

# Závěrečná zpráva k OAI9, Cern, Ženeva, Švýcarsko, 2015

*Vlastimil Krejčíř, Ústav výpočetní techniky, Masarykova univerzita*

Zpráva je rozdělena na část stručnou (1 strana), která jen shrnuje v bodech program OAI9, a podrobnou, která se věnuje jednotlivým přednáškám do větších detailů. Workshop OAI9 byl třídní (17. - 19. 6. 2015), 16. 6. 2015 se konalo setkání uživatelů DSpace (jakožto pre-conference k OAI9).

## **Stručné shrnutí workshopu**

### ***Dspace User Group Meeting (16. 6. 2015)***

*Setkání uživatelů DSpace se věnovalo metadatům, dlouhodobé archivaci obsahu, OA politikám a licencím (a jejich strojovému hledání), statistikám v DSpace a budoucímu vývoji DSpace (nové funkce, systém vývoje, Roadmap).*

*Zajímavé přednášky:*

- *Arkivum (Nik Stanbridge) – dlouhodobé uchování dat.*
- *Repository user stories (Sönmez ÇELİK) – repozitáře a OA v Turecku.*
- *DSpace Strategy (Tim Donohue) – plány vývoje DSpace.*

### ***OAI9 (17. - 19. 6. 2015)***

*Workshop byl rozdělen do několika sezení dle témat. Druhý a třetí den byly přednášky natáčeny na video a jsou k dispozici na odkazech uvedených u jednotlivých přednášek v této zprávě.*

***První den*** se zabýval převážně technickými tématy, začal tutoriály (běžely v paralelních sekcích, já byl na tutoriálu Herberta van de Sompela o protokolu Memento, který má umožnit snadné prohlížení webu tak, jak vypadal v různých časech v minulosti). Odpoledne byl blok věnovaný tématice sémantického webu a Linked data (představen byl nový koncept vyhledávání v Linked Data), problematice distribuovaného uložení dat se strojovým zpracováním a také se řešila problematika mizejících odkazů (řešení – protokol Hiberlink). V dalším bloku věnovaném Open Access byl představen projekt Open Access Button, dále jsme mohli slyšet motivační přednášku mladé vědkyně na téma „proč chci OA“ a závěrem o propojení OA publikování a Wikipedie (a Wikimedia).

*Zajímavé přednášky prvního dne:*

- *Memento Tutorial (Herbert van de Sompel) – protokol Memento, prohlížení www stránek „v minulosti“.*
- *Sustainable, Query-Able Access to Linked Data (Ruben Verborgh) – vyhledávání v Linked Data, nový koncept.*
- *Early career researchers and OA, why and where (Dr. Erin McKiernan) – motivační přednáška „proč chtít OA“.*

***Druhý den*** se věnoval dopoledne problematice publikování a klasického vědeckého cyklu, kritice současného stavu a nastínění možností, jak to řešit. Představeny byly projekty CHORUS a SHARE. Další blok byl zaměřen na peer review procesy, jejich nedostatky a ukázky projektů, které se pokouší jít na peer review odlišným způsobem. Byla prezentována otevřená publikační platforma PLOS. Poslední blok byly tzv. Breakout Groups, běžely paralelně v různých tematických blocích, já si vybral blok o OA politikách, kde se diskutovalo o současném stavu a úspěšnosti ukládání vědeckých prací do institucionálních repozitářů.

Zajímavé přednášky druhého dne:

- *Managing peer review at a large scale (Damian Pattinson) – platforma PLOS.*
- *Publons – credit for peer review (Andrew Preston) – obecný systém pro peer review.*
- *Breakout Groups – politiky OA ve světě, depozit do IR a jak na něj a jeho úspěšnost.*

**Poslední den** se v prvním přednáškovém bloku zaměřil publikování vědeckých prací přímo na institucích (obejdeme velké nakladatele) – byly ukázány způsoby, jak by to mohlo jít a jaký software je k dispozici. Druhý přednáškový blok se zabýval ukládáním vědeckých dat a s tím souvisejícími problémy (jak ukládat data, aby byla znovupoužitelná).

Zajímavé přednášky:

- *Preserving Complex Scientific Objects - Process Capture and Data Identification (Andreas Rauber) – systém pro ukládání vědeckých dat včetně výpočetního prostředí.*
- *Research Data Curation - Crafting (Open) Data Services at CERN (Patricia Herterich a Tibor Šimko) – jak jsou na ukládání dat v CERNu.*

## **Podrobné shrnutí přednášek**

16. 6. 2015

### **DSpace User Group Meeting (OAI9 Pre-Conference Dspace)**

Den před samotným workshopem OAI9 se uskutečnilo setkání uživatelů DSpace, které organizovala a sponzorovala firma @mire. Ta se zabývá komerční podporou a hostováním instalací DSpace. Setkání bylo věnováno především nastávajícím změnám v dalším plánovaném vývoji DSpace. Jako další témata byla diskutována metadata, OA politiky a související nástroje pro jejich automatickou detekci, statistiky v DSpace a dlouhodobé uchovávání obsahu.

Všechny prezentace ze setkání jsou k dispozici na

<https://atmire.com/website/?q=content/dspace-user-group-geneva-presentations>

#### **Publisher OA Policies and you: the quest for full automation**

*Richard Jones (Cottage Labs, UK)*

<https://atmire.com/dspace-labs3/handle/123456789/22959>

*Richard Jones* představil automatické nástroje pro práci s Open Access politikami. V praxi je někdy obtížné zjistit, zda-li a jakou OA politiku podporuje nakladatel, jaké jsou licence pro konkrétní článek apod. Tento úkol se snaží Richard Jones vyřešit strojovým způsobem: nástroje *Article Gauge* (<http://howopenisit.org/>) a *Lantern* (<https://lantern.cottagelabs.com/>) se pokouší o automatickou detekci a extrakci informací o licencích u článků a licenčních podmínkách nakladatele. Oba nástroje pracují na mírně odlišných principech. Nástroj *Article Gauge* se pokouší o detekci licencí uvedených u článku přímo na stránkách vydavatele (používá metodu parsování webové stránky, na které je článek vystaven). Služba *Lantern* je stále ve vývoji a integruje několik metod a zdrojů (mj. Article Gauge, DOAJ, Sherpa/RoMEO aj.), pomocí nichž se snaží potřebných informací o licencích a OA politikách dobrat. Vše integruje do jednoho webového uživatelského rozhraní. Je tam možné nahrát soubor ve formátu CSV se seznamem DOIs nebo PMIDs a služba se k nim pokusí dohledat relevantní informace (OA, licence, embargo – více na <http://compliance.cottagelabs.com/docs>). V praxi mají zatím oba nástroje spíše horší výsledky, nakladatelé totiž často neuvádějí informace

o licencích na svých stránkách. Na závěr byl představen ještě protyp (není zatím dokončen) aplikace *Jisc Monitor*, která má sloužit jako zastřešující nástroj pro management publikování v OA (řeší administraci kolem APC – Article Processing Charge, licencí atp.). Instituce do této aplikace vloží info o APC a ona jim na základě toho vypracuje různé statistiky. K dispozici jsou pouze demo (viz poslední slajd přednášky).

