Závěrečná zpráva k OAI 8, Cern, Ženeva, Švýcarsko, 2013

Vlastimil Krejčíř

18. 6. 2013

DSpace User Group Meeting (OAI8 Pre-Conference)

Setkání uživatelů DSpace se zúčastnilo asi 30 lidí z různých koutů Evropy. Mj. byli přítomni i tři vývojáři DSpace: Bram Luyten, Robin Taylor a Richard Jones.

Stručné shrnutí DSUG: Systém ORCID – identity a svázání identit s daty ve vědecké komunitě, implementace protokolů SWORDv2 a ResourceSync v DSpace, shrnutí nových vlastností a funkcí v DSpace 3.x s následnou diskusí ohledně budoucího vývoje DSpace.

Shrnutí v angličtině od Brama Luytena:
http://atmire.com/website/?q=content/dspace-today-and-tomorrow&utm_source=gplus&utm_medium=social&utm_campaign=oai8report

First Session: Invited Speakers

"Interoperability between ORCID and DSpace repositories", by Laurel L. Haak (ORCID) and Todd Vision (NESCent). (US)

Paní *Lauren Haak* představila systém ORCID – v podstatě se jedná o velice propracovaný a sofistikovaný systém pro identifikaci osob (správu identit) a objektů. Částečně je podobný například ResearchID. ORCID kromě samotného jednoznačného identifikátoru pro osoby však zároveň nabízí i propojení tohoto identifikátoru s digitálními objekty (například články, monografiemi, ale i samotnými výzkumnými a jinými daty, napojení na grantové přihlášky apod.). Jedná se o mezinárodní projekt, do kterého se může zařadit kdokoli, včetně celých organizací (systém je zdarma, navíc poskytuje i API pro práci s ním).

Systém ORCID spolupracuje také přímo s řadou nakladatelů a organizací (NATURE, CrossRef, PUBMED, NIH). Dokáže zároveň propojovat již užívané systémy identifikátorů (např. zmiňovaný ResearcherID).

V následné diskusi padla otázka „Na co další systém identifikátorů?“ - odpověď byla, že ORCID je asi nejkomplexnější a nemá nahradit stávající systémy, ale sdružit identity.

Todd Vision pak ukázal integraci ORCID a DSpace v systému Dryad (datadryad.org). Dryad je v podstatě komerční oragnizace, která vyvíjí za úplatu nové funkce a pluginy do systému DSpace. Nabízí možnost si v jejich repozitáři ukládat data s celou řadou dalších přidaných služeb, které naprostá většina jiných subjektů (údajně) nenabízí:

- napojení na publikovaný obsah (například u nakladatele)
- snadné propojení článků s výzkumnými daty
- vámi uložený článek (či jiná data) procházejí ještě přes tzv. curator, tedy člověk, co to následně zkontroluje (fungují linky apod.)
- verzování uložených dat
- podporují peer-review, uložený článek je ve spolupráci s nakladatelem poskytnut pro peer-

review proces (jen pro autentizované uživatele)
- jmenné autority, ORCID
- …

Není to ovšem zadarmo – pokud si tam chcete vkládata data, pak za to musíte platit (ceny se pohybují v desítkách USD).

"Can Standards Save Us? (aka ResourceSync and SWORDv2 and what they mean for DSpace)", by Richard Jones (Cottage Labs)

Richard Jones je jedním z hlavních vývojářů DSpace (už od počátků vývoje). Ve svém technicky zaměřeném příspěvku mluvil především o protokolech SWORDv2 a ResourceSync a jejich implementaci v DSpace. Základní filozofií při budoucím vývoji je dodávat funkce do DSpace pomocí pluginů, nejinak je tomu i pro tyto dva protokoly.

Protokol SWORDv2 slouží jako základní mechanismu pro interoperabilitu mezi různými repozitáři. Umožňuje především manipulovat s obsahem repozitáře (např. vkládat objekty) standardním způsobem. V podstatě se jedná o jednoduché (neúplné) REST API – tedy neumožňuje některé komplexnější operace s větším množstvím objektů. Richard rozebíral současné problémy protokolu v DSpace (např. jak se chovat k objektům typu WorkspaceItem) a mluvil o méně známých možnostech protokolu (výměna jednotlivých souborů, mazání, výpis kolekcí).

Protokol ResourceSync je v podstatě stále v procesu návrhu a bylo mu věnováno hodně prostoru i v rámci samotného OAI8 (za návrhem protokolu stojí Herbert van de Sompel). Tento protokol by měl sloužit obecně k synchronizaci zdrojů na Internetu a být náhradou za OAI-PMH. Výhodou ve srovnání s OAI-PMH je zejména to, že není zaměřen jen na metadata (v terminologii DSpace mohu sklízet přímo Bitstreams) a umožní zcela jiný přístup ke sklízení množin (Sets), které budou moci být mnohem flexibilnější (byť finální verze v tomto směru ještě není jasná). Richard Jones podrobně rozebíral technická úskalí implementace protokolu ResourceSync v DSpace – problémy s podporou jednoznačných URI pro objekty (a zejména samotná metadata) a konkrétní formáty, udržování historie změn, ověřování autorizací aj. Mj. rozebíral i systém, jakým je ResourceSync implementován v rámci vylepšeného systému pluginů pro novější verze DSpace (verze 4 a vyšší, stále ve vývoji). Poměrně zajímavá byla informace o tom, že při vývoji kódu pro ResourceSync byla využita řada podpůrných tříd, které se již užívají pro jiné funkce v DSpace (což do značné míry ukazuje na kvalitní objektový návrh programového kódu DSpace).

Second Session: New and Upcoming DSpace features

DSpace: Hidden Gems, by Bram Luyten (@mire)

Bram Luyten mluvil o vlastnostech a funkcích DSpace, které jsou v komunitě málo známy, případně které se připravují pro novější verze DSpace. Zde vybírám některé z nich:

- Nasazení GIThub a JIRA pro vývojáře DSpace přineslo transparentnost. Podpora Wiki.
- Vylepšené rozhraní DSpace pro mobilní zařízení by mělo být ve verzi DSpace 4.
- Submission forms je možné lépe konfigurovat a například je provázat dle typu vkládaného souboru (jiné formuláře pro obrázky, jiné pro dokumenty apod.). Měla by být možnost používat funkci AutoComplete, závislosti apod. Celé workflow vkládání je plně konfigurovatelné.

- Integrace s externími soubory jmenných autorit – na DSpace Wiki je návod na napojení autoritní báze OCLC LoC (https://wiki.duraspace.org/display/DSPACE/Authority+Control+of+Metadata+Values).
- Byl podrobněji popsán systém pro embarga (moving wall) v DSpace 3.0, nicméně v budoucnosti bude ještě vylepšen, tak aby přímo podporoval embargo pro specifikcý soubor a omezil například přístup jen pro vybranou skupinu (Group) uživatelů.
- Nový Browse & Search by měl být postaven na discovery (SOLR), nějaké informace by opět měly být na Wiki.
- Další ze zajímavých funkcí by mělo být napojení kolekce na externí zdroj (ze kterého se kolekce bude plnit).
- V DSpace 3.x je rozhraní OAI-PMH napsáno zcela od základu znovu – a poskytuje mnohem lepší možnosti konfigurace (např. lepší definování množin).
- Mluvilo se také o různých implementacích REST API pro DSpace – Bram doporučoval pro zájemce WIJITI případně New REST API Work (vše na Wiki).
- Statistiky – v budoucnu by měl být připraven nový stadardní modul, který bude lépe konfigurovatelný a statistiky tak budou (i autorizovaně) dostupné přes rozumné UI (přímo přes www).
- Implementace podpory oficiálního REST API, DOI, SWORDv2 v DSpace 4.0.

