

# Závěrečná zpráva k OAI10, Cern, Ženeva, Švýcarsko, 2017

*Vlastimil Krejčíř, Ústav výpočetní techniky, Masarykova univerzita*

Zpráva je rozdělena na část stručnou (1 strana), která jen shrnuje v bodech program OAI10, a podrobnou, která se věnuje jednotlivým přednáškám do větších detailů. Workshop OAI10 byl třídní (21. - 23. 6. 2017), 20. 6. 2017 se konalo setkání uživatelů DSpace (jakožto pre-conference k OAI10).

## **Stručné shrnutí workshopu**

### ***Dspace User Group Meeting (20. 6. 2017)***

*Setkání uživatelů DSpace se věnovalo shrnutí nových funkcí DSpace 6 a zejména pak nově připravované verzi DSpace 7 a toho, co nás v ní čeká a proč to bude, jak to bude.*

*Zajímavé přednášky:*

- *Providing a nation with repositories (Jyrki Ilva) – jak to dělají ve Finsku.*
- *Angular 2 User Interface by DSpace (Tim Donohue) – přednáška k DSpace 7 a novému webovému rozhraní.*
- *Elements update from Symplectic (John Fearn) – přednáška o software na správu vědeckých publikací.*

### ***OAI10 (21. - 23. 6. 2017)***

*Workshop začal sekcí tutoriálů, po nich již následovaly klasické přednášky rozdělené do několika tematických sezení: Tech Session, Copyright and Licencing Session, OA Transformation Session, OA Outside Session, Social Media Session a Future of Repositories Session.*

### ***Všechny přednášky byly nahrávány, zájemci si je mohou poslechnout!***

***První den*** se zabýval převážně technickými tématy, začal tutoriály, které běžely v paralelních sekcích. Já se zúčastnil tutoriálu Herberta van de Sompela, který tentokrát prezentoval nejdříve koncept Sign Posting – usnadnění vyhledávání plných textů pro agregátory a indexátory obsahu. Ve druhé části tutoriálu se věnoval protokolu ResourceSync. V odpoledním programu bylo zahájení konference (přítomen zástupce EU, referoval o programu Horizon 2020), následované technickým blokem, který představil několik koncepcí a zajímavých projektů, mj. věnujících se decentralizované vědecké komunikaci (projekt dokie.li), koncept na podporu FAIRness u vědeckých dat a jejich popisu pomocí dostupných standardů (schema.org). Druhý odpolední blok se věnoval problematice copyrightu a licencování – porovnání systému USA vs. EU spolu s návrhy na zlepšení v EU, diskutována byla situace v UK, a nakonec probírána právní situace ohledně dolování textů a dat (data mining).

***Druhý den*** začal třemi přednáškami, ve kterých se střetly tři různé pohledy na současný stav okolo OA (především Gold) a přístup velkých nakladatelů (tři různé pohledy na věc) – následováno diskusí. Před obědem ještě proběhla Poster Session, nejzajímavější postery budou uvedeny v podrobné zprávě níže. Odpoledne bylo opět věnováno OA, ale spíše ze statistického hlediska (jak OA funguje, ukázka studií současného stavu, pohled různých subjektů v různých zemích a nakonec i jeden konkrétní use case proč publikovat OA). Večer byl poslední blok nazvaný Unconference, kde sami účastníci mohli navrhnout několik témat a ty následně v měsích skupinách diskutovat – zúčastnil jsem se diskuse okolo automatického nahrávání dat do repozitářů.

***Poslední den*** se v prvním přednáškovém bloku věnoval sociálním médiím a měření vědeckého

výkonu (*altmetrics*) – proč ano, proč ne, etická stránka věci, zkreslení, srovnání s tradičními metrikami. Poslední blok celého workshopu rozebíral budoucnost repozitářů, agregátorů, kdo a jak repozitáře používá, jak dělat agregaci do budoucna (nové technologie), interoperabilita, jak se poprat s konkurencí komerčních repozitářů velkých nakladatelů. Závěrečná přednáška porovnávala situaci okolo OA ve fyzice (*arXiv.org*) a biologii.