### **Arkivum: Long-term bit-level preservation of large repository content**

*Nik Stanbridge (Arkivum)*

<https://atmire.com/dspace-labs3/handle/123456789/22964>

Dlouhodobá bezpečná archivace dat byla tématem přednášky Nika Stanbridge, který představil komerční systém *Arkivum*. Arkivum slouží jako bezpečné úložiště pro jakákoli data, zajišťuje jejich bit-level ochranu (tj. že se data neztratí a především, že nedojde k jejich změně, služba se stará jen o raw data, tedy neřeší věci jako je konverze formátů apod.). Systém Arkivum je poskytován za úplaty jako služba, kterou zajišťuje stejnojmenná firma. Honosí se řadou certifikátů a k ukládání používají pouze open-source software. K dispozici mají i exit strategii pro zákazníka, který chce jejich služby opustit. Data jsou ukládána ve třech kopiích, jedna kopie je off-line v jakémisi speciálním bunkru (garantují 25 let uložení). V pravidelných intervalech se provádí kontrola integrity dat (pomocí kontrolních součtů, dva různé algoritmy), migrace software a hardware, který daná úložiště obhospodařuje, na různých úložišti používají hardware různých typů od různých výrobců atp. Arkivum může být napojeno na systém DSpace, jehož data mohou být takto bezpečně uložena (využívá kontrolní součty již generované v systému DSpace).

### **Repository user stories**

V tomto bloku byly představeny tři různé zkušenosti s instalací a provozem DSpace ve světě.

Začala paní *Marina Muilwijk (Utrecht University)*, která mluvila o integraci jejich DSpace s různými externími a interními systémy. DSpace je zde představen jako základní repozitář, který data přijímá, a ze kterého jsou data exportována do různých systémů. Pro import se využívá DSpace LNI (Lightweight Network Interface), byť mimo slajdy bylo konstatováno, že tento způsob není zcela spolehlivý. Import probíhá přímo z výzkumného systému *Pure* univerzity v Utrechtu. Pro export využívají modifikaci formátu MPEG-21 DIDL (Dutch nl\_DIDL) přes OAI-PMH, data se tak dostávají do systému *Narcis*, což něco jako národní holandský Google Scholar (harvester dat z univerzit).

Druhá přednáška byla až z Afriky a přednesl ji *Andrew Mwesigwa z Makerere University* v Ugandě. Jednalo se tak trochu o obecné povídání, jak se jim na univerzitě daří, jak spolupracují s univerzitou v Zimbabwe (kde používají hostovaný DSpace). Obě univerzity používají DSpace s některými rozšířeními, které jim zajišťuje @mire. Považují DSpace za velmi přínosný software.

Mnohem zajímavější přednášku přednesl *Sönmez ÇELİK z Doğuş University Library* v Turecku. Informoval o současném stavu Open Access v Turecku. OA a související věci jsou v Turecku do značné míry centralizovány pod hlavičkou organizace ANKUS (asi trochu jako naše AKVŠ) a její speciální pracovní skupiny OAIR (Open Access and Institutional Repositories). Organizují semináře a školení pro podporu OA a repozitářů v Turecku, překlady (například Driver Guide), technickou podporu a národní doporučení pro OA politiky. Zapojují se do evropských projektů (OpenAIRE, Pasteur4OA). Vše na dobrovolné bázi, o vše se stará asi 50 dobrovolníků. V současnosti je v ANKOS asi 185 univerzit, 139 z nich má institucionální repozitář a z nich 127 DSpace, byť největší repozitáře co do obsahu dat jsou provozovány na software *Mitos* (vyvíjen v

Turecku). 52 z těchto univerzit už zavedlo nějakou formu politiky OA. Cílem je vybudovat národní repozitář všech vědeckých prací, k čemuž mají, zdá se, i poměrně silnou podporu vlády.

### **DSpace Statistics: Challenges and Opportunities**

*Bram Luyten (@mire)*

<https://atmire.com/dspace-labs3/handle/123456789/22960>

Statistiky v DSpace je stále opakované téma. Bram shrnul současný stav a možnosti a v podstatě navrhnul, že by měl být jen jeden systém pro statistiky („one system rules them all“). V DSpace je nyní k dispozici starší *legacy statistics* systém (v jazyce Perl), který již není dále vyvíjen. Novější je modul využívající statistik ze SORL, defaultně však bez vizualizace (nengeneruje grafy a tabulky). Tu nabízí za úplatu @mire jako nadstavbu, která vypadá moc hezky (grafy, tabulky, podrobnosti). Diskutováno bylo i užívání Google Analytics a jeho nevýhody (obtížné sledování dowloadů, nemáme k dispozici raw data, závisíme na libovůli Googlu, neprůhledné počítání, bezpečnost?). Zmíněn byl i systém PIWIK, což je systém pracující na podobném principu jako Google Analytics, ale uživatel jej má zcela pod kontrolou. Nějakých větších vylepšení ohledně statistik se dočkal DSpace 5.2 (filtrace robotů).

Diskutovány byly i možnosti rozšířených statistik (např. kolik stažení měl konkrétní autor v daném roce apod.). To je náročnější úloha (caching, latence asi 50 minut, než se všechna data zpracují), ale statistický modul od @mire to také zvládne. Poslední zmíněnou možností jak získat statistiky je použití dat se systémů třetích stran, které sbírají metriky (PLUM, Altmetrics, ImpactStory), ale tato data jsou značně nepřesná. V závěru Bram ještě ukázal systém IRUS\_UK na centralizované statistiky – repozitáře předávají své lokální logy do centrálního systému, který je následně zpracuje (včetně filtrace robotů) a vytvoří z nich tabulky a grafy.

