

<http://www.techlib.cz/>

NTK

Národní technická knihovna
National Library of Technology

Cestovní zpráva

13th International Digital Curation Conference

Hotel Hesperia, Barcelona, Španělsko

19. – 22. 2. 2018

Účastnice:

Petra Černošávková, kurátorka IDR a NUŠL, odd. 11

březen 2018

Struktura cestovní zprávy:

1. Účel cesty

Cesta byla schválena mimo PHÚ, jelikož 13. ročník mezinárodní konference o digitálním kurátorství (IDCC) věnovaný převážně datům a data management plans (DMP) je pro Národní úložiště šedé literatury (NUŠL) velmi důležitý, kdy research data, typ šedé literatury, nabývají na významu a na jejich zveřejňování je kladen stále větší důraz. Účast na konferenci byla pouze pasivní, protože v současné době nebylo nic, co bychom mohli vzhledem k zaměření konference prezentovat, nicméně i přesto byla pasivní účast z informačního hlediska velkým přínosem. Účelem cesty bylo získání nových poznatků a inspirace od odborníků, kteří spravují datové repozitáře nebo poskytují služby s tím spojené, které mohou posloužit jako podklad pro zvážení možností zapojení NTK v shromažďování výzkumných dat.

2. Časový průběh cesty

- 18. 2. letadlem do Barcelony
- 19. 2. účast na 2 workshopech
- 20. a 21. 2. IDCC konference
- 22. 2. účast na workshopu
- 23. 2. návrat do Prahy letadlem

3. Průběh navštívené akce/jednání

19. 2. účastnice navštívila dva workshopy věnované deklaraci European Open Science Cloud (EOSC) a možnostem, které tato deklarace může přinést. První nesl název „*EOSC as a 'skills commons' for developing research data stewardship skills at scale*“ a byl zaměřen více na oblast vzdělávání v kontextu a možnostech EOSC a na novou funkci/pozici na univerzitách a výzkumných centrech a to data steward(ship). Druhý workshop „*How Research Institutions and Libraries can help deliver the European Open Science Cloud*“ volně navazoval na první, ale tentokrát z pohledu vztahů a postojů výzkumných institucí a knihoven k EOSC.

Konference jako taková trvala dva dny (20. a 21. 2.) a byla nabyta mnoha přínosnými příspěvky. První den konference zahájila jako keynote Sabina Leonelli s příspěvkem „*No (Open) Science Without Data Curation: Five Lessons From the Study of Data Journeys*“, ve kterém pojednává o výsledcích výzkumu ohledně „putování dat“. Studie odhalila klíčové výzvy pro implementaci Open Science pro práci s daty: 1) context-specific curation is key to data reuse, 2) long-term maintenance is key to

trustworthiness, 3) which data and why?, 4) data and materials, 5) role of ethics, humanities & social sciences in data management. Závěrem uvádí, že nelze dosáhnout spolehlivé a efektivní otevřené vědy, pokud není aplikováno datové kurátorství, které zajistí důvěryhodnost dat.

Druhý den zahájil příspěvek „*Blending analytics and curation: data explorations from a library in a cultural organization*“ od Luise Martínez-Urbe, který působí v DataLab při knihovně Fundación Juan March. Příspěvek přináší náhled do aktivit datové laboratoře a ukazuje velké možnosti analýz a využití umělé inteligence při jejich analýze, pokud je o ně staráno digitálním kurátorem od počátku jejich vzniku po celý jejich lifecycle. Byla to krásná ukázka dobré praxe, když je součástí knihovny datová laboratoř, což je věc samozřejmě finančně náročná.

Konferenci uzavíral třetí keynote a to Nancy McGovern s příspěvkem „*Collaborating Across Communities: Leveraging Our Strengths for Sustainable Programs and Services*.“ Příspěvek se snaží motivovat všechny typy institucí, které se zajímají o data, dlouhodobé uchovávání k intenzivní komunikaci a přinést je ke kulatému stolu: archivy, knihovny, muzea, data science, digitální uchovávání, softwarové vývojáře a records management, kdy každý z nich může do diskuze přinést trochu jiný pohled a zkušenost a jako celek jsou silnější, mají společné zájmy, překrývající se cíle a vzájemně se jednotlivé části překrývají. Příspěvek nastiňuje možnosti zlepšení spolupráce a vybudování inkluzivní komunity.

Další nově nabyté poznatky byly získány zejména v následujících oblastech: technická řešení a metadata, reporty a metriky, vzdělávání a data stewardship jako pozice, FAIR principles a data management plans a GDPR.

Za první oblast bych poukázala na příspěvek „*Giving datasets context: a comparison study of institutional repositories that apply varying degrees of curation*“, na němž se podílely čtyři univerzity v USA. Tyto univerzity provedly srovnávací studii institucionálních repozitářů, které aplikují různé úrovně kurátorství. Z jejich výzkumu dochází ke čtyřem doporučením: zvažte institucionální faktory, když hodnotíte kurátorství, zvažte kurátorské praktiky a nabídněte best practice pokyny pro různé scénáře, metadatové schéma by mělo být standardizované za účelem výšení podpory výměny dat mezi repozitáři, komunita by měla evaluovat rozdíly mezi schématy a vyvinout minimální požadavky na metadata pro datasety v instit. repozitářích.

