

Závěrečná správa k Open Repositories 2018, Bozeman, USA

Ivan Masár
4. - 7. 6. 2018

Open Repositories je prestížna medzinárodná konferencia o inštitucionálnych repozitároch a Open Access, ktorej už tradične predchádzajú workshopy týkajúce sa DSpace, Fedora Commons a tento rok napr. aj Wikimedia a SWORD. Tento rok v meste Bozeman, MT v USA host'ovala Montana State University 13. ročník tejto konferencie. Nosná téma konferencie „Sustaining Open“ naráža na to, že už viac ako desaťročie budujeme otvorené repozitáre a naša pozornosť by sa mala obrátiť k tomu, ako otvorenosť udržať.

Tento ročník experimentoval s novým formátom, kde prezentácie neboli organizované do jednotlivých „interest groups“ podľa platformy, ale rozčlenené čisto tematicky.

Oficiálny web konferencie s programom je
<http://or2018.net>

V programe sa nachádzajú odkazy na poznámky (Notes) k jednotlivým častiam konferencie, často vrátane odkazov na prezentácie:

<https://www.conftool.net/or2018/index.php?page=browseSessions>

Neskôr bude sprístupnený archív jednotlivých prezentácií.

Streaming vystúpení je archivovaný na:

<https://www.youtube.com/channel/UCncXCjL2X4QQtZECWbsuUqA>

Neskôr bude sprístupnený videoarchív zostrihaný na jednotlivé prednášky.

Ďalšou novinkou tento rok boli neformálne stretnutia účastníkov už v predvečer konania workshopov, kde sme mali možnosť zoznámiť sa a vzájomne si ukázať, na čom pracujeme.

OR18 Pre-Conference

Pred samotnou konferenciou sa konali workshopy jednotlivých záujmových skupín, kde si na to svoje prišli vývojári aj nevívojári – okrem tradičných platforiem repozitárov tu boli časti zamerané na novú verziu 3 protokolu SWORD, skúsenosti s budovaním národnej dátovej služby (Jisc Research Data Shared Service) alebo workshop o práci s LOD (linked open data) platformou Wikidata.

DSpace Rest API: Getting Started with the DSpace 7 REST API

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=285

Časť zameraná na vývoj DSpace predstavila aktuálne novinky v oblasti REST API, ktoré bolo pre DSpace 7 prepísané nad Spring a rozšírené o mnohé chýbajúce časti. Celková koncepcia sa síce od verzie DSpace 6 zásadne nezmenila, ale spätná kompatibilita sa oficiálne neudržiava. Vďaka

samodokumentujúcemu API je zaujímavou novinkou generovanie interaktívneho HTML rozhrania k tomuto API, ktoré umožňuje softvér HAL browser. Nad novým REST API prebieha tvorba nového webového používateľského rozhrania založeného na Angular, o tom však už podrobnejšie informuje ďalší workshop.

DSpace 7 Angular UI: Getting Started with the DSpace 7 Angular UI

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=279

Hlavná časť konferencie

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=229

Konferenciu oficiálne otvoril keynote od Casey Fiesler s názvom „Growing Their Own: Building an Archive and a Community for Fanfiction“ o skúsenostiach aj útrapách pri budovaní webu Archive of Our Own, kde fanúšikovia môžu zverejňovať tzv. fan fiction, svoju literatúru inšpirovanú inými, často známymi dielami (napr. Star Wars, Harry Potter, ...). Princípy a dobré i zlé skúsenosti budovaním komunity ilustrovala na problémoch s prechodom od konkurenčnej platformy, s autorskými právami, anonymitou a pseudonymitou, stanovením si a dodržiavaním zásad projektu, udržateľnosťou projektu, ochranou používateľov, ich motiváciou a zapojením do vedenia a udržiavania projektu.

Konferenciu tradične sprevádzali prezentácie plagátov v rámci Minute Madness, po ktorých nasledovala samotná recepcia s plagátmi; ideas challenge, v rámci ktorej účastníci konferencie vytvorili malé tímy a prezentovali návrh užitočných riešení zaujímavých problémov z oblasti repozitárov a nechýbala ani večera, ktorá poskytla možnosť konverzácie s ostatnými účastníkmi konferencie.