V následné diskusi jsme se shodli, že současný model pro ukládání metadat v DSpace není ideální a má řadu nedostatků, především není možné metadata strukturovat. Zajímavé bylo vidět, jak toto omezení vesměs všichni programátoři (včetně mě) obcházejí stejným způsobem. Také se diskutovalo o možnosti nějakým způsobem vkládaná metadata validovat (závěr zněl ano, ale moc už se nemluvilo do detailů jak).

Na OR 2013 by se měla sejít skupina okolo DSpace a nastínit vizi vývoje DSpace na dalších 4 až 5 let a případně se dále bavit až v horizontu 15 let. Z diskusí pak vycházejí některá doporučení, která se budou dále dodržovat:

- studovat use cases a na nich stavět nové funkce
- vše pokud možno dělat přes systém pluginů
- podpora DSpace jako hostovaného řešení (tedy více možností konfigurace přes webové rozhraní) – zde Bram zmiňoval například DSpace DIRECT řešení, které je mj. podporováno sdružením DuraSpace, které do toho dává nějaké peníze – bylo zajímavé vidět, kdo všechno a jak DuraSpace sponzoruje, a překvapilo mě, kolik peněz celá organizace získává (mám pocit, že tam byla čísla kolem 400 tisíc USD za rok)

Afternoon session: Repository managers *(Chair: Iryna Kuchma (EIFL.net))*

V rámci této session mluvili uživatelé o svých instalacích DSpace a zkušenostech s provozem. Po prezentacích následovaly diskuse ve skupinách – tématem bylo financování provozu a vývoje DSpace.

Marina Muilwijk (Utrecht University, Netherlands) ukázala systém Igitur postavený (jakožto PHP front-end) na jádře DSpace, který je po MIT DSpace druhou největší instalací DSpace ve světě. Igitur je spravován na komerční bázi (univerzita v Utrechtu si za správu a rozšiřování DSpace platí). Obsahem repozitáře Igitur je kompletní produkce celé univerzity. Hezké bylo vidět, že podobně jako u nás v ČR to mají spojené s řadou jiných systémů. Viz http://igitur-archive.library.uu.nl/search/search.php.

Vlastimil Krejcir (Masaryk University, Brno, Czech Republic) – přenášel jsem o našich dvou projektech DML-CZ (Česká digitální matematická knihovna) a FFdigi (Digitální knihovna Filozofické fakulty Masarykovy univerzity). Pokusil jsem se lehce odlehčenou formou podat naše zkušenosti s vývojem a se spoluprací se dvěma odlišnými skupinami uživatelů. Zároveň jsem prezentoval možnosti interoperability a naše zkušenosti s napojením na EuDML (European Digital Mathematics Library).

Jordan Piscanc and Stefania Arabito (University of Trieste, Italy) ukazovali implementaci NBN-IT, systému identifikátorů užívaném v Itálii (National Bibliography Number) – NBN jako takové používá více státu v Evropě, je na to napojen i rezoluční systém. Svůj DSpace nazvali OpenStarTS a díky podpoře NBN (speciální plugin do DSpace) jsou v rámci Itálie schopni zajistit jednoznačnost vloženého dokumentu. Například ukládané dizertační práce jsou z DSpace sklízeny národními knihovnami (v Itálii mají dvě – ve Florencii a Římě) – sklízení je nařízeno zákonem.

DuraSpace Update session

Následující část byla věnována již záznamům příspěvků, které pro toto setkání připravili američtí kolegové z komunity DSpace. Nejdříve mluvil jeden z klíčových vývojářů **Tim Donohue**, který v podstatě shrnul současný stav okolo DSpace (víceméně totéž, co říkal už Bram Luyten odpoledne), ukázal jak se dělal poslední release DSpace 3 a následně se přes Skype odpovídal online na dotazy. Pak jsme si ještě poslechli **Michele Kimpton**, jednu z hlavních postav DuraSpace, která ale mluvila spíš obecně o DuraSpace jako takové a o spolupráci – měl jsem pocit, že jsem se nic zásadního od ní nedozvěděl.

19. 6. 2013

OAI8 Workshop

Dopolední začátek celého wokshopu začal tutoriály na různá témata. jednotlivé tutoriály běžely paralelně, já si vybral tutoriál:

The NISO/OAI ResourceSync Synchronization Framework.
https://indico.cern.ch/contributionDisplay.py?sessionId=1&contribId=26&confId=211600

Tutoriál vedli *Herbert van de Sompel*, *Robert Sanderson* a *Richard Jones*. *Herbert van de Sompel* (Los Alamos National Laboratory) je hlavním duchovním tvůrcem protokolu **ResourceSync**; *Robert Sanderson* je programátor a výzkumník na University of Liverpool a v Los Alamos National Laboratory, dlouhodobě se věnuje technologiím okolo repozitářů a digitálních knihoven (mj. SRW/SRU, data mining, informational retrieval), na protokolu ResourceSync úzce spolupracuje se Sompelem; *Richard Jones* tvoří první implementaci (prototypově) v systému DSpace. Mj. se na tvorbě protokolu podílí celá řada dalších lidí z různých institucí (RedHat, ExLibris, JISC a další).

ResourceSync má být obecným nástrojem pro synchronizaci ***jakýchkoli*** zdrojů na Internetu zcela nezávisle na zdroji i rychlosti změny dat. Celý protokol má být striktně modulární, každý modul má mít nějakou funkci – uživatel pak podle svých potřeb použije pouze vybrané moduly (nemusí použít vše). Protokol pracuje s konceptem URI – to, co je synchronizováno je právě obsah URI nezávisle na formátu. Definuje dva hlavní pojmy a to je **Source** a **Destination**, tedy zdroj, odkud se synchronizuje, a cíl, kam se synchronizuje. Protokol následně ustanovuje povinné i volitelné funkce, vlastnosti a postupy, které musí Source i Destination mít a umět, aby mohly efektivně celou synchronizaci provádět.

Celý tutoriál trval přes 2 a půl hodiny a protokol byl rozebrán skutečně podrobně včetně technických detailů (jednotlivé formáty XML pro komunikaci, zipování, verzování etc.). V případě zájmu se mohu pokusit tyto informace předat formou podobného workshopu, byť zdaleka nemám tak hluboký vhled do celé problematiky jako autoři protokolu.

Zde jen uvedu bodově základní aspekty protokolu:

- možnost popsat obsah
- možnost balit data do balíčků „zip“
- popisování změn, rozdílové „aktualizace“ (i do historie)
- alternativní zdroje (resp. alternativní reprezentace téhož obsahu), mirroring
- linkování na starší verze obsahu
- discovery ohledně vlastností zdroje

Sychronizace je možná řadou způsobů: iniciální sync všeho, inkrementální sync - audit (co se dělo, povedly se klasické sync), selektivní sync (cíl se nemůže volně volit, co bude synchronizovat – toto je stále ve fázi řešení, poměrně hodně komplikovaný problém a byla kolem toho dost diskuse), metadata harvesting (snižování množství přenesených dat např. přes diff).

Protokol není stavěn na bázi busy waiting, ale samotné zdroje by měli posílat notifikace cílům (v různých definovaných intervalech). Zde je to stále ve velice experimentální fázi, nicméně na několik příkladech z praxe se ukázalo, že to v zásadě nebude problém, protože notifikace představují jen velmi malý objem dat a malý provoz na síti. (Use cases: arXiv.org, Dbpedia.)

