

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

Cestovní zpráva

Digitálna knižnica 2012
Ukladanie a prístupnenie vedeckeho a vzdelávacieho
digitálneho obsahu
Hotel Družba, Jasná pod Chopkom, Slovensko
16. 4. 2012 – 18. 4. 2012

Účastnice:

PhDr. Petra Pejšová, (petra.pejsova@techlib.cz, NUŠL, 51)

Mgr. Hana Vyčítalová, (hana.vycitalova@techlib.cz, NUŠL, 51)

Duben 2012

1. Účel cesty

Cílem zahraniční služební cesty byl 6. ročník konference Digitálna knižnica 2012. Konference nesla letos podnázev „Ukladanie a prístupnenie vedeckého a vzdelávacieho digitálneho obsahu“ a uskutečnila se v prostorách hotelu Družba v Jasné pod Chopkom na Slovensku ve dnech 16. - 18. dubna 2012. Konference měla především informovat odbornou komunitu z oblasti knihovnictví a muzejnictví o trendech v oblasti digitalizace, uchování a zpřístupňování digitálních dat, o nástrojích a technologiích používaných při procesu digitalizace a v prostředí digitálních knihoven. Stejně tak měla představit různé projekty, v rámci kterých vznikají digitální knihovny a archivy. Konference se zúčastnily Mgr. Hana Vyčítalová a PhDr. Petra Pejšová, která zde vystoupila s prezentací o OpenGrey. Konference se zúčastnilo přes 120 zástupců z 12 zemí, včetně účastníků z USA.

2. Časový průběh cesty

Pro cestu na konferenci bylo využito vlakové spojení ze stanice Praha Hl. n. do stanice Liptovský Mikuláš. Z Liptovského Mikuláše do zastávky Jasná pod Chopkom účastnice jely autobusem. Pro zpáteční cestu byl opět využit vlak z Liptovského Mikuláše do Prahy s přestupem ve stanici Žilina.

3. Průběh navštívené akce/jednání

Během zahraniční cesty bylo ubytování zajištěno v hotelu Družba, kde se konal i program konference. Konferenční poplatek a ubytování platil oběma účastnicím pořadatel. Po příjezdu dne 16. 4. 2012 proběhla od 17:00 do 19:00 registrace účastníků a následovala uvítací recepce.

17. 4. 2012

Program konference byl zahájen v úterý dopolední sekcí „Aktuálne trendy v budovanie digitálnych knižnic“.

První příspěvek přednesl Josef Dzivák ze Slovenské chemické knižnice a byl věnován Národnímu projektu Digitálna knižnica a digitálny archiv. Přednášející v krátkosti popsal celý digitalizační projekt, nástroje a infrastrukturu, které budou pro digitalizační proces použity. Zajímavým nástrojem je např. software WITNESS, který umožňuje simulaci digitalizačního procesu a tím pádem i identifikaci rizikových faktorů.

Plánované trvání projektu je od 2/2012 do 6/2015, rozpočet 49,6 mil. EUR, zdigitlizováno má být 250 mil. stran. 7.3.2012 byl projekt podepsán ředitelem NK a ministrem kultury Slovenské republiky. V současné době nakupují jen technologie a služby vytváří sami.

Následovala přednáška z českého prostředí o projektu Česká digitální knihovna, kterou prezentoval Martin Lhoták z Knihovny Akademie věd.

Hovořil o vytvoření agregátoru tzv. centrálním Krameriovi. Agregátor by měl prohledávat jednotlivé digitální knihovny vytvořené v systému Kramerius a v budoucnu by byl schopný harvestovat metadata i z jiných digitálních knihoven. Agregátor bude poskytovat jednotné vyhledávací rozhraní pro uživatele a předávat data do Europeany. Agregátor bude nejen index, ale metadata budou přímo v systému a digitální dokumenty se též stáhnou do systému.

Vývoj Krameria v roce 2012 je zaměřen především na statistiky, VuFind, OAI-PMH harvester, vylepšení administrátorského a uživatelského rozhraní. Dokumentace je dostupná na www.CODE.GOOGLE.COM/P/KRAMERIUS. Aktuální instalace běží v Moravské zemské knihovně.

