

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 2.3
Důvod cesty	<ul style="list-style-type: none">- účast na konferenci Archiving 2010- konzultace k LTP systému s odborníky ze světových knihoven- navrhování nových trendů ohledně obrazových formátů a jejich uchovávání
Místo – město	Haag
Místo – země	Nizozemí
Datum (od-do)	31.5-4.6. 2010
Podrobný časový harmonogram	<ul style="list-style-type: none">- 31.5. cesta letadlem do Amsterdamu, vlakem do Haagu- 1.6. účast na workshopu- 2.- 4.6. účast na konferenci, večer návrat letecky do Prahy z Amsterdamu
Spolucestující z NK	PhDr. Bedřich Vychodil
Finanční zajištění	- VaV.INST.0131
Cíle cesty	- viz důvod cesty
Plnění cílů cesty	<ul style="list-style-type: none">- účast na konferenci- jednotlivé konzultace s odborníky
Program a další podrobnější informace	viz níže
Přivezené materiály	
Datum předložení zprávy	22.10. 2010
Podpis předkladatele zprávy	

Proceedings viz. <http://www.imaging.org/ist/store/physpub.cfm?seriesid=28&pubid=941>
poznámky z konference zahrnují jen relevantní příspěvky

Workshop - jpeg2000 Robert Buckley

R. Buckley vytvářel specifikace JPEG2000 pro kongresovou knihovnu, Welcome library, Švédsko, Norsko aj.

středa 2.6.

Access to Digital Collections: Toward a National Infrastructure, Bas Savenije, Koninklijke Bibliotheek (the Netherlands)

holandská aliance pro digital preservation – domluva s muzei, archivy o společném úsilí a kontaktu s vládou NL

to samé USA – digital preservation stewardship alliance bude ohlášena příští týden

- udělat něco podobného u nás? vytvořit národní výhled o DP společně se všemi knihovnami?
viz prezentace ředitele NL Nizozemí

NL Nizozemí– snaha zdigitalizovat kompletně sbírky do roku 2030, do tří let prvních 10%

- noviny online – 8 mil stran za rok! přístup přes jpeg2000 – image server na jpg

Assessing Digital Preservation Infrastructures: A Framework for Library, engineering and eScience Organisations, Perla Innocenti and Seamus Ross, University of Glasgow (UK);

- projekt Shaman
- využití gridu pro DP, návaznost na dramboru, trac
- shaman dp framework
- high level metodologie > assessment plan na tom postavili
- mapování irods rules na trac a na dramboru

Archivematica: Lowering the Barrier to Best Practice Digital Preservation (Focal), Peter Van Garderen, Artefactual Systems Inc. (Canada)

- vznikla před 2 lety po POC zda to půjde, nebyla ještě rosetta ani sdb
- unesco document – toward the open source repository software
- unix pipeline of oasis micro services
- python and bash
- může běžet jako aplikace na xubuntu – lze dat na usb a mit napr. v sandboxu
- vancouver archive, unesco, - měli pocit, že musí být něco pro malé instituce
- 25toolů je tam – tj. celý SW je postavený na open source nástrojích
- FITS umožňuje ohodnotit který nástroj je pro nás směrodatnější>obejít tak známé problémy lze i v archivematice
- používají uuid pro interní id> tj. nepotřebují žádný globální id – uuid nelze opakovat ;-)
- po ingestu se vytvoří mets se všemi metadaty – fits metadata mapují do premisu a ten pak do mets

mají i normalizaci

- open nástroje na konverzi – mají tam migrační plány?
- rozlišují access a preservation formáty
- media type preservation plans – format policies mají v xml schématu – lze tak menit normalisation policies v xml přímo
- umí i gridy, plánují irods a lockss
- umí bagit
- RDF migrační plány jsou online a může je využít každý
- planets testbed je fajn, ale malé knihovny chtějí rovnou vědět co a jak a čím mají migrovat, nechtějí testovat

