

Open Repositories 2019: all the user needs Hamburg, Německo 10. 6. – 14. 6. 2019

Účastnice:
Petra Černošávková, odd. 812

červen 2019



Struktura cestovní zprávy:

1. Účel cesty

Účelem služební cesty byla účast na mezinárodní konferenci Open Repositories 2019, jejíž letošní ročník byl věnován veškerým uživatelským potřebám. Cílem cesty bylo získání nejnovějších poznatků v oblasti repozitářů, uživatelských potřeb a chování, technologie související s repozitáři a pak také o vývoji softwaru Invenio, na němž repozitáře IDR a NUŠL běží a kde právě probíhá rozsáhlý upgrade.

2. Časový průběh cesty

Harmonogram služební cesty:

9. 6. v 8:30 vlakem z Prahy do Hamburku

10. 6. účast na celodenním workshopu Invenio User Group Workshop 2019

11. 6. – 13. 6. konference

14. 6. návrat do Prahy vlakem

3. Průběh navštívené akce/jednání

Workshop: Invenio User Group Workshop 2019

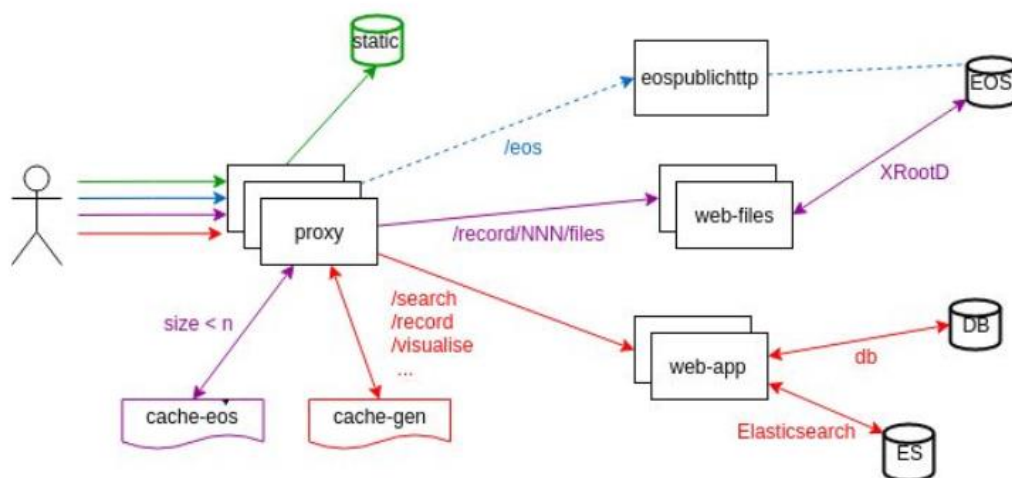
Workshop vedl Lars Holm Nielsen a Jose Benito Gonzalez Lopez z CERN a Kai Wörner z Universität Hamburg. Během workshopu proběhlo tzv. „round table“, kdy každý zájemce dostal příležitost seznámit ostatní s jejich vztahem k softwaru Invenio (v3) a se současným stavem jejich softwaru. Zástupce za NTK rovněž stručně shrnul stav vývoje našich repozitářů.

Lars H. N. představil stručně koncept Invenia 3. Dále se zmínil, že vývojářům nedoporučuje používat kód Zenoda. Nicméně tým Zenoda by měl projít své kódy a podívat se, o co by se mohli podělit, a co by bylo rozumně použitelné pro většinu repozitářů.

Martin Koehler představil záměry pracovní skupiny, která chce zachovávat v jádru Invenia MARC a vyzýval ostatní účastníky k zapojení do této skupiny. Kristi Holmes představila praxi v Chicagu (*Using Invenio RDM to power biomedical research*), kde využili pro vybudování vlastní infrastruktury mapování uživatelského chování (výstup COAR) k požadavkům repozitářů druhé generace (viz <https://github.com/data2health/menRva>). Zároveň uvažují o rozšíření sbírky o šedou literaturu.

Další příspěvek představil CERN Open Data portal a práci CERNu s velkými daty. Repräsentace dat (interactive visualisations) na tomto portálu je velmi zajímavá a mohla by možná být v nějaké podobě aplikována pro browsing našich sbírek. Dokumentace dostupná z: <https://gitlab.cern.ch/invenio/template-openshift>. Tento portál zajímavě řeší přístupy uživatelů a vyhledávání (viz. Obr. 1), kde pokud spadnou web-files, tak web-app stále běží. Obdobně rozdělení web-app na databázi a elastic umožňuje uživatelům databázi alespoň prohlížet, když spadne elastic. Ukázalo se, že toto řešení / infrastruktura je pro velká data dostatečně pružné.