### **Nejzajímavější přednášky:**

- Opening Keynote: Open Research Europe. (*Pohled na OA a publikování z pohledu člena evropské komise pro Open Science, co EU v tomto směru chystá – sw na podporu publikování etc.*)
- Interoperability and Data FAIRness emerges from a novel combination of Web technologies (*Vědecká data a jejich znovupoužití – jak na FAIRness.*)
- Copyright and Education (*Spíše zajímavé – jak copyright (ne)blokuje výuku v různých zemích a co s tím.*)
- Some introductory thoughts on economics aspects of scholarly communication (*Pohled na OA a velké vydavatele a co s nimi.*)
- Discussion - impact of open access on the cost of scholarly communication (*Jiný pohled na OA a Gold OA, jejich výhody a nevýhody.*)
- Can we afford not to go Gold? (*Proč se nevyhneme Gold OA, ekonomická analýza publikování.*)
- How open access opens doors - reflections on my recent 'Megaregions of the United States' paper (*Zajímavá přednáška o prospěchu publikování v režimu OA.*)
- Social media metrics as a source of understanding scientific communication : descriptive and evaluative possibilities (*Studie a porovnání klasických a alternativních metrik měření vědeckého výkonu.*)

### **Podrobné shrnutí přednášek**

20. 6. 2015

### **DSpace User Group Meeting (OAI10 Pre-Conference Dspace)**

Den před samotným workshopem OAI10 se uskutečnilo setkání uživatelů DSpace, které organizovala a sponzorovala firma *@mire*. Ta se zabývá komerční podporou a hostováním instalací DSpace. Letos bylo setkání zaměřeno na změny v DSpace 6 a novinky v DSpace 7, plus zahrnulo některé přednášky uživatelů DSpace.

#### **Current version of DSpace**

*Benoit Weiry (@mire, Belgie)*

*Benoit* byl jedním z organizátorů setkání a v první přednášce shrnul stávající roadmap DSpace a nové funkce a vlastnosti DSpace 6 (zejména z pohledu XMLUI). Zájemce o detaily mohou odkázat na moji přednášku z Otevřených repozitářů 2017, která shrnula totéž.

#### **Contribution to the requirements of DSpace 7**

*Wouter Janssens (@mire, Belgie)*

Wouter se věnoval nadcházející verzi DSpace 7. Nastínil, proč jsou nutné radikálnější změny a popsal situaci okolo stávajících platných use cases. Vyzval uživatele, aby se na úpravě use cases

podíleli, případně se zapojili i do testování nové verze DSpace. Dále popsal, jak pracuje vývojová skupina (testování, rozdělování úkolů) a jak je do celého vývoje zapojeno @mire.

### **Improving ORCID integration in DSpace**

*Robert Peters and Josh Brown (ORCID)*

Oba pánové se věnovali ORCIDu – v první části zejména propagaci (k čemu to je, jak je to skvělé, etc.). Dostali nějaké nemalé peníze od Sloanovi nadace na zkoumání integrace 10 projektů s ORCID (dva z těchto projektů byly postaveny na DSpace): 10 % ORCID uživatelů má DSpace. Druhá část přednášky byla věnována autentizaci osob s ORCID – jak zjistit, že osoba registrovaná pod ORCID je skutečně tou osobou, za kterou se vydává. Poukázali, že řada profilů má značně nedoplněné osobní informace a jen podle jména je identifikace osob obtížná. Navrhli systém, který pracuje na principu vzájemného potvrzování identit *Share My ORCID* – lidé, kteří se osobně znají, mohou potvrdit identitu sobě navzájem. Další možností je identifikace na úrovni instituce (autora ztotožní přímo instituce), což se ale moc často neděje. Jako pozitivní příklad uvedli pražské ČVUT, které má ORCID vázáno na interní identifikaci osob a DSpace UUID.

### **DSpace's strategic course**

*(Debra Hanken Kurtz, DuraSpace), přednáška přes Skype*

Přednáška se zaměřila na strategii rozvoje DSpace – co se bude dít dále (velmi obecně). Shrnula kolik je na světě instalací DSpace (~ 2250), kdo to financuje, DSpace jako SaaS (software as a service). Zmínila cestu k DSpace 7 a snažila se změny zdůvodnit (schizofrenii dvou verzí UI, lepší spolupráce s externími systémy, ...). DuraSpace také hodlá podporovat hostovaný DSpace, který by měl být ve verzi 7 snadněji konfigurovatelný přes webové rozhraní. DSpace 7 by měl být po první půlce roku 2018. Vyzvala, abychom se připojili k pravidelným meetingům ohledně vývoje DSpace 7.

### **Integrating search for mathematical formulas**

*Vlastimil Krejčíř (Masarykova univerzita, ČR)*

<http://www.dspace.cz/media/2922472/mir-dspace.pdf>

Já jsem přednášel o problému vyhledávání a indexování matematických formulí, který řeší výzkumná skupina na MU. V první části přednášky jsem nastínil problémy, které v sobě indexace matematiky nese (nejednoznačnost, proměnné, podformule) a krátce jsem ukázal, jak tyto problémy řešit. Ve druhé části jsem se věnoval integraci nástrojů pro hledání matematických formulí do DSpace a na živo jsem ukázal několik příkladů na prototypu nové verze naší Digitální matematické knihovny.