### **To DSpace and beyond: New and upcoming in DSpace**

*Ignace Deroost (@mire)*

<https://atmire.com/dspace-labs3/handle/123456789/22958>

Přednáška obecně shrnula současný stav okolo DSpace po technické i „politické“ stránce. Ukázána byla hezká mapa s instalacemi DSpace po světě (ČR se s hustotou instalací DSpace rozhodně neztrácí). Uvádí se, že je známo přes 1950 instalací a další stále přibývají. Byly také ukázány možnosti tématu *Mirage 2*, nastíněny možnosti řešení autorit (nově DSpace nabízí ukládat identifikátor autority přímo v tabulce `metadatabalue` ve sloupci `authority`). Integrována je i podpora na úrovni uživatelského rozhraní (dohledávání autorů) a zejména podpora ORCID (vyhledávání v databázi ORCID), přičemž data o autoritách je možné nahrávat i dávkově ve formátu CSV (viz <https://wiki.duraspace.org/display/DSDOC5x/ORCID+Integration>). DSpace také podporuje dohledávání informací v Sherpa/RoMEo, MathML (jazyk pro zápis matematických vzorců a jejich „vykreslování“), generování předsádek dokumentů a další.

### **DSpace Strategy**

*Debra Hanken Kurtz (DuraSpace) a Tim Donohue (University of Illinois)*

<https://atmire.com/dspace-labs3/handle/123456789/22956>

<https://atmire.com/dspace-labs3/handle/123456789/22955>

Spojená přednáška byla formou namluveného slova – přednášející nebyli přítomni, pouze jsme shlédli přednášku na videu, následně jsme se online spojili s přednášejícími a mohli jim klást otázky. Začala přednáška Debry, která shrnula situaci v DuraSpace, co se v poslední době dělo, ukázala nějaká čísla (jak si vede DSpace a Fedora ve světě), a že se podařilo vytvořit fungující

bussiness model postavený na příspěvcích členů a hostovaných službách *DSpaceDirect* (<http://dspace-direct.org/>), *ArchivesDirect* (<http://www.archivesdirect.org/>) a *Chronopolis* (<http://duraspace.org/duraspace-and-chronopolis-partner-build-long-term-access-and-preservation-platform>). Pro zájemce – DuraSpace stále shání spolupracovníky, je možné kontaktova přímo *Debru Hanken Kurtz*.

Druhá přednáška byla namluvena *Timem Donohue*. Ten podrobně rozebral stávající situaci okolo vývoje DSpace a vývoj budoucí. V minulosti byl vývoj DSpace hodně živelný, to se v budoucnu má změnit. Byla ustanovena skupina pro vytvoření roadmapy pro další vývoj DSpace. Byla provedena analýza 130 use cases a na základě ní začala být tvořena roadmapa (<https://wiki.duraspace.org/display/DSPACE/RoadMap>). Na základě toho bylo rozhodnuto, že se kód DSpace nebude přepisovat „na zelené louce“. Roadmapa byla rozdělena na dvě skupiny úkolů dle priority – co musíme a co by bylo dobré udělat. Verze DSpace 6.0 by měla integrovat příspěvky komunity z předešlých verzí. Verze 7.0 již pojedje dle roadmapy, příspěvky komunity budou vítány, ale musí být dle roadmapy. Nejzajímavější body roadmapy:

- jedno uživatelské rozhraní (kompletně nové)
- konfigurace DSpace přímo z uživatelského rozhraní
- zjednodušit hierarchii komunit a kolekcí
- veřejný registr modulů
- snadnější vkládání dat

Z neprioritních úkolů:

- hierarchická metadata
- vylepšené statistiky
- integrovat systém pro zaznamenávání vztahů mezi objekty

Z dalších pak streamování videa, optimalizace search enginů, integrace dalších identifikátorů, autentizace a autorizace přes SSO. Předpokládá se velká diskuse a doladování roadmapy příští rok na konferenci *Open Repositories 2016*.

### **RIOXX: standardizing repository metadata in the UK**

*Neil Jacobs (JISC)*

<https://atmire.com/dspace-labs3/handle/123456789/22957>

Příspěvek byla zaměřen na situaci kolem metadat a Open Access ve Velké Británii, kde je OA velmi silně podporován přímo vládou (včetně finanční podpory). RIOXX je v podstatě metadatový standard/doporučení, které má zajistit, aby používaná metadata byla kompatibilní s politikami pro OA vydanými RCUK (Research Councils ve Velké Británii) – má pomoci zejména organizacím, které OA financují. Na začátku byly shrnuty OA politiky ve Velké Británii a EU. RIOXX dodává některá metadatová pole, které se týkají výlučně procesu publikování v Open Access, například *LicenceRef*, *dateAccepted*, *project*, *version\_of\_record*. Předpokládá se podpora nakladatelů (o přijetí článku informují autora i odpovědné pracovníky univerzity). Pro DSpace by měl být připraven modul RIOXX add-on (v druhé polovině roku 2015). Na konci vznesl Neil otázku, zda by měl někdo zájem o spolupráci (z hlediska zastoupení států).



## OAI9 Workshop

17. 6. 2015

Workshop OAI9 začínal ve středu dopoledne několika paralelními sekcemi – tutoriály. Z nabízených tutoriálů jsem zvolil *T5-Memento: Uniform and Robust Access to Resource Versions*.

### **Memento: Uniform and Robust Access to Resource Versions**

*Herbert van de Sompel (Los Alamos National Laboratory)*

<http://www.slideshare.net/hvdsomp/memento-101>

Herbert van de Sompel probral detailně nově navrhovaný (cca od roku 2013, tedy ne zase tak nový) protokol *Memento*. Memento by mělo sloužit k snadnějšímu prohlížení webového obsahu zpět v čase. V praxi by to mělo fungovat tak, že si vyberu stránku a datum a čas, ve kterém bych chtěl danou stránku vidět, a zbytek zajistí Memento – dostanu stránku, která je k dispozici v čase, který se nejvíce blíží datu a času požadovanému. Technicky jde o rozšíření protokolu HTTP, v principu funguje podobně jako vyjednávání o obsahu (Content negotiation) v HTTP/1.1. Samozřejmým předpokladem dostupnosti historického obsahu je jeho existence. Zde Memento předpokládá, že takový obsah je k dispozici – ať už přímo na daném serveru (obvykle toto umí Content management systémy, které si ukládají i starší verze obsahu) nebo ve webových archivech či někde jinde na webu. Protokol předpokládá obě možnosti a v obou případech si s nimi dokáže poradit. V prvním případě dává uživateli požadovaný obsah, ve druhém využívá mechanismu, který odkazuje na příslušné webové archivy (tzv. *TimeGate*). Odkazovat může přímo na archiv nebo i na server zvaný *Memento Agregator*, který pouze shromažďuje informace o webových archivech a vybírá z nich ten, který požadavku uživatele vyhoví nejlépe, v ideálním případě dokonce rekonstruuje stránku z více zdrojů (webový portál *Time Travel*). Předpokládá se, že klient archivy nebude, a že se již využijí stávající. Přesto byla zmíněna možnost automatického budování archivů, tzv. *transactional archives*, kdy se na serveru uloží snapshot stránky v každém momentě, kdy k této stránce kdokoli přistoupí – kdykoli si uložím cokoli do svých bookmarků, pak jde takový obsah do internetového archivu, například Zotero apod.