Invenio with resilience



11

Obrázek 1 Printsreen prezentace Diego Rodriguez Rodriguez

https://indico.cern.ch/event/818650/contributions/3447485/attachments/1859298/3055061/or2019-Invenio_CERN_Open_Data_portal.pdf

Další příspěvek popisoval migraci knihovního systému v CERN (ILS) na Invenio 3, z MARC na JSON (autorka Karolina Przerwa, její github <https://github.com/kprzerwa>). Pracovní skupina na migraci ILS vyvinula vlastní nástroj pro čištění dat před migrací v reakci na problémy způsobené při čištění metadat o videích až po migraci (pravděpodobně součástí tohoto repo <https://github.com/kprzerwa/invenio-app-ils>). „If we don't clean before then won't be able to do it after.“ I oni hádají význam některých polí podle jejich obsahu, jelikož se znalost ztratila s fluktuací lidí.

Velmi inspirativním příspěvkem byl CERN Enterprise search od Pablo Panero a to zejména v řešení rozhraní a parsování query. Zároveň stručně představil další velkolepé plány. V jeho prezentaci¹ jsou odkazy na veškerou dokumentaci.

Cristina Mossoni vyvíjí na Inveniu 3 digitální knihovnu rukopisů. Při upgradu Invenia 3 na verzi 3.1 bylo pro ni snazší nainstalovat ho celé znova. Píše knihu o svých zkušenostech s Inveniem 3.

Chiara Bigarella představila COUNTER-nástroj pro statistiky využívání datasetů (aplikovaný v Zenodu), který zohledňuje verzování DOI podle verzí datasetů. Dále tento nástroj zohledňuje různé relace, aby se eliminovaly snahy autorů uměle si zvýšit statistiky, ať už ručně nebo strojově.

Lars v další prezentaci představil nadcházející roční projekt CERNu ve spolupráci s 10 multioborovými institucemi na vývoji Invenia (cíl: „turn-key open source research data management platform called InvenioRDM“). InvenioRDM se bude zakládat na Zenodu a Inveniu v3 Framework. Cílem projektu je i integrace RDM s OpenAIRE a GitHub. Poslední prezentace se zaměřila na budování (uživatelské) komunity – doporučení aplikovat „stakeholders analysis“. Závěr workshopu sloužil ke shromáždění všech poznámek od účastníků ohledně dalšího rozvoje Invenia a doporučení, na co by se CERN měl primárně zaměřit. Zároveň byli účastníci dotazováni, jakým způsobem by chtěli přispět k rozvoji Invenia.

¹ dostupná z

https://indico.cern.ch/event/818650/contributions/3447475/attachments/1859318/3055100/CERN_Search_OR19.pdf

Keynote Jeff Gothelf

Keynote speech byla věnována důležitosti user experience (UX) a srovnání s user interface (UI). Odborník na UX Jeff Gothelf vysvětloval, jak dosáhnout lepší UX a zdůrazňoval kritiku vlastní práce – to, že jsme něco vymysleli a vyvinuli, neznamená, že je to dobře a že jsme to měli vyvíjet. Poukázal na metodu Agile, z níž je patrné, že UX a vývoj systémů je nekonečný běh. Upozornil na očekávání uživatelů, kteří od všech systémů vyžadují stejné chování jako od těch největších, které používají každý den jako Google, Amazon, Netflix, ...

Téměř po celý zbytek konference probíhalo několik paralelních sekcí. Dále jsou popsány navštívené sekce, u nichž byl shledán největší informační přínos pro rozvoj repozitářů v NTK.

P2B designing interfaces with UX

Nejen z této sekce byla patrná rostoucí popularita vizualizace dat v repozitářích. Tato sekce představila tři projekty, jejichž součástí byl vývoj systému/repozitáře. Během vytváření UI se dotyční řídili UX a reflektovali své kroky s uživateli systémů – „i knihovníci jsou uživatelé a i oni si zaslouží pěkné rozhraní“ zaznělo během příspěvku Experts and Novices: redesigning user interfaces for the White Rose Repositories. Velmi zajímavým výstupem z projektu je francouzský systém Polaris OS, který využívá ElasticSearch a Docker. Jako NUŠL disponuje bilingvním rozhraním i metadaty. Github systému: <https://github.com/MyScienceWork/PolarisOS>

Posterová sekce byla velmi rozsáhlá a nebylo možné detailně shlédnout ani polovinu posterů. Během této sekce jsem se spojila se zástupci OpenAIRE a společně jsme dořešili věci ze strany harvestování NUŠL a jeho správy v OpenAIRE.

Panel „Towards next generation repositories“

V panelu byli zástupci z COAR, Invenia, DuraSpace, Antleaf a 4Science. Každý z nich stručně představil, odkud je a na čem pracuje v souvislosti s repozitáři příští generace. Zaznělo, že současné repozitáře jsou již vlastně repozitáři příští generace. Jedná se o repozitáře, které jsou schopné komunikovat nejen s dalšími repozitáři, ale i s jinými systémy; jejich záběr je mnohem celistvější a komplexnější, hlavním úkolem není pouhé ukládání dokumentů. Jen vlastně stále není jasné, jaké jsou základní komponenty, které by repozitáře nové generace měli mít, jelikož každý repozitář se snaží naplnit jiné uživatelské potřeby.