Pan Lhoták rovněž zmínil několik novinek o Registru digitalizace. Je vyvíjen systém pro digitalizační workflow Registru digitalizace ČR. Cílem je vytvořit jednotlivé instalace pro knihovny, aby si mohly spravovat své kroky v rámci digitalizačního workflow. KNAV usilovala o mezinárodní projekt - ČR, Rakousko, Slovensko, Polsko... Projekt nebyl podpořen, hledají dál.

Dále vyvíjí následující nástroje:

RDflow – digitalizační workflow management systém

ProArc – produkční a archivační systém – dvě funkce:

- produkce dig. dokumentů dle standardů NK ČR,
- ALTP - systém pro archivaci.

Cíle – sběr digitálního obsahu, jednotné uživatelské rozhraní a odesílání dat do Europeany.

Všechny systémy jsou/budou volně k dispozici dalším knihovnám. Zaměření na open source technologie, aby mohli být šířeny, dívají se také ve světě, co bylo vyvinuto a šlo by využít.

Po českém příspěvku vystoupila paní Christen Madsen z Oxfordské univerzitní knihovny. Její prezentace se týkala projektu Digital Bodleian - tvorby digitální knihovny v prostředí univerzitní knihovny. Tento, pouze devítiměsíční, projekt je podporován JISC a má za cíl zpřístupnit digitální dokumenty ze sbírek Bodleian Library jak uživatelům, tak strojům pro vyhledávání. Výsledkem by mělo být zhruba 300 000 digitálních objektů ve 40 kolekcích. Byly zmíněny použité nástroje, metadata i výstupní formáty digitálních objektů.

Byl využit open source SW Greenstone a připojení k Europeaně. Poučení z pohledu řešitelů je, že data se nemění, ale uživatelské rozhraní ano. Projekt je pod licenci CC pro nekomerční užití.

Po krátké přestávce vystoupil Michael Popham, rovněž z Oxfordské univerzitní knihovny, s příspěvkem, ve kterém informoval o dalších dvou digitalizačních projektech, které probíhají v rámci budování Oxfordské digitální knihovny. Jeden z projektů byl zaměřen na speciální kolekci z Oxfordské univerzitní knihovny, digitalizace byla podpořena JISC. Druhý projekt proběhl ve spolupráci se společností Google, jednalo se o dokumenty z různých sbírek, které už nejsou chráněny copyrightem.

Dopolední sekci zakončil Dariusz Paradowski z Polské národní knihovny. Zaměřil se na porovnání vyhledávacích rozhraní několika evropských digitálních knihoven, jako jsou Gallica a další. Zdůraznil současné trendy, které v současné době ovlivňují vzhled a funkčnost uživatelských rozhraní. Trendy v digitálních knihovnách se zaměřují na uživatelské rozhraní, na získání výsledku hned na první straně bez nutnosti „klikat“ víckrát. Vyhledávání stále častěji není spojeno s daty.

Odpolední sekce nesla stejný název jako dopolední a obsahovala osm příspěvků. Hned po obědě prezentoval pan Aleksander Nowiński z Varšavské univerzity příspěvek s názvem Otevírání vědy v Polsku. Nastínil nejrůznější iniciativy, které se v Polsku rozběhly na podporu sdílení vědeckých informací a Open Access. Zmínil zejména projekt CeON – Centre for open science a projekt Open a book (digitalizace knih). Tyto iniciativy jsou podporované rovněž ze strany polské vlády.

Zmíněná řešení:

YADDA – virtual Library SW vytvořeno v ICM - http://yaddainfo.icm.edu.pl/?page_id=46

PBN – Polish Scholarly Bibliography

SYNAT platforma pro vědecký obsah

CeON – vytvořili platformu pro vytváření digitálních repozitářů a knihoven – velice modulární – pouze metadata, ne digitální dokumenty

YADDA2 - platforma = portál integrující ostatní repozitáře v Polsku.

Jako další vystoupila Petra Pejšová z NTK s přednáškou o OpenGrey (evropská databáze šedé literatury). Prezentace je přílohou této zprávy.

Před přestávkou přednesl Łukasz Mesek z Jagellonské knihovny v Polsku příspěvek o Jagellonské digitální knihovně, kde se zaměřili především na ochranu a digitalizaci dokumentů, které byly poškozeny fyzicky uživateli, častým užíváním, díky špatnému materiálu (kyselý papír), případně špatným podmínkám, ve kterých byly dokumenty uloženy.