Technicky se pro popis toho, co zdroj nabízí používá formát pro Sitemap a Siteindex, který dané XML rozšiřuje o řadu dalších elementů (nový namespace *rs*).

Plenary 1: Technical Session

Shrnutí: První technická sekce se zabývala sémantikou na webu – technicky z hlediska protokolů i více lidsky z hlediska obecného. První přednáška ukazovala stručně přehled standardů a příkladů sémantických aplikací a webů a dávala návody, jak z dat sématiku dostat. Druhá přednáška se zabývala standardem Open Annotations. Třetí přednáška byla o jménech a odkazech na webu, o konceptu jako takovém a jak řešit problém mizejících odkazů.

How semantic representations can support scholarly communication
Presented by Mr. Paul GROTH (VU University Amsterdam, Netherlands)
https://indico.cern.ch/contributionDisplay.py?sessionId=3&contribId=0&confId=211600

Přednáška na začátku shrnula sémantické formáty nebo lépe řečeno způsoby přístupu k sémantice, které jsou používané na webu: *microformats* (http://microformats.org/), *OpenURL CoinS* (ContextObjects in Spans), *Open Graph Protocol* (http://ogp.me/) a *W3C RDF*. Byl doporučen zajímavý nástroj *Google Structured Data Viewer*, který slouží ke snadnému prohlížení sémantických dat na webové stránce. Jako příklad stránek se sémantickými informacemi byl ukázán projekt *Pinterest* (věda pro děti). Autor ukázal, že technicky sémantiku umíme a umíme ji dostatečně. Byly ukázány zajímavé příklady sémantického obohacování (například využití Google Maps – vezmou se data z článku a graficky se promítnou do mapy). Díky správné kombinaci přístupů lze získat poměrně zajímavé informace například o tom, kdo co studuje (ve vědecké komunitě), odkud na to získává peníze, … - předvedeno v rámci projektu *VIVO*, které ukazuje vědecké profily, na základě nich může sestavovat grafy jak které obory spolupracují. Dalším zmiňovaným projektem byl *W3C: PROV* – standard poskytující informace o původu dat, entit,

profilů aj., kontrolované slovníky pro komunikaci apod. *Open PHACTS* – agreguje informace z řady aplikací a nad nimi poskytuje API, které lze využít na budování jiných aplikací – ukázka například z oblasti chemie.

W3C Open Annotation effort: Status and Use Cases
Presented by Mr. Robert SANDERSON (Los Alamos National Laboratory, USA)
https://indico.cern.ch/contributionDisplay.py?sessionId=3&contribId=1&confId=211600

Robert Sanderson představil stanndard Open Annotations, který vydává organizace W3C. S pomocí současných webových technologií má pomoci vytvářet různé vazby mezi objekty na webu (Annotation = množina sémanticky spojených objektů) a tím jej obohacovat – pokud se to bude dělat standardizovaně, pak všichni budou schopni tyto informace číst a zpracovávat. Například: uživatel si něco na webu čte a chce k tomu přidat komentář, popsat to, prolinkovat na jiný článek, klást otázky, …). Byl představen model anotace a rozebrány technické problémy s tím související včetně návrhu řešení a podrobného popisu celého modelu (poměrně do detailů). Při tom R. S. ukázal, jakými různými způsoby je možné tento standard využívat a k čemu to může být dobré. OpenAnnotations je vlastně framework postavený na technologii JSON (JavaScript), pomocí které se vytváří RDF vazby mezi objekty. Příklady použití: peer-review procesy, bookmarkování, tagging (tedy i klasické „poznámky na okraji“).

Naming on the Web: What scholars should want, and what they can have
Presented by Mr. Henry THOMPSON (University of Edinburgh, UK)
https://indico.cern.ch/contributionDisplay.py?sessionId=3&contribId=2&confId=211600

Musím říci, že toto byla asi nejzábavnější a svým způsobem nejzajímavější přednáška celého dne. Autor se v ní zamýšlel nad tím, jak pojmenovávat věci (a ne jen na webu) a jak nám nejednoznačnost komplikuje život (a jak z toho ven). Vlastně tak nějak shrnul i řadu problémů, s kterými se potýkáme denně a přitom si tak úplně neuvědomujeme, že to problémy jsou. Prof. Thompson se zabývá perzistentními identifikátory a věcmi s tím spojenými, mj. je to také bývalý kolega Tima Bernes-Lee (autor první webové stránky a www serveru) a jak pan profesor na začátku řekl, T. Bernes-Lee nikdy nesouhlasil s jeho názory a proto je mu pan profesor za spolupráci velmi vděčný. Na začátku přednášky nastínil zdánlivě černou budoucnost knihovníků, protože „co není na webu, tak dnes neexistuje“. Nicméně věda závisí na kvalitních referencích a ty současný web moc udržovat neumí – odkazy rychle mizí a zpětně něco dohledat je často nemožné (ukazoval docela hezké příklady URL citací, které byly v podstatě zpětně nedosažitelné). Tedy koncept URL je špatný a v současnosti persistence nemá na webu jednoznačné řešení. Zde mu bylo z publika předhazováno např. DOI apod. - což pan profesor po bouřlivější diskusi tak trošku smetl ze stolu – závěr byl takový, že to, co potřebujeme, není technické řešení (to v podstatě máme), ale spíše něco, čemu pan profesor říkal „social contract“, tedy prostě se na nějakém systému mezi lidmi dohodnout a ten hlavně dodržovat. *Slajdy z přednášky jsou docela bohaté – doporučuji k prohlédnutí.*

Plenary 2: Metrics

Shrnutí: Sekce se se věnovala způsobům měření vědy (výhody, nevýhody, jak se to kde dělá). První přednáška diskutovala alternativy k impakt faktoru a měření dle chování uživatelů. Druhá přednáška se věnovala měření vědy na sociálních sítích. Poslední přednáška se zabývala (ne)transparentností současného peer-review procesu v kontextu OA a navrhuje postupy k měření transparentnosti tohoto procesu.

An overview of scholarly impact metrics
Presented by Mr. Johan BOLLEN (Indiana University, USA)
https://indico.cern.ch/contributionDisplay.py?sessionId=4&contribId=3&confId=211600

Klíčovou otázkou je, co měřit a v jakém rozsahu. Příklad trojrozměrné krychle, ve které do sebe všechny její části zapadají – ze které osy měřit? Můžeme zkusit i něco jiného než je klasický impakt faktor - něco má například Google (page rank i pro časopisy), zmíněny byly i jiné podobné metody (např. eigenfactor.org). Jako inovaci měření navrhuje měřit dle chování uživatelů – co lidé skutečně stahují a čtou? Kdo stahuje, odkud, kdy – to vše se dá zjistit a analyzovat = projekt MESUR na měření traffic flow ve vědecké komunitě (data od nakladatelů, agregátorů, konsorcií). Altmetrics – alternativní měření na základě chování uživatelů na sociálních sítích. Uživatelé sami vědí, co je dobré a kdo je dobrý, o tom se na sociálních sítích mluví (facebookovou terminologií „má více liků“). Například studie provedená na Twitteru zmiňuje korelaci mezi počty citací a počty stažení článků. Na závěr přednášky bylo řečeno, že stejně nevíme, co to vlastně ten impakt faktor je – že je to jako vyrábět mapy ve středověku. Celá přednáška byla zajímavá, ale nejsem si moc jistý, jak moc by se to dalo uplatňovat v praxi.