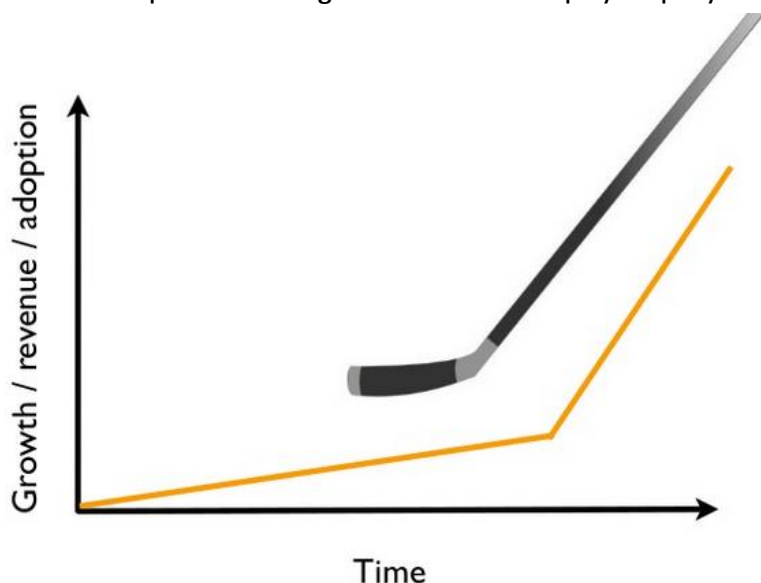
Sekce Reuse

Sekce Reuse byla v konceptu 24x7. Příspěvek z CERN o FAIR principech zdůrazňoval, že FAIR začíná již na začátku výzkumu, jinak dramaticky klesá možnost znovu využití výzkumných dat. Upozorňuje na výsledky výzkumu [Monyi Baker](#), kdy kolem 50 % vědců nebylo schopno reprodukovat vlastní výzkum. Další z příspěvků prezentoval možnosti sdílení obrázků skrze Wikimedia Commons a zvýšení tak jejich dopadu. Posledním příspěvkem této sekce byla integrace ORCID v repozitáři Cadmus (software DSpace). Pro realizaci integrace ORCID a automatického přiřazování identifikátorů vědcům v instituci, bylo potřeba naučit spolu komunikovat asi tři systémy – repozitář, ORCID, interní systém(y) s evidencí vědců a jejich délkou působnosti v instituci.

Closing keynote Heather Piwowar

Konferenci uzavřela příspěvkem „The Breakout Moment for open repositories is now“

Heather Piwowar, která poukazovala na fakt, že repozitáře nikdy nebyly využívány tak hojně jako nyní. Domnívá se, že se repozitáře nacházejí ve zlomovém bodě tzv. Hockeystick graph viz obr. 2 níže. Zároveň představila svoji práci na unpaywall, který pro koncového uživatele představuje pouze doplněk do prohlížeče, ale za nímž je obsah právě velkého množství databází a repozitářů s legálně volně dostupnými plnými texty vědeckých článků.



Obrázek 2 Printscreen prezentace Heather Piwowar; dostupné z: <https://www.slideshare.net/hpiwowar/open-repositories-2019-closing-keynote-heather-piwowar>

4. Závěry/shrnutí

Účast na konferenci byla velmi přínosná, a to jak účast na workshopu k Inveniu 3, tak na vybraných částech obecného programu. Všechny poznatky budou hodnotným příspěvkem v průběhu upgradu softwaru Invenio 3 na obou repozitářích NTK. Velmi důležité jsou kontakty na vývojářskou komunitu kolem Invenia i na provozovatele repozitářů s tímto softwarem, které umožní případné sdílení zkušeností nebo řešení konkrétních problémů, které mohou při upgradu vzniknout. Z dalšího programu konference bych vyzdvihla především část věnovaná tvorbě a vylepšení uživatelského rozhraní systémů, která bude rovněž velmi užitečná při tvorbě nového rozhraní pro NUŠL i IDR.

5. Dokumentace

- Program konference na webu <https://or2019.blogs.uni-hamburg.de/program/>
 - prezentace a papery jsou dostupné v detailech jednotlivých sekcí
- Videozáznamy by měly být výhledově dostupné z: <https://lecture2go.uni-hamburg.de/en/web/vod> pod licencí CC-BY-SA 4.0
- Program workshopu a prezentace jsou dostupné z: <https://indico.cern.ch/event/818650/timetable/#20190610.detailed>

6. Anotace ZC

Konference Open Repositories se snaží účastníkům přiblížit nejnovější trendy v oblasti digitálních repozitářů, jedná se tedy o velmi důležité téma vzhledem k tomu, že NTK provozuje dva digitální repozitáře. Účastnice získala přehled o nejnovějším vývoji technologií v oblasti repozitářů a měla možnost navázat kontakty s provozovateli podobných repozitářů po celém světě. Nejprínosnější byla sekce věnovaná softwaru Invenio, který NTK využívá pro oba repozitáře a pracuje na upgradu na verzi 3.