Metadata mastery

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=233

Časť venovaná práci s metadátami predstavila nástroj na tvorbu hlásení ohľadne obsahu repozitára DSpace, ktorý je štandardnou súčasťou REST API DSpace od verzie 5 a jeho webové rozhranie ponúka komfortné možnosti aj bez znalosti SQL či programovania. Nasledovalo zhrnutie skúseností s migráciou z DSpace na Hydra. Migráciu autor využil zároveň ako príležitosť vyčistiť samotné metadáta a predstavil nástroje, ktoré na to vyvinul. Ďalšia prednáška bola rýchlym letom nad technológiami, ktoré sa často používajú na prácu s metadátami, ich porovnanie a skúsenosti s ich používaním.

Research Data & Digital Preservation

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=249

V prísne sledovanom formáte 24 snímok odprezentovaných za 7 minút (24x7) sa v tomto bloku vystriedalo 8 prednášajúcich. Prvý z nich sa venoval ťažkému problému archivácie výskumných dát vo formáte relačných databáz, kde ako riešenie navrhol archiváciu celých virtualizovaných prostredí s predinštalovaným softvérom a dátami. Ďalší popisoval bežné aj nečakané hrozby a reálne skúsenosti so stratou dát pri práci s veľkými repozitármi (Zenodo). Predstavená boli aj možnosti certifikácie dátových repozitárov. Nasledovalo zamyslenie nad tým, aké nástroje a postupy

používame pri budovaní repozitára, pretože časom strácame odstup a udržateľnosť vyžaduje, aby sme tieto aspekty občas prehodnotili. Prezentujúci z Hamburgu poukazoval na to, že pri výbere platformy repozitára u nich nakoniec došlo nielen k zohľadneniu vopred stanovených „tvrdých“ dát a kritérií, ale aj k prihliadnutiu na vzhľad a použiteľnosť systému z pohľadu koncového používateľa.

Carnival of clever tips & tools

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=292

Sekcia zameraná na vývojárov predstavila ako sa pozrieť „pod kapotu“ a pracovať priamo so Solr, ktorý mnohé platformy využívajú. Na dynamické zobrazovanie veľkých obrázkov v súlade so štandardom IIIF predstavila výkonný server Cantaloupe a jeho výhody, najmä jeho jednoduché použitie. W3C Web Annotations popisujú štandardizovaný spôsob ako zobrazovať vo webovom rozhraní textové popisy k napr. multimediálnemu obsahu a ich implementáciu v systéme Islandora. Posledný prednášajúci sa venoval vývoju a použitiu hlasových počítačových rozhraní a predstavil ich možnú aplikáciu v prostredí repozitárov.

Open Source Code Sustainability

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=248

Prezentácia o agilnom vývoji kontrastovala tradičnú predstavu o tíme dedikovanom na vývoj jedného produktu s reálnym svetom, kde členovia tímu často pracujú na niekoľkých projektoch zároveň a na žiaden z nich nie je k dispozícii celý úväzok. Praktické odporúčania sa zamerali na oblasť komunikácie a plánovania, stanovenie jasných očakávaní a dodržiavanie pravidelnosti práce na rozdiel od nárazovej práce. Ďalší prezentujúci predstavil vlastného klienta pre API Fedora Commons napísaného v jazyku Python a zameraného na výkon. Prednášajúci z Corpus Christi prezentoval svoju cestu od nevívojára k osvojeniu si XQuery, čo mu umožnilo samostatne a flexibilne pracovať s metadátovým obsahom repozitára, ktorý spravuje. Prezentácia nástroja SwampIrk ukázala backendový nástroj, ktorý knižnica používa na správu obsahu repozitára založeného na Fedora 3.8 s rozhraním Islandora.

Savvy systems & architecture alchemy

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=248

Prvý prednášajúci argumentoval v prospech prevádzkovania repozitárov nie na vlastnom hardvéri, ale vo verejných cloudových službách. Skúsenosti z Caltech Library hovorili o použití štandardných riešení, a rozvíjaní vlastných systémov na ich „okraji“ spôsobom, ktorý do nich nezasahuje. To má následne kladný vplyv na udržateľnosť. Nasledovala živá demonštrácia systému ReDBox na správu výskumných dát umožňujúcu okrem iného zdieľanie dát medzi výskumníkmi. Je napísaná v Node.js s ukladaním dát v MongoDB a Apache Solr.

