

Zpráva ze zahraniční služební cesty

Jméno účastníka cesty	Mgr. Jan Hutař
Pracoviště – instituce, adresa	Národní knihovna ČR, Klementinum 190, Praha 1, 110 00
Pracoviště – zařazení	Odbor digitální ochrany 1.5
Důvod cesty	<ul style="list-style-type: none">– účast na konferenci International Conference on Preservation of Digital Objects 2009- konzultace s přítomnými odborníky a firmami- účast na PREMIS workshopu
Místo – město	San Francisco
Místo – země	USA
Datum (od-do)	4.-8.10. 2009
Podrobný časový harmonogram	<ul style="list-style-type: none">– 4.10. let Praha-Frankfurt-San Francisco– 5-6.10. iPres konference– 7.10. PREMIS workshop
Spolucestující z NK	Libor Coufal
Finanční zajištění	- VaV.INST.0131
Cíle cesty	- viz důvod cesty
Plnění cílů cesty	- v souladu s cíly
Program a další podrobnější informace	viz níže
Přivezené materiály	
Datum předložení zprávy	24.11. 2009
Podpis předkladatele zprávy	

poznámky z nejdůležitějších prezentací:

<http://www.cdlib.org/iPres/confsched.html>

1. den konference

David Kirch *"The Public Interest in Private Digital Records and Why We Should Care if Corporations have the Right to be Forgotten."*

Prezentace byla velmi obecná, spíše o strategiích velkých firem a o tom, jak záznamy při krachu společnosti mizí, ačkoliv mají povinnost je uchovávat.

- i velké firmy by měly mít archivy a records... není pravidlem v USA (viz firmy, kt. v krizi zkrachovaly...) – nikdo se nikdy nedozví proč, co bylo špatně apod.
- navíc ty archivy jsou s mnoha formáty, vypovídají o firmách, kt. už neexistují apod.
- v USA na to není zákon, v Evropě je to ošetřené

měl by se vytvořit nějaký speciální archiv dokumentů

- návrhy na řešení 5 návrhů

-----x-----

Brian Lavoie / Abby Smith / Martha Anderson / Patricia Cruse. *Perspectives on the Economics of Sustainable Digital Preservation (Panel discussion).*

- DP (digital preservation) jako ekonomický problém
- DP jako soběstačná a udržitelná ekonomická aktivita – zajistit zdroje dlouhodobě

- task force – LOC, mellon, jisc atd. - v lednu 2009 vznikla závěrečná zpráva
- příklady z reálného provozu

sustainibilita je když:

- rozpoznání přínosu
- donucení k jednání těch, kteří rozhodují
- výběr dokumentů k ochraně - prioritace
- mechanismy pro podporu rozdělení zdrojů
- vhodná organizace a dohled

je to o omezování risku – hlavně technických risků a finančních hrozeb také

- snížit ekonomický risk pro DO (digitální objekty)
- požadovat aktivní investice od managementu

panel:

Cruse – jak DP ovlivňuje její univerzitu? \UCLA – ona je za CDL

- curation services- pomáhají studentům a hlavně lidem, kt. dělají výzkum s celým do lifecyklem DO
- nakonec to pomůže celé univerzitě a získají kredit pro své výzkumy

Marta Anderson LOC

- lidé v kongresu chápou problém digi informací
- nejlepší je jim ukázat jakou cenu mají tyto informace pro národ a zemi
- chceme ochránit výsledky výzkumu, kt. jste podpořili- to chápou
- ještě lepší je jim říct, že je to potřeba pro education
- chtějí ale vidět, že peníze, kt. na DP dají se využijí dobře – tj .dobře připravený projekt

Paul

- nikdo nedělá DP kvůli ochraně samotné, spíše z nějakého jiného důvodu

předpokladem OAccess je DP

budou mít seznam otázek pro každou instituci, kt. si pak bude moci udělat test zda mají sustainable DP

-----x-----

Henry Lowood. *Memento Mundi: Are Virtual Worlds History?*

- culture and entertainment
- closed virtual worlds – otázka online her, kt. přestaly být provozovány – kde jsou ta data?
- tj. virtuální světy jsou historie
- je to vlastně kontext historie a kultury, kt. ztrácíme
- když se to stane, čeho si všimneme jako prvního?
- deathswitch
- etický problém – Andy Rauber

