

Závěrečná správa k Open Repositories 2013, Charlottetown, Kanada

Ivan Masár
8. - 12.7.2013

Open Repositories je prestížna medzinárodná konferencia o inštitucionálnych repozitároch a Open Access, tradične sprevádzaná stretnutiami používateľských skupín DSpace, Fedora, a Eprints. Tento rok sa konal 8. ročník tejto konferencie v Charlottetowne na Ostrove Princa Edwarda, kde ju hostovala University of Prince Edward Island, kde vznikla platforma na tvorbu repozitárov s názvom Islandora.

DSpace User Group Meeting (OR13 Pre-Conference)

<http://or2013.net/sessions/dspace-developers-meeting>

Stretnutia používateľov DSpace sa zúčastnilo asi 30 ľudí z celého sveta. Okrem iných boli prítomní aj títo členovia skupiny DSpace committers: Andrea Bollini, Peter Dietz, Tim Donohue, Keith Gilbertson, Bram Luyten, Richard Jones, Hardy Pottinger, Richard Rodgers, Andrea Schweer, Graham Triggs a Mark Wood a členovia DSpace Community Advisory Team: Bram Luyten, James Evans, Maureen Walsh, Sarah Potvin, Sarah Shreeves a Valorie Hollister.

Stručné zhrnutie DSUG: Bol prezentovaný dokument vízie vývoja DSpace na nasledujúcich 3-5 rokov. Pripomienkovaný bol návrh DCAT ohľadne zlepšenia podpory metadátových štandardov. Prebehla diskusia o modularite DSpace, o podpore konceptu SHARE v DSpace (SHARE sa týka USA, ale jednotlivé body sú všeobecne relevantné) a diskusia vydania DSpace 4.0 a nových vlastností v ňom.

Zhrnutie v angličtine:

https://docs.google.com/document/d/1QzrmHDvpJlpsVo6e8tJvWk2rXZm1p_F3TaUcEKleODU

Reuse and Reproducibility: Opportunities and Challenges

<http://or2013.net/content/opening-plenary-victoria-stodden>

Otváracia prednáška vedkyne Victorie Stoddenovej s titulom associate professor v oblasti štatistiky a zakladateľkyne RunMyCode, webovej služby umožňujúcej komukoľvek spúšťať kód sprevádzajúci vedecké publikácie a tak reprodukovať výsledky vedy, sa niesla v duchu komentára k vývoju vedy. Boli popísané tradičné vedecké vetvy - deduktívna, založená na matematike a formálnej logike a empirická, založená na štatistickej analýze riadených experimentov. Nasledovala predpoveď a želanie, aby sa rozvinula vetva výpočtová, založená na

počítačových simuláciách a ich reprodukovateľnosti. Predpokladom k tomu je publikácia nie len vedeckých výsledkov, ale aj vstupných dát a algoritmov, na základe ktorých boli dosiahnuté.

An Open Access Triptych: IsItOpenAccess, the Open Citations Corpus and IDFind

<http://or2013.net/sessions/open-access-triptych-isitopenaccess-open-citations-corpus-and-idfind>

V tomto bloku sme sa dozvedeli o troch nových službách užitočných pri zisťovaní stavu open access článkov. IsItOpenAccess je služba, ktorá nahlási licenčné podmienky a stav open access diel na základe dodanej množiny identifikátorov. Momentálne podporuje DOI a PubMed ID. Open Citations Corpus je open access bibliografická databáza, ktorá navyše spája jednotlivé záznamy prostredníctvom citácií. IDFind je koncept služby, ktorá pri zadaní identifikátora zistí, aký identifikátor to je (ISSN/ISBN/DOIs/PMID/ORCIDs/UUID/...).

From Open Access Metadata to Open Access Content: Two Principles for Increased Visibility of Open Access Content

<http://or2013.net/sessions/open-access-metadata-open-access-content-two-principles-increase-visibility-open-access-co>

Autor začal kritikou dostupnosti plných textov v režime open access, čo podložil štatistikou. Napriek tomu, že väčšinou sú dostupné metadáta, samotné plné texty je často ťažké dohľadať. Ako riešenie navrhol dva princípy, ktoré nie sú prevratné, ale pragmatické - vždy v metadátovom zázname v príslušnom formáte uvádzať odkaz na plný text a umožniť strojový prístup k plnému textu na rovnakej úrovni ako majú ľudia. Dôsledky oboch týchto princípov ďalej rozvádza.