Discussions of scholarly articles online: who, why and where
Presented by Mr. Euan ADIE (Altmetric LLP, UK)
https://indico.cern.ch/contributionDisplay.py?sessionId=4&contribId=4&confId=211600

Zde předesílám, že shrnutí bude stručnější – přednášejícímu bylo velmi špatně rozumět (dost mumlal), takže jsem možná ne všechno pochytil přesně. Víceméně se opět věnoval měření vědy v sociálních sítích a ukazoval projekt altmetrics.org. Mělo by to být jen pomocné kritérium – například dle altmetrics.it je na Twitteru pouze 14 % vědců, 54 % článků bylo tweetováno jen jednou apod. Altemetrics spolupracují například s Pubmedem. Účast vědců v projektu velmi závisí i na komunitě – například sociální vědy nebo matematika a fyzika se moc neúčastní (cca 10 %). Tedy jsou to zajímavá čísla, která ale klasické měření nenahradí.

Assessing the transparency of peer review in (Open Access) journals
Presented by Mr. Jelte M. WICHERTS (Tilburg University, Netherlands)
https://indico.cern.ch/contributionDisplay.py?sessionId=4&contribId=6&confId=211600

Přednáška byla do značné míry kritikou současného peer-review procesu, navrhuje postupy, jak ověřit kvalitu a na příkladech porovnává, jak jsou tyto postupy úspěšné.. Peer-review proces je vzhledem ke své netransparentnosti velmi nedobrý pro OA časopisy, které tak nejsou pod žádnou kontrolou. V OA musí nakladatel především vydělávat a udělat si jméno, proto publikuje hodně, z čehož má hodně peněz. A aby toho dosáhl, snižuje kvalitu peer-review. V tomto bodě by to velmi chtělo transparentnost. V BioMed Central transparentní peer-review vyzkoušeli a fungovalo to. Autor příspěvku navrhuje jakousi stupnici hodnocení transparentnosti peer-review (5 stupňů) a dává doporučení, jak by měl celý transparentní proces vypadat (jaké otázky si klást, chceme-li zjistit transparentnost časopisu) – to znamená zejména poskytování podrobných informací o nakladateli, licencích, financování, instrukcích pro autory. Recenzenti by měli být nezávislí a nakladatel by si toto měl vynucovat, ale zároveň by měl akceptovat, pokud autor nechce, aby mu nějaký konkrétní člověk jeho článek recenzoval apod. Na základě těchto kritérií nastavili vlastní nástroj OA tool a provedli měření – poté se dotazovali cca 20 vědců (kteří byli ochotni spolupracovat) z řady odborných časopisů a zjištěné informace z OA tool porovnali s realitou (informacemi přímo do vědců). Vycházela čísla kolem 80 % shody (alespoň doufám, že jsem to správně pochopil). Ve druhé podobné studii během meetingu v Rotterdamu (nakladatelé, vědci, knihovníci, lidé z financování) bylo dotazníkem na účastníky takto hodnoceno asi 42 časopisů (17 etablovaných, 12 OA dobrých a

13 OA predátorských) a shoda s OA tool byla 90 %. Třetí představená studie brala 400 časopisů (náhodně) z DOAJ a nechala je hodnotit 20 knihovníky z nizozemských univerzit. Opět vyšla spolehlivost OA toolu cca 80 %. Závěr zněl, že zjišťování transparentnosti lze dělat i automatizovaně – celý OA tool zatím však není k dispozici online (nicméně měl by v budoucnu být).

20. 6. 2013

Plenary 3: Data and Document Semantics

Shrnutí: První přednáška představila současné i budoucí možnosti sémantického vyhledávání v PubMedu a strojového indexování. Druhá přednáška se zabývala strojovým přiřazováním sémantiky slovům. Ve třetí přednášce mluvila přednášející o systému automatické detekce důležitých informací ve vědeckých publikacích (vyhledávání tvrzení, závěrů apod.). Poslední přednáška se věnovala sdílení a obohacování vědeckých dat (které jsou základem výzkumu).

Semantic indexing in PubMed
Presented by Mr. Olivier BODENREIDER (U.S. National Library of Medicine)
https://indico.cern.ch/contributionDisplay.py?sessionId=6&contribId=8&confId=211600

Příspěvek ukazoval možnosti, které nabízí nebo by v budoucnu měl nabízet vyhledávací systém v PubMedu. Kromě klasického vyhledávání experimentují vývojáři i s vyhledáváním sémantickým přes MeSH. Na základě analýzy se pak celé vyhledávání chová chytřeji, například to asociuje slova s termíny v MeSH (například anglicky zadaný dotaz „heart attack“ převede na dotaz téže diagnózy v termínech MeSH, tedy „myocardical disfunction“). Dokáže dodávat i synonyma (na základě systému UMLS – Unified Medical Language System), pak na jednoslovní dotaz hledá celou další množinu slov – vůči uživateli transparentně, ale lze si nechat i rozvedený dotaz vypsat). Toto sémantické vyhledávání se dá již používat – za dotaz je třeba napsat řetězec „[Mesh]“. Druhý koncept představený během přednášky se týkal automatického indexování – strojového vyhodnocování klíčových slov z názvu a abstraktu (což nemusí být finální index, ale může sloužit jako pomůcka pro knihovníky). V závěru byla zmíněna i možnost dalších vylepšení – například automatické tvorby RDF trojic z textů – normalizací termínů (využívá opět MeSH a UMLS). Výsledkem je Semantic PubMed, který pak umí pěkné grafy vykreslující vztahy mezi termíny, a dokonce umí i nějakým způsobem odpovídat na otázky. K dispozici je zatím jen jakási demo verze na http://skr3.nlm.nih.gov/SemMedDemo/

Transformation of keyword indexed collections into semantic repositories
Presented by Mr. Javier LACASTA (University of Zaragoza, Spain)
https://indico.cern.ch/contributionDisplay.py?sessionId=6&contribId=9&confId=211600

Přednášející z mého pohledu vyhrál první cenu v kategorii „nejhůře srozumitelná angličtina“ (dost huhlal) a snažil se ukazovat složité a maličké grafy na slajdech – navíc tento obor moc neznám, proto následující shrnutí může být nepřesné. Přednáška se zabývala sémantickým hledáním – snižování nejednoznačnosti, ukazovala způsoby, jak dát klíčovým slovům nějakou sémantiku spojením se slovy z Wordnetu, který zároveň nějak spojují se systémem DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering). Významové nejasnosti se řeší přidáváním váhy a pravděpodobností – každý význam pak získává nějaké skóre, na základě kterého jej lze danému klíčovému slovu přiřadit. Systém například rozlišuje i mezi druhy slov. Systém nakonec tvoří RDF trojice s nimiž pracuje pomocí jazyka SPARQL.

Detecting knowledge-level claims in research articles
Presented by Mrs. Ágnes SáNDOR (Xerox Research Centre Europe, France)
https://indico.cern.ch/contributionDisplay.py?sessionId=6&contribId=10&confId=211600

Přednášející se věnovala problému hledání odborných faktů a tvrzení v textech, jejich extrakci a strojovému zpracování jejich sémantiky (dělá výzkum u firmy Xerox). Na začátku položila několik klíčových otázek – co to vlastně takové „tvrzení" nebo „idea" (vědecká) jsou a pak se pokusila ukázat, jak jejich zvýraznění v textu může pomoci. Základním postupem je hledat v textu známé fráze typu „... a little is known...", „... inconsistent with previous hypothesis..." nebo „... role... has benn elusive...". Na základě toho se dají uhádnout podstatné části textu (které se zároveň i klasifikují: myšlenka, otevřená otázka, tvrzení apod.) – ty můžeme zvýraznit nebo vytáhnout a získáme tak rychle výsledky daného výzkumu. Speciálně pak byly probrány společenské vědy, kde je situace pro strojové zpracování složitější, přesto i s tímto si nějakým způsobem poradili. Tyto algoritmy pak byly použity v některých evropských projektech jako pomocné indikátory hodnocení kvality. Pomoci to může i při peer-review procesech, při studiu a mj. také při psaní článků v angličtině (jaké fráze a obraty používat apod.). Závěrem bylo řečeno, že těch příkladů využití to najde v budoucnu možná i více a že se stále nové možnosti objevují.