-----x-----

Reinhard Altenhöner. *E-infrastructure and Digital Preservation: Challenges and Outlook.*

- data
- 7th framework
- ecosystem of eu digital repositories
- 2008 national hosting strategy

- GRID
- velký zájem o řešení v rámci Německa
- kde jsme, v jaké situaci pokud se týká řešení DP?
- hodnocení EU projektů
- jaké SW jsou hotové pro DP? dias, rosetta etc.
- velmi důležitý je výstup z projektu Parseinsight (jejich roadmapa)
- stále je patrný nedostatek společného postupu pro řešení DP

planning strategy level – spolupráce na této úrovni funguje

stále je ovšem minimální infrastruktura pro DP

technology level – spolupráce funguje

slide 10 velmi důležitý – kdo z těch knihoven, které nějak reálně uvádějí, že řeší DP dělá např. DP planning? Kdo dělá migrace? Nutno přiznat, že nikdo. Klíčovým zůstávají tedy kooperace a integrace snah o řešení DP – slide 13!

dp4lib

bit stream preservation v Německé národní knihovně je outsourcovaný

Úkoly k řešení v NK Německa - rozšíření Kopalů – stávající Kopal berou jako projekt na testování (např. Ingestu)

Je potřeba více mezinárodní spolupráce, více úsilí na technickém řešení a integrace workflow spojené úsilí pro služby – Pronom apod.

-----x-----

Stephen Arbams. *An Emergent Micro-Services Approach to Digital Curation Infrastructure.*

more stuff, smaller budget – více materiálu, více formátů, méně peněz.

- změna technologie, user expectations
- data sustainability plan – velmi důležité
- data curation není potřeba mít centralizované
- bezpečnost díky vícenásobnému uložení

Důraz je kladen na obsah, ne na systém ve kt. je obsah uložený. Lepší malé a jednoduché než velké a komplikované.

micro services- lehké na vývoj a údržbu a případné nahrazení

- interoperable
- úsilí na jejich vývoj nesmí být velké, lze se toho pak lépe zbavit

Rozlišuje mezi preservation (kontext a uložení) a curation (interoperabilita, aplikace – služby, kt. přidají hodnotu – vyhledávání, indexace apod.).

Následující slide je zajímavý – lifecycle of curation > preservation

(<http://www.cdlib.org/iPres/presentations/Abrams.pdf>).

Popisuje micro services v CDL – vlastně příkazy k operacím nad úložištěm, systém uložení ve file systému a popisu jednotlivých úrovní složek.

-----x-----

Pam Armstrong and Johanna Smith. *Preserving the Digital Memory of the Government of Canada: Influence and Collaboration with Records Creators.*

2004 kanadská NK a národní archiv byly spojeny

- projekt na trusted digital repository 2008-2010 zaměřený na electronic records
- value management framework
- zabývají se i web 2.0, databázemi
- 2003 audit on record keeping – začalo se to řešit
- budou mít vlastní LTP systém

Alone we can go faster, together we can go further .

Mají local digital format registry – tj. seznam co berou a co ne a případně co s tím dělají (např. word do pdf/a apod.)

-----x-----

Robert Sharpe. *Are You Ready? Assessing Whether Organizations are Prepared for Digital Preservation.*

Informace o projektu Planets.

Proběhl průzkum 96 národních archivů a NK v Evropě: zda jsou připraveni používat nástroje DP.

Ukázalo se, že knihovny nemají kontrolu nad formáty dat, které přijímají, naopak archivy mají skoro úplnou. 77% knihoven je připraveno.

-----x-----

Adam Farquhar. *Implementing Metadata that Guides Digital Preservation Services.*

metadata v BL (Britská knihovna)

- keep preservation simple, not too simple
- learning from experience
- mají tiffy, neví jestli je dobrý nápad je převádět do jpeg2000 (lossy or lossless)
- musí na to mít zajištěno dlouhodobé financování, jinak do toho nepujdou

DO mají charakteristiky, environment taky má charakteristiky a nároky (viewer, hw, sw nároky apod.).

Akce samotné mají vlastnosti (charakteristiky) – náklady, rychlost, HW nároky apod.