### **Streamlining workflows in the DSpace**

*Sarah Maloy (Queens University London, UK)*

Přednáška byla vlastně ukázka případu užití DSpace na Queens University v Londýně. DSpace mají od roku 2009, mají několik úprav, nic extra zajímavého – jen jim to údajně urychlilo práci (což ale mohlo být i jinými faktory). Víceméně to krmí ručně, starají se o to 4 lidi, ale je jako část jejich úvazku.

### **DSpace in National Library of Finland**

*Jyrki Ilva (National Library of Finland)*

Jyrki představil obecně Finskou národní knihovnu a věnoval se DSpace a tomu, jak a pro koho jej provozují. Mají 48 zákazníků (kteří si hostování DSpace platí), DSpace však používají i pro sebe. Ukázal několik dalších zajímavých projektů jako Valto (repozitář pro vládu) nebo Theseus. Theseus je celonárodní repozitář kvalifikačních VŠ prací, postavený na DSpace (verze 5.6), v současnosti

má kolem 120 tisíc položek, mají poměrně velký provoz a zátěž (je to dobře viditelné z Google). Vzhledem k velké zátěži řešili výkonové problémy (více serverů, load balancing).

## **Angular 2 User Interface**

*Tim Donohue (DSpace Tech Lead, DuraSpace, USA)*

Tim je v podstatě šéfprogramátorem DSpace a poměrně podrobně se věnoval DSpace 7 a zejména změně technologie webového prostředí. Nejdřív představil nové logo DSpace a pak už se vrhnul na technickou část přednášky. Ukázal jak se cesta k nově použité technologii Angular 2 vyvíjela, jak postupovala skupina DSpace 7 UI working group, proč nebylo možno použít staré REST API, na jakých technologiích se budovalo REST API nové (HATEAS + ALPS + HAL), ukázal demo. Zdůvodnil, proč byl nakonec vybrán Angular (vhodný pro prohlížeče i roboty, rychlý – v první fázi posílá HTML vyrobené na serveru, až si prohlížeč natáhne příslušné Javascript knihovny, pak se teprve přenáší zátěž na prohlížeč; pro roboty však i nadále posílá čisté HTML – ověřeno, že Google to zvládá indexovat dobře). V současnosti je hotovo browsování a stránky komunit, kolekcí a items, ale ještě zbývá velmi mnoho práce... Snažil se sehnat další vývojáře a lákal :-).

## **Elements update from Symplectic**

*John Fearn (Symplectic, UK)*

John představil informační systém určený pro vědce a výzkumníky – primárně pro zpracování publikační činnosti a její distribuci do některého repozitářového systému (primárně do DSpace). DSpace mají napojen přes vlastní proprietární API. Systém Symplectic natáhne k dané osobě všechny možné informace a články (z PubMedu, Scopusu a Web of Science), které pak protlačí do DSpace. Následně přijde vědec, který již vidí připravený seznam svých publikací, včetně informace, je-li k nim dostupný plný text ve formě OA, a může jen potvrdit, které z nich jsou skutečně jeho. OA plné texty je následně možné automaticky přenést do DSpace, případně lze ručně plné texty nahrát a následně přenést do DSpace. Systém vypadal v ukázce moc hezky, i konfigurovatelnost přes webové rozhraní vypadala dobře. Mají i speciální modul pro implementaci OA politik. Integrace s různými jinými systémy je možná, ale prý nejlepší je DSpace (má nejlepší API).

## OAI10 Workshop

21. 6. 2017

Workshop OAI10 začínal ve středu dopoledne několika paralelními sekcemi – tutoriály. Z nabízených tutoriálů jsem zvolil *T.2. Interoperability for Discovery and Navigation: ResourceSync and Signposting*.

### **Interoperability for Discovery and Navigation: ResourceSync and Signposting**

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

<https://indico.cern.ch/event/405949/contributions/2486374/>

Tutoriál byl rozdělen na dvě části. První byla věnována technologii SignPosting a druhá se zaměřila na praktické aspekty protokolu ResourceSync a část druhé části byla přednášena Petrem Knothem z Knowledge Media Institute, The Open University (UK).

