 2.0 v odborných knihovnicko-informačních zdrojích diskutovalo spíše teoreticky, zaznamenal tento trend velmi rychlý rozvoj do praxe. Kromě níže uvedených příkladů bylo z konferenčních příspěvků patrné, že se vytvářením rozhraní s přidanou hodnotou zabývá (případně plánuje zabývat) většina prezentovaných projektů – byť na této konferenci nebyly z tohoto hlediska prezentovány (např. BibApp, INSPIRE apod.). Z těchto příspěvků jsem nabyla dojmu, že vytváření rozhraní s přidanou hodnotou nad repozitáři je vnímáno již jako samozřejmost, obdobně jako snaha o integraci a národní zastřešení individuálních repozitářů.

Německá DINI (Deutsche Initiative für Netzwerkinformation - <http://www.dini.de>) rozvíjí řadu projektů integrace jednotlivých repozitářů s cílem sjednotit dosud převážně heterogenní repozitáře (po obsahové i metadatové stránce) a umožnit vzájemnou kolaboraci jednotlivých repozitářů a vytvořit národní portál pro přístup ke všem OA dokumentům. Projekt OA Network (<http://www.dini.de/projekte/oa-netzwerk/>) je zastřešující iniciativa, která velmi úzce spolupracuje s dalšími projekty v oblasti OA: OA Statistics (<http://www.dini.de/projekte/oa-statistik/>), DOARC - *Distributed Open Access Reference Citation Services* (<http://www.isn-oldenburg.de/projects/doarc2/>), CARPET – *Community for Academic Reviewing, Publishing and Editorial Technology* (<http://www.dini.de/projekte/carpet/>).

Výlučně sestavováním statistik využívanosti článků v repozitářích se zabývá projekt PIRUS - *Publisher and Institutional Repository Usage Statistics* (<http://www.jisc.ac.uk/whatwedo/programmes/pals3/pirus.aspx>), resp. PIRUS2 (<http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php>). Cílem projektu je vyvinout statistiky využívání jednotlivých článků kompatibilní se statistikami COUNTER a implementovatelný na jakýkoli repozitář, resp. databázi tak, aby bylo možné posoudit globální využívanost individuálních článků napříč více repozitáři.

Konečně alternativou ke statistikám může být i služba Google Analytics. Do roku 2009 měl Google Analytics problém s rozpoznáváním struktury repozitářů a s identifikováním zobrazených záznamů z jiného místa než ze samotného repozitáře a další technické nedostatky, které implementaci do repozitářů v podstatě znemožňovaly. S novou verzí Google Analytics sice ještě několik drobných problémů přetrvává, ovšem v podstatě implementace do repozitáře možná je. Výhodou je i možnost vizualizace statistik prostřednictvím Google Visualization API.

Nové projekty implementace OA a budování IR

Projekty budování institucionálních repozitářů byly na konferenci zastoupeny jen minimálně. Za zmínku stojí prezentace z jedné ze závěrečných sekcí (*Open Access Policy*).

V oboru atmosférických věd ve Spojených Státech Amerických přijali politiku, že všechny recenzované publikace, vzniklé na základě grantů z veřejných prostředků, musí být povinně zveřejněné v OA repozitáři. Jedná se tedy o oborový repozitář pokrývající výsledky práce okolo 150 amerických i mezinárodních institucí. Administrace projektu se ujala knihovna Národního centra pro atmosférický výzkum. Významem pro rozvoj repozitáře je také podpora Národní nadace pro výzkum (National Science Foundation). Autoři popisují tři základní kroky, které museli pro prosazení OA repozitáře podniknout:

1. získat podporu v dané komunitě odborníků (odborných organizací),
2. získat podporu vydavatelů (vyjednávání trvala přibližně rok než dvě klíčové organizace – vydavatelé více než 50% odborných časopisů v oboru – na dohodu přistoupili. Klíčové pro budování repozitáře bylo odstranit embargo na autoarchivaci a opětovné využití článku autory)
3. naplnění repozitáře (nařízení o povinném ukládání finální verze textu do repozitáře).

V současnosti se repozitář naplňuje, dostupný by měl být na podzim roku 2010.

Na univerzitě v Liège (Belgie) spustili na podzim 2008 institucionální repozitář ORBi - (<http://orbi.ulg.ac.be>), který se do února 2010 rozrostl o cca 30.000 bibliografických záznamů a cca 20.000 plných textů. Příčinu úspěchu vidí autoři v:

1. silné institucionální politice pro OA – projekt navazuje na zavedené projekty repozitářů diplomových prací a otevřený portál elektronických časopisů vydávaných na univerzitě. Po uvedení repozitáře vyšlo v platnost nařízení o povinném ukládání publikací vydaných od roku 2001. Další nařízení z roku 2009 stanovilo, že v rámci hodnocení VaV budou evaluovány pouze záznamy uložené v repozitáři ORBi.
2. zapojení více stran do projektu – vývoj repozitáře (konceptualizace, workflow, integrace, sklizení dat, školení apod.) byl v kompetenci týmu odborníků z univerzitní knihovní sítě (často ve spolupráci s jednotlivými fakultami, pokud šlo o odborné záležitosti). Ovšem plné texty ukládají do repozitáře sami autoři, resp. ve sdíleném modu všichni spoluautoři. Na tomto procesu se knihovna již nepodílí.
3. uživatelsky přívětivém prostředí – vzhledem k vkládání dat samotnými uživateli byl kladen velký nárok na jednoduchost celého procesu. Prostředí DSpace bylo upraveno a byly přidány další funkce - např. vkládání záznamu spoluautorem a jejich sdílení tak, aby i autor viděl, že záznam byl již vložen, bylo vytvořeno dvanáct variant formuláře pro různé typy dokumentu. Každý spoluautor musí podepsat smlouvu, ve které dává univerzitě právo zveřejnit záznam a plný text článku. Až poté je záznam zveřejněn. Více informací o autorskoprávní problematice je autorovi poskytováno prostřednictvím FAQ, kde je celý proces podrobně vysvětlen.

Vedle těchto funkcí je repozitář dále propojen na informace o autorovi (kontaktní údaje, seznam publikací apod.), statistiky využitosti a citovanosti jeho publikací (včetně Web of Science, IF apod.) a autorům umožňují snadno generovat seznamy vlastních publikací.

Posledním příspěvkem této sekce byly informace o mezinárodní platformě www.open-access.net, vytvářené v Německu, která by měla sloužit jako hlavní informační portál o politice Open Access zaměřený na autory, vedení univerzity a knihovníky. Všechny informace jsou zde dostupné v angličtině.

Závěrečná diskuse pak shrnula základní směry nejbližšího vývoje v oblasti repozitářů – komplexnost, integrace, kolaborace, interoperabilita, vytvoření provázané infrastruktury, implementace jednotných a otevřených standardů, jednoznačná a trvalá identifikace autorů a textů a udržitelnost a dlouhodobé uchování dokumentů.

V Praze, 19.7.2010