Dále bych ráda zmínila příspěvek z Německa, který naráží na stálé opakování a kladení důrazu na FAIR principy avšak opomíjení problému dlouhodobého uchovávání data setů při různorodosti formátů („*Are research data sets FAIR in the long run?*“). Jeho autoři jsou D. Wehrle a K. Rechert, kteří testovali datasety z 92 repozitářů (cca 2TB dat), které pak 12 nástroji mnoho hodin analyzovali. Po všech úpravách skončili na 145 formátech, nezahrnovalo verze jednotlivých formátů, z nich pouze 32 označili za velmi vhodné pro LTP, 10 středně a zbylých 103 bylo označeno „s nízkou pravděpodobností udržitelnosti“. V závěru označují FAIR principy za důležitý koncept LTP, i když je velmi abstraktní. Jedná se o multidimenzionální problém, na který neexistuje jedno řešení.

V oblasti research data plan a management (RDM) a data stewardship se ukázalo, že jsou nejdále

Nizozemci, zejména pak TU Delf. Hovořili také o datovém repozitáři DANS, který je v provozu již řadu let. Příspěvek „*Data Stewardship TU Delft*“ od Ellen Verbakel prezentuje krátce historii jejich práce v oblasti dat, úspěchy v r. 2017, RDM výzkum (425 respondentů) a plány na r. 2018. Prezentace rozbírá zejména výsledky výzkumu, který ukazuje, že většina respondentů nezálohuje svá data okamžitě, skoro polovina se není jistá, komu patří jejich data, atd. Doufají, že tyto výsledky jim pomohou získat pozornost a podporu pro implementaci politiky ohledně RDM.

22. 2. byl navštíven třetí workshop – „*A Research Data Shared Service in a localised, international context—what could/should it mean for you?*“ věnovaný platformě pro data management, který nabízí uložení do repozitáře s okamžitým uložením kopie k dlouhodobé archivaci v AIP balíčku. Tento systém byl vyvinut neziskovou organizací JISC pro oblast UK. Workshop byl věnován možností, jak z modelu udělat mezinárodně fungující službu, která by zajistila i vyžadovanou udržitelnost dat. Označila bych tento workshop za ukázkou dobré praxe a odvedeného kusu práce na vývoji datového repozitáře s přidanými hodnotami pro velké užití.

4. Závěry/shrnutí

Zahraniční cesta byla pro záměr budování datového repozitáře velmi přínosná, přinesla prvotní podrobný vhled do problematiky a do nových trendů na mezinárodní úrovni. Zároveň umožnila navázat kontakty mezi odborníky v dané oblasti, které už vlastní repozitáře spravují. Příspěvky jsou také inspirací pro další ročníky mezinárodní konference o šedé literatuře a repozitářích pod záštitou NTK. Z poznatků nabytých na konferenci se domnívám, že pro NTK nyní není nejvhodnější okamžik zahájit budování datového repozitáře. NUŠL nyní vyžaduje mnoho úprav a vylepšení a přesun na vyšší verzi Invenia a není zdaleka připraven řešit velké objemy dat a pestrost jejich formátů z hlediska udržitelnosti (viz příspěvek „*Are research data sets FAIR in the long run?*“ od D. Wehrle a K. Rechart). Zároveň vybudování samotného datového repozitáře je otázka pár let, ale potřeba ukládání datových setů již existuje a je velmi ovlivněna Horizontem 2020. Po dlouhé úvaze nad celým obsahem konference a stavem repozitářů v NTK jsem dospěla závěru, že ve výhledné době by bylo lepší se věnovat data stewardship – tj. poskytovat podporu ohledně research data management, kam data ukládat, jaké nástroje jsou vhodné, jak je to s ochranou osobních údajů, autorskými právy atd.

5. Dokumentace

Prezentace jsou dostupné z webu konference: <http://www.dcc.ac.uk/events/idcc18/programme>

Flash disk s plnými texty příspěvků je v kanceláři účastnice, Petry Černohlávkové.

6. Anotace SC

Cestovní zpráva popisuje průběh účasti na 13. ročníku mezinárodní konference o digitálním kurátorství (IDCC) věnovaném převážně datům a data management plans. Účast na konferenci byla klíčová pro plány vybudování nového repozitáře. Zahraněční cesta přinesla prvotní podrobný vhled do problematiky a do nových trendů na mezinárodní úrovni. Zároveň umožnila navázat kontakty mezi odborníky v dané oblasti, které už vlastní repozitáře spravují. Poznatky a texty příspěvků mohou být využity jako podklad pro zvážení možností NTK v oblasti research data and data stewardship, zároveň jsou inspirací pro další ročníky mezinárodní konference o šedé literatuře a repozitářích pod záštitou NTK.