SignPosting ([signposting.org](http://signposting.org)) je metodika/protokol/technologie navržená pro snadnější práci těch, kteří sklízí weby (harvestery, roboti, indexátoři obsahu jako je Google) a chtějí se dostat snadněji k podstatným datům (například fulltextům). Současná situace je taková, že webové stránky jsou značně nepřehledné s velkým množstvím odkazů a je velmi obtížné oddělit „zrno od plev“. Technologie SignPosting dává administrátorům webových stránek nástroje, pomocí kterých je možné „vyznačit“ podstatný obsah na stránce. V praxi využívá HTTP hlaviček, ve kterých je možné specifikovat odkazy na podstatné dokumenty na dané webové stránce přístupné (například `Link: <URL>; rel='item'; type='pdf'` – takto se každý snadno dostane k PDF, které webová stránka odkazuje). V současnosti je to spíše specifická záležitost pro konkrétní roboty a klienty. Nicméně je to již oficiální součástí hlaviček HTTP.

ResourceSync je protokol pro synchronizaci zdrojů na webu a je mj. zamýšlen jako náhrada protokolu OAI-PMH. Během posledních dvou let konečně začaly práce na praktické implementaci programátorských knihoven (zatím k dispozici v jazyce Python), které umožní vývojářům integraci tohoto protokolu do repozitářů a digitálních knihoven. Zájemce o ResourceSync mohou odkázat například na článek <http://www.inforum.cz/sbornik/2014/37/>. S pomocí protokolu ResourceSync je budován systém CORE (<http://core.ac.uk>), který agreguje celosvětový open access obsah (a dělá to pokud možno strojově). Mj. se snaží sklízet i velké nakladatele jako Elsevier, Springer, Wiley aj. (pro ně je však nutné pro každého zvlášť napsat speciální software na harvestování). Při sklizení takto velkých objemů dat narazili tvůrci CORE na limity protokolu OAI-PMH a i to byla motivace pro přechod na ResourceSync. Vše je v současné době ještě ve vývoji, ale kód je dostupný na GitHubu.

### **Opening Keynote: Open Research Europe**

*Dr. Jean-Claude Burgelman (Open data and science cloud)*

<https://indico.cern.ch/event/405949/contributions/2487849/>

Přednášející byl zástupce komise EU pro Open Science – shrnul priority, které EU (jejich komise) vzhledem k výzkumu v EU má (cca 8 bodů, typu Open data, Altemetrics, all publications freely available atd.). Shrnul Horizon 2020, věci okolo OA (údajně je 60 – 68 % publikací v režimu OA – ale nezmínil, kde a jak to číslo vzal). Zdůraznil, že je nutné trvat na OA, pokud je výzkum financován z veřejných peněz a nejtít na ruku Elsevierum a jiným nakladatelům. Především pak byly představeny ORE (Open Research Europe) a European OpenScience Cloud. Tyto dva nástroje mají

tvořit jakousi publikační platformu s vlastním peer review, možností ukládat preprinty. Mělo by to být platforma, na které budou moci vědci publikovat (za poměrně malý peníz – většinu nákladů zaplatí EU), cíl je pomoci vědcům plnit program Horizon 2020. EU to chce outsourcovat (výběrové řízení poběží od listopadu 2017, snad by to mohlo být k dispozici někdy v roce 2018. V závěru ještě rozmýšlel, jak přesvědčit vedoucí a vlády, aby braly v potaz altmetriky. V diskusi padl názor „Na co další platforma, když už jich máme tuny... - bude to další monopolní těžkopádný systém, který nebude fungovat!“.

## Technical session

### **Making a Case for Decentralised Scholarly Communication**

*Sarven Capadisli (University of Bonn, Švýcarsko)*

<https://indico.cern.ch/event/405949/contributions/2487819/>

Přednáška představila webovou aplikaci/systém [dokie.li](https://dokie.li), který má sloužit jako prostředek pro mírně decentralizovanou komunikaci vědecké komunity. Umožňuje psaní článků, jejich ukládání/export do jiných systémů, dále umožňuje dělení sdílených anotací a poznámek, vkládání dat, grafů a multimédií atp. Celý systém je volně dostupný na GitHubu a vypadá celkem hezky (v podstatě moderní webová aplikace). Zmíněny byly i technologie a standardy, na kterých systém staví (s důrazem na otevřenost). Přednáška se následně dotkla i tématu autentizace uživatelů (jak skutečně určit, kdo je kdo) a problému dlouhodobé archivace digitálních dokumentů.