Infrastructures & new open technologies and standards

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=268

Ďalší blok vo formáte 24x7 predstavil možnosti linkovania z webového rozhrania PubMedu priamo do repozitára s preprintami. Vďaka informáciám o identitách autorov dokáže repozitár poskytnúť nielen ich publikácie, ale aj ich prácu v iných oblastiach, napr. ocenenia, ktoré autor dostal, čo o

ňom napísali iní, v akých komisiách bol členom či jeho nekrológ. Tým môže repozitár poskytnúť ucelený pohľad na človeka nielen ako na autor, ale aj ako na osobu. Petr Knoth vyvinul klienta protokoly ResourceSync a provnal jeho výkonnosť a možnosti so zavedeným protokolom OAI-PMH. Neapolská národná knižnica popisala migráciu kultúrneho dedičstva vo svojej správe do systému Dspace-GLAM. OpenAIRE ponúka poskytovateľom obsahu službu sledovania štatistík prístupov. Nie všetok obsah repozitárov je možné otvorene zverejniť a v takom prípade príde vhod možnosť autentifikácie na úrovni rozhrania IIF na prehliadanie veľkých obrázkov. Doplnkový modul pre DSpace založený na Woocommerce umožňuje predaj digitálnych objektov uložených v repozitári.

Developing like you live in the Future - Re-architecting Avalon, Accessing AV, and minimizing dev workspaces

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=321

Zaujímavý panel venovaný nástrojom na automatizáciu vývojových pracovných prostredí pomocou virtualizácie (Vagrant), kontajnerov (Docker), ale aj webového integrovaného vývojového prostredia (Codeenvy) predstavili vývojári z DSpace a ORCIDu. Hlavnými výhodami je možnosť rýchlo vytvoriť takéto prostredie a zaručiť jeho opakovateľnosť. To, že nemusia strácať čas s nastavovaním prostredia ocenia noví vývojári, ktorí sa tak môžu rýchlejšie začať venovať kódu, ale aj bežní vývojári, ktorým uľahčí vytvoriť predvídateľné testovacie, prípadne produkčné prostredie.

Community, governance & new audiences

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=294

University of Toronto inforovalo o zaujímavom prístupe k získavaniu postprintov, ktoré je notoricky ťažké vydolovať od autorov. V rámci workflow, ktoré si vytvorili získavajú postprinty priamo od spolupracujúcich vydavateľov. Je zároveň potrebné vystihnúť okamih žiadosti, pretože vydavatelia často archivujú postprinty len po obmedzenú dobu po vydaní článku. Martin Lhoták poskytol stručný prehľad českého ekosystému okolo systému Kramerius – jeho vývoj, národné repozitáre na ňom vybudované, vyvíjané nástroje na archiváciu a dlhodobú prezerváciu. Pedro Principe predstavil nové OpenAIRE Guidelines for Literature Repository Managers 4.0, momentálne vo verzii návrhu. Rozširujú svoj záber nad rámec projektov 7.RP a snažia sa obsiahnuť Open Access publikácie vo všeobecnosti a integrovať ich do infraštruktúry portálu OpenAIRE. Upravená schéma matadát vychádza z Dublin Core and DataCite a podporované sú identifikátory článkov, autorov, organizácií, zdrojov financovania.

50 Shades of DSpace

https://www.conftool.net/or2018/index.php?page=browseSessions&form_session=326

Veľmi pozitívny ohlas vzbudila prezentácia chystanej verzie DSpace 7. Hlavnými novinkami sú prepracované REST API, nad ktorým je postavené nové rozhranie založené na Angular, ktoré v DSpace 7 nahradí ako jediné oficiálne rozhranie JSPUI aj XMLUI. V štruktúre rozhrania sa nekonajú radikálne zmeny a nájdeme v ňom rovnaké časti ako doteraz, samozrejmosťou je responzívny dizajn, novinkou je funkčnosť v Javascripte na strane klienta v prehliadačoch, ktoré to podporujú, pričom ostatné prehliadače a crawlery vyhľadávačov dostanú hotovú HTML stránku, o ktorej zostavenie sa postará interpreter Javascriptu na strane servera. Zaujímavé vylepšenia dostali submission forms s možnosťou nahráť neštruktúrované PDF, z ktorého sa predvyplnia metadátové polia nového záznamu. Zároveň bolo ohlásené oficiálne rozhodnutie, že DSpace 7 bude obsahovať

flexibilnú podporu tzv. entít, čím sa myslí okrem doterajších záznamov (items) napr. autor, organizácia alebo časopis. Plánom je DSpace do konca roka stabilizovať a začiatkom roka 2019 vydať DSpace 7.0.

Týmto by som chcel poďakovať Asociácii knihooven vysokých škôl České republiky za finančnú podporu mojej účasti na tejto zahraničnej ceste.

Ing. Ivan Masár