Technicky je pro Memento protokol HTTP rozšířen o dvě nové hlavičky *Accept Datetime* a *Memento Datetime*. Pro další funkcionalitu také využívá stávající hlavičky. Pro koncové uživatele je k dispozici plugin do prohlížeče Google Chrome. Pro vývojáře je k dispozici generický TimeGate server, případně je možné využít modulu *sitestory* serveru Apache. Další nástroje (API pro JavaScript, JSON apod.) pro práci s Mementem jsou přímo na webu protokolu (<http://timetravel.mementoweb.org/>).

Protokol Memento je velice zajímavý, pokud to bude možné, zkusím připravit podrobnější přednášku na nějaké vhodné akci.

### **Opening Keynote**

*Dr. Michael Nielsen (nezávislý propagátor OA)*

<https://indico.cern.ch/event/332370/contribution/82>

V zahajovací přednášce se Michael Nielsen dotkl hlavního problému elektronického publikování – zásadním způsobem se změnila média, ale postupy publikování se nemění a nereflktují změny technologií. Jako příklad moderního článku ukázal článek v aplikaci *IPython*. *IPython* je nástroj pro interaktivní programování. V něm mohu psát článek a do něj vkládat kód, který se automaticky hned interpretuje a může ukazovat výsledky. Čtenář si navíc může ten kód upravovat, případně

měnit vstupní parametry výpočtu a sledovat výsledky. Samozřejmě je nutné umět trochu programovat, ale koncept je to velmi zajímavý, napomáhá výrazně pochopení vykládaného textu. Otázka je, jak se s těmito novými médii celý publikační proces vypořádá.

## Technical session

### Sustainable, Query-Able Access to Linked Data

Ruben Verborgh (Ghent University)

<https://indico.cern.ch/event/332370/session/6/contribution/16>

Přednáška diskutovala technologie související se sémantickým webem, přesněji s *linked data*, a navrhoval nové efektivnější koncepty pro přístup k nim. Bylo to velmi technické – shrnuto pro ne IT čtenáře by měly navržené koncepty umožnit například dotaz „Jaké knihy v naší knihovně jsou od laureátů Nobelovy ceny?“. Níže uvádím technické detaily:

Přednášející ukazoval, že současné možnosti přístupu k *linked data* nejsou v zásadě špatné, ale mají řadu technických omezení. Získávání informací z databází obsahujících *linked data* je v současnosti možné dvěma způsoby – využitím jazyka SPARQL a nebo stahováním celých množin dat. Jazyk SPARQL je pohodlný pro klienta, který chce informace získat, ale klade vysoké výkonové požadavky na server (přednášející tvrdil, že to je do značné míry důvod, proč není moc SPARQL serverů a proč mají existující časté výpadky). Na druhou stranu stahování celé množiny dat není náročné pro server, ale je velmi náročné pro klienta a navíc jednou stažená data rychle zastarají a přestanou být aktuální. Přednášející proto představil nový koncept, který by měl eliminovat nevýhody výše zmíněných přístupů a jít tak „zlatou střední cestou“. Tento nový přístup zároveň umožní kombinovat různé množiny *linked data* (které mohou být uloženy na různých serverech). Obtížné je třeba i kombinování více dat z různých serverů (což by umožnilo komplikované dotazy). Byl představen obecný koncept *Linked Data Fragments*, který definuje jakési jednotné rozhraní zastřešující různé přístupy (interface) k linkovaným datům. Obecně se dá říct, že *Linked Data Fragment* je nějakou podmnožinou konkrétních *linked data*, tato podmnožina je pak určena selektorem (metodou, která z řekne, jakou část *linked data* chceme), metadaty (informace o charakteru a rozsahu dané podmnožiny) a tzv. *controls* (popisující jak se dostat k jiným podmnožinám). V praxi může být takovým fragmentem i celá množina *linked data* nebo podmnožina určená SPARQL dotazem (který zde funguje jako selektor). Nově navrženým speciálním případem *Linked Data Fragment* je představený *Triple Pattern Fragment*, jehož selektor je navržen jako deklarativní dotaz na klasickou trojici subjekt:predikát:objekt (?subject ?predicate ?object, známo z RDF). Příkladem selektoru může být následující '? dbpedia-owl:birthPlace dbpedia:Italy', který z *linked data*, která poskytuje projekt [DbPedia](#) vybere podmnožinu osob, které se narodili v Itálii.

Výše uvedený koncept přenáší část zátěže ze serveru na klienta – selektor pro *Triple Pattern Fragment* může být poměrně obecný a může kombinovat různé množiny *linked data* (které se nachází na různých serverech). Předpokládá se, že klient (software) bude natolik inteligentní, že i komplikovaný dotaz dokáže „rozbít“ na několik dotazů jednoduchých a jejich výsledky následně inteligentně spojit. Celý tento koncept je velice mladý (drafty na stránkách konsorcia W3C jsou ze začátku května 2015) a v praxi málo ověřený (je otázka nakolik bude životaschopný a jestli jeho výhody stran výkonu budou natolik veliké, aby se uchytil). Zájemcům doporučuji následující dokumenty:

<http://linkeddatafragments.org/concept/>

<http://www.hydra-cg.com/spec/latest/linked-data-fragments/>

<http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/>

## **A Decentralized Network for Publishing Linked Data**

*Tobias Kuhn (CERN)*

<https://indico.cern.ch/event/332370/session/6/contribution/17>

Přednáška se zabývala spolehlivostí uložení vědeckých dat v sémantickém webu. V prostředí sémantického webu, kde se předpokládá strojové zpracování informací, je problém neplatných odkazů palčivější, než v případě www stránek. Přednášející definoval, jaké by měla mít spolehlivě uložená data vlastnosti (mezměnitelnost, ověřitelnost apod.), tyto vlastnosti by měly mít taktéž definované tzv. *nanopublications*, což mají být kousky informací vztahující se k autorovi publikace (na bázi RDF). Nanopublikace pak mají dosáhnout kýžených vlastností pomocí techniky hashování (kontrolních součtů), což zajistí integritu, nezměnitelnost i jednoznačnou identifikaci (hash přímo v URL, viz *trustyuri.net*). Všechny nanopublikace budou uloženy na nezávislých serverech po světě, bude se automaticky budovat i index těchto publikací a vlastní nanopublikace budou se sémantických dat budovat tzv. *Science bots*. Přednáška mnohem více naznačovala, řada z popsaných věcí nebyla vyzkoušena pod skutečnou zátěží. Druhou otázkou, která byla v následné diskusi probírána, byl neexistující bussiness model – to nedokázal přednášející obhájit. Celé je to velmi zajímavé a je otázka budoucnosti, jestli to zcela zapadne nebo se z toho stane užívaná technologie.

## **Reference Rot and Link Annotation**

*Martin Klein (UCLA, USA)*

<https://indico.cern.ch/event/332370/session/6/contribution/18>

Martin Klein představil další řešení problému neplatných odkazů, technologii *Hiberlink*. Nejdříve uvedl nějaké statistiky ohledně neplatných odkazů (cca 29 %, u referencí na vědecké články až 49 %) a jak jsou na tom v daném směru Elsevier, arXiv a Pubmed (včetně zajímavých grafů). Definoval typy článků (immunní – bez odkazů, zdravé – všechny odkazy platné a nemocné – aspoň jeden odkaz neplatný). Uvedl, že cca 1/5 článků je nemocných (infected). Doporučil ukládat články do archivu (a dělat to často spolu s datem dané verze). V HTML odkazech pak tzv. dekorovat tyto odkazy přidáním atributů *data-versionurl* a *data-versiondate*. Ty mají primárně odkazovat na verzi daného článku v archivu spolu s datem, kdy tam byla uložena. Pro uživatele v prohlížeči by se měla po najetí na takový link objevit informace o dostupnosti daného článku v archivu.

## **Barriers and Impact Session**

### **Open Access Button**

*Joseph McArthur (Right to Research Coalition, UK)*

<https://indico.cern.ch/event/332370/session/8/contribution/19>

*Open Access Button* je v podstatě mobilní aplikace, která má pomáhat uživateli s orientací ve světě OA. Hlavním cílem celého projektu *Open Access Button* je pomoci lidem, kteří mají pocit, že OA naprosto nefunguje. Do prohlížeče Firefox je možné si nasíntalovat *OA Button* plugin, který se při brouzdání některými weby (např. ScienceDirect) pokouší o právě prohlíženém článku dohledat další informace k OA a ukázat je uživateli. U každého identifikovaného článku, u kterého není moc informací nebo je nedostupný, pak tento plugin umožňuje odeslat indikaci přání dostupnosti tvůrcům *OA Button* (tvořil tzv. *wishlist*). Přednášející ukázal nějaké statistiky po 2 letech provozu *OA Button* (kterými se tak trochu chlubil). V závěru vybídl účastníky k účasti a šíření informací o konferenci [OpenCon2015](#), která se má problematikou OA zabývat do hloubky. V následné diskusi vznikla pře ohledně *ResearchGate* (tu zmiňoval přednášející během přednášky), kde byly zastávány



dva názory: ResearchGate rozhodně ne (je to neprůhledné a tvůrci na tom jen vydělávají) versus ResearchGate ano, protože to prostě vědcům vyhovuje (mají pocit, že publikují mezi svými).

### **Early career researchers and OA, why and where**

*Dr. Erin McKiernan (Wilfrid Laurier University, Kanada)*

<https://indico.cern.ch/event/332370/session/8/contribution/20>

Hezká přednáška mladé vědkyně o jejích zkušenostech s OA. Sama začínala v USA, kde jsou převážně velké a bohaté univerzity a ty nevidí v současném systému publikování a placení za publikace žádný problém. Erin se však dostala i na menší univerzity v Portoriku a Mexiku, kde peníze na nákup zdrojů nejsou – její sdělení, že tam v podstatě nenakupují téměř žádné elektronické zdroje (balíky od Elsevieru, Springeru apod.), ale jsou odkázáni na několik tištěných exemplářů v knihovně bylo pro řadu lidí v auditoriu zřejmě překvapivé. Například uváděla, že v Mexiku mají 139 titulů dostupných elektronicky (pouze 88 plně, zbytek s omezením) a řadu časopisů jen tištěně (Nature, Cell apod.). Erin následně oznámila, co vše ona bude dělat pro boj za OA – včetně toho, že nebude publikovat v uzavřených časopisech. Zmiňovala publikování v [PeerJ](#) (OA časopis pro biovědy, medicínu a informatiku), brojila proti uzavřenému peer review a proti impakt faktorů (který příliš mnoho institucí zbožňuje). Přednáška to byla velmi entuziastická, ale je otázka, jak dobře jí to půjde v praxi...

### **Integrating Open Access and Wikimedia**

*Dr. Daniel Mitchen (National Institute of Health, Bethesda, USA)*

<https://indico.cern.ch/event/332370/session/8/contribution/21>

Přednáška se zabývala spojením odborných článků a Wikipedie (přesněji „dokumentové“ části Wikipedie zvané Wikimedia). Přednáška byla dosti nepřehledná, přednášející ukazoval věci online a dosti skákal. Na začátku uvedl motivaci – Wikimedia a OA mají řadu společných prvků i cílů, proto je vhodné je integrovat a tuto integraci pokud možno maximálně automatizovat. Ukázal řadu příkladů, kdy jsou tyto dva světy propojeny, zmínil politiky OA na Wikimedia (vztahují se kromě článků i na software a multimédia) a technické problémy OA na Wikimedia. Předvedeny byly nějaké statistiky citovanosti časopisů na Wikipedii a také automatizované propojování položek na Wikipedii s DOI (ukazoval nějaké výsledky, které vytvořili automatizovaní boti – zdálo se, že toto by mohlo fungovat).