Small Data, or: Bridging the Gap Between Smart and Dumb Research Repositories
Presented by Mrs. Anita DE WAARD (Elsevier)
https://indico.cern.ch/contributionDisplay.py?sessionId=6&contribId=11&confId=211600

Příspěvek se zabýval problémem sdílení a zpřístupnění vědeckých dat – zejména takových, na kterých leží zásadní tvrzení výzkumu. Údajně až 90 % takových dat leží někde zahrabáno na pevných discích a pouze 1 až 2 % jsou sdílena, 8 % leží v zapadlých repozitářích. O zlepšení se pokouší např. Elsevier svým programem Research Data Services, v rámci soutěže o nejlepšího digitalizátora dat je možné vyhrát 5 tisíc USD) – podporuje navíc celou řadu pilotních projektů. Množství klíčových dat není digitalizováno – byly ukázány projekty, které teď masivní digitalizaci dělají (digitalizace map mořského dna, vyhráli soutěž Elsevieru, http://www.marine-geo.org/index.php). Doporučuje se zároveň obohacovat metadata o datech – pokud je to však děláno špatně, tak to spíše škodí. Další klíčovou věcí je, aby vědec data našel, pokud možno na jednom místě – řešením by měla být interoperabilita mezi repozitáři. Poslední velkou otázkou vznesenou na přednášce bylo „jak to celé financovat?", tedy najít vhodný ekonomický model - zde je ještě řada otázek.

Posters session

Během poster session jsem měl na starosti své postery (DML-CZ, EuDML), bohužel mi tak nezbylo mnoho času na postery ostatních. Několik vyfocených posterů přikládám na konci zprávy. Nicméně dva postery bych rád zmínil [jejich náhledy přiloženy na konci zprávy]:

LIBRE – liberating research (http://www.libreapp.org/)
Poster a jeho autoři kritizovali současný neprůhledný model peer-review procesu a nabízejí aplikaci, která vědecké komunitě umožní toto zásadním způsobem zlepšit – aplikace zcela podporuje peer-review proces. Momentálně hledají nějakou vědeckou komunitu, která by toho šla.

Citation Finder.
Poster představil komplexní nástroj na vyhledávání a parsování citací. Výsledky vypadaly velice dobře.

Plenary 4: Research Data

Shrnutí: Sekce se nesla ve znamení práce s vědeckými daty. První přednáška se zabývala politikami pro uchovávání vědeckých dat a ukázala řadu příkladů. Druhá přednáška se věnovala obohacování, ukládání a integraci vědeckých dat z různých oborů a jejich využití v praxi (projekt iMarine). Další přednáška se zabývala obecně kvalitou dat. Poslední přednáška ukazovala jakým způsobem se zpracovávají velké objemy dat (desítky PB) v CERNu.

Research Data Policies: Seachange or Zeitgeist?
Presented by Dr. Wolfram HORSTMANN (Oxford, UK)
https://indico.cern.ch/contributionDisplay.py?sessionId=8&contribId=12&confId=211600

Přednáška přinášela řadu doporučení (definovala politiky) ohledně ukládání vědeckých dat. Celá se vesměs nesla v příkladech – v Německu doporučuje DFG (Deutsche Forschungsgemeinschaft – pokud jsem to dobře pochopil, tak je to nějaká nadace nebo grantovka) ponechat vědecká data z projektu uchovaná minimálně 10 let. Human genome project – data musí být zveřejněna do 24 hodin od jejich získání (nařízení US NIH). Jiným příkladem je program CRC (Collaborative Research Centers) – jedná se o program na podporu speciálního výzkumu, který může být veden na univerzitách (ty se mohou do tohoto programu zapojit). Zapojená univerzita musí dodržovat určité podmínky, které se vztahují i k uchovávání vědeckých dat (infrastrukturu si musí na to vybudovat sama). EPSRC (Engineering and Physical Sciences Research Council, britská grantová agentura) jde ještě dál a přímo požaduje dosažení určitých cílů a procesů (požadují roadmap) v oblasti uchovávání vědeckých dat. Jako příklad institucionálních pravidel byla uvedena University of Edinburgh (specielně mají zakotveno právo o znovuvyužití dat). Přednášející dále zmínil některá pravidla Evropské komise, která především chce vše v režimu Open access. Na úrovni vlád byl ukázán příklad Británie prosazující politiku Open Data. V mezinárodním měřítku pak politika zemí G8 – co je veřejně financováno, musí být i veřejně dostupné. Celkový závěr přednášky: nic komplexního a pořádně zpracovaného nemáme v ruce, v jednotlivých doporučeních a politikách je dost chaos a řada z nich nejsou ani jednoznačně definovaná. Přednášející doporučuje podívat se na politiku Harvardské univerzity a klidně ji použít (je údajně poměrně propracovaná a přenositelná).

Interoperability of Research Data
Presented by Mrs. Donatella CASTELLI (CNR-ISTI, Italy)
https://indico.cern.ch/contributionDisplay.py?sessionId=8&contribId=13&confId=211600

Přednáška začala kritikou současné vědy, která byla nazvána skrytou. K výsledkům se lze dostat, k datům pro věření těchto výsledků už mnohem hůře. Cílem by neměla být nejen interoperabilita repozitářů s výsledky (články), ale totéž i pro vědecká data. Tedy zveřejňovat hlavně smysluplná a dobře popsaná data (jinak je to k ničemu). Na druhou stranu je nutné ponechat větší díl odpovědnosti na straně, která chce data přebrat a znovu využít (odpovědnost na straně konzumenta). Příkladem, jak to dělat, je projekt iMarine infrastructure, který má podporovat ekologické rybářství - k tomu, aby se to dalo dělat pořádně je nutné velké množství dat z různých vědeckých oblastí, tato data je třeba nějakým způsobem integrovat). Cílem je něco, co paní Castelli nazvala Virtual Research Environment, čili data z několika nezávislých repozitářů a v různých formátech dostat do jediného systému, ve kterém se to smíchá a ze kterého to budu poskytovat na přání podle toho, jak se to komu hodí. K tomu je nutné definovat řadu standardů, používat kontrolované slovníky, pluginy aj. V praxi je těch standardů ovšem strašně moc, kontrolované slovníky jsou statické (realita se mění rychleji – příkladem jsou názvy zemí). Velkou překážkou jsou nejasné licenční podmínky využívání dat. Závěr zněl, že je to velmi složité.

Quality and curation of Research Data
Presented by Mr. Kevin ASHLEY (University of Edinburgh, UK)
https://indico.cern.ch/contributionDisplay.py?sessionId=8&contribId=14&confId=211600

Přednáška byla zaměřena na kvalitu dat a jejich kontrolu – nicméně byla taková velmi obecná, spíše s radami typu, že potřebujeme globální politiky na správu a zpracování dat, data musí být kvalitní, zpracování se má dělat strojově atp. Zajímavá byla poznámka ohledně aktuálnosti dat: pro určité obory lidské činnosti mohou být data důležitá v kontextu ubíhajícího času – např. ekonomická real-time data (už 15 minut zpoždění u ekonomických data může mít dalekosáhlé finanční důsledky).

