

# Open Repositories 2008, Southampton, Velká Británie, 1. - 4. dubna 2008

Zpráva ze zahraniční cesty  
Mgr. Pavla Rygelová (ÚK VŠB-TUO)

Konference Open Repositories 2008<sup>1</sup> představuje významné mezinárodní fórum pro výměnu zkušeností s budováním institucionálních repozitářů. Letos proběhl 3. ročník s podtitulem „Practice and Innovation“. Pořadatelem byla fakulta School of Electronics and Computer Science of University of Southampton<sup>2</sup>, která se podílí na vývoji open source software EPrints.

Čtyřdenní konference byla rozdělena do dvou částí, první dva dny probíhala společná hlavní sekce, věnující se budování repozitářů z nejrůznějších hledisek, druhá část konference se dělila do sekcí podle typu využívaného softwaru. Paralelně probíhaly tři sekce: setkání uživatelů systému DSpace, EPrints a Fedora. Konferenci uzavíral poslední odpolední blok, v němž byla představena nově se rozvíjející technologie OAI-ORE.

Konferenci otevřel profesor Peter Murray-Rust (University of Cambridge<sup>3</sup>) svou přednáškou s názvem „*Repositories for Scientific Data*“. Peter Murray-Rust je odborníkem v oblasti molekulární informatiky, zabývá se využíváním informačních technologií v přírodních vědách. Významně se podílel na vývoji značkovacího jazyka CML (Chemical Markup Language). Ve svém příspěvku shrnul podstatu směřování dalšího vývoje využívání repozitářů. Vymezil rozdíl mezi pojmy „institutional repositories“ a „data repositories“ (dále jen DRs), přičemž první pojem představuje dnes již tradiční způsob pojetí repozitářů, kdy repozitáře slouží jako úložiště publikační činnosti vědeckých pracovníků, resp. jako úložiště výsledků jejich výzkumné činnosti. Vývoj by však měl směřovat k budování DRs, které budou přímo zapojeny do vědecké práce jednotlivého pracovníka, jeho pracoviště nebo celých výzkumných týmů spolupracujících institucí. Tvorba DRs znamená využít např. technologií Active Directory, Samba nebo SVN, které umožňují vzdálený přístup k souborům a kolaborativní postupy práce, např. na nejrůznějších typech projektů. Budování DRs má velký význam pro vědecké pracovníky v tom, že přímo podporují jejich běžné, zavedené pracovní postupy a nenutí je k nadbytečné práci nad rámec jejich vědeckého zájmu.

Příspěvky byly rozděleny do několika tematických okruhů (interoperabilita, modely, architektura a struktura repozitářů, sociální síťování, Web 2.0, využívání a vliv repozitářů, právní otázky a další). V sekci Interoperabilita zazněl příspěvek Julie Allinson (UKOLN<sup>4</sup>, University of Bath and University of York) s názvem „*WORD: simple web service offering repository deposit*“. Jedná se o projekt zahájený v roce 2007 a zaštiťuje jej organizace Joint Information Systems Committee<sup>5</sup> (JISC). Cílem projektu je standardizovat proces přispívání do repozitářů, tedy podpořit vstupní a zároveň klíčový krok, určující zásadním způsobem

---

<sup>1</sup> <http://or08.ecs.soton.ac.uk/>

<sup>2</sup> <http://www.ecs.soton.ac.uk/>

<sup>3</sup> <http://www.cam.ac.uk/>

<sup>4</sup> <http://www.ukoln.ac.uk/>

<sup>5</sup> <http://www.jisc.ac.uk/>

životaschopnost každého repozitáře. Bez efektivního naplňování (přispívání, vkládání) pozbývá samotný repozitář svůj účel a nemůže být rozvíjena žádná z jeho dalších funkcí. Výsledkem projektu je webová služba SWORD, která umožňuje mimo jiné vkládání digitálního obsahu do jednoho či více repozitářů najednou. SWORD implementovaly čtyři softwarové platformy: EPrints, DSpace, Fedora a IntraLibrary a představuje další nástroj pro užší interoperabilitu repozitářů.

Velice podnětné byly příspěvky týkající se využívání repozitářů, resp. měření informačního dopadu jejich obsahu ve vědeckém světě. Johan Bollen a Herbert Van de Sompel (Digital Library Research & Prototyping Team, Research Library, Los Alamos National Laboratory<sup>6</sup>) promluvili o projektu MESUR, který řeší problematiku evaluace vědy a výzkumu v digitální éře. Vstupní motivací celého projektu byl fakt, že zavedené metody evaluace vědecké práce postavené na citačních statistikách (impakt faktor, h-index a další) byly vyvinuty a představovaly dostatečně spolehlivé a platné měřítka v době převážně tištěných médií, ovšem pro dnešní narůstající elektronické publikování vědeckých výsledků je potřeba hledat i jiné způsoby měření informačního dopadu. Autoři projektu MESUR zvolili hledisko využívání jako ekvivalent k citačním metrikám. Přitom se však nejedná o prosté měření využívání, ale „usage evaluation“ kombinuje celou řadu hledisek - využívání, citovanost, bibliografická data. Spolehlivost výsledků měření zajišťuje obrovské množství zpracovávaných dat, získaných z nejrůznějších zdrojů. Hledisko využívání digitálních médií nese několik nových aspektů, které pomocí klasických měření nelze vyjádřit, např. údaje o využívání („usage data“) v sobě nesou širší informační kontext, který lépe zachytí propojenost digitálních zdrojů, vzájemné vztahy jejich autorů, uživatelů apod. Projekt MESURE kombinuje data o využívání a data získaná tzv. sítíovou metrikou („network-based metrics“), která využívá základních principů teorie sociálních sítí. Výsledky projektu MESURE mapují nejen vědecké časopisy, ale i nejširší spektrum vědecké publikační činnosti vznikající vně tradičního publikačního prostoru (šedou literaturu a netextové typy dokumentů: preprinty, blogy, software, chemické vzorce apod.) a v neposlední řadě na rozdíl od citačního impakt faktoru lze „usage impact factor“ daného digitálního dokumentu sledovat bezprostředně po jeho zveřejnění.