Po přestávce zazněla přednáška Juraje Žabky ze společnosti HP o nově vzniklé digitální knihovně Paula McCartneyho. Obsahem této knihovny jsou zdigitalizované materiály, týkající se Paula McCartneyho – nahrávky, videa, fotografie, novinové články atd. Digitalizaci zajišťovala firma HP. Digitální knihovna je dostupná v celém světě 24 hodin denně, 7 dní v týdnu.

<http://h10124.www1.hp.com/campaigns/enterprise/mccartney/us/en/paul-mccartney-digital-library-experience.html#nid=home>

Následující příspěvek Norberta Végha ze Slovenské knižnice pre nevidiacích byl věnován technikám vytváření zvukových knih pro nevidomé (při tvorbě těchto knih se využívají digitalizované texty). Velmi zajímavá byla ukázka několika typů hlasových syntetizátorů – open source SW NVDA, placeného JAWS a syntetizátoru, který vzniká v prostředí Ústavu informatiky Slovenské akademie věd http://www.ui.sav.sk/speech/projlist_s.php. Problémem použití syntetizátorů je především samotný slovenský jazyk s jeho fonetickou stránkou.

Na podobné téma hovořila i další přednášející – paní Mária Šimková ze Slovenské akademie věd. Hovořila o tvorbě Slovenského národního korpusu, který vytváří Ústav pro lingvistiku SAV.

Další příspěvek byl spojený s digitalizací muzejních sbírek v prostředí Muzea Slovenského národního povstání. Digitalizace muzejních sbírek je rozhodně náročným úkolem, za cíl si v Muzeu SNP dali, aby digitalizované objekty mohly být zkoumány online stejně dobře jako samotné objekty. Ján Jurkovi upozornil, že na Slovensku zcela chybí muzejní informatici.

Poslední prezentaci uvedli Alojz Androvič a Ivan Cíglan, týkala se Centrálního datového archivu pro Slovenskou republiku.

Poslední dvě prezentace představily projektu vznikající v rámci dotačního programu OPIS (Operační program informatizácia spoločnosti) viz <http://www.opis.culture.gov.sk/>.

18. 4.

Poslední sekce konference nesla název „Technologie pre budovanie digitálnych knižnic“. Úvodní příspěvek přednesl Christoph Becker z Technické univerzity ve Vídni. Hovořil o plánování ochrany dokumentů a projektu SCAPE (Scalable Preservation Environments) podpořenému Evropskou komisí.

Následovaly dvě komerční prezentace zástupců firmy IBM a Treventus, obě zaměřené na technologie pro digitalizaci.

Po přestávce přišly na řadu dva příspěvky z Národní knihovny České republiky. Jiří Polišenský přednášel o budování Virtuální depozitní knihovny, která bude sloužit pro potřeby depozitních knihoven a v určitých případech se může jednat i o unikátní dokument z jiné knihovny, ale v tom případě dostane status „archivní konzervační jednotky“.

Za tímto účelem je vytvořen modul připojený k Registru digitalizace, který eviduje záznamy o stavu jednotky. Příspěvek doplnil Pavel Kocourek z firmy INCAD, který představil systém PROKOP – framework pro sklizení bibliografických záznamů a kontrolu duplicit záznamů a chybně vytvořených záznamů. Cílem nástroje PROKOP je udržovat aktuální stav záznamu ve všech zapojených systémech. Nyní se testuje na bázích NK ČR, v budoucnu budou sklízet data i z jiných knihoven. Framework – sklizení dat – algoritmické operace – analýzy – vyhodnocení – scénáře řešení – výstupem je soubor (text, statistika, SQL příkazy).

DNK je stále v plánu, již přešli na řešení skenování dokumentovým skenerem = rozřezání multiplikátů. Také vznikl matematický vzorec kolik je potřeba exemplářů pro uchování titulů na určitý počet let.

Druhá prezentace Tomáše Foltýna se týkala nástrojů pro optimalizaci digitalizačních procesů – zejména OCR. Představil dva nástroje pro zlepšování kvality OCR – DICON a CODEG.

DICON – vazby nastavení obrazových souborů ke skeneru = zda je obrázek v pořádku = zda je nastavení správně. Umožňuje 16 kalibračních parametrů a na základě naskenování

vyhodnotí, zda obrázek odpovídá parametrům. Vyvinutá aplikace bude zdarma k dispozici, organizace si koupí pouze kalibrační tabulku.