Compliance with Funders' Mandates - FACT

<http://or2013.net/sessions/compliance-funders-mandates-%E2%80%93-fact>

Tento príspevok pojednával o kontrole, či licenčné podmienky vydaných článkov zodpovedajú povinnostiam, ktoré si stanovili inštitúcie financujúce daný článok. Zároveň predstavil čerstvo spustenú službu SHERPA/FACT, ktorá interpretuje dáta SHERPA/RoMEO a po zadaní sponzora a vydavateľa článku zobrazí dostupné spôsoby licencovania, resp. či časopis podmienky sponzora spĺňa.

Blok o riadení autorít bol rozdelený do troch samostatných častí:

Implementing Name Authority Control into Institutional Repositories: A Staged Approach

<http://or2013.net/sessions/implementing-name-authority-control-institutional-repositories-staged->

[approach](#)

Prvá časť popisovala praktické skúsenosti University of North Texas (UNT) Libraries pri implementovaní autorít do IR, ktoré zhrnula do štyroch krokov: zverejnenie riadených slovníkov ako five-star open linked data, ich sprístupnenie tvorcom metadát na aktívne využitie, ukladanie odkazov namiesto reťazcov a zmysluplné zobrazenie dát používateľom.

ORCID Implementation in Open Access Repositories and Institutional Research Information Management Systems

<http://or2013.net/sessions/orcid-implementation-open-access-repositories-and-institutional-research-information-manage>

Táto prednáška bola predstavením registra identifikátorov autorov ORCID a zhrnutím možností jeho integrácie do IR a CRIS. Popísala postup registrácie individuálneho autora, vyplnenie profilu a prepojenie so súvisiacimi službami, štruktúru organizácie zaštitujúcej ORCID a informácie o členstve, aktuálne štatistiky využívania a rozhranie API. ORCID v súčasnej dobe nepodporuje afiliácie autorov, ale kvôli veľkému záujmu túto vlastnosť pripravuje. V súvislosti s touto prednáškou boli spomenuté aj minulé a súčasné aktivity na podporu ORCID v DSpace - v rámci Developer's Challenge vytvorili Andrea Schweer a Ryan Scherle koncept integrácie ORCID do tzv. submission process v DSpace; v súčasnosti ohlásil výzvu na rozsiahlejšiu integráciu Andrea Bollini s cieľom získať grant od nadácie Alfreda P. Sloana. Moja otázka k prednáške sa týkala nutnosti členstva inštitúcie v ORCID kvôli možnosti dávkového nahrávania autorov danej inštitúcie; odpoveďou bolo, že máme odporúčať autorom individuálnu registráciu, ktorá je dostatočne jednoduchá.

Using SKOS and FOAF for Name Authority in the Institutional Repository

<http://or2013.net/sessions/using-skos-and-foaf-name-authority-institutional-repository>

Oregon State University Libraries tiež riešili problém autoritných záznamov pre IR. Rozhodli sa ísť cestou tvorby autoritných záznamov pomocou Linked Data s využitím štandardov SKOS (Simple Knowledge Organization System) and FOAF (Friend of a Friend). Popisovali zvolený dátový model, implementáciu spolupracujúcu s DSpace, workflow, výzvy automatizovaného dohľadávania - zdroje dát, konflikty mien, varianty, afiliácie; a ich algoritmické riešenie.

Use of Solr and Xapian in the Invenio document repository software

<http://or2013.net/sessions/use-solr-and-xapian-invenio-document-repository-software>

Táto veľmi zaujímavá prednáška priamo od autorov Invenia z CERNU sa zamerala na konkrétny problém - experiment s hľadaním najlepšieho indexovacieho nástroja pre Invenio. Dvaja vybraní

kandidáti, pre ktorých bola vytvorená fungujúca implementácia, boli Solr a Xapian. Takéto zovšeobecnenie viac ako jedného backendu do jediného modulu IR je v rámci dnes dostupných IR jedinečné. Bol popísaný proces indexovania a následného vyhľadávania v systéme Invenio jeho problémy, najmä tie súvisiace výkonom v prípade obrovského počtu záznamov. Škálovateľnosť oboch riešení bola demonštrovaná grafmi počtu dokumentov vs. času vyhľadávania, kde sa ukázala vhodnosť využitia Solr pri množine výsledkov vyhľadávania v ráde tisícov a vhodnosť Xapianu pri väčšej množine. Moja otázka na autora sa týkala získania kompletnej množiny výsledkov od Solr, čo štandardne nie je možné. To autori vyriešili napísaním vlastného pluginu pre Solr, ktorý túto množinu vracia vo forme bitového poľa, kde jeden bit zodpovedá jednému dokumentu a indikuje jeho prítomnosť v množine výsledkov.