nebude nikdy hotové, stále se vyvíjí, jak se vyvíjí technologie, nyní verze 0.6 May 2010

je to opravdu free – GNU GPL

free puppy ;-)

peter(at)artefactual.com

The eArchive Service—A Practical, Cost effective and Sustainable Approach for Preserving Authentic electronic Records, Mark Pickering, AstraZeneca PLC, and Mark Evans, Tessella, Inc. (USA)

mají sdb nebo alespon jejich konzultace
důraz na transformace před ingestem

- asi by bylo přínosné mít jen omezenou množinu preservation formátů a do nich vše dávat – ušetří nám to práci s migrací, staráním o příliš mnoho formátů, řešení problémů ke každému typu apod. v budoucnu
- používají image magick pro transformace i validity checking

**Gathering the Requirements for a Next Generation Digital Preservation Framework,
Brian Aitken, Perla Innocenti, and Seamus Ross, University of Glasgow (UK)**

- v rámci shamanu partneři měli definovat požadavky na funkcionalitu ltp systému budoucího
- snaha obohatit a rozšířit oasis model
- viz prezentace nebo článek – popis požadavků pro jednotlivé moduly oasis
- + pre-ingest requirements – není v oasis – to co se musí udělat než se dokumenty pošlou do ingestu
- ingest preparation policies
- validace ingest workflows

data management a archival storage – možnost disaster recovery z backupu – přidat do tendru?

jinak v podstatě popis toho co umí tessella a rosetta

Examining the Aptitude of Tools for File Format Migration from TIFF to JPEG2000 Using the Planets Testbed, Sven Schlarb¹, Edith Michaeler¹, Brian Aitken², Andrew Jackson³, Andrew Lindley⁴, and Max Kaiser¹; ¹Austrian National Library (Austria), ²University of Glasgow (UK), ³British Library (UK), and ⁴Austrian Institute of Technology (Austria)

stáhnout sw pack

testbed – web based aplikace

- zkusit migrace
- zkusit emulátor – najít grate emulator
- zkusit preservation plan – např. pro naše jpegy co by bylo nejlepší? preservation plan vzniká v plato nástroji
- lze zkusit mass experiments – test migrace
- lze otestovat charakterizaci

XCL na zapsání vlastností obrázků a migrací, používají to pro automatické porovnání výsledků migrace – xcl pro původní a pro výsledný obraz (dokument)

openplanetsfoundation.org

čtvrtek 3.6.

Image Quality Assurance for the Real World, Paul W. Jones and Chris W. Honsinger, Certifi Media Inc. (USA)

- techniky na kontrolu obrazů
- iso 1223, 14524, 15739 a další
- firma certifi http://www.certifi-media.com/images/Certifi_Pedigree_quality_metrics.pdf
- <http://www.certifi-media.com/>

- okolo 300msekund na jeden obraz
- integrovaná image processin QA musí být
- <http://www.certifi-media.com/QPDownload.aspx>
- metadata o kvalitě se pak ukládají s objektem

A Practical Approach: The Quest for a Large-Scale Newspaper Digitization Workflow, Edwin Klijn, Koninklijke Bibliotheek (the Netherlands)