Working with large data sets
Presented by Dr. Tim SMITH (CERN, Switzerland)
https://indico.cern.ch/contributionDisplay.py?sessionId=8&contribId=15&confId=211600

Poměrně zajímavá přednáška o zpracování velkých objemů dat v CERNu. LHC (Large Hadron Collider) produkuje během svého provozu ohromné množství dat – v současné době nelze veškerá data, která z LHC lezou ukládat (uvádí se cca 10 PB/s), proto se tato data filtrují, aby se neukládaly zbytečnosti, ale jen zajímavý obsah. Výsledkem je proud cca 6 GB/s, což už je dobře zpracovatelné. Za tři roky provozu LHC je uloženo cca 100 PB (peta byte) dat (fyzicky cca 80 tisíc disků, pro zpracování pak 88 tisíc procesorových jader – v konečném důsledku jde ale většina dat na magnetické pásky, což je stále nejvýhodnější médium pro tyto účely). Takové množství dat se musí replikovat (nelze mít pouze na jednom místě kvůli riziku poškození), což je další velký problém – kam to ukládat (jen v CERNu mají dvě kopie), jak to přenášet. Na tato data existuje celosvětový LHC grid, přenos je speciálně optimalizován na data z LHC, v přenosech je speciální logika. Data LHC rok 2012 – po lokálních sítích se prohnaly 3 EB (exa byty) dat, po globálních sítích asi 300 PB. Systémy používané pro ukládání dat – Invenio, InSPIRE, DPHEP, CASTOR (Cern Advantage STORage manager). Provádí se zároveň verifikace uložených dat (ztrátovost cca 0,000065 %). Závěr: nejlépe si data ohlídá jejich vlastník; je těžké dopředu odhadovat, kolik dat bude; udržování uložených dat v dobrém stavu je nákladné (časově i finančně) a vyžaduje znalosti z mnoha oborů (IT, knihovníci, manažeři, vědci).

Breakout Groups

Formou diskuse se měla řešit různá aktuální témata – probíhalo 5 paralelních sekcí. Já jsem navštívil sekci **Open Annotations**, která mi jako programátorovi byla nejbližší. Diskusi vedl Robert Sanderson. Zúčastnilo se poměrně velké množství lidí – ze začátku se hlavně probírali možné způsoby využití standardu Open Annotations, které do značné míry vycházely ještě ze středeční přednášky. Pak to už malinko sklouzlo do technických detailů (a trošku se to vylidnilo :-)). Řešili jsme hodně, jak odkazovat části dokumentů – prakticky je to tak, že já mohu mít zájem např. udělat poznámku jen k nějaké části obrázku (a nikoli k obrázku jako celku), případně jen k části dokumentu (ať už na stránce nebo PDF) apod. Zkoušeli jsme, co vlastně Open Annotations snese, jak se ty odkazy dají udělat. Např. pro obrázek lze uvést souřadnice podobrázku, pro text v HTML to je možné řešit přes <a name="..."> tagy a mřížkové odkazy, u PDF jsme nic moc použitelného nevymysleli :-(. Dále se probírala dostupnost knihoven pro implementaci OpenAnnotations, technické detaily okolo JSONu, přesný formát vnitřního zápisu Open Annotations. (Sám si musí celý standard ještě jednou prostudovat a některé věci rozmyslet, ale řekl bych, že většina účastníků diskuse na to byla velmi podobně. Je otázka, kam se to bude vyvíjet a jak to půjde využít.)

21. 6. 2013

Plenary 5: Arts, Humanities, and Social Sciences

Shrnutí: Poslední den byl věnován Open Access a věcem s ním spojeným. Dopoledne se přednášky zabývaly problematikou Open Access v oblasti humanitních věd, byla představena řada projektů a zajímavých zkušeností. Mj. jak otevřená data využívá Google, projekt OpenBooks a sociální síť MLA Commons.

OA Research Monographs in HSS: Opportunities & Challenges
Presented by Mr. Rupert GATTI (Open Book Publishers, UK)
https://indico.cern.ch/contributionDisplay.py?sessionId=11&contribId=16&confId=211600

Přednáška začala kritikou současného modelu publikování, který zamezuje přístup ke znalostem (současný finanční model publikování je závislý na NEPŘÍSTUPU ke znalostem – ten kdo platí, způsobuje zvým způsobem znepřístupnění znalostí ostatním). A hned bylo konstatováno, že OA je pro HSS (Humanities and social science) spasitelem :-). Britský RCUK (Research Council UK) je fondem/grantovkou poskytující značné prostředky na vědu v Británii. Pokud poskytne prostředky, pak vyžaduje, aby výsledky byly publikovány buď v režimu Gold OA nebo tak, aby po 6 měsících (nebo 12 měsících pro HSS) byla archivní kopie uvolněny pro kohokoli (neplatí pro monografie). Výhodou tohoto nařízení je, že to podporuje vláda a tlačí to na vydavatele i univerzity (aby na Gold OA přešli). Špatné na tom je, že RCUK požaduje Gold OA všude, což znamená, že APC (article processing charge) jsou v podstatě jediná možnost financování, peníze ale mnohdy chybějí. Ideální není ani to, že to v podstatě svým způsobem udržuje současný status quo na trhu a nebourá to moc bariéry pro netradiční publikování (tedy stávající byznys běží vesele dál). V UK na OA reagovali vydavatelé (21 vydavatelů historických časopisů) poměrně striktně: podkope to současný systém vědy v UK a zneužije se to pro plagiátorství a komercionalizaci třetími stranami – čili to striktně odmítli. Velkým argument pak je odlišnost HSS od tvrdé vědy (v HSS je s OA méně zkušeností, monografie mají odlišný charakter, není to tak moc financováno z grantů, méně se spolupracuje, výsledky jsou hůře měřitelné, …). Nicméně lze na to jít i jinak – OpenBook Publishers – dát nějakou přidanou hodnotu: interakce se čtenářem (online komentáře – učitel-student-autor), multimédia, napojení na primární zdroje. Kniha v elektronické podobě může být i výrazně levnější! Je zde řada nevyřešených problémů: jak se to bude financovat, jak se do toho zapojí knihovny a akademici (a přijmou to vůbec?) - potřebuje to podporu institucí, což je v současnosti nejslabší místo OpenBooks. Ukázaný finanční model pro OpenBooks má čísla příjmy/výdaje plus mínus vyrovnány – na straně příjmů jsou to peníze z prodeje knih (~75 %) a grantů (~25 %). Možnosti financování jsou různé: přímo z institucí nebo knihoven, fondů, grantovek, případně využít například nabídky programu SpringerOpen (Springer nabízí vědcům, že jim pomůže nějakým způsobem s náklady na OA publikování – nebylo upřesněno jakým způsobem a za jakých podmínek). Na straně knihoven je to problém – stále mají proces akvizice nastaven standardním způsobem a s „adoptováním“ OA publikací moc nepočítají – přitom elektronické knihy mohou mít nezanedbatelnou přidanou hodnotu. Knihovny zároveň vyzývá, aby jej kontaktovaly, pokud mají o danou problematiku a produkci OpenBook zájem (rupert.gatti@openbookpublishers.com). Financující organizace pak vyzývá vzletnými frázemi typu: „Otevřená věda je dobrá pro veřejné blaho – podobně jako je například veřejné osvětlení.“. Uvádí také zajímavé srovnání licencí, nakladatelů a nákladů autora v oblasti OA (viz slajdy ke stažení). Závěrem navrhuje především vydělávat na přidaných službách – tedy ať je všechno volně dostupné a zároveň se k tomu za peníze dodává nadstandard.