Minute madness a prezentácia plagátov

<http://or2013.net/content/minute-madness>

<http://or2013.net/sessions/poster-reception>

Cieľom minute madness bolo v extrémne krátkom čase jednej minúty umožniť stručnú prezentáciu jednotlivých posterov. Z nich ma najviac zaujali automatická tvorba metadát na základe vyhľadávacích dotazov Google získaných z logov HTTP servera; prezentácia Českej digitálnej knižnice Knižnice Akadémie vied založenej na systémoch Fedora Commons a kramerus; a poster CERNu o systéme Zenodo na spoluprácu pri výskume.

Sisyphian Task or the Holy Grail? Satisfying Demands of Higher Education through Community-Driven Open Source Projects

<http://or2013.net/sessions/panel-sisyphian-task-or-holy-grail-satisfying-demands-higher-education-through-community-driven-open-source-projects>

Panelová diskusia o radiaciach štruktúrach v open source projektoch bola zhrnutím dostupných modelov a skúseností vysokoškolských inštitúcií v tejto oblasti; osobne ma však neoslovila ničím novým.

24 x 7

<http://or2013.net/sessions/24-x-7-330-0>

24 x 7 bol blok prezentácií vo formáte do 24 snímok a do 7 minút. Za všetky spomeniem prezentácie DSpace Grahama Triggsa o spájaní dát o výskume inštitúcie z rôznych systémov a Simeona Warnera o protokole ResourceSync, ktorý je koncipovaný ako zovšeobecnený nástupca OAI-PMH a Marka Jordana o platforme easyLOD na jednoduché sprístupnenie dát vo forme Linked Data.

Invisible Repositories, Re-Use and Reproducible Research

<http://or2013.net/sessions/invisible-repositories-re-use-and-reproducible-research>

Kritika súčasného stavu, kde inštitucionálne repozitáre sú viditeľné a inštitucionálny výskum je

“neviditeľný”, predstavila víziu, že inštitucionálne repozitáre by mali byť “neviditeľné” a viditeľný by mal byť samostatný obsah - inštitucionálny výskum, obsahovala veľké množstvo ukážok vzhľad a rozhraní jednotlivých repozitárov a spôsobov ako sa k ich obsahu dostať, pričom tento spôsob nebol vždy pre bežného používateľa zjavný.

Developer Challenge Show & Tell

<http://or2013.net/content/developer-challenge-show-tell>

Developer Challenge je výzva vytvoriť za dobu cca 2 dní konferencie funkčný a užitočný koncept softvéru alebo služby súvisiacej s OR. Výzvu vyhral tím skladajúci sa z jediného vývojára, ktorým bol Kevin Bowrin so softvérom na kontrolu stupňa validity dokumentu so štandardom PDF/A.

OR2014 Welcome

<http://or2013.net/content/or2014-welcome>

Tento bol bol pozvánkou na ďalší ročník konferencie Open Repositories 2014, ktorý sa bude konať v Helsinkách. Miesto konania zaujímavou a vtipnou formou popísali prezentujúci z Univerzitetnej knižnice Helsinky a z Fínskej národnej knižnice. Nechýbalo ani pútavé video.

DuraSpace Plenary

<http://or2013.net/sessions/duraspace-plenary>

Plenárne zasadnutie neziskovej organizácie DuraSpace zhrnulo jej hospodárenie, výsledky, súčasnú prácu a plány.

Fedora Roadmap/Building Fedora Futures

<http://or2013.net/sessions/fedora-roadmapbuilding-fedora-futures>

Charizmatický Edwin Shin hovoril o tom ako malá, sústredená skupina vývojárov dokázala od základu prepísať novú verziu Fedora Commons zohľadňujúcu nové ciele, najmä škálovateľnosť a výkon, pričom zachovali 80 % pôvodnej funkcionality. Tento úspech pripisuje najmä možnosti na týždeň túto malú skupinu dostať fyzicky na jedno miesto, aby mali možnosť sa projektu naplno venovať.