18. 6. 2015

## **Open Science workflows, CHORUS and SHARE**

### **CHORUS Advancing Public Access to Research**

*Howard Ratner (CHOR Inc., USA)*

<https://indico.cern.ch/event/332370/session/10/contribution/22>

CHORUS je organizace na podporu publikování Open Access ve Spojených státech amerických. Snaží se o spolupráci a podporu nakladatelů, agentur i vědeckých institucí s cílem dostat maximum vědeckých publikací do režimu Open Access – každý má v tomto směru na nakladatele jiné požadavky (jiné mají vědci, jiné grantové agentury a vlády po světě apod.), CHORUS se snaží toto vše sladit a každému v procesu publikování pomáhat a na příslušné strany i tlačit (zejména na nakladatele, například aby slučovali licence). Spolupracují s většinou velkých nakladatelů (včetně Springeru, Wiley a Elsevieru), podporu mají přímo na vládní úrovni (úzká spolupráce s US Department of Energy). Financování obstarávají členské instituce (akademické instituce na

dobrovolné bázi, členy je řada amerických univerzit, peníze plynou zejména od nakladatelů). Cílem je snižovat náklady na OA pro nakladatele i instituce a vytvářet příznivé prostředí pro OA.

CHORUS definuje několik bodů, ve kterých se v oblasti OA angažuje: přístup (Access), ukládání (Preservation), hledání (Discovery) a identifikace (Identification). Shrnuli si požadavky na OA z různých stran a na základě nich navrhli příslušná řešení. V praxi nabízejí služby *Dashboard* (administrativní rozhraní pro správu věcí okolo OA) a *Search* (speciální discovery služba na bázi Google + Crossref, není to fulltextové hledání, pouze strojově zpracovávaná metadata). Mají k dispozici i API na napojení dalších služeb na jiné portály.

V závěru přednášky byla ukázána řada statistických údajů – tabulky, grafy – kolik toho pokrývají atp. Skoro jsem měl pocit, že dělají úplně všechno :-).

### **Opening up research infrastructures and workflows in higher education**

*Dr. Tyler Walters (Virginia Tech, USA)*

<https://indico.cern.ch/event/332370/session/10/contribution/23>

Tyler Walters představil projekt SHARE. Projekt se snaží postihnout celý vědecký cyklus a sbírat z něj data (tedy ne jen finální články, ale i šedou literaturu). Cílem je poskytnout informace všem, kteří to potřebují (vědci – co se děje v oboru, jaké jsou projekty; administrativa a knihovny – co vlastně naše univerzita produkuje; grantové agentury – co vlastně financují). Zájemci se mohou do SHARE zaregistrovat a SHARE pak bude zpracovávat i jejich data (v současné době je registrováno 42 přispěvatelů – mj. PubMed, ArXiv, univerzitní repozitáře). Celé řešení je postaveno na Open Source softwaru (Python, Django) a snaží se o vylepšování nasbíraných dat (a metadat). Provádí se deduplikace, normalizace atp., hlavní problém ale není technický, největší brzdou jsou různé licence, které celý proces velmi omezují. Snaží se o integraci podpory SHARE do různých repozitářů (mj. DSpace ve spolupráci s DuraSpace).

### **Opening up research infrastructures and workflows in higher education**

*Bianca Kramer a Jeroen Bosman (Utrecht University, Nizozemí)*

<https://indico.cern.ch/event/332370/session/10/contribution/24>

Přednáška se zaměřila na vědecký proces malinko obecněji – byla kritikou současného stavu a navrhovala, jakými způsoby to změnit a zlepšit. Byl popsán současný vědecký cyklus (preparation, discovery, analysis, writing, publication, outreach, assesment) a potíže, které provázejí jeho jednotlivé body (opakované psaní grantů, odmítání zaslání článku, ...) a vnáší do něj velkou neefektivitu. Open Access samotný nestačí – je třeba změnit ekonomická a licenční schémata, softwarové prostředky a i samotný způsob dělání vědy (aby se do toho necpali vlády, nevstupovalo do toho tolik lidí a organizací). Pokud jde o software, v současnosti existuje více než 500 různých nástrojů na podporu některé části vědeckého cyklu (a kdo se v nich má vyznat). Byl srovnán model publikování klasický a v Open Access a byly předloženy statistiky z výzkumu shrnujícího oblíbenost užívaných softwarových nástrojů. Obecně pro přístup k literatuře se povětšinou používají institucionální repozitáře, ResearchGate, Open Access Button a kontaktování autora mailem. Na archivaci dat pak GitHub a Figshare. Rádi by do dotanikového šetření (skončí v únoru 2016) zapojili více lidí, dotazník je dostupný na <http://101innovations.wordpress.com>. V následné diskusi se opět dostalo na ResearchGate (napadána jako netransparentní a uzavřená), ale jak je ze statistik patrné, lidem se líbí a prostě ji používají. Zajímavý byl ještě dodatek, že speciální discovery služby (EDS, Summon, Primo, OCLC Worldcat) vědci moc nepoužívají, mnohem více jdou přes Google případně speciální oborové vyhledávače.

## Panel discussion

Po předešlých třech přednáškách následovala „panelová diskuse“ (uvozovky uvádím zcela záměrně, protože měla cca 10 minut), které se zúčastnili přednášející z předešlého bloku. Vzhledem k délce diskuse se vlastně ani moc o diskusi nejednalo a diskutující do značné míry jen shrnuli své názory z přednášek: je potřeba nástroje, které umožní sdílení čehokoli kdykoli a kdekoli, všechno se nakonec stejně točí kolem peněz (nejefektivnější se zdá být membership driven model, protože generuje peněz dostatek, potřebujeme výjimku z copyrightového práva, abychom mohli efektivně provádět data mining).

## Poster session

Vzhledem k omezeným prostorům (krátká, úzká a špatně osvětlená chodba) nebylo letos posterů mnoho (ve srovnání s OAI8). Pokusil jsem se je nafotit, ale světelné podmínky a možnosti mého fotoaparátu bohužel poznamenaly výsledek. Nemohu říci, že by mě některý z posterů obzvláště zaujal. Pěkný byl poster na CERN Document Server (DSC00077.JPG), zajímavý byl poster o Zea E-books (DSC00089.JPG) – na University of Nebraska mají oddělení, kam může kdokoli poslat text a oni mu zadarmo vyrobí knihu (jsou tam na to univerzitou placeni 4 lidé), zatím nemají moc požadavků, jsem celkem zvědavý, jestli budou zvládat, když jim toho bude chodit více.