### **Interoperability and Data FAIRness emerges from a novel combination of Web technologies**

*Dr. Mark Wilkinson (Center for Plant Biotechnology and Genomics UPM-INIA, Španělsko)*

<https://indico.cern.ch/event/405949/contributions/2487823/>

Přednáška se zabývala „férovostí“ (**FAIRness = Findable, Accessible, Interoperable, Reusable**) vědeckých dat – tedy tím, aby vědecká data byla k dispozici k opětovnému použití (autor uvedl vlastní studii, podle které 64 % dat není „fér“). V přednášce představil některé technické koncepty, které by měly fairness dat zlepšit. Kritizoval nadužívání dalších a dalších API standardů a doporučoval použít stávající (W3C, Linked Data Platform). Navrhl FAIR Accessor – jakýsi kontejner v XML, který umožní jednotně popsat metadata vztahující se k datům, který nevyžaduje žádné API a je snadno strojově zpracovatelný (například Googlem). Ukázal i praktický příklad. Následně ukázal další technologie, které by měly pomoci zvýšit interoperabilitu (a tedy FAIRness) vědeckých dat (schéma RLM, RDF apod.). Byly popsány i slabiny výše zmíněného a ukázáno, že automatizace celého procesu dělení vědeckých dat FAIR je náročná a ještě vyžaduje zkoumání.

### **The Web is not a Glorified USB Stick**

*Phil Archer (W3C, UK)*

<https://indico.cern.ch/event/405949/contributions/2490466/>

Zástupce konsorcia W3C shrnul doporučení ohledně ukládání dat na webu. Dále velmi propagoval W3C schéma **schema.org** (údajně pokrývá předmětovými hesly všechny možné obory). Ukázal nějaké W3C převodníky jak dostat například CSV na web (ve formátu JSON). Dodržování standardů W3C umožňuje lepší zpracování ukládaných dat (byť vědci stále používají proprietární formáty).

## Copyright and Licensing Session

### Copyright and Education

*Natalia Mileszyk (Centrum Cyfrowe, Comunia Association, Polsko)*

<https://indico.cern.ch/event/405949/contributions/2487874/>

Přednášející se věnovala praktickým dopadům licencí a copyrightu na výuku v různých zemích. Uvedla příklad doktorského studenta (Dieogo Gomez z Kolumbie), který publikoval cizí práci, protože ho štvalo, že je těžké se k plným textům dostat. Autor práce jej však zažaloval přímo v Kolumbii a dle tamních zákonů to vypadá na 3 až 4 roky ve vězení. Problematiku copyrightu mírně řeší CC licence, ale ne úplně. Pomoci by snad měla nová legislativa EU (podobná právní ustanovení jako jsou pro novinářské zprávy by se mohla vztahovat i na školní práce) – měla by zmírnit dopady copyrightu ve výuce. Ukázala několik scénářů (například jsem učitel a chci ukázat video z Youtube) a jejich dopady v 15 zemích EU – legislativa se velmi liší a například Francie má poměrně striktní zákony a téměř vše je zakázáno. Naproti tomu ČR (a obecně východní blok) z celého porovnání vyšla téměř nejlépe (copyright nejméně blokuje výuku). Legislativa by se měla napříč EU více sjednotit a máme prý tlačit na naše evropské poslance :-).

### Rights as the foundation of scholarly communication

*Chris Banks (Imperial College London, UK)*

<https://indico.cern.ch/event/405949/contributions/2487876/>

Přenáška o tom, jak to mají s copyrightem a OA ve Velké Británii. Problém – OA, institucionální politiky a copyright nejsou kompatibilní s požadavky a podmínkami těch, kdo financují výzkum (a především Gold OA) – a je otázka, co s tím. Samotné licence nikdo nechte (aby člověk přečetl všechny licence, se kterými běžně přichází do styku, potřeboval by asi 76 dní ročně). Nelze stanovit obecná pravidla, která by odpovídala vědcům na otázky okolo OA – je nutná individuální konzultace každého případu. V UK se pokusili zavést model, který používá Harvardská univerzita a docela se to osvědčilo (na všichni nakladatelé na to však reagovali pozitivně). Ještě do celé doladují.

### Legal Aspects of Text and Data Mining (A current European Outlook)

*Jonas Holm (Stockholm University, Sweden)*

<https://indico.cern.ch/event/405949/contributions/2527697/>

Shrnutí situace okolo (ne)legálnosti dolování textů a dat. Shrnul současnou právní situaci (z evropského hlediska je dolování v podstatě legální, jiná je situace ve Švýcarsku a USA). V USA je to nicméně obecně jednodušší (nemají výjimky, ale obecná ustanovení – byť to přebíjí konkrétní licence). V EU se připravuje na podzim 2018 nová právní úprava.