The Humanities in and for the Digital Age
Presented by Mrs. Kathleen FITZPATRICK (Modern Language Association, USA)
https://indico.cern.ch/contributionDisplay.py?sessionId=11&contribId=17&confId=211600

Přednáška byla trošku filozofická o vztahu HSS a digitálních technologií (nemají se moc rádi). Oborem Digital humanities (užitím počítačů v humanitních vědách, digitalizační projekty aj.) se zřejmě zabývá řada lidí (a platí to vláda USA) – dokonce existují asociace (celosvětová a evropská), které se danou problematikou zabývají, časopis Digital Humanities Quartely aj. V přednášce byla řada ukázek z různých projektů a pak také spousta různých klišé o tom, jak počítače ovlivňují HSS. Z příkladů mě zaujal projekt TILE, software, který se zabývá zpracováním digitalizovaných textů (rozpoznávání textu mezi obrázky, poloautomatické vyznačování řádků textu apod.). Pak začala být přednáška celkem zábavná – přednášející vyprávěla svoji anabázi o tom, jak se pokoušela dostat do formy knihy svoji disertační práci (hnaná ironickým poňoukáním vlastní maminky „Ona na tom chce někdo vydělávat?“ :-)). Nejdříve tedy rozeslala desítky nabídek do vydavatelství, pak jí něco schválili, ona připravila knihu (cca 2 roky), přišla ekonomická krize a oni ji následně odmítli. Začala o tom psát blog (http://www.plannedobsolescence.net/kathleen-fitzpatrick/), kde celou anabázi popsala – blog se stal poměrně populárním načež se jí ozvalo několik nakladatelů a výsledkem je vydaná kniha, která celé to mnohaleté martýrium popisuje :-). Tedy nakonec pomohly právě digitální technologie (původní disertačku nakonec také vydala jako knihu) – čili technicky není problém, problém je pouze sociální. Na závěr byla lehce rozebrána sociální síť MLA Commons – něco jako Facebook pro vědce.

Empowering Development: Why Open is Right for Development
Presented by Mr. Carlos ROSSEL (The World Bank)
https://indico.cern.ch/contributionDisplay.py?sessionId=11&contribId=18&confId=211600

Přednášející začal ukázkou klasických rozevírajících se nůžek mezi chudým „jihem“ a bohatým „severem“ – přičemž to vztáhnul k vědě a OA – kdo nemá na předplatné, nedostane se k vědeckým informacím a toto onu propast dále prohlubuje. Kritizována byla netransparentnost financování, ukazoval tam nějaká čísla a povídal, jak to dělá Světová banka. Ta od roku 2010 zveřejňuje veškerá svá data, články, publikace a jiný obsah. (To pak například používá Google pro různé své pomocné aplikace – například na dotaz „population Vietnam“ ukáže Google nějaké grafy a data má právě z WB). Od července 2012 běží WB v režimu OA (licence CC BY) – jde-li o data přímo z WB, pokud jde o externí přispěvatele, pak je požadován Green OA (opět CC BY) – vše by mělo být v Open Knowledge Repository (OKR, https://openknowledge.worldbank.com/ - mimochodem mají to v DSpace, dělá to pro ně @mire), včetně profilů autorů (napojení na ORCID), citace v Google Scholar, … Mají podporu pro autory, kteří chtějí publikovat i řadu smluv s předními nakladateli (Elsevier, Taylor & francis), takže pokud váš výzkum sponzoruje WB, tak mát u těchto vydavatelů usnadněnu práci. Na konci ještě shrnul, proč je OA dobré.

Panel Session: Gold OA Infrastructure [přiloženo foto na konci zprávy]

V rámci panelové diskuse o Gold OA diskutovali a na dotazy odpovídali:

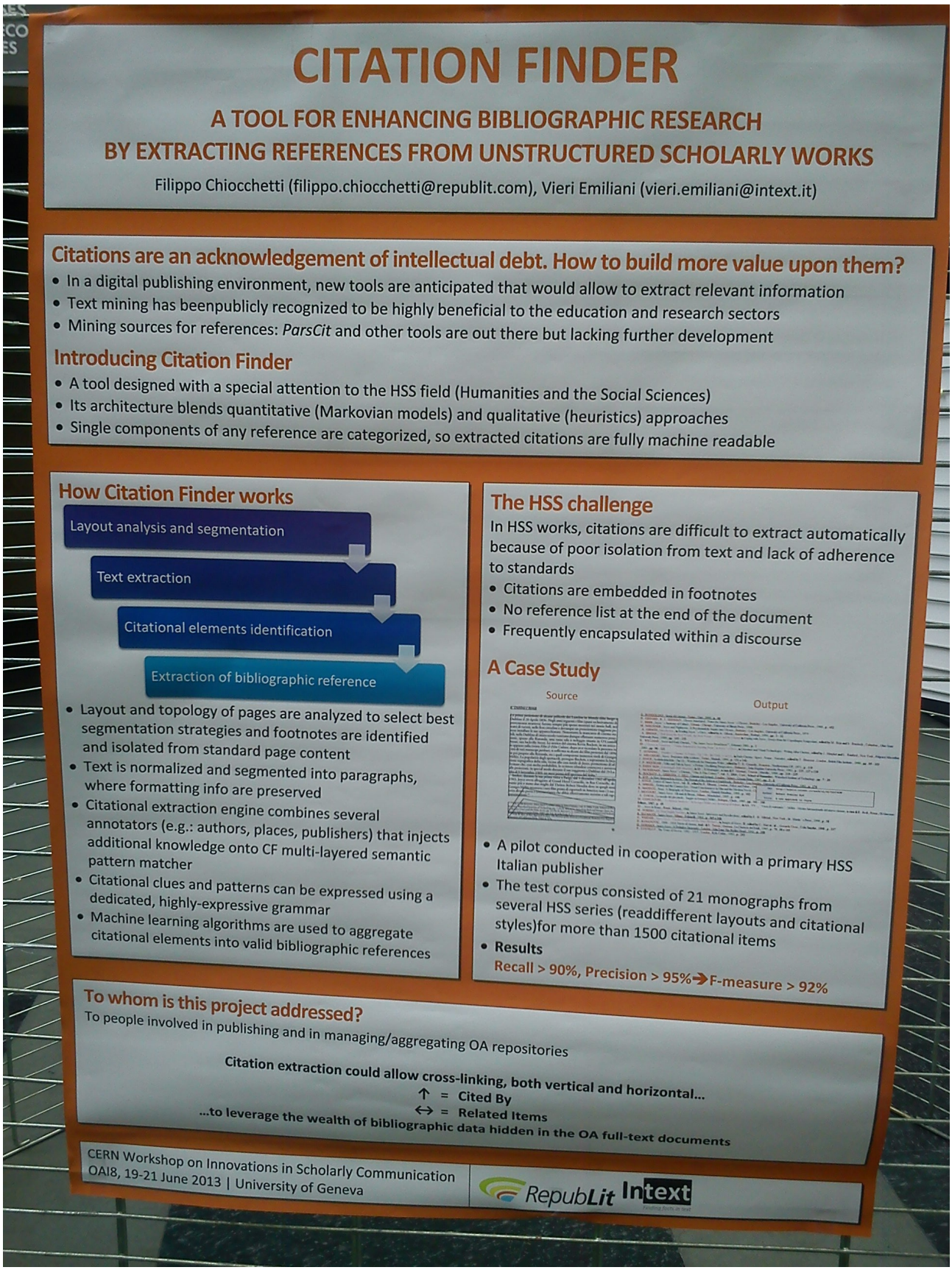
1. Mr. Lars BjøRNSHAUGE (SPARC Europe, Denmark),
2. Mr. Johannes FOURNIER (German Research Foundation (DFG)),
3. Mr. Simon THOMSON (OAK, Ireland),
4. Mr. Geoffrey BILDER (CrossRef, UK).