- digitalizace novin
- skenují do tiffu lzw – ušetří 90% místa, z toho dělají preservation master v šedé škále v jp2 a uživatelskou kopii barevnou v jp2
- celé workflow je ccs a probíhá externě, včetně metadat, úpravy článků, nadpisů (rumusko, kambodža, laos)
- dohromady s místním workflow a postupy to ccs funguje dobře
- při případné migraci v budoucnu budou migrovat mastery i user copies
- v prohlížeči je užíván jp2 a pomocí luratech image serveru (vieweru) se uživateli dodává jpeg
- hlavním problémem je doba zpracování dat a výroba metadat – to nejvíce zdržuje – dělají asi 10.000 stran(den) – nemají robotické skenery, zeutschel ruční
- OCR- používají alto, ale ne v podobě mets, dávají to do mpeg21- důvodem je jen policy knihovny, žádná výhoda oproti mets v tom není – mají to tedy uložené po člancích v jednoduchém xml s hlavičkou
- nejsou preservation projekt, proto mají mastery v šedé škále
- probíhá vizuální kontrola v rámci workflow – oproti tabulkám pak v dávkách a každá dávka je nebo není přijata v KB
- skenují z papíru i z mikrofilmu – z mikrofilmu raději, je to levnější
- keep it simple stupid!!!!!! kiss principle
- kontrola QA jednou týdně dávkově (50.000 obrazů)
- <http://kranten.kb.nl/about/Overzicht>
- http://kranten.kb.nl/documents/DDD_invitationtotender_eng.pdf

+ rozhovor s E. Klijnem

- 1.2 mil pages, 8mil v 2012
- free text search, article level
- article type
- kiss – keep it simple stupid
- problémem jsou výjimky, pokud je jich hodně, je to problém
- dělat něco pokud je to moc komplikované? ne, raději slevit z nároků
- start at the end – začnět od webu, jak má vypadat a co chceme aby dělal? Dle toho to celé nastavit
- master, mix, mpeg21 with DC do LTP systému, vše ostatní jen pro web (alto xml, user copy etc.)
- přenos naskenovaných dat na externích hdd a po kontrole kvality to kb převezme
- plan processing time! kopírování apod., indexace
- users want content! i za cenu menší kvality, to je vůbec nezajímá nějaká kvalita ;-)
- 2013 – 73mil stran knih, 2030 všechny sbírky hotové

Managing Visually Lossless Compression, Christy Henshaw and Dave Thompson, Wellcome Library; and Robert Sharpe, Tessella plc (UK)

- zkušenosti z welcome library
- skeny do tiffu původně, bylo by potřeba 2PB, rozhodli se pro jp2 – visually lossless – 90% storage ušetří
- studii pro ně dělal buckley a tanner
- multiple resolution levels to make access faster
- lowest resolution image first to get thumb fast
- uk working group for jp2

- SDB – wellcome je jako první zákazník ,kt. bude používat systém pro zdigitalizované věci
- sdb vytváří kroky ve workflow pro jpeg2000, rozpoznávají lossless a non-lossless apod.
- color histogram nástroj ve workflow

Hilde Van Wijngaarden (rozhovor)

Head Acquisitions and Processing Division at Koninklijke Bibliotheek, National Library of the Netherlands

- preservation pro všechno, ale na různé úrovni – sbírky budou posuzovány podle důležitosti, míry využití teď a v budoucnu, nutnosti zachovat to a jak a taky podle typu formátu – u klasických formátů je ochrana „levná a lehká“, u komplikovaných není- dle toho se budou řadit
- tj. různé typy úrovní ochrany podle nákladů, podle toho to půjde i na různé typy HW – pásky, disky – počítají s tiery (různými typy medií na uložení)
- pošle do 3 týdnů KB requirements slíbila, já poslat exlibris ido peled prezentaci a pak výstup z POC

Pátek 4.6.

Mass CD/DVD Migration: A Novartis Case Study, Daniela Bienz and Rudolf Gschwind, University of Basel; and Mario F. Pozzaand and Ludwig Gantner, Novartis (Switzerland)

- 3000 cd a dvd a zippek, research data
- nestrukturovaná data, bez návodů a kontextu – nic neměnili, jen to dali na úložiště

- rozhodli se pro hybrid ISO image- má to porovnání, logy, kontr. součty, seznam obsahu, fotku původního média v tiffu
- iso se dá připojit jako optický disk, platform independent atd

- problémy – více než 5% chyb při původním zápisu, celé cd nečitelné apod – měli několik mechanik a sw na opravu, minimální úspěšnost