22. 6. 2015

## OA Transformation - From Subscription to OA

### **Some introductory thoughts on economics aspects of scholarly communication**

*Dr. Thomas Krichel (Open Library Society, Rusko)*

<https://indico.cern.ch/event/405949/contributions/2487829/>

Poměrně zajímavá (a svým způsobem i kontroverzní) přednáška o současných modelech vědeckého publikování. Základní předpoklad – současná ekonomika je přehlcena informacemi a je nutné přitáhnout pozornost – a zde model předplatného selhává (na rozdíl od ekonomiky informací, kde je informací méně, ale jsou kvalitní). Nikdo nechce přeplácet zbytečný odpad, ale nakupování článků formou APC to neřeší (vyšlo by to až 20x draž než předplatné, takže všichni stále jedou na předplatném i s tím zbytečným odpadem). V ekonomice, která chce přitáhnout pozornost, dává OA mnohem větší smysl (snadněji se dostane ke čtenáři). Thomasova vize je, že OA musí nakonec zvítězit, protože dává větší smysl z hlediska autora a snad i z hlediska nakladatele! Předplatné je podle něj dnes špatně až do morku kostí (neprůhledné náklady, vysoké ceny, spousta odpadu, nemorální byznys). Knihovny a nakladatelé jsou vlastně jen prostředníky v komunikaci mezi vědci a jako takoví mohou být nahrazeni technologiemi. Předpovídá, že budoucnost bude patřit oborové komunikaci a oborovým repozitářům (uvádí ekonomický RePEc, u fyziky SCOAP3, Pubmed) – dojde tedy k fragmentaci (byť ji brzdí EU se svými nápady jako Open Research Environment). Přednáška následně vyvolala poměrně zajímavou a bouřlivou diskusi :-).

### **Discussion - impact of open access on the cost of scholarly communication**

*Dr. David E. Shulenburg (Association of Public and Land Grant Universities, USA)*

<https://indico.cern.ch/event/405949/contributions/2519395/>

Přednáška předestřela možná řešení současné situace okolo nákladů vědeckého publikování a plusy a mínusy Gold OA. Shrnuje situaci okolo předplatného – proč se situace vyvinula tak, jak se vyvinula (řada tabulek, grafů etc.). Ekonomické metriky ukazují, že zvyšování cen předplatného je dáno tím, že se trh nakladatelů monopolizuje (Harvardova univerzita už v roce 2002 deklarovala, že si všechna předplatná nemůže dále dovolit) – nicméně i z hlediska trhu nemohou ceny růst do nekonečna. Podle Davida bude nakonec všechno free v modelu Gold OA. Pokud budou autoři nuceni platit APC, tak si dají větší pozor, co publikují (ubude odpadu) a nakladatelé budou zvyšovat ceny jen tolik, co trh snese. Na druhou stranu to může vyřadit ze hry knihovny – a autor zůstane proti nakladateli sám (dosud za něj bojovaly knihovny) – porovnal situaci na trhu s léčivými v USA, kde nejsou ceny regulované a jsou nejvyšší na světě (na rozdíl od zemí, kde za pacienta bojují vlády). Navíc nelze odhadnout, jestli bude model Gold OA (APC) levnější než předplatné – což je z tržního pohledu problém, protože kdo musí nějaký model obhajovat, tak ten v tržní konkurenci prohrává. Cílový stav by měl být v tom, že překonáme dysbalanci v tom, kolik můžeme za předplatné platit a kolik chtějí nakladatelé, nicméně je diskutabilní, jestli je Gold OA řešením.

### **Can we afford not to go Gold?**

*Dr. Jeff MacKie-Mason (University of California, Berkeley, USA)*

<https://indico.cern.ch/event/405949/contributions/2601831/>

Přednáška diskutovala model Gold OA a ve srovnání s předchozí přednáškou jej obhajovala jako jediný smysluplný a ekonomicky udržitelný model do budoucna. Vycházela hlavně z ekonomického pohledu – celý proces publikování něco stojí a někdo jej musí zaplatit. Nakladatelů, kteří tento servis dělají se prostě nezbavíme. Podle přednášejícího je Green OA naprostý nesmysl (nevidí v



tom udržitelný byznys model). To, co dnes dává nakladatelům sílu je copyright, který jim autor odevzdává – ale to v Gold OA není, nebo nemusí být! Hlavní hybnou silou na trhu je tedy autor a jeho možnost pracovat copyrightem, autor by tedy měl vyrovnávat sílu nakladatelů.

