



Zpráva ze zahraniční služební cesty

Jméno a příjmení účastníka cesty	Libor Coufal
Pracoviště – dle organizační struktury	OAW 8.1.1
Pracoviště – zařazení	knihovník
Důvod cesty	Jednání pracovní skupiny IIPC Preservation a konference The future of the past of the web
Místo – město	Londýn
Místo – země	Velká Británie
Datum (od-do)	6. – 7. 10. 2011
Podrobný časový harmonogram	6.10. dopoledne – přilet, jednání PWG 7.10. konference, odlet
Spolucestující z NK	J. Hutař, R. Kreibich
Finanční zajištění	VaV.0121
Cíle cesty	- účast na jednání pracovní skupiny Preservation (PWG) - účast na konferenci
Plnění cílů cesty (konkrétně)	Viz zápis v příloze
Program a další podrobnější informace	Viz zápis v příloze
Přivezené materiály	
Datum předložení zprávy	12.10.2011
Podpis předkladatele zprávy	
Podpis nadřízeného	
Vloženo na Intranet	
Přijato v mezinárodním oddělení	

Čtvrtek 6.10.2011

Jednání skupiny Preservation Working Group

Clement Oury (BnF) – Arc to Warc conversion mapping

BnF připravila dokument, který obsahuje návrh mapování jednotlivých polí při konverzi z formátu Arc na formát Warc. Dokument popisuje jednotlivá pole v Arc a Warc, údaje, které se do nich zapisují a jak jsou tato pole mapována mezi Arc/Warc. Dokument byl podrobně projednán. Konečným cílem by mělo být vytvoření určitého standardu, kterého by sloužil jako vodítko pro instituce při migraci z Arc na Warc. Dokument bude postupně dále upřesňován. (Pozn.: NK ČR v současnosti také pracuje na této problematice v rámci institucionálního výzkumu – naše závěry budou použity pro srovnání; dále se ukazuje jako nutné vzít v úvahu strukturu a možnosti zapisování do Warců při sklizení v budoucnu – struktura těchto „původních“ a konvertovaných Warců by si měla co nejvíce odpovídat).

Amanda Spencer (TNA) – informace o Pronom/Droid

TNA pracuje na nových verzích nástrojů Pronom a Droid

Pronom

- měl by být konvertován do formátu LinkedData
- připravují požadavky, poté najmou externího programátora (příští rok)

Droid

- proběhne veřejná konzultace k dalšímu vývoji, výsledky by se měly promítnout do příští verze Droid
Steve Knight (NLNZ) – zdá se, že některé požadavky/potřeby TNA jsou v rozporu s tím, co potřebují ostatní instituce – bylo by vhodné více koordinovat a najít společné zájmy, které by se měly vzít v potaz při vývoji

Barbara Sierman (KB) a Sven Schlarb (ONB) – informace o evropském projektu SCAPE

- 2012 – 2016, 16 partnerů
- 5 podprojektů (Platform, Preservation and watch, Preservation components, Take up, Testbed)
- k dispozici první výsledky (zpráva Evaluation of characterisation tools - Part 1: Identification dostupný ze stránek projektu Open Planets
http://www.openplanetsfoundation.org/system/files/SCAPE_PC_WP1_identification21092011_0.pdf a zpráva Identification and selection of large-scale migration tools and services http://www.scape-project.eu/wp-content/uploads/2011/09/SCAPE_D10.1_KEEPS_V1.0.pdf)

Pátek 7.10.2011

Konference The future of the past of the web

Podrobný program konference je na http://www.dpconline.org/events/details/35-future_past_web?xref=35

Výběr z nejzajímavějších příspěvků:

H. Van der – mluvil o projektu Memento; tento projekt byl představen už na dřívějších akcích IIPC, naposledy na valném shromáždění IIPC v Haagu v květnu t.r. V současnosti byly vytvořeny další aplikace pro Firefox, Android, mcurl a platforma pro Wikimedia. V druhé části příspěvku přednesl svou vizi o nutnosti archivovat akademické a vědecké publikace na webu. Tato nutnost vyplývá z povahy těchto publikací, jejichž obsah, kontext a reference se, na rozdíl od tradičních tištěných publikací, často mění. Jeho vize řešení tohoto problému vychází z rozšíření metodologie použité v projektu Memento na tyto typy publikací, tak aby jejich uživatelé byli schopni využívat webové vědecké publikace v původním kontextu tak, jak vypadal při jejich vzniku.

Maureen Pennock a Lewis Crawford z BL představili nové možnosti práce s britským webovým archivem – 3D vizualizace archivovaných verzí stránek, generování tag cloudů z nejčastěji se vyskytujících klíčových slov, n-gramy – srovnání četnosti výskytu 2 klíčových slov nebo frází v čase – vše založeno za aplikaci Big Sheets od IBM. Vyhledávání v archivu umožňuje také klustrovat výsledky podle 3 typů – dokumenty, obrázky a média.

Celkově se celá konference nesla v duchu budoucího využití archivů – možností práce s obrovskými objemy dat a data miningu.