Setkání uživatelů systému DSpace proběhlo ve dvou dnech a bylo sestaveno z mnoha příspěvků, které názorně demonstrovaly nejrozmanitější způsoby budování repozitáře pomocí open source software DSpace. DSpace má širokou uživatelskou základnu, rozsetou doslova po celém světě, příspěvky v bloku případových studií pocházely z např. z Řecka (projekt Pandektis - archiv řeckých kulturně-historických materiálů, Jamajky (projekt eJamaica - veřejně prospěšný repozitář, Číny (projekt Virtual Olympic Museum).

Největší pozornost se ovšem soustředila na vystoupení programátorů softwaru DSpace, kteří prezentovali novou verzi systému, tj. verzi 1.5, která představuje podstatný posun ve vývoji systému DSpace. Softwarové jádro systému zaznamenalo akcelerující růst, DSpace získal nové rysy a implementoval několik nových technologií. Budoucí rozvoj softwaru bude řízen pomocí nástroje Maven, založeného na konceptu „project object model“ (POM). Pomocí Maven lze řídit softwarové projekty postavené na Java platformě, umožňuje řídit

---

<sup>6</sup> <http://library.lanl.gov/>

distribuovaný rozvoj softwaru, má usnadňovat zapojení nových rysů a posílit větší modularitu celého systému.

Součástí DSpace, verze 1.5 se stal šablonovací modul Manakin, vyvíjený Texas A&M University<sup>7</sup>. Manakin nabízí institucím možnost úprav uživatelského rozhraní DSpace podle specifických požadavků repozitáře, komunity nebo kolekce. Úpravy vzhledu či celého designu institucionálního repozitáře pomocí Manakinu jsou snadnější a snáze přenositelné do vyšších verzí systému DSpace, než úpravy pomocí skriptů JSP. Manakin daleko účinněji odděluje složku prezentační od logické.

Další dva rysy systému DSpace rozšiřují možnosti procesu vkládání a vyhledávání. Původní vkládací workflow obsahoval několik kroků v neměnném pořadí, novou vlastností se stala možnost konfigurace jednotlivých kroků vkládacího workflow podle vlastních potřeb. Jako názorný příklad lze uvést posunutí kroku o licenčním ujednání na první místo, čímž je uživateli umožněno, aby se nejprve seznámil s licenčními podmínkami a rozhodl se je akceptovat dříve, než projde jednoduchým, ale určitou pozornost vyžadujícím krokem vkládání metadat, popřípadě vložením souboru/ů do repozitáře. Širší funkcionalitu nabízí také systém vyhledávání, byly posíleny zejména funkce prohlížení, systémový administrátor získal nástroj pro snadnější přidávání potřebných indexů k prohlížení a nově je přidáno řazení výsledků podle relevance, autora, data vydání nebo data vložení do archivu apod.

Nově jsou implementovány do DSpace technologie SWORD, LNI (Light Network Interface) - využívající protololy WebDAV a SOAP pro vzdálené řízení obsahu repozitáře a tzv. „event system“ pro automatické řízení událostí (změn) v repozitáři.

Konferenci zakončoval celý blok prezentací věnovaných technologii OAI-ORE (Open Archives Initiative - Object Reuse and Exchange). Blok zahajoval Herbert Van de Sompel<sup>8</sup>, který promluvil o motivacích a historickém kontextu vzniku této technologie. OAI-ORE reprezentuje další funkční nadstavbu repozitářů, respektive slouží k opětovnému využívání informačního obsahu v nich uloženého. Životní cyklus digitálních objektů by neměl končit uložením do repozitáře, ale právě by měl tímto krokem začínat. OAI-ORE má umožnit výměnu a opětovné využívání digitálních objektů napříč repozitáři. Model OAI-ORE funguje na základě standardů, známých z prostředí webu (XML, Atom, RDF, URI) a zatím je stále ve stádiu testování.

Všechny prezentace jsou dostupné v konferenčním archivu<sup>9</sup>. Další ročník konference Open Repositories proběhne v květnu 2009 v Atlantě<sup>10</sup>.

Mgr. Pavla Rygelová, 13. 5. 2008

---

<sup>7</sup> <http://www.tamu.edu/>

<sup>8</sup> <http://library.lanl.gov/>

<sup>9</sup> <http://pubs.or08.ecs.soton.ac.uk/>

<sup>10</sup> <http://or09.library.gatech.edu/>