Fotografie posterů: <https://kic-internal.ics.muni.cz/~krejvl/oai9-postery.zip>

## Quality assurance

### Managing peer review at a large scale

*Damian Pattison (PLOS, Cambridge, UK)*

<https://indico.cern.ch/event/332370/session/11/contribution/26>

Přednáška se věnovala systému PLOS, přesněji se jedná o neziskového nakladatele OA publikací. V současné době má přes 140 tisíc článků a poměrně značný provoz (hledání, stahování – měsíčně kolem 2 milionů stažení, je ovšem otázka, jak k těmto číslům došli). Přednášející se hned ze začátku hodně chlubil a hodně chválil :-). Následně konstatoval, že tlak na vědce, aby publikovali, má za následek, že publikace trpí na plagiátorství, zmanipulované výsledky a statistiky, předsudky, krádež dat, falešné peer review, takže nemalá většina publikovaných výstupů má velmi nízkou kvalitu, což do značné míry souvisí s nastavením procesu peer review. Vědci provádějící peer review jsou často tlačeni k odmítnutí, navíc se věda stává stále rozmanitější a tedy je i obtížnější dělat posudky – ty jsou pak velice různé (v kvalitě, rozsahu), obvykle negativní a trpí předsudky. Přesto je závěr, že peer review není „rozbitý“ proces, ale vede k otázce: „Je tento článek vhodný pro daný časopis?“, přičemž správně by mělo být: „Je tento článek vhodný pro nějakého čtenáře?“. Snad by to mohlo řešit otevřené peer review (příkladem byly uvedeny zákaznické recenze produktů na Amazonu nebo komentáře na Stack Overflow). V následné diskusi se ozvali kolegové z Ruska, kteří se o otevřené peer review v ruském národním repozitáři CyberLeninka pokusili a moc dobře to nedopadlo – důvody: předně je té produkce tolik, že vědci nemají čas psát recenze (velké množství článků nebylo vůbec hodnoceno) a také, že se bojí říct svůj názor (a jít proti kolegům, které často znají – titíž pak mohou hodnotit je).

### Peerage of Science - community driven initiative to improve peer-review

*Dr. Janne-Tuomas Seppänen (Peerage of Science, Finsko)*

<https://indico.cern.ch/event/332370/session/11/contribution/27>

Přednáška začala kritikou současného peer review procesu, především privilegií, které získá časopis, když do něj vědec pošle článek (zejména časová – časopis není nucen dodržovat nějaké

zásadní harmonogramy a může celý proces protahovat). Toto se přednášejícímu nelíbilo, proto přišel s vlastním projektem na peer-review a nazval jej *Peerage Science*. Jedná se o komunitní portál, kde se mohou vědci registrovat a následně tam posílat své články, které mohou další registrovaní zcela otevřeně hodnotit. Následně může autor článěk opravit a poslat je tam znovu (tento cyklus je bez omezení) – pouze cca 10 % článků je bez recenzí. Recenzovat je možné i recenze. Každý autor má profil, ve kterém se dá sledovat, jak si v peer review vede (ale může to autor schovat). Registrace na portálu podléhá schválení autora systémem (zájemce musí prokázat, že je skutečně vědec). Portál je prozatím poměrně malý (~1930 vědců ze 728 institucí, 70 článků, 469 recenzí, 1086 recenzí recenzí).

### **Publons – credit for peer review**

*Andrew Preston (Publons, UK)*

<https://indico.cern.ch/event/332370/session/11/contribution/28>

Andrew Preston představil systém *Publons.com*, který má sloužit na urychlení a zlepšení peer review procesu. Jedná se o službu, která sdružuje recenzenty (jsou za recenzování placeni). Zároveň spolupracují s nakladateli (Wiley, Sage, ...). Z recenzí se budují reporty a statistiky, které se následně promítají do profilů recenzentů a jejich hodnocení – mohou tak například prokazovat svou aktivitu na poli recenzních řízení. Podporují jakékoli typy recenzních řízení (otevřené i uzavřené, tajné i veřejné). V současné době mají kolem 38 tisíc recenzentů, 100 tisíc recenzí a asi 6 tisíc časopisů. Zároveň mají i nástroje pro podporu editorů a vyhledávání nových recenzentů.

### **Breakout Groups**

*Breakout Groups* běžely paralelně v 5 sekcích, já se zúčastnil sekce 2 s názvem *OA Policy*. Moderátorkou skupiny byla *Alma Swan (Sparc Europe, Nizozemí)*. Skupina byla zaměřena na OA politiky ve světě. Byla prezentována data z projektů PASTEUR4OA a ROARMAP, po nich proběhla diskuse týkající se zkušeností účastníků s OA politikami na jejich institucích.

PASTEUR4OA je program Evropské unie na financování OA (s dodržáním pravidel pro vědce i ty, kdo budou OA financovat, pro Horizon 2020). Cílem je vytvořit síť expertů na OA v jednotlivých zemích EU, dle žádostí z těchto zemí se budou tvořit příslušné materiály, analýzy a obhajoby OA politik. Dále se mají tvořit návody (např. jak udělat takovou a takovou smlouvu), organizovat regionální workshopy (září 2015 až duben 2016). V současnosti všichni pracují samostatně, ale cílem je všechny OA experty spojit.

ROARMAP je databáze OA politik z různých zemí a institucí (včetně podrobných informací o každé politice). V loňském roce byla databáze ručně aktualizována a doplněna a dá se hezky prohledávat. V současnosti se předpokládá, že uživatelé budou politiky už doplňovat sami (k dispozici formulář). Byli jsme vyzváni k zapojení. Z ROARMAP také vypadly zajímavé informace. 122 institucí má povinné uložení do repozitáře (ne však nutně zveřejnění v režimu OA!, ve většině případů se jedná pouze o depozit, který má až nekonečné embargo). Tyto byly do přehledu zahrnuty. I když mají povinné uložení, tak nejlepší dosahovali sotva 50% úspěšnosti (a jen 3 instituce se mohli chlubit čísly kolem 70 a 80 %: University of Minho, University of Liege a Queensland University). Na Queensland University na to mají na plný úvazek jednoho člověka a silnou podporu vedení, přesto udávali, že je velmi obtížné (až nemožné) dostat se nad 80% úspěšnost. Důležité je mít nějaké „velké zvíře“, které to tlačí a podporuje, silně zainteresovanou knihovnu, strategii na fakultách (silná podpora vedoucích), školení, propagaci (OA week, pohlednice, postery), nástroje (jaký má OA impakt) a lidi, kteří jsou placeni za to, že to ukládání za vědce celé udělají. Pak je možné dosáhnout relativně dobrých výsledků.