Problém je s hodnotami - počet stran, počet obrazů apod. DO může mít více hodnot, kt. se pak melou (např. 1cm vs. 0.01 metru). Důležité je odkud hodnoty pocházejí atd.

migrace – může změnit objekt i prostředí

objekty, akce a prostředí je potřeba popsat (jhove, pronom, planets testbed apod.)

Některá metadata některá nemusí být uchovávána napořád!

-----x-----

Rebecca Guenther and Robert Wolfe. *Integrating Metadata Standards to Support Long-Term Preservation of Digital Assets: Developing Best Practices for Expressing Preservation Metadata in a Container Format.*

PREMIS a METS

PIM tool – umí udělat z mets s premisem samotný premis a naopak – lze stáhnout z premis webu id.loc.gov- bude tam slovník věcí, kt. se do elementů dají vyplnit – kontrolované slovníky

2. den konference

Martha Anderson. *The National Digital Stewardship Alliance Charter: Enabling Collaboration to Achieve National Digital Preservation.*

Obecné povídání o vývoji NDIIPP. Selecting, collecting and preserving.

Cíle- identifikovat dokumenty v ohrožení> vyvinout národní strategii k jejich ochraně. Dnes to řeší mnohé instituce v usa, nejen knihovny, ale i komerční sektor, vydavatelé apod.

DP network je potřeba – nutné sdílet expertní znalosti (oblasti: infrastruktura, hodnocení risků, metadata, management, access etc etc)

- vlastně popsala vývoj NDIIPP po jednotlivých fázích.
- dalším důrazem v pozdějších fázích projektu byly nástroje a služby pro DP

Sítí musí být stále aktivní – něco nového aby o ní bylo vědět. Obecné problémy DP je lepší řešit v rámci nějaké skupiny věnované určitému typu dokumentu nebo obsahu

- content based

Kooperace na úrovni všech států usa a jejich státních knihoven.

Sítě – distribuované nebo centralizované – jejich porovnání.

trust is build by working together to common target/aim

- každopádně členové sítí musí být aktivní

-----x-----

Tyler Walters, Liz Bishoff, Emily Gore, Mark Jordan and Thomas Wilson. *Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments.*

- panelová diskuze

preservation systems složené z jednotlivých uzlů, kt. dělají ochranné akce

- lockss a private lockss network (PLN)
- chronopolis – based on SRB <http://chronopolis.sdsc.edu/publications.html>

„technology iceberg“ - technologie v DP je to co je vidět nad vodou, to důležitější – organizace, organizační změny, sustainibilita jsou pod vodou, to vidět není, ale my všichni se musíme dát do plavek a jít řešit problémy pod vodu ;-) Liz Bischof

PLN – je jich v usa 7 – vždy za nějakým účelem, je to ale vlastně jen bitstream preservation

-----x-----

Ardys Kozbial. *Chronopolis: Preserving our Digital Heritage.*

Určité státy USA nesmí mít svá data mimo hranice státu.

Mají 50TB 3x replikované, data ze 3 univerzit - většinou vědecká data.

Pro přenos metadat a dat používají bagit metodu.

Chtějí projít TRAC certifikací.

-----x-----

Christopher A. Lee. *Mainstreaming Preservation through Slicing and Dicing of Digital Repositories: Investigating Alternative Service and Resource Options for ContextMiner Using Data Grid Technology.*

Hovořil např. o automatickém downloaderu na služby typu youtube, twitter, apod. – lze nastavovat contextminer.

blogy ,flickr

Mají to nějak napojené na iRODS.

-----x-----

Esther Conway. *Towards a Methodology for Software Preservation.*

DP pro software- velmi široké téma, tj. musí se omezit.

Na co je dobrá ochrana sw? Na ochranu research dat mimo jiné, pro budoucí využití výstupů z nějaké vědecké činnosti (využití dat s původním SW, který třeba vznikl jen pro určitý projekt).

Vztah k oasis modelu. Vše to vychází z casparu, takže mrknout na jejich stránky.

-----x-----

Jens Ludwig. *Into the Archive: Potential and Limits of Standardizing the Ingest.*

Je možné mít jeden standard pro ingest long-term preservation data?

- nekonečná variabilita typů dat
- main tasks are not technická ale organizační

Jeden standard není možný – tak co s tím?

Jedině říci vydavatelům a producentům jak a co mají dodávat+ následovala prezentace příručky, kt. za tímto účelem vznikla.