### **Poster session**

Postery byly letos docela zajímavé, bylo na ně dost prostoru i času. Ze zajímavých posterů bych vybral: *Normalization and Authority Control in a Repository Landscape* (IMG\_20170622\_123939.jpg), *A snapshot of the situation of flipping journals published by large publishers between 2013 and 2017* (IMG\_20170622\_124153.jpg) a *Machine accessibility of Open Access scientific publications from publisher systems via ResourceSync* (IMG\_20170622\_124442.jpg).

Fotografie posterů: [https://kic-internal.ics.muni.cz/~krejvl/oai10\\_posters.zip](https://kic-internal.ics.muni.cz/~krejvl/oai10_posters.zip)

## **OA Outside academy session**

### **Who needs access to research? an overview of available evidence**

*Mr. ElHassan ElSabry (National Graduate Institute for Policy Studies, Japonsko)*

<https://indico.cern.ch/event/405949/contributions/2487831/>

Autor se věnuje tématu OA a jeho společenským dopadům ve své doktorské práci – z ní celá přednáška vysílaná po Skype (s řadou přerušení kvůli technickým problémům) vycházela. V podstatě jen ukázal řadu tabulek a grafů, jak je na tom OA v rámci disciplín apod. Vypadla z toho celkem zajímavá data – podstatný z toho celého byl slajd „Conclusion“, vybírám zajímavé body:

- OA má pozitivní dopad i na průmysl
- stejně tak na oblast zdravotnictví
- řada vědců fenoménu OA moc nerozumí
- ne moc zajímavý pro neziskové organizace

Výzkum nadále pokračuje.

### **Is open access helping or hindering the international development agenda? Reflections from a consortium of developing country NGOs**

*Dr. Nilam Ashra-McGrath (University of Leeds, UK)*

<https://indico.cern.ch/event/405949/contributions/2493509/>

Kratší přednáška o tom, jak se na OA dívají různé typy organizací – neziskovky, novináři, nezávislí vědci, think tanky apod. Jeden ze závěrů byl, že OA podporuje demokratické procesy a je dobrý pro udržitelný rozvoj...

### **How open access opens doors - reflections on my recent 'Megaregions of the United States' paper**

*Alasdair Rae (Publons, UK)*

<https://indico.cern.ch/event/405949/contributions/2493512/>

Přednáška byla v podstatě case study jak je OA dobrý a užitečný. Alasdair Rae ukazoval na svém vlastním článku, jak moc mu pomohlo, že jej publikoval v režimu OA – všichni to přebírali, budovali na tom vlastní výzkum etc. Měl to i podložené daty. Pěkná motivační přednáška.

## Unconference

Speciální část programu – kdokoli chtěl, mohl vést svůj vlastní blok na vybrané téma, které se vybíralo v průběhu poster sessions hlasováním (dělali jsme čárky na papír :-)). Nakonec bylo vybráno 5 témat, která bylo možné navštívit. Já jsem byl na tématu **Populating repositories in an automatic way**.

Několik účastníků tam sdílelo své zkušenosti s automatizací plnění institucionálních repozitářů plnými texty. Thomas Krichel se pokoušel strojově plnit DBLP – to je bibliografická databáze pro informatiku. Zjistil, že pro relativně úspěšné dohledání plného textu stačí zadat Googlu autora a název a následně nějakou relativně snadnou heuristikou vybrat z výsledků nabídnutých Google ten, ze kterého se dá získat plný text (resp. parsoval stránky, na kterého ho Google odkázal – často osobní stránky autorů článku, kde byl vystaven i plný text článku).

Další dvě zajímavé diskutované varianty – zkoušet dohledávat Googlem přes DOI. Případně zkoušet hledat OA články přes dostupné API Scopusu.

23. 6. 2015

## Social Media session

### **Ethical Challenges of Social Media Data: Insights from Academia and Industry**

*Dr. Peter Bath, Wasim Ahmed (University of Sheffield, UK)*

<https://indico.cern.ch/event/405949/contributions/2488942/>

Přednáška se věnovala sociálním sítím a problémům souvisejícím s jejich využíváním (anonymita, informovaný souhlas apod.). Uvedla studii UK National Centre for Social Research – Research using Social Media: uživatelé na sociálních sítích přehánějí a chovají se jinak než ve skutečnosti, jejich profily jsou odtrženy od reality. Uvedli odkazy na některé výzkumné skupiny na Twitteru (#PhDChat v oblasti sociálních věd).

### **The ethical and practical challenges of using social media data in research**

*Anne Alexander (University of Cambridge, UK)*

<https://indico.cern.ch/event/405949/contributions/2488943/>

Přednáška diskutovala etický rozměr sociálních sítí ve výzkumné komunitě. V zásadě uvedla řadu zajímavostí jako je problém vlastnictví médií nadnárodními společnostmi a miliardáři, rozsah užívání sociálních médií ve světě (například v Číně hodně, v Indii vůbec). Nakonec shrnula spíše obecná doporučení jak a proč se na sociálních médiích (ne)chovat.