19. 6. 2015

## Institution as publisher

### **Putting the Academy at the heart of scholarly publishing**

*Catriona MacCallum (PLOS, Cambridge, UK)*

<https://indico.cern.ch/event/332370/session/0/contribution/0>

Jednalo se o další přednášku o publikování a o PLOSu. Byla probrána historie publikování (současný model byl optimální v minulém století) a shrnuty některé nástroje, které boří současný status quo. Mění se systém publikování, více se diskutuje a slovo publishing už není přesné a na místě. Nestáčí něco pouze zveřejnit – musí se to ještě propagovat, prokázat spolehlivost zveřejněného, znát lidi, komunikovat, ... A s tím vším nám PLOS pomůže :-).

### **Putting the Academy at the heart of scholarly publishing**

*Dr. Rupert Gatti (Open Book Publishers, Cambridge, UK)*

<https://indico.cern.ch/event/332370/session/0/contribution/1>

Historicky bylo publikování vědeckých výstupů doménou univerzit, což se v průběhu let měnilo a publikování přecházelo do obchodní sféry. Současné technologie však umožňují návrat publikování zpět na univerzity. Přednášející definoval, co je to vlastně centrum výzkumu, jaké má cíle a na koho se zaměřuje. Následně konstatoval, že pokud nebudeme vědce vzdělávat na poli OA, tak budou automaticky dál posílat své práce nakladatelům (kteří za ně udělají zveřejnění). Což ovšem znamená, že exkluzivní práva přechází na nakladatele, který nadále vlastně rozhoduje o tom, kdo bude cílová skupina čtenářů. Toto sice řeší OA, ale má řadu bariér – jak právních tak technických. Od těch je nutné vědce odstínit. Přednášející doporučil jako příklad instalovat Open Journal System (nutná podpora IT na univerzitě!), ale zdůraznil, že na správu celého procesu užívání OJS jeden člověk nestačí (není to zadarmo). Na závěr předvedl některé příklady za praxe. V diskusi padla otázka na udržitelnost – odpověď byla, že nikdo neví, co budou chtít/potřebovat vědci za 20 let, takže ta otázka je bezpředmětná :-).

### **University-based open access publishing in Europe: challenges and opportunities in the Humanities**

*Victoria Tsoukala (National Hellenic Research Foundations, Řecko)*

<https://indico.cern.ch/event/332370/session/0/contribution/33>

Přednáška byla krátká a shrnula současný stav OA v oblasti humanitních věd a provedla srovnání s vědami tvrdými.

## Digital curation and preservation of large and complex scientific objects

### **Data Life-Cycle Management - The Swiss Way**

*Pierre-Yves Burgi (University of Geneva)*

<https://indico.cern.ch/event/332370/session/3/contribution/9>

Přednáška byla o švýcarském projektu, který má za cíl připravit komplexně řešení pro management,



zpřístupnění a dlouhodobé ukládání vědeckých dat ve Švýcarsku. Projekt vypadá celkem megalomansky (36 milionů CHF) a zatím není schválen (už jednou byl zamítnut). V přípravné fázi si udělali průzkum (dotazováno bylo asi 50 vědců), rádi by v projektu řešili software i hardware, administrativu, vydávali doporučení a manuály, dělali školení atp. Chtějí využít již hotových struktur a softwaru, např. SCALE (což je švýcarský národní cloud), openBIS (anotace v biologických vědách), SALSAH (linkování a anotace v humanitních vědách), Eprints, Rosetta, Invenio, ...

### **Preserving Complex Scientific Objects - Process Capture and Data Identification**

*Andreas Rauber (Vienna University of Technology)*

<https://indico.cern.ch/event/332370/session/3/contribution/31>

Pravděpodobně nejzajímavější přednáška na OAI9, jejímž tématem bylo zachování vědeckých dat z výzkumu tak, aby bylo možné je znovuvyužít a experimenty opakovat a ověřovat. Samotná vědecká data pro přesné opakování experimentu nestačí, výsledky často závisí na použitém software (od operačního systému až po konkrétní specializované aplikace). Byly uvedeny příklady, kdy změna operačního systému při zachování stejné aplikace dávala rozdílné výsledky na stejných vstupních datech. Totéž pro změny mezi verzemi stejné aplikace. Opakovatelnost v takovém případě ztrácí smysl. V Rakousku na to šli tak, že kromě samotných vstupních dat uchovávají kompletní informace o SW prostředí, ve kterém byly spouštěny výpočty nad danými daty. Vyvinuli komplexní sledovací systém, který zaznamenává kompletní průběh výpočtu od verzí aplikací až po úroveň procesů operačního systému, o formátech i hardwaru – vše je tedy zdokumentováno a může být uloženo společně s daty. Za těchto podmínek je snadnější výpočet přesně opakovat.

Druhá část přednášky se věnovala způsobu, jak komplexní vědecká data citovat. Nestačí je jen někde uložit, je nutné je jednoznačně identifikovat, distribuovat spolu s metadaty a zaručit jejich integritu. V závěru byla dána doporučení, jak se danými problémy vypořádat.

### **Research Data Curation - Crafting (Open) Data Services at CERN**

*Patricia HERTERICH a Tibor ŠIMKO (CERN)*

<https://indico.cern.ch/event/332370/session/3/contribution/32>

Poslední přednáška workshopu byla popisem toho, jak se s ukládáním dat, článků a jiných materiálů vypořádávají v CERNu. Začala popisem systému Invenio, který používá jako svůj institucionální repozitář CERN (jak řeší ukládání tabulek, provázanost s daty apod., jen pro zajímavost: interně na metadata používají mírně upravený MARC21). Stručně zmínili ukládání dat z urychlovače a představili CERN Open Data portál, kde jsou uložena primární data včetně virtualizačních nástrojů i ke kódu, kterým se generovala. Mají vlastní virtualizační software (CERN Virtual Machine), na kterém je možné si opakovaně spouštět výpočty nad uloženými daty (v současnosti spíše dávkově, zpracování přeci jen nějakou dobu trvá). Výše zmíněný software je open source a k dispozici zdarma.

### **Closing keynote**

Závěrečné slovo zmiňuji jen proto, že jeho součástí byl seznam spoznorů akce, mezi nimiž nechyběli EBSCO, Elsevier, Exlibris, ProQuest, Springer, Wiley a v podstatě všichni velcí hráči na trhu, na které se dštíla v mnoha přednáškách síra :-).