DSpace RoadMap and 3-to-5 Year Vision

<http://or2013.net/sessions/dspace-roadmap-and-3-5-year-vision-0>

Vízia ďalšieho vývoja DSpace v horizonte 3 až 5 rokov, na ktorej pracovala organizácia DuraSpace s pomerne malou skupinou zástupcov severoamerických inštitúcií podporujúcich vývoj DSpace tu bola prezentovaná s výzvou o ďalšie komentáre od celosvetovej komunity. Vízia obsahuje silné a slabé stránky DSpace, prípady použitia a smery, v ktorých by bol možný a žiadúci rozvoj v tomto časovom horizonte.

Introducing DSpace 3.1

<http://or2013.net/sessions/introducing-dspace-31>

Táto prednáška predstavila novinky v aktuálnej verzii DSpace 3 a jej existujúce implementácie. Prezentoval ju kolega Hardy Pottinger a je obdobia mojej prezentácie z workshopu AKVŠ. Začínáme s (otevřeným) inštitucionálnym repozitárom, 2012.

Implementing ResourceSync for DSpace

<http://or2013.net/sessions/implementing-resourcesync-dspace>

Richard Jones prezentoval svoju prácu na implementácii fungujúceho konceptu modulu ResourceSync pre DSpace, ktorý má byť alternatívou k modulu OAI-PMH a má umožniť harvestovanie DSpace pomocou tohto nového protokolu, vrátane implementácie klientskeho softvéru. Výsledkom tejto snahy sú praktické skúsenosti pri implementácii (mapovanie množín, reprezentácia metadát a súborov, ...), javovské knižnice servera ResourceSync, PHP knižnice klienta ResourceSync a modul servera pre DSpace.

DSpace REST API

<http://or2013.net/sessions/panel-dspace-rest-api-630>

Táto panelová diskusia hostovala členov pracovnej skupiny, ktorá sa zaoberá výberom/vývojom vhodnej implementácie REST API pre DSpace. V súčasnej dobe existujú dve rozličné implementácie REST API v produkčnej kvalite. Ďalej členovia tejto pracovnej skupiny vyvíjajú vlastnú, modernú implementáciu a skúmajú možné prípady využitia tohto API. Diskusia sa zaoberala okrem spomínaného aj ďalším postupom pri výbere REST API a možnosťou zahrnúť REST API do vydania DSpace 4.0.

Moving BIG DATA into DSpace

<http://or2013.net/sessions/moving-big-data-dspace>

Pridávanie veľkého objemu dát do DSpace je aktuálnym problémom pre inštitúcie, ktorým DSpace slúži ako úložisko pre výskumné dáta. University of Exeter sa rozhodla tento problém riešiť integráciou služby Globus, ktorá sa práve na prenos veľkých dát prostredníctvom internetu špecializuje. V tejto prezentácii predstavili svoju prácu v tejto oblasti.

Slice and Dice your repository usage: Google Analytics for DSpace workshop

<http://or2013.net/sessions/workshop-slice-and-dice-your-repository-usage-%E2%80%A8google-analytics-dspace-workshop-1000>

Ako člen programovej komisie som uvádzal túto prednášku určenú pre správcov obsahu repozitárov, ktorá im predstavila možnosti Google Analytics pri sledovaní využívania obsahu IR. Bola koncipovaná ako workshop, takže sme prešli všetkými krokmi nastavenia DSpace do GA a vyskúšali sme si možnosti štandardných aj vlastných správ a kampaní. Pretože Google Analytics je koncipovaný pre webstránky všeobecne a nie špeciálne pre IR a preto, že má určité obmedzenia na množstvo obsahu, vhod prišlo aj porovnanie s ďalšími možnosťami zhromažďovania a prezentovania štatistík dostupných v DSpace.