### **Social media metrics as a source of understanding scientific communication : descriptive and evaluative possibilities**

*Dr. Rodrigo Costas (CWTS-Leiden university, Nizozemí)*

<https://indico.cern.ch/event/405949/contributions/2488944/>

Nejzajímavější přednáška celého bloku – studie, která porovnávala alternativní metriky měření výkonu (získávané například ze sociálních sítí) s metrikami tradičními. Vycházela z toho poměrně zajímavá čísla. Provedli například porovnání altmetrik s daty ve Web of Science, jen 19 % článků ve WoS má nějaký impakt v sociálních médiích (z čeho je 87 % na Twitteru). Totéž zkusili pro Medelay (akademická sociální síť) – nicméně celkově je korelace altmetrik a klasických citací velice nízká! Ukázalo se, že ke zvýšení citovanosti na Twitteru je především třeba být sprostý a šokovat. Při přístup k altmetrikám je třeba být opatrný. Zcela zavrhnout sociální sítě netřeba – poskytují i řadu zajímavých informací: předvedeny byly speciální mapy témat, ze kterých lze vyčíst, čím se některé geografické komunity aktuálně zabývají apod. Nicméně – pouze 2 % vědecké komunity jsou na Twitteru a ten spíše slouží jako médium pro popularizaci vědy.

## Future of repositories session

Celou session uvedl *Eloy Rodriges* a položil základní otázku – je potenciál repozitářů plně rozvinut? Odkázal také na dokument skupiny Work group of next generation repositories (pod COAR).

### **World landscape of repositories and repository networks: achievements, challenges, opportunities.**

*Dominique Babini (CLASCO – Latin american council of social sciences, Argentina)*

<https://indico.cern.ch/event/405949/contributions/2483220/>

Přednáška byla jen shrnutím jaké repozitáře se kde na světě používají, jací jsou agregátoři, jaký obsah apod. Stále je v repozitářích (vzhledem k celkovému objemu mimo ně) málo obsahu. Mluvila také o některých projektech a jejich doporučeních (FORCE11, Open & Collaborative Science).

### **COAR NGR user stories and current developments**

*Petr Knoth (Knowledge Media Institute, The Open University, UK)*

<https://indico.cern.ch/event/405949/contributions/2483221/>

Petr mluvil o globální repozitářích, které aby se prosadili musí především umět dobře spolupracovat (sdílet data) a hlavně musí být konkurenceschopné (lepší než) komerční řešení. Tvrdí, že nedostatek interoperability je brzdou rozvoje. Potřebujeme transparentní sociální síť napříč repozitáři – kromě metadat a obsahu potřebujeme nástroje, které umožní uživatelům interakci s repozitáře: anotace, komentáře, sdílení, peer review... Repozitář mi má jednoduchým způsobem nabídnout to, co chci a potřebuji.

### **Governance and social interoperability**

*Leslie Chan (University of Toronto Scarborough, Kanada)*

<https://indico.cern.ch/event/405949/contributions/2636814/>

Přednáška se věnovala tomu, jak překonat komerční konkurenci repozitářů – množstvím peněz nelze konkurovat, je nutné se zaměřit jinam (například maximální jednoduchost instalace apod.). Leslie ukázal příklady, kde a jak může nasazování repozitáře selhávat a co lze dělat pro to, aby se podobné scénáře neopakovali.

### **Závěrečná diskuse (panel)**

Problém repozitářů je ten, že obsahují velké množství balastu a tudíž to lidi moc nezajímá. Přičemž

plný text je stále klíčový – a jeho získání se dá dělat i strojovou cestou.

**Closing keynote: Promoting a culture of preprinting in biology**

*Dr. Jessica Polka (ASAPBio, USA)*

<https://indico.cern.ch/event/405949/contributions/2487847/>

Závěrečná přednáška se věnovala oborovým repozitářům – zejména pak situaci v biologii. Přednášející porovnávala situaci okolo repozitářů ve svém oboru biologie se situací ve fyzice a zejména s repozitářem ArXiv.org (fyzikové jsou na to mnohem lépe než biologové). Ale i v biologii se situace mění k lepšímu (mají vlastní repozitář BioRxiv). Snažila se vysvětlit situaci okolo preprintů a proč jsou preprinty důležité, a proč je důležité, aby praxi zveřejňování preprintů podporovali ti, kdo financují výzkum.