CODEC – příprava dokumentů (Sirius) – OCR (Fine Reader) – data (SIP) = kontrola a ruční oprava. Využito v projektu ANL+ a IMPACT. Oba nástroje vyvinula firma Elsynt Engineering. Plánují též lematizaci ve spolupráci s UK, která má český jazykový korpus.

Následovaly tři kratší, ale zajímavé prezentace, všechny z prostředí Žilinské univerzity. První se týkala analýz webu a médií, které souvisí s digitálními knihovnami – např. počet přístupů, charaktery dotazů uživatelů, o které pole ze záznamu se uživatelé nejčastěji zajímají apod., ale také sledování sociálních médií. Pro tyto analýzy byly představeny nástroje Klábosení, Google Alerts, Buzzboot, nebo Newton Media a Altaxo Social Média (poslední tři nástroje jsou placené). Zajímavý je též nástroj Clic Tale http://www.clicktale.com/default_b. V rámci prezentace byl představen průzkum, jak využívají knihovny statistické nástroje. Do průzkumu bylo zařazeno 113 knihoven z celého světa. Zjištěno bylo, že 37% knihoven nepoužívá statistické nástroje vůbec, 59% knihoven používá Google Analytics a 2% knihoven používá dva statistické nástroje. Na závěr prezentace vyzvaly přednášející k zamyšlení nad novými kompetencemi v knihovnách, zda by v knihovnách měla vzniknout funkce webových analytik.

Druhá prezentace byla o možnostech podpory vědecké spolupráce. Téma je zaměřeno na vývoj formální i neformální spolupráce. Hlavní funkce spolupráce je produkce znalostí pro praxi a následné vzdělávání účastníků. Metodika se vyvíjí s opakováním cyklů, které se formují v následujících krocích: 1. vytváření podmínek pro spolupráci, 2. výzkum 3. závěr, pokud je špatný, tak se celý cyklus opakuje.

Poslední prezentace byla zaměřena na uživatelskou analýzu kvantity metadat v databázích.

Témata, která zazněla v druhé a třetí prezentaci, jsou řešena v rámci projektu Paměť Slovenska.

4. Závěry/shrnutí

Zahraněční cesta na konferenci Digitálna knižnica 2012 se konala od 16. dubna do 18. dubna 2012. Na konferenci byl 17. 4. 2012 prezentován příspěvek s názvem OpenGrey Petrou Pejšovou. Hlavní náplní tohoto ročníku bylo seznámení s různými digitalizačními projekty, s technologiemi a nástroji, které jsou určeny pro budování a fungování digitálních knihoven. Součástí tohoto ročníku bylo také několik komerčních prezentací firem HP, IBM a Treventus. Trendy v digitálních knihovnách se zaměřují na uživatelské rozhraní, na získání výsledku hned na první straně bez nutnosti „klikat“ víckrát. Vyhledávání stále častěji není spojeno s daty. V muzeích chybí muzejní informatici a v knihovnách se nabízí vznik funkce webového analytika.

Byly získány nové kontakty:

Lubica Jedličová SPU Nitra – plánují instalovat Invenio pro institucionální repozitář a zajímají se o připojení k OpenGrey – mají asi 8 tis. VŠKP, 10 let zpět sborníky z konferencí, nyní též sborníky z konferencí pořádaných další institucí.

Dariusz Paradowski NK Poland – předá kontakt na oddělení šedé literatury.

5. Dokumentace

Webové stránky akce <http://www.schk.sk/digilib/> - budou zde vystaveny prezentace přednášejících.

Program akce

Prezentace OpenGrey

6. Anotace

Účelem zahraniční cesty byla účast PhDr. Petry Pejšové a Mgr. Hany Vyčítalové na konferenci Digitálna knižnica 2012, která se konala od 16. dubna 2012 do 18. dubna 2012 na Slovensku v hotelu Družba v Jasné pod Chopkom. PhDr. Petra Pejšová zde 17. 4. 2012

přednesla příspěvek s názvem OpenGrey, ve kterém seznámila účastníky konference s evropskou databází šedé literatury, možnostmi zapojení do OpenGrey a procesu zapojení Národního úložiště šedé literatury do této databáze.