Integrate external bibliographic services in DSpace submission process to make self-deposit easy and improve metadata quality and presence of full-text

<http://or2013.net/sessions/integrate-external-bibliographic-services-dspace-submission-process-make-self-deposit-easy->

Andrea Bollini vysvetlil zlúčenie svojho zamestnávateľa - konzorcií CILEA do konzorcia CINECA. Tieto konzorciá sú významným vývojárom DSpace. Andrea prezentoval, čo CINECA robí a ako do toho zapadá DSpace. Ďalej hovoril o potrebe integrovať externé (nielen bibliografické) služby ako Scopus, PubMed, CrossRef, ArXiv, ale aj napr. SHERPA/RoMEO a o ich využití v rámci submission process.

Usage Statistics powered by Elastic Search

<http://or2013.net/sessions/usage-statistics-powered-elastic-search>

Peter Dietz prezentoval svoje nové štatistiky založené na platforme ElasticSearch, ktoré sú novinkou v DSpace 3. Niektoré zlepšenia budú dostupné v DSpace 4. Hlavnou výhodou využitia ElasticSearch je šlákovateľnosť, ktorá v prípade štatistík založených na Solr vykazovala problémy pri súčasnom zápise a čítaní štatistík. Pútavé sú grafy, ktoré sa vykresľujú pomocou Google Chart API.

Implementing an altmetrics reporting service into DSpace using Altmetric.com

<http://or2013.net/sessions/implementing-altmetrics-reporting-service-dspace-using-altmetriccom>

Tradičné metriky (štatistiky) využívania IR sú postavené na počte zobrazených stránok a stiahnutých fulltextov. Tradičné hodnotenie výskuných článkov je založené na citáciách. V súčasnej dobe existujú mnohé iné spôsoby využívania týchto článkov, ktoré sa v týchto tradičných metrikách buď neprejavujú, alebo majú ďalšie aspekty, ktoré by bolo užitočné sledovať. Prominentnou službou, ktorá sa na zhromažďovaní týchto "alternatívnych" metrik zaoberá, je Altmetric.com. Táto prednáška hovorí o implementácii služby Altmetric.com do

DSpace, konkrétne na úrovni tém rozhrania XMLUI, aby sa pri jednotlivých článkoch zobrazoval tzv. "odznak" Altmetric.

DSpace SWORDv2: A Technology How-To

<http://or2013.net/sessions/dspace-swordv2-technology-how>

SWORD je overená technológia na zakladanie nových záznamov v DSpace. Richard Jones vyvinul novú verziu modulu SWORDv2 servera pre DSpace, ktorá by mala byť dostupná v DSpace 4. Táto prednáška popisuje technické základy využitia tohto protokolu, operácie CRUD a ich aplikáciu v prostredí DSpace.

Introduction to Github: Modern source code management for DSpace

<http://or2013.net/sessions/workshop-introduction-github-modern-source-code-managment-dspace-230>

Moja samotná aktívna účasť bola naplánovaná až na úplný záver konferencie. Z dôvodov spomínaných v mojej prihláške do výberového konania som tento workshop spoluprezentoval s kolegom Bramom Luytenom. Zúčastnilo sa ho cca 30 záujemcov, ktorí mali rôzne stupne technických znalostí, ako ukázal krátky prieskum na začiatku workshopu. Očakávali sme, že sieť Wi-Fi na konferencii súčasny nápor workshopu nezvládne, preto sme pripravené dáta a programy rozdistribuovali na niekoľkých flash diskoch. Teoretická časť prednášky popísala čo je Git a GitHub, na čo slúžia, princípy a výhody distribuovaného modelu vývoja a praktický prínos, ktorý tento model mal pre DSpace a prejavil sa na množstve nových príspevkov v DSpace 3. Praktická časť workshopu spočívala prevedenie účastníkov inštaláciou softvéru git na ich laptopy, vytvorením účtu na webe GitHub, získaním kódu DSpace a základnými konceptami gitu ako commit, vetva, pull a push. Záverečný prieskum ukázal, že všetci tieto základné úlohy zvládli.

Mojím stanoveným cieľom bolo najmä oboznámiť potenciálnych prispievateľov do projektu DSpace s aktuálnym procesom prispievania pomocou GitHubu, čo by sa malo odzrkadliť na množstve nových vlastností, vylepšení a opráv chýb zaslaných na zaradenie do DSpace verzie 4 pred koncom tohto roka.

O konferencii Open Repositories 2013 budem prezentovať na konferencii Bibliotheca academica 2013. Zároveň chcem poďakovať AKVŠ, ktorá mi poskytnutím finančných prostriedkov umožnila účasť na tejto prestížnej konferencii.