Pokusím se alespoň v hlavních bodech shrnout některé z názorů, které tam padly. Bohužel nejsem expert na OA, takže je možné, že jsem všechno nechytil úplně přesně.

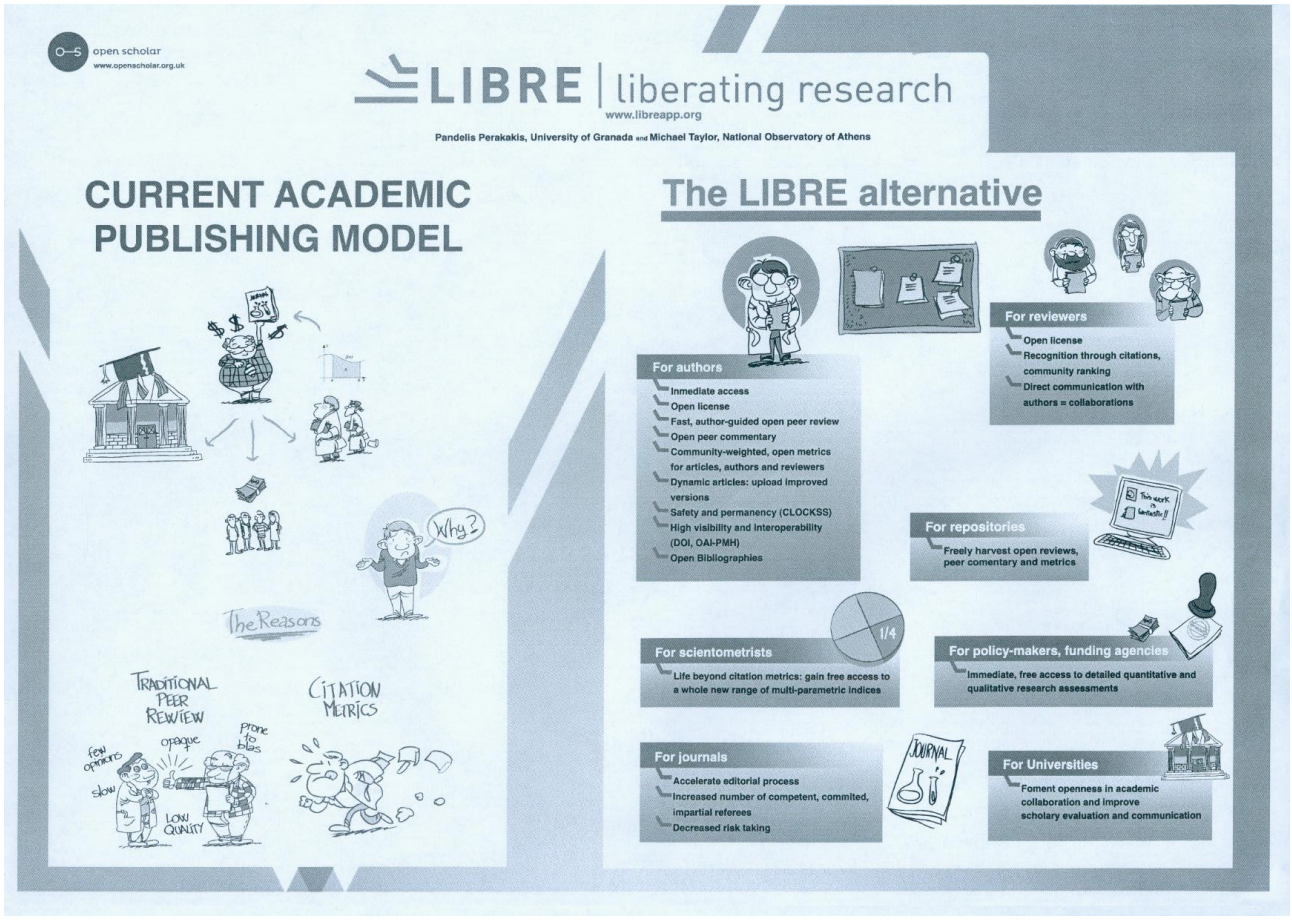
- Musíme změnit sociální chování uživatelů – jsou zvyklí na něco a je třeba jim ukázat, že GoOA je lepší. Mít uživatele na své straně. Musíme jim rozumět a vědět, co chtějí, musíme je znát.
- Potřebujeme vyřešit financování – v současnosti zde zeje velká mezera. Chybějí programy financování, nadace, které by to podporovali. Zlepšit průhlednost (proč to stojí nyní právě tolik?) Budou to financovat knihovny? Padlo doporučení pro začátek použít financování stropem (dát na GoOA nějaký balík peněz, 2. uváděl, že oni mají 2 tisíce EUR ročně).
- Ohledně financování se také mluvilo o možnosti jak ušetřit – neplatit individuálně za článek, ale platit nějak více hromadně (instituce by měla nějaké smlouvy s vydavateli) – mohlo by to být levnější. (Něco o tom říkal i prof. Deketelaere v Closing Keynote, viz konec zprávy).
- Chybějící infrastruktura – spíše byl závěr nebudovat novou, ale využít nějak lépe stávající (ale úplně se neshodli co s tím) – jako problém vidí heterogenitu prostředí (různí vědci z různých univerzit, finance přes různé granty s jinak nastavenými pravidly), což celý proces velice komplikuje.
- Chybějící nástroje pro GoOA – nemáme software ani pro vědce, ani pro vydavatele (SW pro samotné procesy i pro financování aj.). SW musí být především modulární, aby zvládal heterogenní procesy.
- Netrvat na konkrétních licencích („dostaneš peníze, ale jen když použiješ tu t tu licenci...“), odstraňovat bariéry.
- Podporovat OA vydavatele (zejména ty malé), aby se staly silnějšími. Mluvit s vydavateli a přesvědčovat je, že OA je dobré.
- Do (Go)OA musím jít cílevědomě i se všemi riziky, které to nese (nedělat to napůl).

Closing Keynote
Prof. Deketelaere, Kurt (LERU - League of European Research Universities).

Prof. Deketelaere je generální sekretář LERU, což je organizace zastupující 21 významných evropských univerzit (mj. Cambridge a Oxford). Snaží se ovlivňovat nastavení pravidel a politik na národních úrovních (protože bez podpory na vládní úrovni není celé OA dost silné, potřebuje mít za zády dobrou legislativu) i na úrovni EU (lobuje v Evropském parlamentu za vědu, OA a věci související). Byl představen program Horizon 2020 – do tohoto roku si dali za cíl mít všechno v režimu OA. LERU podporují spíše Green OA, ale zvažují, že za určitých příznivějších podmínek (nejsou na to teď peníze) půjdou i na Gold OA – zejména když to zaplatí někdo jiný než vědec nebo univerzita. Možná by bylo zajímavé, aby se přímo univerzity staly Gold OA vydavatelem. Prof. Deketelaere a LERu navrhují se spojit (univerzity) a společně zatlačit na 5 největších vydavatelů – nedělejme desítky různých doporučení a politik, ale mějme jen jednu. Zkusme vyjít z doporučení (byť nezávazných EU) => vpřed za světlými zítřky :-).



. Poster Citation Finder



. Poster Liberating research

. Kolega Lhoták mi udělal fotku na důkaz, že jsem se snažil :-)

. Panelová diskuse ke Gold OA (panelisté zprava 1. - 4.)